Towards the Development of Innovative Quantitative Structure-Activity Relationship Models for Human and Ecological Risk Assessment of Chemicals and their Mixtures

Edoardo Carnesecchi – 2021

Towards the Development of Innovative Quantitative Structure-Activity Relationship Models for Human and Ecological Risk Assessment of Chemicals and their Mixtures

PhD thesis

Copyright © Edoardo Carnesecchi, 2021

The copyright of published articles has been transferred to the respective publishers. No part of this thesis may be reproduced, stored or transmitted in any way or by any means without permission of the author.

The research described in this thesis was performed at Istituto di Ricerche Farmacologiche Mario Negri (Milan), European Food Safety Authority (EFSA, Parma) and Institute for Risk Assessment Sciences (IRAS), Faculty of Veterinary Medicine, Utrecht University.

Towards the Development of Innovative Quantitative Structure-Activity Relationship Models for Human and Ecological Risk Assessment of Chemicals and their Mixtures

Kwantitatieve structuur-activiteitsrelaties in de humane en ecologische risicobeoordeling van chemische stoffen en mengsels

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
dinsdag 29 juni 2021 des middags te 12.15 uur

door

Edoardo Carnesecchi

geboren op 27 Februari 1991

te Massa Marittima, Italië

Promotor:

Prof. dr. ir. J. Legler

Copromotoren:

Dr. N.I. Kramer

Dr. E. Benfenati

Dit proefschrift werd mede mogelijk gemaakt met financiële steun van de Europese Autoriteit voor Voedselveiligheid (EFSA) en de Europese Commissie, contractnummers NP/EFSA/AFSCO/2016/02 (Edoardo Carnesecchi), OC/EFSA/SCER/2018/01 (Istituto di Ricerche Farmacologiche Mario Negri IRCCS, S-IN Soluzioni Informatiche, Molecular Networks GmbH and Altamira) en LIFE16 ENV/IT/000167 (LIFE-VERMEER).

Assessment committee:

Prof. dr. R. Gehring

Prof. dr. E.H.M. Moors

Prof. dr. A.M.J. Ragas

Prof. dr. Q. Chaudhry

Dr. E. Leinala

"Vita sine proposito vaga est"

Seneca, I d.c.

Contents

Chapter 1 General introduction

Basic principles of chemical risk assessment applied to human health and the environment

Humans and the environment are exposed to thousands of chemicals of anthropogenic and natural origin, which may cause adverse health effects (Escher et al., 2020). The risks of adverse effects occurring depends on both the toxicological properties of the chemical (i.e. hazard) and exposure levels. The famous adage by the fifteenth century alchemist Paracelsus, "the dose makes the poison", neatly captures the dose dependency of risk in relation to exposure levels (Frank and Ottoboni, 2011).

Scientists have put considerable efforts into developing sound methodologies for evaluating risks resulting from exposures to chemicals, so called "chemical risk assessment", which aims at protecting human health, animal health and the environment through the characterisation of safe exposure levels (WHO, 2009). Worldwide, different regulatory bodies such as the United States Environmental Protection Agency (US EPA), the Food and Drug Administration (FDA), the European Food Safety Authority (EFSA), the European Chemicals Agency (ECHA), Health Canada, the Food and Agriculture Organisation (FAO) and the World Health Organisation (WHO) to cite but a few, derive safe levels for Human Health Risk Assessment (HHRA) and estimate the nature and probability of adverse health for regulated products (plant protection products – PPPs, food additives, feed additives, food contact materials) and environmental contaminants (WHO, 2009; EFSA, 2009). Ecological Risk Assessment (ERA) follows the same principles and aims to identify and characterise the likelihood of an agent to adversely affect living organisms, their natural habitats and whole ecosystems (Belanger et al., 2015; USEPA, PMRA, CALDPR, 2014). Generally speaking, both ERA and HHRA consist of the following steps: i) problem formulation ii) hazard identification, iii) hazard characterisation, iv) exposure assessment, and v) risk characterisation (WHO, 2009; EFSA, 2009; USEPA, PMRA, CALDPR, 2014). While problem formulation is an iterative process that involves exchange and agreement between risk assessors and risk managers and defines the need for and the level of detail of the risk assessment (RA), risk characterisation provides conclusions about the nature and magnitude of the risks (i.e. irrespective of whether they are considered acceptable or not) (EFSA, 2009, 2013).

In current RA practice, animal tests are performed to derive dose-response relationships for individual chemicals, and safety factors are applied to no observed adverse effect levels (NOAEL) or ideally benchmark dose limits (BMDL) to account for the uncertainty and variability in sensitivity of species and individuals to the chemical (WHO, 2009). To give an example of how HHRA is performed at the EU level, particularly in the food safety area, Acceptable Daily Intake (ADI) for regulated products and Tolerable Daily Intake (TDI) for contaminants are illustrated and are both defined as "health-based guidance values" (EFSA, 2009; Dorne et al., 2013). ADIs and TDIs represent the

maximum amount of a chemical that can be ingested daily and over a lifetime with no appreciable risk for consumers (WHO, 1987). The derivation of ADIs and TDIs involves the application of the default uncertainty factor (UF) of a 100-fold to the reference point or point of departure (i.e. NOAEL or the BMDL) from sub-chronic/chronic toxicity studies in test species (i.e. rat, mouse, dog or rabbit). This default UF of a 100-fold takes into account interspecies differences (10-fold) and human variability (10-fold). This approach has been refined to take into account toxicokinetic (TK) and toxicodynamic (TD) differences for both interspecies differences and human variability and derive chemical-specific adjustment factors (CSAFs) when chemical-specific data are available (EFSA, 2012).

In the ERA of the active substances in PPPs, risk assessors generally apply a tiered approach, which includes a simple and cost-effective first tier and more complex higher tier studies under semi-field or field conditions. This approach is applied to both target species such as crop pests and non-target species such as honey bees (EFSA, 2013). The first-tier triggers are derived by comparing the hazard quotient (HQ) or exposure toxicity ratio (ETR) with a threshold trigger value (set by risk assessors according to the available scientific evidence). The HQ or ETR is the ratio between the Predicted Environmental Concentration (PEC, i.e. the concentration of a chemical in the environment calculated using modelling or third-party data available in databases) and a standard index of the PPP's toxicity to bees (e.g. the LD_{50}) (EFSA, 2013). For example, risk characterisation of PPP active substances for honey bees following oral exposure involves the comparison of an estimated HQ or ETR value and conclusions on acceptable risk would require that such comparative ratios are lower than the proposed trigger value (i.e. HQ or ETR < 0.2 for oral exposure) (EFSA, 2013). For instance, if a PPP active substance of concern presents a PEC = 50 and LD_{50} = 100, the resulting HQ is 0.5. Consequently, since the HQ exceeds the suggested trigger value (0.2), risk assessors would conclude that the substance may pose a potential risk to honey bees.

Opportunities for refining traditional chemical risk assessment approaches

Traditionally, scientific advisory bodies such as EFSA, ECHA and the US EPA have used classical test species to derive safe exposure levels for HHRA and ERA of chemicals (Buonsante et al., 2014; Balls et al., 2018). This involves a suite of costly and animal-intensive toxicity tests which need to be performed for pre-market assessment of regulated chemicals. This approach is often hazard-based and sometimes may disregard the actual expected exposure levels (Barlow et al., 2015; Gwinn et al., 2017). For most chemicals such active substances in PPP or biocidal product, *in vivo* studies are performed using laboratory animals and are conducted in acute, sub-chronic and chronic exposure settings at varying dose levels for key exposure routes (i.e. oral, dermal or inhalation). This allows the differentiation between toxic and non-toxic doses in order to derive reference points or points

of departure such as NOAELs, BMDLs in HRRA or no-observed effect concentration (NOEC) in ERA. Default UFs are then applied to these reference points to estimate the safe dose or concentration for the species of relevance to the assessment (WHO, 2009; UNEP, IPSC, 1999). These toxicity tests are considered black-box approaches to RA, as the molecular mechanisms by which chemicals cause adverse health effects and the extent to which effects and dose levels are relevant to other species and individuals remains often, to date, poorly characterised or unknown, hence the need for default UFs (Leist et al., 2008). Concerns on their relevance and ethics were raised such as the number of animals, cost and time required for testing (Rovida and Hartung, 2009). In 2017, 9.39 million animals were used for scientific purposes, 69 % of which were employed in research, while 23 % were used for regulatory purposes; this means that roughly 2.18 million animals were used in EU laboratories to meet legislative requirements in order to ensure chemical safety for human health and/or the environment (EC, 2020a; ECHA, 2020).

Another important issue related to such traditional chemical RA approaches is their reliance on data from toxicity tests with single chemicals. In the EU, EFSA's remit requires RA of chemical and biological hazards, plant and animal pests associated with food and feed in a range of scientific areas such as animal and plant health, chemicals and nutrition as well as communication of the outcome of such risk assessments. Within its remit, EFSA performs ERA and HHRA of regulated products such as active substances used in PPPs before they reach the EU market. Traditional RA of PPPs active substances is mainly based on the analysis of the effects of a single chemical tested on non-target species (through *in vivo* laboratory tests), together with an in-depth evaluation of exposure scenarios of that substance in the environment or in humans (EFSA, 2009). However, there is a general recognition that the assessment of chemicals on an individual basis does not reflect conditions in the environment or in humans, where exposure is typically to various chemicals either simultaneously or over a period of time (Kienzler et al., 2014). The "real world" is extremely complex, as multiple chemicals and stressors (e.g. pathogens, climate change, and habitat destructions/loss) may affect human and animal health and the environment (Topping et al., 2020; EEA, 2018). Hence, a crucial challenge is to assure that scientific research and RA provide robust conclusions about the possible relation to real world conditions under which organisms are exposed, as well as the potential adverse effects multiple chemicals and stressors may pose.

Consequently, across the globe, emphasis has been shifting from the historical and traditional single-chemical RA approach to consideration of risk scenarios that integrate multiple sources, stressors, pathways and effects on a community-relevant scale, to move towards more realistic and protective approaches (NEJAC, 2004; Hynes and Lopez, 2007; NRC, 2009). However, due to the huge amount of chemicals and their potential (infinite) combinations occurring in the "real world", it is virtually impossible to test experimentally all possible mixtures in species of ecological relevance

such as bees, birds, fish and amphibians (OECD, 2018; More et al., 2019). Nonetheless, it is expected that global chemical production will double by 2030, thus requiring huge costs and time for performing the required *in vivo* tests for new chemicals and their mixtures (UNEP, 2019). Concurrently, regulators and industry seek safer alternatives to hazardous chemicals (so-called substitution), for which sound scientific methodologies and innovative testing capacity are required (Bopp et al., 2018). Hence, there is an urgent need to develop and apply smart strategies such as New Approach Methodologies (NAMs), which rely on alternative methods to animal testing, including "animal-free methods" such as *in silico* tools. In this context, it is foreseen that these methods and tools will enable the scientific community to perform chemical RA in a more efficient, sustainable and green manner (Bopp et al., 2019).

NAMs respond to the needs stipulated in the European Chemicals Strategy for Sustainability as part of the European Green Deal (EC, 2020b). The document points at the double-edged sword of chemical manufacturing. Chemicals play a major role in modern society, contributing to our general well-being and increasing life expectancies. Chemicals are produced to protect human health (e.g. pharmaceuticals, disinfectants, repellents), as well as to guarantee animal welfare (e.g. veterinary medicines, biocidal products) and plant health (e.g. PPPs). Despite their benefits, nowadays, 84% of Europeans have concerns regarding the impact of chemicals present in everyday products on their health, while 90% are concerned about the impact of chemicals on the environment (Hartmann and Klaschka, 2017).

The European Chemicals Strategy specifically stresses the need to account for the cocktail effect of chemicals when assessing risks from chemicals posed to humans and the environment. In this context, EFSA has recently published the MIXTOX guidance document to support the use of harmonised methodologies for human and ecological RA of combined exposure to multiple chemicals (More et al., 2019). In addition, this guidance document also provides examples illustrating the assessment of combined toxicity to multiple chemicals on human health, farm animals and honey bees (More et al., 2019). The framework is based on the RA steps (problem formulation, exposure assessment, hazard identification and characterisation, and risk characterisation including uncertainty analysis), with tiered and stepwise approaches for both whole mixture and component-based approaches. If a chemical mixture is poorly defined, a whole mixture approach is usually applied, thus the mixture is essentially evaluated in the same way as for a single chemical substance (More et al., 2019; Bopp et al., 2018). In contrast, if components in the mixture are chemically defined, the component-based approach is generally applied and the risk is assessed based on exposure and effect data of its individual components. Key recommendations include the need to further develop and implement generic *in silico* models such as quantitative structure–activity relationship (QSAR) in support of component-based approaches to predict combined toxicity for a

range of species including a broader range of species of ecological relevance such as honey bees. It is foreseen that such models will support the integration of toxicity and mechanistic data for hazard assessment of single chemicals as well as for component-based approaches for mixture risk assessment (MRA).

Combined exposure to multiple chemicals and bees: "cocktail effects"

Pollinators such as honey bees (*Apis mellifera*), bumble bees (*Bombus* spp.) and solitary bees (e.g. *Osmia* spp.) play a key-role as ecosystems service providers (ESP) contributing to the maintenance, reproduction of wild plant communities and biodiversity as well as bridging agriculture, the food chain and the ecological communities, thereby ensuring food production and security (Breeze et al. 2011; Schulp et al. 2014; Rose et al., 2015). In terms of economy, pollination services from managed and wild bees contribute to at least 22 billion Euros of the European agriculture sector on a yearly basis (EC, 2016). In a similar fashion, honey bees and their hive products (i.e. honey, pollen, beebread) also represent sentinel species and bioindicators respectively, to monitor environmental contamination from 1. regulated products (e.g. PPPs and veterinary residues), 2. anthropogenic chemicals (e.g. persistent organic pollutants, heavy metals, particulate matter) and, 3. natural contaminants (mycotoxins, plant alkaloids) (Negri et al., 2015; Bargańska et al., 2016; Tosi et al., 2018). However, concerns have arisen in recent years due to the potential harmful effects of PPPs including insecticides and their "cocktail effects" on ecosystems, particularly towards non-target species such as bees (Douglas et al., 2020; EFSA, 2018; Tosi and Nieh, 2019; Sanchez-Bayo and Goka, 2016; Simon-Delso et al., 2015; Tosi et al., 2018).

Over the last decade, significant losses of honey bee colonies have been reported, particularly in North America and Western Europe (Jacques et al., 2016; Sanchez-Bayo and Goka, 2016; Steinhauer et al., 2014; Van der Zee et al., 2012). Scientific evidence shows that the weakening or death of bee colonies is mainly caused by the combined effects of multiple stressors rather than by one-off sudden attacks by a single factor (Goulson et al., 2015; EFSA, 2014a; Potts et al., 2010; Rortais et al., 2017). Such interactions can occur principally between (i) biological factors (Nazzi et al., 2012; Nazzi and Pennacchio, 2014), (ii) environmental factors (Di Pasquale et al., 2016; Goulson et al., 2015; Le Conte and Navajas, 2008), (iii) chemical and nutritional stressors (Tosi et al., 2017; Tong et al., 2019), (iv) chemical and biological factors (Williamson et al., 2013; Klein et al., 2017; Alaux et al., 2010; Vidau et al., 2011; Pettis et al., 2012; Renzi et al., 2016) and (v) multiple chemicals (EFSA, 2013b; Robinson et al., 2017; Han et al., 2019; Sanchez-Bayo and Goka, 2016). In particular, the latter is raising concerns among scientists and regulatory bodies since bees can be exposed to a wide range of multiple chemicals, "chemical mixtures", including compounds from

anthropogenic (e.g. PPPs or veterinary drugs) or natural origin (e.g. mycotoxins, flavonoids, plant toxins) (Johnson, 2015; Tosi and Nieh, 2019; EFSA PPR Panel, 2012; EFSA, 2014a).

The RA of PPP active substance in pollinators is mainly focused on the analysis of toxicity data resulting from *in vivo* laboratory tests on single chemicals and in non-target species, where the honey bee (*Apis mellifera*) is employed worldwide as surrogate species for *Apis* and non-*Apis* bees (EFSA, 2013; USEPA, PMRA, CALDPR, 2014). However, due to the great variety of chemicals and their combinations occurring in the environment, it is unfeasible to experimentally test all possible mixtures in species of ecological relevance such as bees. Hence, it is necessary to rely more on the "*a priori*" knowledge of chemical structure in order to better define toxicological mode of actions (MoAs) and target site of PPPs active substances in honey bees, thus enabling risk assessors and scientists to refine chemical grouping and toxicity extrapolations for component-based approaches for MRA (More et al., 2019; Carnesecchi et al., 2019, 2020a).

Towards New Approach Methodologies in chemical risk assessment

The introduction of the "3Rs" principle, *reduction, refinement and replacement* has played a key role in the development of alternative methods to animal testing (Russell and Burch, 1959). Since then, animal-free methods have been developed and validated with the main scope of reducing and possibly replacing *in vivo* testing for toxicological hazard assessment and ultimate for RA of chemicals (Fabian et al., 2019). The so-called Tox21 strategy is shifting the toxicological assessments away from traditional animal studies to target-specific, mechanism-based and biological observations mainly obtained using NAMs (Thomas et al., 2018; ECHA, 2016). In fact, NAMs include *in silico* approaches, *in chemico* and *in vitro* assays, as well as the inclusion of information from the exposure of chemicals in the context of hazard assessment (Raies et al., 2016). In a similar fashion, NAMs aim at embedding a variety of new testing tools, such as "high-throughput screening" and "high-content methods" e.g. genomics, proteomics, metabolomics; as well as some "conventional" methods that aim to improve understanding of toxic effects, either through improving toxicokinetic (TK) or toxicodynamic (TD) knowledge for substances (ECHA, 2016). Consequently, NAMs aim to facilitate the replacement of animal testing with combinations of predictive *in silico* models (e.g. of human exposure structure-related toxicities), *in vitro* assays (e.g. human cell-based systems), and computational models of external and internal exposure (e.g. *in vitro* to *in vivo* extrapolation (IVIVE), physiologically based pharmacokinetic (PBPK) modelling) (Thomas et al., 2019).

In the field of chemical RA, *in silico* models have been recognised as valuable methods in order to simulate biological processes for a range of applications, species and level of biological organisation (e.g. cellular, molecular, species, population, ecosystem and landscape) (ECHA, 2017; EFSA, 2014b). In particular, QSARs are mathematical models employed to predict the physicochemical (e.g. lipophilicity, flammability), biological (e.g. toxicity, clearance) and

environmental fate properties of compounds from the knowledge of their chemical structure (OECD, 2014). The term "quantitative" indicates that the molecular descriptors are quantifiable on a continuous scale, thus allowing a quantitative relationship with substance toxicity (Benfenati et al., 2017). The predictive toxicity is usually derived from the molecular descriptors of chemicals, which include their inherent physicochemical properties such as atomic composition, structure, sub-structures, hydrophobicity, surface area charge, and molecular volume (OECD, 2014). QSAR models can be divided into classification-based, when a relationship between the descriptors and the categorical values of the response variable(s) is established, or regression-based, whenever the aim is to find a relationship between the descriptors and the quantitative values of the response variable(s) (Ambure et al., 2019). In practice, while the first ones allow discriminating chemicals in categories (e.g. toxic vs non-toxic, active vs non-active) according to a threshold set "a priori", the second ones provide users with a quantitative predictions of a given endpoint (e.g. LD_{50} value) for the chemical of concern. Similarly, (Q)SARs can also be used directly or stepwise to predict the MoA of chemicals (Carnesecchi et al., 2020a; Verhaar et al., 1992; Russom et al., 1997; Kienzler et al., 2019). In the human health and animal health areas, MoA refers to the major steps leading to an adverse health effect following interaction of the chemical with biological targets at the sub-cellular level, whilst not necessarily implying the full understanding of the mechanism of action at the molecular level (WHO, 2009; Boobis et al. 2006; OECD, 2017; EFSA PPR, 2013). Similarly, in ecological RA, MoA has been defined as a functional change at the cellular level triggered by the substance entering the organism, which then involves levels of biological organisation from organisms, multiple species, to populations all the way to ecosystems (Kienzler et al., 2019; Segner, 2011).

In a regulatory context, particularly under the REACH legislation (European Parliament, European Council, 2006), registrants must prove that they have considered alternative methods such as QSARs in their testing proposals before carrying out any animal experiments. Although an increased number of *in silico* tools were used in REACH dossier submissions in 2017 compared to 2010 (ECHA, 2017), QSARs are still mainly employed to provide supporting evidence for read-across strategies or to add credence to experimental results of unknown or limited validity (e.g. missing Good Laboratory Practices compliance) in hazard assessment (Chinen et al., 2020; Thomas et al., 2019). Consequently, QSARs are typically used in combination with other non-testing and testing (e.g. *in vitro*) methods in the context of integrated testing strategies (IATA) and Weight-of-Evidence assessments (OECD, 2020; Hardy et al., 2017). Nonetheless, the majority of QSAR models available to date are designed for predicting (missing) information on single chemicals (e.g. VEGA-HUB *in silico* platform), thus not accounting for chemical mixtures toxicity and limiting their use to one-by-one compound assessment (Khan et al., 2020; Carnesecchi et al., 2020a,b; Benfenati et al., 2017).

Hence, further demonstration of how *in silico* models can play a key role in the RA of emerging chemicals and their mixtures is of crucial importance to move the field forward.

Objective of the thesis

In the light of the principles described above and in line with the European Green Deal (EC, 2020b), the main objective of this thesis is to develop and apply alternative methods to animal testing such as QSAR models for human and ecological RA of single and multiple chemicals. Five different case studies are presented to illustrate how NAMs such as QSARs can provide the means to move towards a more mechanistic understanding of toxicity, thereby paving the way for replacement of traditional *in vivo* experiments with non-animal alternatives for regulated chemicals, emerging contaminants and their mixtures.

- Chapter 2 presents the evaluation of the ability and applicability of freely available QSAR models (VEGA-HUB platform) to estimate human health-related properties (i.e. genotoxicity and carcinogenicity) for screening purposes using three different EFSA databases.

- Chapter 3 provides the first comprehensive meta-analysis on combined toxicity of binary mixtures in bees, highlighting the mechanistic basis and implications for RA, and thereby providing an open-source database to develop QSARs models to predict PPP mixture toxicity.

- Chapter 4 reports the development of classification-based QSAR model to predict acute contact toxicity of single PPP active substances in honey bees (*Apis mellifera*) for regulatory purposes.

- Chapter 5 presents three innovative QSAR models using CORAL software to predict the acute contact toxicity of PPP mixtures ($LD_{50\text{-}mix}$) in honey bee, and the nature of combined toxicity (synergism / non-synergism) within a weight of evidence approach (WoE).

- Chapter 6 aims to address the challenge of integrating MoA information in computational models with the first integrative honey bee QSAR models for PPPs using open source databases i) to predict acute contact toxicity (LD_{50}) and ii) to profile the MoA of active substances. In addition, the current study explores the development of harmonised MoA classification schemes to relate the structure toxicological information with the target sites of PPPs active substances for a range of applications, including toxicity predictions and refining the grouping of chemicals for component-based RA of multiple chemicals.

- Chapter 7 discusses how the findings presented in the previous chapters may contribute to realising the European Chemicals Strategy for Sustainability (as part of the European Green

Deal) specifically, and toxicity testing in the twenty-first century in general (NRC, 2007; Thomas et al., 2018).

References

Ambure, P., Halder, A.K., González Díaz, H., Cordeiro, M.N.D.S., 2019. QSAR-Co: An Open Source Software for Developing Robust Multitasking or Multitarget Classification-Based QSAR Models. J. Chem. Inf. Model. https://doi.org/10.1021/acs.jcim.9b00295

Alaux, C., Brunet, J.L., Dussaubat, C., Mondet, F., Tchamitc Han, S., Cousin, M., Brillard, J., Baldy, A., Belzunces, L.P., Le Conte, Y., 2010. Interactions between Nosema microspores and a neonicotinoid weaken honeybees (Apis mellifera). Environ. Microbiol. l12 (3), 774-782.

Balls, M., Combes, R., Worth, A., 2018. The History of Alternative Test Methods in Toxicology, The History of Alternative Test Methods in Toxicology. https://doi.org/10.1016/c2016-0-00193-6

Bargańska, Ż., Ślebioda, M., Namieśnik, J., 2016. Honey bees and their products: Bioindicators of environmental contamination. Crit. Rev. Environ. Sci. Technol. 46, 235–248. https://doi.org/10.1080/10643389.2015.1078220

Barlow, S.M., Boobis, A.R., Bridges, J., Cockburn, A., Dekant, W., Hepburn, P., Houben, G.F., König, J., Nauta, M.J., Schuermans, J., Bánáti, D., 2015. The role of hazard- and risk-based approaches in ensuring food safety. Trends Food Sci. Technol. https://doi.org/10.1016/j.tifs.2015.10.007

Belanger, S.E., Sanderson, H., Embry, M.R., Coady, K., Dezwart, D., Farr, B.A., Gutsell, S., Halder, M., Sternberg, R., Wilson, P., 2015. It is time to develop ecological thresholds of toxicological concern to assist environmental hazard assessment. Environ. Toxicol. Chem. https://doi.org/10.1002/etc.3132

Benfenati E, Como F, Marzo M, Gadaleta D, Toropov A and Toropova A, 2017. Developing innovative in silico models with EFSA's OpenFoodTox database. EFSA supporting publication 2017:EN-1206. 19 pp. doi:10.2903/sp.efsa.2017.EN-1206

Bopp, S.K., Barouki, R., Brack, W., Dalla Costa, S., Dorne, J.L.C.M., Drakvik, P.E., Faust, M., Karjalainen, T.K., Kephalopoulos, S., van Klaveren, J., Kolossa-Gehring, M., Kortenkamp, A., Lebret, E., Lettieri, T., Nørager, S., Rüegg, J., Tarazona, J. V., Trier, X., van de Water, B., van Gils, J., Bergman, Å., 2018. Current EU research activities on combined exposure to multiple chemicals. Environ. Int. https://doi.org/10.1016/j.envint.2018.07.037

Bopp, S.K., Kienzler, A., Richarz, A.N., van der Linden, S.C., Paini, A., Parissis, N., Worth, A.P., 2019. Regulatory assessment and risk management of chemical mixtures: challenges and ways forward. Crit. Rev. Toxicol. https://doi.org/10.1080/10408444.2019.1579169

Breeze, T.D., Bailey, A.P., Balcombe, K.G., Potts, S.G., 2011. Pollination services in the UK: How important are honeybees? Agric. Ecosyst. Environ. https://doi.org/10.1016/j.agee.2011.03.020

Buonsante, V.A., Muilerman, H., Santos, T., Robinson, C., Tweedale, A.C., 2014. Risk assessment's insensitive toxicity testing may cause it to fail. Environ. Res. https://doi.org/10.1016/j.envres.2014.07.016

Carnesecchi, E., Svendsen, C., Lasagni, S., Grech, A., Quignot, N., Amzal, B., Toma, C., Tosi, S., Rortais, A., Cortinas-Abrahantes, J., Capri, E., Kramer, N., Benfenati, E., Spurgeon, D., Guillot, G., Dorne, J.L.C.M., 2019. Investigating combined toxicity of binary mixtures in bees: Meta-analysis of laboratory tests, modelling, mechanistic basis and implications for risk assessment. Environ. Int. 133. https://doi.org/10.1016/j.envint.2019.105256.

Carnesecchi, E., Toma, C., Roncaglioni, A., Kramer, N., Benfenati, E., Dorne, J.L.C.M., 2020a. Integrating QSAR models predicting acute contact toxicity and mode of action profiling in honey bees (A. mellifera): Data curation using open source databases, performance testing and validation. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.139243

Carnesecchi, E., Toropov, A.A., Toropova, A.P., Kramer, N., Svendsen, C., Dorne, J. Lou, Benfenati, E., 2020b. Predicting acute contact toxicity of organic binary mixtures in honey bees (A. mellifera) through innovative QSAR models. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2019.135302

Chinen, K., Malloy, T., 2020. QSAR Use in REACH Analyses of Alternatives to Predict Human Health and Environmental Toxicity of Alternative Chemical Substances. Integr. Environ. Assess. Manag. https://doi.org/10.1002/ieam.4264

Di Pasquale, G., Alaux, C., Le Conte, Y., Odoux, J.-F., Pioz, M., Vaissière, B.E., Belzunces, L.P., Decourtye, A., 2016. Variations in the availability of pollen resources affect honey bee health. PLoS One 11, e0162818. https://doi.org/10.1371/journal.pone.0162818.

Dorne, J.L.C.M., Fink-Gremmels, J., 2013. Human and animal health risk assessments of chemicals in the food chain: Comparative aspects and future perspectives. Toxicol. Appl. Pharmacol. https://doi.org/10.1016/j.taap.2012.03.013

Douglas, M.R., Sponsler, D.B., Lonsdorf, E. V., Grozinger, C.M., 2020. County-level analysis reveals a rapidly shifting landscape of insecticide hazard to honey bees (Apis mellifera) on US farmland. Sci. Rep. https://doi.org/10.1038/s41598-019-57225-w

ECHA (European Chemicals Agency), 2020. The use of alternatives to testing on animals for the REACH Regulation. Fourth report under Article 117(3) of the REACH Regulation. ECHA-20-R-08-EN. Available at https://echa.europa.eu/documents/10162/0/alternatives_test_animals_2020_en.pdf/db66b8a3-00af-6856-ef96-5ccc5ae11026

ECHA (European Chemicals Agency), 2016. New Approach Methodologies in Regulatory Science. Proceedings of a scientific workshop Helsinki, 19–20 April 2016. Available at https://echa.europa.eu/documents/10162/22816069/scientific_ws_proceedings_en.pdf

ECHA (European Chemicals Agency), 2017. The use of alternatives to testing on animals for the REACH regulation. ECHA Article 117(3) of the REACH Regulation Report No.: ECHA-17-R-02-EN. Helsinki (FI): ECHA. Available at https://echa.europa.eu/documents/10162/13639/alternatives_test_animals_2017_en.pdf

EEA (European Environment Agency), 2018. Chemicals for a sustainable future: report of the EEA Scientific Committee Seminar. Copenhagen, Denmark. Available at https://www.eea.europa.eu/about-us/governance/scientific-committee/reports/chemicals-for-a-sustainable-future

EFSA (European Food Safety Authority) Scientific Committee, 2016. Scientific opinion on recovery in environmental risk assessments at EFSA. EFSA Journal 2016; 14(2):4313. 85 pp. doi:10.2903/j.efsa.2016.4313

EFSA (European Food Safety Authority), 2009. Guidance of the Scientific Committee on Transparency in the Scientific Aspects of Risk Assessments carried out by EFSA. Part 2: General Principles, 2009. . EFSA J. https://doi.org/10.2903/j.efsa.2009.1051

EFSA (European Food Safety Authority), 2013. International Framework Dealing with Human Risk Assessment of Combined Exposure to Multiple Chemicals. EFSA Journal 2013;11(7):3313. [69 pp.] doi:10.2903/j.efsa.2013.3313

EFSA (European Food Safety Authority), 2018. Evaluation of the data on clothianidin, imidacloprid and thiamethoxam for the updated risk assessment to bees for seed treatments and granules in the EU. EFSA supporting publication 2018:EN-1378. 31pp.doi:10.2903/sp.efsa.2018.EN-1378

EFSA (European Food Safety Authority), 2014a. Towards an integrated environmental risk assessment of multiple stressors on bees: review of research projects in Europe, knowledge gaps and recommendations. EFSA J. 12. https://doi.org/10.2903/j.efsa.2014.3594.

EFSA (European Food Safety Authority), 2014b. Modern methodologies and tools for human hazard assessment of chemicals. EFSA Journal 2014;12(4):3638, 87 pp. doi:10.2903/j.efsa.2014.3638.

EFSA (European Food Safety Authority) Panel on Plant Protection Products and their Residues (PPR), 2012. Scientific Opinion on the science behind the development of a risk assessment of Plant Protection Products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA J.10, 2668. https://doi.org/10.2903/j.efsa.2012.2668.

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Scientific Opinion on the identification of pesticides to be included in cumulative assessment groups on the basis of their toxicological profile (2014 update). EFSA Journal 2013;11(7):3293, 131 pp. https://doi.org/10.2903/j.efsa.2013.3293

EFSA (European Food Safety Authority), 2012. Guidance for establishing the safety of additives for the consumer. EFSA J. 10, 2537. https://doi.org/10.2903/j.efsa.2012.2537

Escher, B.I., Stapleton, H.M., Schymanski, E.L., 2020. Tracking complex mixtures of chemicals in our changing environment. Science (80-.). https://doi.org/10.1126/science.aay6636

European Commission (EC), 2020a. Report from the Commission to the European Parliament and the Council: 2019 report on the statistics on the use of animals for scientific purposes in the Member States of the European Union in 2015-2017. COM/2020/16 final. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0016&from=EN

European Commission (EC), 2020b. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: The European Green Deal. COM/2019/640 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2019%3A640%3AFIN

European Commission (EC), 2016. Bee health – EU Actions. Available at https://ec.europa.eu/food/animals/live_animals/bees_en (accessed 27 February 2019).

European Parliament, European Union Council, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and

Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Luxembourg (LU). Available at http://eur-lex.europa.eu/legal-content/EN/TXT/?uri¼CELEX%3A32006R1907

Fabian, E., Gomes, C., Birk, B., Williford, T., Hernandez, T.R., Haase, C., Zbranek, R., van Ravenzwaay, B., Landsiedel, R., 2019. In vitro-to-in vivo extrapolation (IVIVE) by PBTK modeling for animal-free risk assessment approaches of potential endocrine-disrupting compounds. Arch. Toxicol. https://doi.org/10.1007/s00204-018-2372-z

Frank, P., Ottoboni, A., 2011. The Dose Makes the Poison: A Plain-Language Guide to Toxicology: Third Edition, The Dose Makes the Poison: A Plain-Language Guide to Toxicology: Third Edition. https://doi.org/10.1002/9780470918449

Gwinn, M.R., Axelrad, D.A., Bahadori, T., Bussard, D., Cascio, W.E., Deener, K., Dix, D., Thomas, R.S., Kavlock, R.J., Burke, T.A., 2017. Chemical risk assessment: Traditional vs public health perspectives. Am. J. Public Health. https://doi.org/10.2105/AJPH.2017.303771

Goulson, D., Nicholls, E., Botias, C., Rotheray, E.L., 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science 347, 1255957. https://doi.org/10.1126/science.1255957.

Han, W., Yang, Y., Gao, J., Zhao, D., Ren, C., Wang, S., Zhao, S., Zhong, Y., 2019. Chronic toxicity and biochemical response of Apis cerana cerana (Hymenoptera: Apidae) exposed to acetamiprid and propiconazole alone or combined. Ecotoxicology 28, 399–411. https://doi.org/10.1007/s10646-019-02030-4.

Hardy, A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, H.K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J.R., Silano, V., Solecki, R., Turck, D., Benfenati, E., Chaudhry, Q.M., Craig, P., Frampton, G., Greiner, M., Hart, A., Hogstrand, C., Lambre, C., Luttik, R., Makowski, D., Siani, A., Wahlstroem, H., Aguilera, J., Dorne, J., Fernandez Dumont, A., Hempen, M., Valtueña Martínez, S., Martino, L., Smeraldi, C., Terron, A., Georgiadis, N., Younes, M., 2017. Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J. https://doi.org/10.2903/j.efsa.2017.4971

Hartmann, S., Klaschka, U., 2017. Interested consumers' awareness of harmful chemicals in everyday products. Environ. Sci. Eur. https://doi.org/10.1186/s12302-017-0127-8

Hynes, H., Lopez, R., (2007). Boston University School of Public Health. Cumulative Risk and a Call for Action in Environmental Justice Communities. Journal of Health Disparities Research and Practice 1(2):29-57.

Jacques, A., Laurent, M., Ribiere-Chabert, M., Saussac, M., Bougeard, S., Hendrikx, P., Chauzat, M., 2016. Statistical analysis on the EPILOBEE dataset: explanatory variables related to honeybee colony mortality in EU during a 2 year survey. EFSA Support. Publ. 13. https://doi.org/10.2903/sp.efsa.2016.EN-883.

Johnson, R.M., 2015. Honey bee toxicology. Annu. Rev. Entomol. 60, 415–434. https://doi.org/10.1146/annurev-ento-011613-162005.

Khan, P.M., Kar, S., Roy, K., 2020. Ecotoxicological QSARs of mixtures, in: Methods in Pharmacology and Toxicology. https://doi.org/10.1007/978-1-0716-0150-1_19

Kienzler, A., Berggren, E., Bessems, J., Bopp, S., Van Der Linden, S., Worth, A., 2014. Assessment of mixtures - review of regulatory requirements and guidance. JRC Science and Policy Report. European Commission, Joint Research Center, Ispra, Italy. doi:10.2788/84264

Kienzler, A., Connors, K.A., Bonnell, M., Barron, M.G., Beasley, A., Inglis, C.G., Norberg-King, T.J., Martin, T., Sanderson, H., Vallotton, N., Wilson, P., Embry, M.R., 2019. Mode of Action Classifications in the EnviroTox Database: Development and Implementation of a Consensus MOA Classification. Environ. Toxicol. Chem. https://doi.org/10.1002/etc.4531

Klein, A.M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., et al., 2007. Importance of pollinators in changing landscapes for world crops. Proc. R. Soc. Lond B 274, 303-313.

Le Conte, Y., Navajas, M., 2008. Climate change: impact on honey bee populations and diseases. Rev. Sci. Tech. l'OIE 27, 485–510. https://doi.org/10.20506/rst.27.2.1819.

Leist, M., Hartung, T., Nicotera, P., 2008. The dawning of a new age of toxicology. ALTEX. https://doi.org/10.14573/altex.2008.2.103

More, S.J., Bampidis, V., Benford, D., Bennekou, S.H., Bragard, C., Halldorsson, T.I., Hernández-Jerez, A.F., Koutsoumanis, K., Naegeli, H., Schlatter, J.R., Silano, V., Nielsen, S.S., Schrenk, D., Turck, D., Younes, M., Benfenati, E., Castle, L., Cedergreen, N., Hardy, A., Laskowski, R., Leblanc, J.C., Kortenkamp, A., Ragas, A., Posthuma, L., Svendsen, C., Solecki, R., Testai, E., Dujardin, B., Kass, G.E., Manini, P., Jeddi, M.Z., Dorne, J.C., Hogstrand, C., 2019. Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals. EFSA J. 17. https://doi.org/10.2903/j.efsa.2019.5634.

Nazzi, F., Brown, S.P., Annoscia, D., Del Piccolo, F., Di Prisco, G., Varricchio, P., Della Vedova, G., Cattonaro, F., Caprio, E., Pennacchio, F., 2012. Synergistic parasite-pathogen interactions mediated by host immunity can drive the collapse of honeybee colonies. PLoS Pathog. 8, e1002735. https://doi.org/10.1371/journal.ppat.1002735.

Nazzi, F., Pennacchio, F., 2014. Disentangling multiple interactions in the hive ecosystem. Trends Parasitol. 30, 556–561. https://doi.org/10.1016/j.pt.2014.09.006.

Negri, I., Mavris, C., Di Prisco, G., Caprio, E., Pellecchia, M., 2015. Honey bees (Apis mellifera, L.) as active samplers of airborne particulate matter. PLoS One. https://doi.org/10.1371/journal.pone.0132491

NEJAC (National Environmental Justice Advisory Council), 2004. Ensuring Risk Reduction in Communities with Multiple Stressors: Environmental Justice and Cumulative Risks/Impacts (Cumulative Risks/Impacts Workgroup Report). https://www.epa.gov/sites/production/files/2015-04/documents/ensuringriskreducationnejac.pdf.

NRC (National Research Council) (2009), Science and Decisions: Advancing Risk Assessment. Washington, DC: The National Academies Press. https://www.nap.edu/catalog/12209/science-anddecisions-advancing-risk-assessment

OECD (Organisation for Economic Co-operation and Development), 2018. Considerations for Assessing the Risks of Combined Exposure to Multiple Chemicals, Series on Testing and

Assessment No. 296, Environment, Health and Safety Division, Environment Directorate. Available at: https://www.oecd.org/chemicalsafety/risk-assessment/considerations-for-assessing-the-risks-of-combined-exposure-to-multiple-chemicals.pdf

OECD (Organisation for Economic Co-operation and Development), 2014. Guidance on grouping of chemicals, second edition. Series on Testing and Assessment, No. 194, ENV/JM/MONO(2014)4, 14-Apr-2014.

OECD (Organisation for Economic Co-operation and Development), 2020. Overview of Concepts and Available Guidance related to Integrated Approaches to Testing and Assessment (IATA), OECD Series on Testing and Assessment, No. 329, Environment, Health and Safety, Environment Directorate, OECD. Available at: http://www.oecd.org/chemicalsafety/risk-assessment/concepts-and-available-guidance-related-to-integrated-approaches-to-testing-and-assessment.pdf

Pettis, J.S., VanEngelsdorp, D., Johnson, J., Dively, G., 2012. Pesticide exposure in honey bees results in increased levels of the gut pathogen Nosema. Naturwissenschaften 99, 153–158. https://doi.org/10.1007/s00114-011-0881-1.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E., 2010. Global pollinator declines: trends, impacts and drivers. Trends Ecol. Evol. 25, 345–353. https://doi.org/10.1016/j.tree.2010.01.007.

Raies, A.B., Bajic, V.B., 2016. In silico toxicology: computational methods for the prediction of chemical toxicity. Wiley Interdiscip. Rev. Comput. Mol. Sci. https://doi.org/10.1002/wcms.1240

Renzi, M.T., Amichot, M., Pauron, D., Tchamitchian, S., Brunet, J.L., Kretzschmar, A., Maini, S., Belzunces, L.P., 2016. Chronic toxicity and physiological changes induced in the honey bee by the exposure to fipronil and Bacillus thuringiensis spores alone or combined. Ecotoxicol. Environ. Saf. 127, 205–213. https://doi.org/10.1016/j.ecoenv.2016.01.028Rortais, A., Arnold, G., Dorne, J.-L., More, S.J., Sperandio, G., Streissl, F., Szentes, C., Verdonck, F., 2017. Risk assessment of pesticides and other stressors in bees: principles, data gaps and perspectives from the European Food Safety Authority. Sci. Total Environ. 587–588, 524–537. https://doi.org/10.1016/j.scitotenv.2016.09.127.

Robinson, A., Hesketh, H., Lahive, E., Horton, A.A., Svendsen, C., Rortais, A., Dorne, J. Lou, Baas, J., Heard, M.S., Spurgeon, D.J., 2017. Comparing bee species responses to chemical mixtures: Common response patterns? PLoS One 12, 1–21. https://doi.org/10.1371/journal.pone.0176289.

Rose, T., Kremen, C., Thrupp, A., Gemmill-Herren, B., Graub, B., Azzu, N., Antunes, V., Bruteig, I.E., Buchori, D., Donaldson, J., Dhyani, P.P., Garibaldi, L., Getz, A., Goss, M., Iqbal, J., Kasina, M., Kinuthia, W., Kwapong, P., Manetto, S., Martins, D., Nyamasio, G., Nyamongo, D.O., Odhiambo, C., América Suarez De Oliveira, D., Owusu, E.H., Pandey, B., Poole, C., Roubik, D.W., Roy, P., Waghchoure, E., Lusike, W., 2015. Policy analysis paper: mainstreaming of biodiversity and ecosystem services with a focus on pollination. United Nations Food and Agriculture Organization (FAO) with the contribution of participants at the "Policies for Pollination Management" Worksh. Available at: http://www.fao.org/3/a-i4242e.pdf.

Rovida, C., Hartung, T., 2009. Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals - A report by the transatlantic think tank for toxicology (t4). ALTEX. https://doi.org/10.14573/altex.2009.3.187

Russell, W., Burch, R., 1959. The Principles of Humane Experimental Technique. John Hopkins Bloom. Sch. Public Heal. https://doi.org/10.5694/j.1326-5377.1960.tb73127.x

Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A., 1997. Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (pimephales promelas). Environ. Toxicol. Chem. https://doi.org/10.1897/1551-5028(1997)016<0948:pmotaf>2.3.co;2

Sanchez-Bayo, F., Goka, K., 2016. Impacts of pesticides on honey bees. In: Beekeeping and Bee Conservation - Advances in Research. InTech. https://doi.org/10.5772/62487.

SCCP, SCHER, SCENIHR, 2009. Risk assessment methodologies and approaches for genotoxic and carcinogenic substances. Eur. Comm. Available at https://ec.europa.eu/health/ph_risk/committees/04_scher/docs/scher_o_113.pdf

Schulp, C.J.E., Lautenbach, S., Verburg, P.H., 2014. Quantifying and mapping ecosystem services: Demand and supply of pollination in the European Union. Ecol. Indic. https://doi.org/10.1016/j.ecolind.2013.07.014

Simon-Delso, N., San Martin, G., Bruneau, E., Delcourt, C., Hautier, L., 2017. The challenges of predicting pesticide exposure of honey bees at landscape level. Sci. Rep. 7, 3801. https://doi.org/10.1038/s41598-017-03467-5.

Steinhauer, N.A., Rennich, K., Wilson, M.E., Caron, D.M., Lengerich, E.J., Pettis, J.S., Rose, R., Skinner, J.A., Tarpy, D.R., Wilkes, J.T., VanEngelsdorp, D., 2014. A national survey of managed honey bee 2012–2013 annual colony losses in the USA: results from the Bee Informed Partnership. J. Apic. Res. 53, 1–18. https://doi.org/10.3896/IBRA.1.53.1.01.

Thomas, R.S., Paules, R.S., Simeonov, A., Fitzpatrick, S.C., Crofton, K.M., Casey, W.M., Mendrick, D.L., 2018. The US Federal Tox21 Program: A strategic and operational plan for continued leadership. ALTEX. https://doi.org/10.14573/altex.1803011

Thomas, P.C., Bicherel, P., Bauer, F.J., 2019. How in silico and QSAR approaches can increase confidence in environmental hazard and risk assessment. Integr. Environ. Assess. Manag. https://doi.org/10.1002/ieam.4108

Tong, L., Nieh, J.C., Tosi, S., 2019. Combined nutritional stress and a new systemic pesticide (flupyradifurone, Sivanto) reduce bee survival, food consumption, flight success, and thermoregulation. Chemosphere 237, 124408. https://doi.org/10.1016/j.chemosphere.2019.124408.

Topping, C.J., Aldrich, A., Berny, P., 2020. Overhaul environmental risk assessment for pesticides. Science (80-.). https://doi.org/10.1126/science.aay1144

Tosi, S., Costa, C., Vesco, U., Quaglia, G., Guido, G., 2018. A 3-year survey of Italian honey bee-collected pollen reveals widespread contamination by agricultural pesticides. Sci. Total Environ. 615, 208–218. https://doi.org/10.1016/j.scitotenv.2017.09.226.

Tosi, S., Nieh, J.C., 2019. Lethal and sublethal synergistic effects of a new systemic pesticide, flupyradifurone (Sivanto), on honeybees. Proc. R. Soc. B Biol. Sci. 286, 20190433. https://doi.org/10.1098/rspb.2019.0433.

Tosi, S., Nieh, J.C., Sgolastra, F., Cabbri, R., Medrzycki, P., 2017. Neonicotinoid pesticides and nutritional stress synergistically reduce survival in honey bees. Proc. R. Soc. B Biol. Sci. 284, 20171711. https://doi.org/10.1098/rspb.2017.1711.

United Nations Environment Programme (UNEP), 2019. Global Chemicals Outlook II: From Legacies to Innovative Solutions - Implementing the 2030 Agenda for Sustainable Development, The Global Chemicals Outlook. ISBN No: 978-92-807-3745-5. Available at https://www.unenvironment.org/resources/report/global-chemicals-outlook-ii-legacies-innovative-solutions

United Nations Environment Programme (UNEP), International Programme on Chemical Safety (IPCS), 1999. Chemical Risk Assessment: Human Risk Assessment, Environmental Risk Assessment, and Ecological Risk Assessment Available at https://apps.who.int/iris/bitstream/handle/10665/66398/WHO_PCS_99.2_eng.pdf?sequence=1

USEPA, PMRA, CALDPR, 2014. Guidance for Assessing Pesticide Risks to Bees. U.S. Environmental Protection Agency (USEPA), Health Canada Pest Management Regulatory Agency (PMRA) and California Departament of Pesticide Regulation, Office of Pesticide Programs USEPA, Health Canada PMRA, CDPR. Available at https://www.epa.gov/pollinator-protection/pollinator-risk-assessment-guidance

Verhaar, H.J.M., Van Leeuwen, C.J., Hermens, J.L.M., 1992. Classifying environmental-pollutants. 1: Structure-activity-relationships for prediction of aquatic toxicity. Chemosphere. https://doi.org/10.1016/0045-6535(92)90280-5

Vidau, C., Diogon, M., Aufauvre, J., Fontbonne, R., Viguès, B., Brunet, J.-L., Texier, C., Biron, D.G., Blot, N., El Alaoui, H., Belzunces, L.P., Delbac, F., 2011. Exposure to sublethal doses of fipronil and thiacloprid highly increases mortality of honeybees previously infected by Nosema ceranae. PLoS One 6, e21550. https://doi.org/10.1371/journal.pone.0021550.

Van der Zee, R., Pisa, L., Andonov, S., Brodschneider, R., Charrière, J.-D., Chlebo, R., Coffey, M.F., Crailsheim, K., Dahle, B., Gajda, A., Gray, A., Drazic, M.M., Higes, M., Kauko, L., Kence, A., Kence, M., Kezic, N., Kiprijanovska, H., Kralj, J., Kristiansen, P., Martin Hernandez, R., Mutinelli, F., Kim Nguyen, B., Otten, C., Özkırım, A., Pernal, S.F., Peterson, M., Ramsay, G., Santrac, V., Soroker, V., Topolska, G., Uzunov, A., Vejsnaes, F., Wei, S., Wilkins, S., 2012. Managed honey bee colony losses in Canada, China, Europe, Israel and Turkey. J. Apic. Res. 51, 100–114. https://doi.org/10.3896/IBRA.1.51.1.12.

Williamson, S.M., Wright, G.A., 2013. Exposure to multiple cholinergic pesticides impairs olfactory learning and memory in honeybees. J. Exp. Biol. 216, 1799–1807. https://doi.org/10.1242/jeb.083931.

World Health Organisation (WHO), 2009. Principles and Methods for the Risk Assessment of Chemicals in Food Chapter 5: Dose-response assessment and derivation of health-based guidance values. Princ. Methods Risk Assess. Chem. Food. Available at: https://apps.who.int/iris/bitstream/handle/10665/44065/WHO_EHC_240_8_eng_Chapter5.pdf;jsessionid=C8BEFE468F684EACAEFD5C821378311F?sequence=8

World Health Organisation (WHO), 1987. Principles for the safety assessment of food additives and contaminants in food. Environmental Health Criteria 70. Available at http://www.inchem.org/documents/ehc/ehc/ehc70.htm

Chapter 2 **Evaluation of non-commercial models for genotoxicity and carcinogenicity in the assessment of EFSA's databases**

Edoardo Carnesecchi [a,b], Giuseppa Raitano [b], Alessio Gamba [b], Emilio Benfenati [b], Alessandra Roncaglioni [b,*]

[a] Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3584 Utrecht, the Netherlands

[b] Laboratory of Chemistry and Environmental Toxicology, Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, 5Milano, Italy

[*] Corresponding author

Published in SAR and QSAR in Environmental Research, 2020, 31.1: 33-48

Abstract

Over the past years, the European Food Safety Authority (EFSA) released to the public domain several databases, with the main objectives of collecting and storing hazard data on the substances considered in EFSA's risk assessment and secondly to serve as a basis for further development of *in silico* tools such as quantitative structure-activity relationship (QSAR) models. In this work, we evaluated the ability of freely available QSAR models to estimate genotoxicity and carcinogenicity properties and their possible use for screening purposes on three different EFSA's databases. With an accuracy close to 90%, the results showed good capabilities of QSAR models to predict genotoxicity in terms of bacterial reverse mutation test, while statistics for in vivo micronucleus test are not satisfactory (accuracy in the predictions close to 50%). Interestingly, results on the carcinogenicity assessment showed an accuracy in prediction close to 70% for the best models. In addition, an example of the potential application of in silico models is presented in order to provide a preliminary screening of genotoxicity properties of botanicals intended for use as food supplements.

Keywords: Genotoxicity; carcinogenicity; (Q)SAR; Ames test; micronucleus *in vivo*; PPPs

1. Introduction

In the EU, the European Food Safety Authority (EFSA) has the mandate to evaluate the risks associated with the food in different scientific areas such as animal and plant health, Genetically Modified Organism (GMO), biological hazards, chemical contaminants and nutrition. In particular, the EFSA has the authority to perform the risk assessment of regulated products such as Plant Protection Products (PPPs), food and feed additives, food contact materials and food ingredients. In this context, EFSA assesses the nature and probability of adverse health effects in humans by performing ad-hoc exposure and hazard assessment of chemicals which humans may be exposed to through the consumption of food. However, the amount of available data characterizing health effects of compounds evaluated by EFSA is largely dependent on the intended use of the substance under evaluation (e.g. PPPs vs food additives). In particular, the availability of hazard information on substances under assessment depends on specific legislation requirements, such as in the case of food contact materials where the greater the exposure is to the substance through migration from packaging to food, the more toxicological are needed (Silano et al., 2008). In addition, the complex nature of chemical composition of some substances addressed by EFSA such as novel food ingredients, botanicals, feed and food additives, might be reflected in the lack of hazard data, causing greater difficulty in filling data gaps. In this context, the use of New Approach Methodologies (NAMs) such as quantitative structure–activity relationship (QSAR) models or read-across might help to gain a more comprehensive view on hazard properties of such substances which are not fully characterized in terms of toxicological properties, and in order to prioritise compounds requiring closer inspection to ensure their safe use. Over the past years, several databases have been released to the public domain by EFSA, with the main objective of collecting and storing hazard data on substances assessed in the previous years, and secondly to serve as a basis for further development of in silico tools (Bassan et al., 2018; Metruccio et al., 2017). In fact, such databases have been designed to be QSAR friendly, thus containing structural information linked to toxicological data, which can be used for computational chemistry and toxicology purposes. One of the key aspects of food safety aims to assess the genotoxicity and carcinogenicity of potential harmful substances. Indeed, potential toxic effects of metabolites or impurities of pesticides, natural complex substances present in food and contaminants generated in food processing or released from food packaging are of primary concern (EFSA Scientific Committee, 2019). As a consequence, EFSA provides a database that reports genotoxicity assessment resulting from different studies on pesticides and related metabolites (Metruccio et al., 2017). Subsequently, this database was extensively investigated in terms of QSAR and read-across capability to address these specific endpoints (Benigni et al., 2019). These previous evaluations showed that acceptable models are currently available for site point mutation while other genotoxic endpoints cannot be characterized by good predictive models.

Currently, most in silico tools applied to EFSA's databases have investigated the application of QSAR models to predict the toxicity of pesticides and their metabolites (Benigni et al., 2019; Benfenati et al., 2017). Hence, the aim of this study is to further evaluate current free in silico software, available in the online VEGA platform (https://www.vegahub.eu/), in order to explore the capability of in silico methods to predict carcinogenic and genotoxic properties of chemicals present in the food chain such as botanicals, food additives, food contact materials and pesticides. Currently, VEGA provides more than 70 QSAR models able to predict (eco) toxicological, environmental and physico-chemical properties of a chemical compound using its structure. Depending on the endpoint and the type of model, the predictions can be qualitative (e.g. mutagenic or not mutagenic) or quantitative (e.g. fish acute toxicity, given in mg/L). In order the measure the reliability of each prediction, VEGA provides an Applicability Domain Index (ADI) value, calculated taking into account the algorithm of the model and of the dataset used to develop the model.

2. Materials and methods

Three different databases provided by EFSA were used to investigate the applicability of (Q)SAR models to fill data gaps for genotoxicity and carcinogenicity of substances relevant for the food safety. Each database stores different types of information and therefore we performed different investigations depending on the data available (e.g. experimental data). For two databases, we tested the predictive abilities of some QSAR models, developed on different datasets, to estimate the experimental test result in genotoxicity assays or the outcome of the expert evaluation process for carcinogenicity. In addition one database containing natural occurring substances present in food but not experimentally characterized for genotoxicity was predicted with (Q)SAR models to set the basis for a screening exercise. Depending on the hazard information stored in the databases, we used different models related to genotoxic and carcinogenic properties available in the VEGA platform. An overview of the work conducted on all databases and the tested models is reported in Table 2.1.

Table 2.1. Overview of the datasets, endpoints and models assessed.

Database	**Dataset distribution**	**Endpoint** (binary classification)	**Models assessed**
EFSA Genotoxicity Database	673 compounds (29 positive and 644 negative)	Bacterial reverse mutation test	• Consensus VEGA (v1.0.2) • SARpy/IRFMN - 18K model (v0.9.1) • KNN/18K model (v1.0.0) • CONSENSUS - 18K model (v1.0.2)
	171 compounds (16 positive and 155 negative)	In vivo micronucleus test	• SARpy56 (56 rules) • SARpy80 (80 rules)
OpenFoodTox Database	'Cleaned' dataset: 628 negative, 97 positive	Carcinogenicity	• CAESAR (v2.1.9) • ISS (v1.0.2) • IRFMN/Antares (v1.0.0) • IRFMN/ISSCAN-CGX (v1.0.0) • Carcinogenicity oral classification model (IRFMN) (v1.0.0)
	'Extended' dataset: 628 negative, 171 positive		
EFSA Compendium of Botanicals	934 compounds	Bacterial reverse mutation test	• CAESAR (v2.1.13) • SARpy/IRFMN (v1.0.7) • ISS (v1.0.2) • KNN/Read-across (v1.0.0) • Consensus (v1.0.2)

2.1 EFSA genotoxicity database

In 2014, EFSA commissioned the collection of genotoxic data for pesticide residues (Metruccio et al., 2017). The database includes information for active substances and their metabolites for about 1000 compounds. The data have been extracted from regulatory toxicological reports as produced by the Rapporteur Member State (RMS) on the pesticide peer review at European Level. Results from different genotoxicity assays are showed, e.g. bacterial reverse mutation (Ames), in vivo mammalian erythrocyte micronucleus, in vitro mammalian chromosome aberration, and in vitro sister chromatid exchange assay in mammalian cells. However, this database contains only experimental study results, and it does not include the assessment of the overall genotoxic potential of the active substance or its metabolite, as in the final EFSA conclusion and in OpenFoodTox database.

2.1.1 Dataset preparation

We used the supplementary files of Benigni et al., (2019) as a basis for the datasets to be predicted with QSAR models since they contain the summary outcome in different genotoxicity test systems of the experimental results stored in a detailed form by the EFSA Genotoxicity Database. We focused on bacterial reverse mutation data and in vivo mammalian erythrocyte micronucleus endpoints.

Compared to the version of the database used by Benigni et al., (2019). for their analysis, we could rely on the publicly available portion of the database only. For the bacterial reverse mutation data we could use 673 structures instead of 918 (73% of the substances, with a prevalence of positive data of 4% in both datasets). For the in vivo micronucleus assay, we included 171 chemicals out of the 219 previously assessed (78% of the substances, with a prevalence of positive data of 9% and 4% respectively). Therefore, the statistics for the VEGA models previously assessed by Benigni et al., (2019) might be different from those reported here.

2.1.2 Tested models

We used several models for mutagenicity (Ames test) assessment, most of them recently developed by our group and therefore not previously assessed for predictivity on this database. The Consensus model as implemented in the latest version of VEGA (since some of the individual models were previously evaluated (Benigni et al., 2019)) and three new models trained on a bigger dataset of more than 18,000 compounds were employed. The last three models are not yet implemented since most of their training data are confidential and cannot be disclosed (Honma et al., 2019; Raitano et et al., 2019). From VEGA v1.1.5 b22 the Consensus model was used. The overall assessment of the VEGA Consensus model is based on the predictions of four individual VEGA models (CAESAR v2.1.13 (Ferrari and Gini, 2010), SARpy v. 1.0.7 (Ferrari et al., 2013), ISS v.1.0.2 (Benigni and Bossa, 2008), and KNN v.1.0.0 (Manganaro et al., 2016)) available in the current version of VEGA, and on their reliability, given as ADI values. The VEGA Consensus algorithm combines the four predictions (mutagenic/non mutagenic) assigning them different coefficients depending on the quality of the estimation as encoded in the ADI. The Consensus model output is a predicted activity (mutagenic/non mutagenic) and a related score that ranges from 0 (the worst case) to 1, when one or more experimental data were present in the dataset. In addition, we used the following new models for bacterial reverse test, which have been developed by our team during the Ames/QSAR international collaborative study, leaded by DGM/NIHS and launched on 2014 (Honma et al., 2019):

- SARpy/IRFMN – 18K model (version 0.9.1) is based on a larger database of compounds (18,338 compounds, 5025 mutagens and 13,313 non-mutagens) (Raitano et al., 2019) compared to those at the basis of the earlier VEGA models. For the predictions, it uses a stepwise approach that includes more than 900 rules (370 active/positive and 567 inactive/negative rules). The possible outcomes of this model show a greater level of detail compared to those of the VEGA Consensus model described before and are: 'mutagenic', 'non-mutagenic', 'possible mutagenic' and 'possible non-mutagenic'.
- KNN/18K model (version 1.0.0) (Honma et al., 2019) was developed applying an in-house software (istKNN) (Manganaro et al., 2016) on the same database of SARpy/IRFMN – 18K

model. The k-Nearest Neighbors algorithm predicts the property of a substance taking into account the experimental data on the most similar compounds in the training set.

- Consensus – 18K model (version 1.0.2) is based on the same algorithm of the Consensus in VEGA where the former SARpy and KNN models are replaced by the two new ones. The models taken into account for the predictions of the Consensus – 18K model are: SARpy/IRFMN – 18K, KNN/18K model, CAESAR v 2.1.13 and ISS v.1.0.2.

For the prediction of the in vivo mammalian erythrocyte micronucleus assay, two sets of rules (SARpy56 and SARpy80) were applied on the EFSA genotoxic database. These fragments were automatically extracted by applying SARpy software to a dataset of about 700 molecules (58% classified as negative and 303 as positive) characterized by experimental results on the in vivo mammalian erythrocyte micronucleus assay. Data were collected from the literature (Morita et al., 2016) and from different databases accessed through the OECD QSAR Toolbox v3.2 (https://qsartoolbox.org/). The model predicts the *in vivo* micronucleus activity by taking into account the presence of structural features (encoded into rules) within the compounds included in the experimental compilation of data. SARpy56 contains 24 active/positive rules and 32 inactive/negative rules and it contains a sub-set of SARpy80 characterized by 32 active/positive rules and 48 inactive/negative rules. The reduced set of 56 rules showed the best accuracy on the training set. To compare the performance of the models mentioned above, we used statistical parameters such as accuracy, sensitivity, specificity (Cooper et al., 1979) and Matthew's correlation coefficient (MCC) (Dao et al., 2011) as specified in the equations below:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \quad \text{(Eq. 1)}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad \text{(Eq. 2)}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}} \quad \text{(Eq. 3)}$$

$$\text{MCC} = \frac{\text{TP}*\text{TN}-\text{FP}*\text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \quad \text{(Eq. 4)}$$

Where TP, TN, FP and FN are defined as true positive, true negative, false positive and false negative, respectively. These valueswere calculated by comparing experimental activity in the EFSA databases with the outcome predicted by each model. MCC values range from −1 to +1, where +1 is the ideal value, −1 complete opposite predictions and 0 random prediction. The Receiver

Operating Characteristics (ROC) plot was used to condense in a graphical format the information on the quality of the models and to simultaneously evaluate sensitivity and specificity. The diagonal line discriminates between better than random models (top left panel) and worse than random (bottom right panel) models. The closer the model is to the upper left corner the better it is (Swets, 1988).

2.2 OpenFoodTox database

Genotoxicity data have been collected from the EFSA chemical Hazards Database 'OpenFoodTox' on 03/09/2019 (Bassan et al., 2018). OpenFoodTox stores summary data resulting from the assessment of different units and Panels of EFSA (e.g. PPPs and their Residues, Food Contact Materials, Contaminants in the Food Chain, etc.). In particular, OpenFoodTox provides data on the chemical entity, the hazard identification, and the hazard characterisation/risk characterisation (e.g. reference points/values, genotoxicity assessment, etc.), and the link to the original EFSA outputs (Dorne et al., 2017). Nowadays, OpenFoodTox stores toxicological data on over 4750 chemical substances including nearly 2115 flavourings, 1150 pesticides, 270 food additives and 200 food contaminants (Bassan et al., 2018). Recently, the applicability of EFSA's OpenFoodTox has been further explored to develop in silico tools such as QSAR models for ecological and human risk assessment (Benfenati et al., 2017; Como et al., 2017; Toropov et al., 2018) as well as to extract reference points (e.g. LD_{50}, LC_{50}) for chemical mixtures toxicity assessment (Carnesecchi et al., 2019).

2.2.1 Dataset preparation

In order to test the predictivity of the QSAR models, the data of the OpenFoodTox database (version updated to January 2019) were filtered according to the following procedure:

- Only compounds with 'positive' or 'negative' entry in the carcinogenicity call were included;
- Inorganic and organometallic compounds were removed;
- Complex substances or substances where a defined structure was not available were discarded too.

For negative compounds, a further selection procedure was then applied by considering only compounds with qualifier: 'as such' and sub_type: 'single chemical entity'. Once duplicated structures were removed, a total of 628 negative compounds were included. For positive compounds, a first extraction was done with the same procedure adopted for negative ones, leading to 97 chemicals. Similarly, 74 additional compounds were evaluated for possible inclusion, being unique structures associated to components within the same assessment group (i.e.: compounds that exhibit similar toxicological properties). These extra compounds are belonging to five chemical clusters: necine bases, polycyclic aromatic hydrocarbons (PAHs), mycotoxins, polychlorinated biphenyls (PCBs) and polybrominated biphenyls (PBBs). We calculated statistical parameters to test the ability of QSAR

models in predicting experimental data considering two datasets. Both datasets contain the same group of negative compounds while for positive ones the first dataset (namely 'cleaned' in the result section) contains only the first subset of positive compounds while a second dataset called 'extended' includes the additional 74 positive compounds, leading to less unbalanced distribution between positive and negative compounds.

2.2.2 Tested models

We used five models for carcinogenicity included in the most recently released version of VEGA (v.1.1.5b22) to obtain the predicted carcinogenic outcome for the compounds included in the datasets described above and compare it with the one based on human expert review:

- The CAESAR model (Fjodorova et al., 2010) is a CP-ANN neural network developed using data for carcinogenicity in rat extracted from the CPDB database.
- The Carcinogenicity model (ISS) implementing the same structural alerts (SAs) as the Carcinogenicity (genotoxic and non-genotoxic) and mutagenicity Toxtree 2.6.13 ISS rulebase (54 SAs in total, 22 for non-genotoxic carcinogenesis) (Benigni et al., 2013).
- The Carcinogenicity IRFMN/Antares model (Golbamaki et al., 2016) is built as a set of rules (127 structural alerts), extracted with the SARpy software from a dataset of about 1550 chemicals obtained from the carcinogenicity database of ANTARES project. This database is a collection of chemical rat carcinogenesis data (presence of carcinogenic effects in male or female rats) obtained from the CAESAR database and the 'FDA 2009 SAR Carcinogenicity – SAR Structures' database from Leadscope. The performance of this model in predicting carcinogenicity, as tested on the external test set obtained from eChemPortal inventory (258 compounds), produced an accuracy of 67%, with a sensitivity of 62%, and a specificity of 70%.
- The Carcinogenicity IRFMN/ISSCAN-CGX Model (Golbamaki et al., 2016) is based on a set of rules (43 structural alerts), extracted with SARpy software from a dataset of almost 1000 compounds. The dataset was obtained from the combination of two data bases: i) The long-term carcinogenicity bioassay on rodents (rat and mouse) ISSCAN data set; ii) The carcinogenicity (different species) data set provided by Kirkland et al., (2005). The leave-one-out evaluation of the ruleset produced an accuracy of 74%, with a sensitivity of 79% and a specificity of 66%.
- Oral Slope Factor (OSF) is a classification model for the prediction of oral carcinogenicity (in terms of slope factor, SF), developed using data from RAIS [available at https://rais.ornl.gov/cgi-bin/tools/TOX_search?select=chem). Values for OSF (expressed as mg/kg-day−1) were retrieved for almost 750 compounds. Compounds having SF values

equal to zero were flagged as negative, while the remaining were flagged as positive. A CART tree model for OSF based on seven descriptors was obtained with 'rpart' package of R. Accuracy, sensitivity, and specificity were above 0.75 on both training and evaluation sets.

All these models provide a categorical output (carcinogenic or not) with slightly different wording (e.g. potential non-carcinogenic) and with different ADI values associated to the corresponding reliability, while all the predictions were summarized into a positive or negative call to be compared with the endpoint outcome. In these models the carcinogenic effects are evaluated with a mix of different levels of evidences. In fact, CAESAR and ANTARES models take into account evidences in rat experiments while ISS and ISSCAN-CGX consider database where a classification has been made by expert judgement on the overall evidences either in animals or human beings. The OSF model relies instead on data arising from the risk assessment process. This observation is reflected in some inconsistencies observed in the experimental values assigned to compounds in the different datasets and in the concordance with the value reported in the OpenFoodTox database. The same statistics described above for the EFSA genotoxicity database were applied here to the OpenFoodTox database.

2.3 The EFSA Compendium of Botanicals

EFSA's Compendium of Botanicals is a publicly available database of information for hazard identification of botanicals, which contain naturally occurring substances of possible concern for human health when present in food (EFSA, 2012). Some details on adverse health effect and health-based guidance values of substances of concern are contained in the Compendium, but the substances are not fully characterized for hazard assessment; therefore, the application of in silico tools such as QSARs is of considerable importance for preliminary screening of those chemicals of concern and their similar compounds potentially present in the botanical preparation. To date, the Compendium provides hazard data on 1134 botanicals. In the present study, 934 single substances for which SMILES were available have been used (i) to perform a preliminary screening for mutagenicity (Ames) assessment, (ii) to fill data gaps for missing experimental data and (iii) to compare predicted toxicological profiles for botanicals present in food and in cosmetics, respectively.

2.3.1 Tested models

The mutagenicity (Ames) models used for the analysis of the Compendium are the 5 models implemented in VEGA: CAESAR model v.2.1.13 (Ferrari and Gini, 2010), SARpy/IRFMN model v.1.0.7 (Ferrari et al., 2013), ISS model v.1.0.2 (Benigni and Bossa, 2008), KNN/Read-across model, v.1.0.0 (Manganaro et al., 2016) and the Consensus model v 1.0.2.

3. Results

3.1 Genotoxicity prediction for the EFSA Genotoxicity database

Results for bacterial reverse test with the new models are reported in Table 2.2 and in Figure 2.1. The two new models are characterized by a very poor sensitivity (below 40%) and very high specificity (above 90%). Conversely, the new Consensus – 18K model has better sensitivity (above 60%) although 10% lower than the VEGA Consensus. Although from the ROC curve in Figure 2.1 the new Consensus – 18K model does not outperform the VEGA Consensus, it demonstrates its higher performance when evaluating the statistics and in particular, the MCC reported in Table 2.2. The number of FP in the Consensus – 18K is lower than VEGA Consensus (48 and 181 respectively), while the number of TP is very similar in the two models (due to the small number of positive compounds). This amplifies the difference in sensitivity (the VEGA Consensus with 21 TP is characterized by 72% of sensitivity while for the Consensus – 18K with 18 TP the sensitivity is 62% only). Similarly to the models evaluated previously (Benigni et al., 2019) for the same endpoints, the SARpy models for in vivo micronucleus confirmed that no satisfactory models are currently available for this endpoint. Accuracy, sensitivity and specificity are all around 50% as reported in Table 2.3; MCC values (in Table 2.3) and the position of the models in the ROC plot in Figure 2.2 indicates that the results achieved so far are very close to random.

Table 2.2. Results for the bacterial revers test on the Genotoxicity database obtained with the newly developed models.

	TP	TN	FP	FN	Acc.	Sens.	Spec.	MCC
Consensus VEGA	21	463	181	8	0.72	0.72	0.72	0.20
SARpy/IRFMN - 18K model	11	589	51	18	0.88	0.34	0.90	0.16
KNN/18K model	10	581	63	19	0.90	0.38	0.92	0.21
CONSENSUS - 18K model	18	596	48	11	0.91	0.62	0.93	0.37

Table 2.3. Results for the micronucleus test in vivo on the Genotoxicity database obtained with the newly developed SARpy models.

	TP	TN	FP	FN	Acc.	Sens.	Spec.	MCC
SARpy56 (56 rules)	8	54	85	6	0.41	0.57	0.39	-0.02
SARpy80 (80 rules)	8	85	66	8	0.56	0.50	0.56	0.04

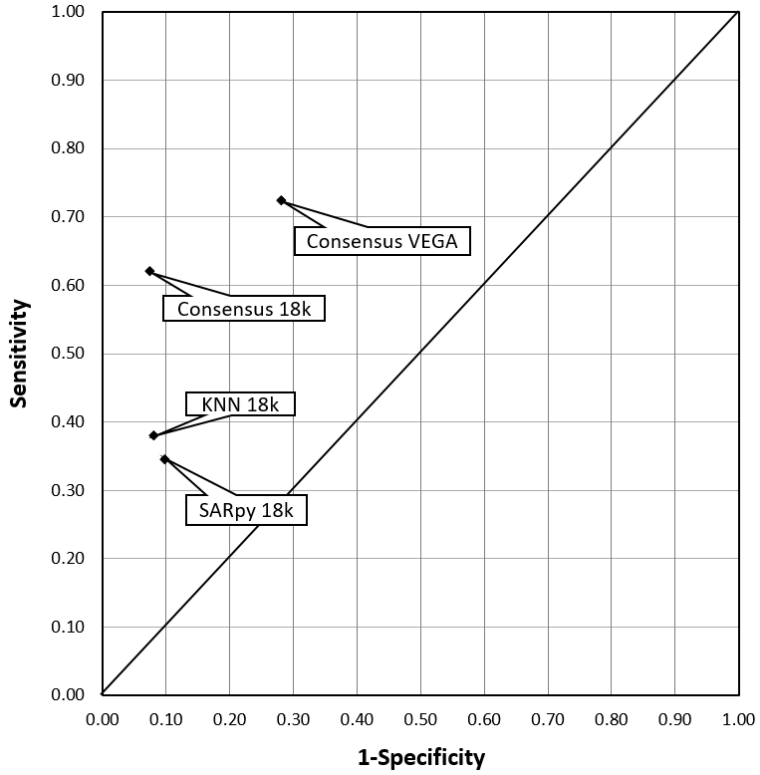


Figure 2.1. ROC curve for the models on bacterial revers test applied to the EFSA Genotoxicity database.

3.2 Carcinogenicity assessment of the OpenFoodTox database

Results of the assessment with the five VEGA models for carcinogenicity are reported in Table 2.4 for the 'cleaned' and 'extended' datasets. In this table only the predicted values are evaluated, regardless the assessment provided by the software in terms of applicability domain or for the presence of experimental values. Most of the models show an accuracy close to 70% (but ISSCAN-CGX model) and a sensitivity around 70–80% (but CAESAR) with a couple of models suffering limited specificity, below 60% (Antares and ISSCAN-CGX). The inclusion of the additional positive compounds present in the 'extended' dataset does not modify substantially the results apart from CAESAR model that increases its sensitivity, resulting the second best model in terms of MCC behind ISS model. However, the results shown in Table 2.4 may be optimistic, due to the presence of some of the compounds here considered for validation purposes, which were also used to develop the models. In particular, Table S2.1 describes the overlapping between OpenFoodTox data and the datasets at the basis of the models. A limited amount of data (from 5% to 15%) are overlapping for negative compounds and additional positive data, while a larger amount of compounds (around 40%-50%) is overlapping when considering the 'cleaned' positive chemicals. Few of these compounds are also present with a different toxicity label.

Table 2.4. Results for the carcinogenicity assessment of the OpenFoodTox database for the 'cleaned' and 'extended' datasets according to the predictions obtained from the VEGA models (regardless the presence of experimental values).

	Predictions for the whole 'cleaned' OpenFoodTox carcinogenicity dataset				*Predictions for the whole 'extended' OpenFoodTox carcinogenicity dataset*			
	Acc.	**Sens.**	**Spec.**	**MCC**	**Acc.**	**Sens.**	**Spec.**	**MCC**
CAESAR	0.69	0.62	0.70	0.22	0.70	0.72	0.70	0.35
ISS	0.68	0.79	0.66	0.31	0.69	0.79	0.66	0.37
Antares	0.60	0.75	0.58	0.23	0.61	0.74	0.58	0.26
ISSCAN- CGX	0.48	0.78	0.43	0.15	0.50	0.76	0.43	0.16
OSF	0.69	0.68	0.69	0.26	0.67	0.61	0.69	0.26

Table 2.5 reports the statistics for the models on both 'cleaned' and 'extended' datasets, once the compounds in common are removed. We have no information about the substances used to obtain the rules for the ISS model, thus for this model we have only excluded substances originally contained in the ISS database (http://old.iss.it/ampp/?lang=1&id=233&tipo=7). No major changes are observed apart from a decrease in CAESAR sensitivity for the 'cleaned' dataset. Figure 2.3 reports the ROC curve for this series of data. ISS model confirmed to be the better performing one, being the closest to the upper left corner of the plot. A small subset of the carcinogens is also classified to be not mutagenic and not genotoxic in the OpenFoodTox database. Although the very limited amount of compounds falling in this category is too small to make any statistical analysis, it is noteworthy that ISS and OSF are the two models capable to identify the larger number of chemicals with these characteristics (see Table S2.2). Non-genotoxic carcinogens are problematic to assess since their hazardous properties are not addressed in standard genotoxicity tests. The ability of some QSAR models to correctly predict also compounds that are active through other mechanisms of action increases the value of these models, since they can be able to raise concern about potential toxic effects difficult to detect.

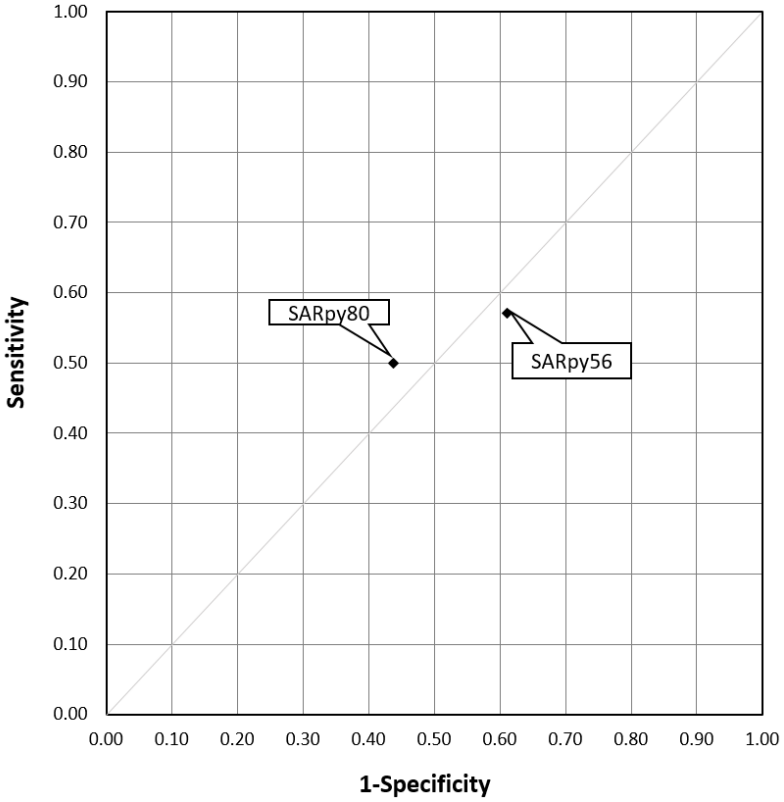


Figure 2.2. ROC curve for the models on in vivo micronucleus test applied to the EFSA Genotoxicity database.

Table 2.5. Results for the carcinogenicity assessment of the OpenFoodTox database for the 'cleaned' and 'extended' datasets according to the assessment obtained from the VEGA models (excluding compounds already present in the training sets).

	New assessments for the 'cleaned' OpenFoodTox carcinogenicity dataset				*New assessments for the whole 'extended' OpenFoodTox carcinogenicity dataset*			
	Acc.	**Sens.**	**Spec.**	**MCC**	**Acc.**	**Sens.**	**Spec.**	**MCC**
CAESAR	0.67	0.52	0.69	0.13	0.69	0.69	0.69	0.30
ISS	0.67	0.78	0.65	0.25	0.68	0.78	0.66	0.33
Antares	0.58	0.74	0.56	0.16	0.59	0.72	0.56	0.21
ISSCAN-CGX	0.44	0.78	0.42	0.10	0.47	0.73	0.42	0.11
OSF	0.69	0.63	0.69	0.19	0.67	0.55	0.69	0.19

3.3 In silico screening of the EFSA compendium of botanicals

Results for the Compendium are represented by the predicted hazard profile for the botanicals considered here, in terms of screening for mutagenicity in Ames test, similarly to the approach previously adopted for botanicals used in cosmetic products (Raitano et al., 2019). For mutagenicity, we used a simplified approach, relying on VEGA models only, compared to the combination

previously used to assess botanicals in cosmetics (Raitano et al., 2019). Results on the four VEGA individual models are included in the VEGA Consensus evaluation together with a consensus score accounting for the final reliability of the single estimations. In Figure 2.4, the predicted mutagenic activity for the Compendium and the botanicals used in cosmetics is compared. A very similar trend is observed in the prevalence of the categories and in particular in both datasets the non-mutagenic predictions are overall representing around 75% of the chemicals. Similarly, a very low number of experimental values (represented by the outcomes with very good reliability) is observed in both datasets, indicating the relatively poor experimental characterization of these naturally occurring compounds.

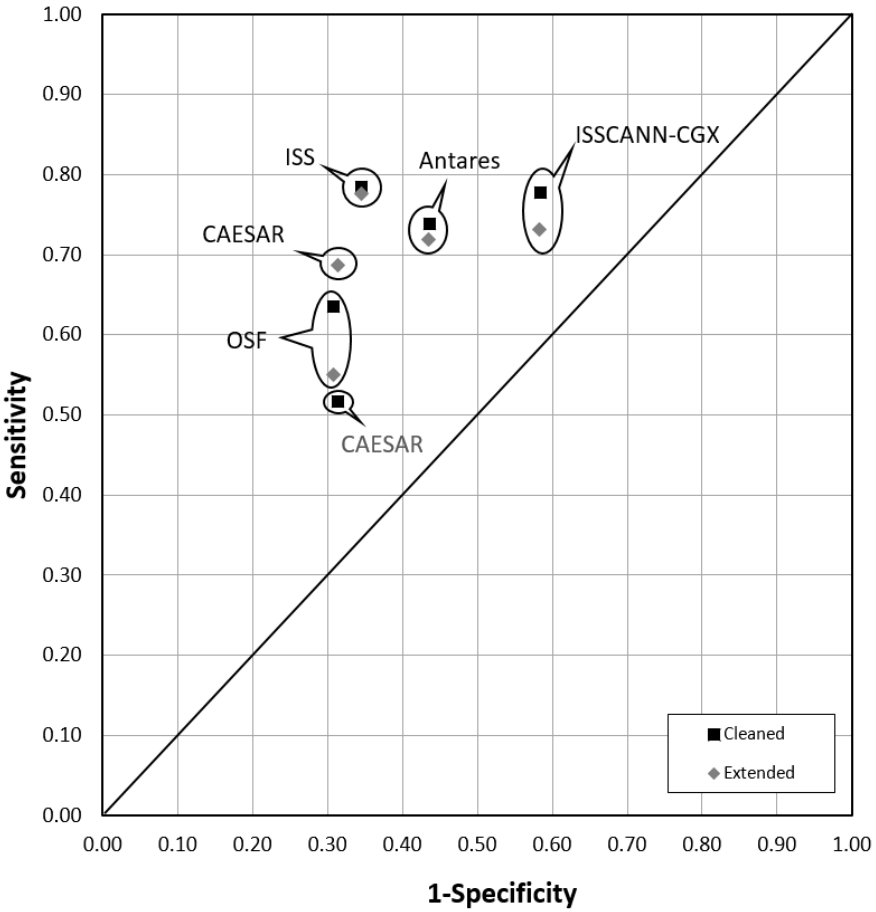


Figure 2.3. ROC curve for the VEGA models for carcinogenicity applied to the OpenFoodTox database.

4. Discussion

Genotoxicty, as described by the Ames test, has been often assessed with in silico methods and modelling results often demonstrated that it is possible to obtain models with an accuracy similar to the intrinsic experimental reproducibility of the Ames test (Benfenati et al., 2018). Compared to the models previously assessed the new models tested here – which are based on a very large compilation of data – are characterized by higher specificity suffering for a reduced sensitivity.

However the combination of these new, highly specific models and the former ones (e.g. CAESAR) with higher sensitivity improved the accuracy of the Consensus limiting detrimental effects on the sensitivity. Overall in silico models demonstrated to be useful for characterizing PPPs metabolites and impurities which are not fully investigated as the active substances. Furthermore, read-across and a weight of evidence approach, coupling QSAR and read-across, could be of valuable support to individual substance assessment (Benigni et al., 2019). On the other hand, results for the in vivo micronucleus test were previously reported to be of low quality and similar results are observed here by employing new models. One possible reason of this low accuracy may be the increased complexity of the endpoint under inspection compared to the most common Ames test assay. Ames test is the most widely used assay to detect gene mutations and it is usually the first tier test to be performed during genotoxic assessment. In vivo micronucleus test is used to detect numerical and structural chromosomal aberrations and therefore allows identifying aneugens and clastogens.

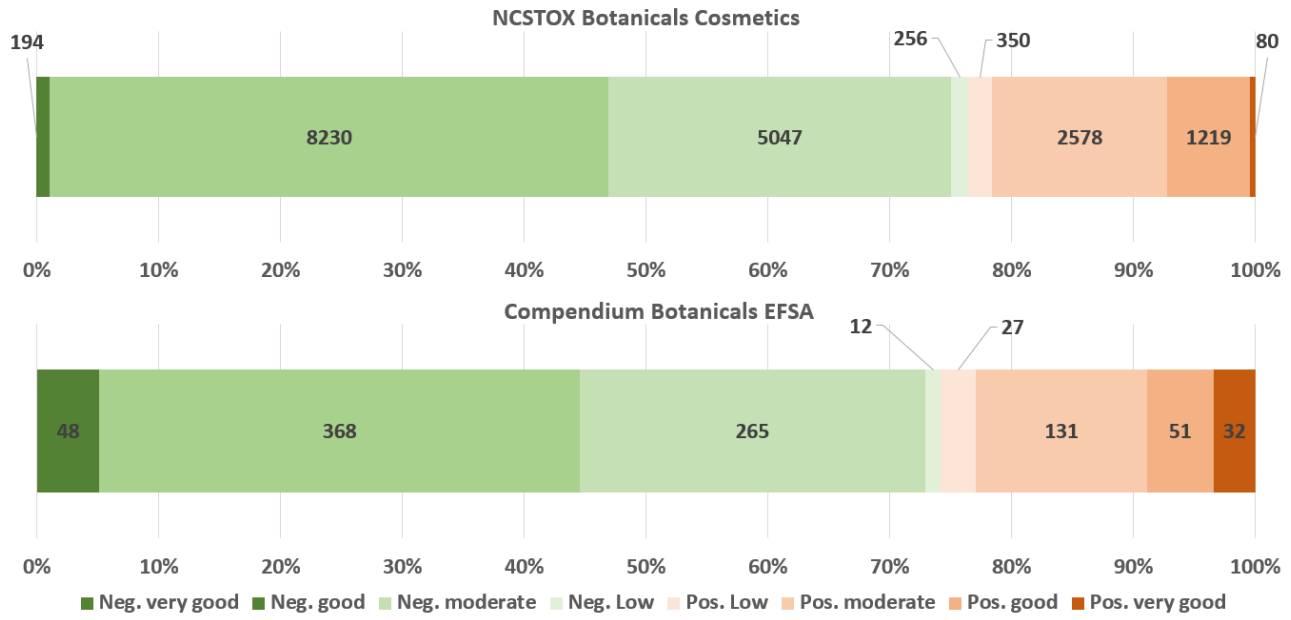


Figure 2.4. Results of the screening of the EFSA Compendium of botanicals (B) and the NCSTOX botanicals used in cosmetics (A) as investigated in the literature (Raitano et al., 2019). The predictions are provided with different reliability depending on the consensus score (C.S.). Low for C.S. ≤ 0.15, moderate for 0.15 < C.S. < 0.5, good for 0.50 ≤ C.S. <1 and very good if C.S. = 1 (when an experimental value is present).

This issue of the complexity of the test can only partially explain the poor performance observed since the models tested on carcinogenicity performed much better although the endpoint itself is a fairly complex one. It has to be underlined that extensive efforts were devoted in the last decades to modelling carcinogenicity with growing number of freely available and commercial models addressing this endpoint. In general, results are less accurate compared to mutagenesis in Ames

test but still characterized by reasonable statistics when probed for their predictivity (Milan et al., 2011). Our results confirmed previous findings and showed that non genotoxic carcinogens might be detected up to a reasonable extent too. A further issue that can account for the low statistics obtained in the analysis of the micronucleus data, is the availability of large and robust data collections for carcinogenicity (Kirkland et al., 2005; CPDB, available at https://toxnet.nlm.nih.gov/cpdb/) and Ames test (Honma et al., 2019), but not for micronucleus, highlighting the importance of any effort in making curated database publicly available, as supported by EFSA. These results demonstrates that New Approach Methodologies (NAMs), such as in silico methods, can be a useful tool for screening purposes as shown for the analysis of the Compendium and they may play a role in the assessment framework of substances regulated by EFSA. Their application goes beyond the assessment of PPPs active substances which have to be tested experimentally for both carcinogenicity and genotoxicity. On the other hand, in silico tools can support the food safety assessment by investigating impurities and metabolites of PPPs, botanicals present in feed and food, food contaminants and complex mixtures (screening each structurally defined component) to guide strategies for further testing.

5. Conclusion

This work analysed the capability of some non-commercial in silico models in assessing carcinogenicity and genotoxicity (in Ames and in vivo micronucleus tests). Results demonstrated that currently in silico models are good screening tools to prioritize substances requiring closer inspection for carcinogenicity and genotoxicity (in Ames test) while no freely available models are currently available for in vivo micronucleus test. Further work is required to better assess aneugenic and clastogenic substances and to improve models for in vivo genotoxicity test. As discussed before, a larger curated collection of experimental data might help overcoming this issue. In this context, the ongoing EFSA's OpenFoodTox 2.0 'OptiTox' project will provide additional experimental toxicokinetic data (e.g. bioavailability, half-life) on tested chemicals, which might help to better capture in vivo genotoxicity test response using in silico models.

Acknowledgements

This work has been financed by the European Food Safety Authority (EFSA) under the contract number OC/EFSA/SCER/2018/01. Authors would like to acknowledge colleagues in S-IN (Soluzioni Informatiche) for providing the extract of OpenFoodTox database and Jean Lou Dorne for providing the EFSA Compendium on Botanicals.

References

Bassan, A, Ceriani, L., Richardson, J., Livaniou, A., Ciacci, A., Baldin, R., Kovarich, S., Fioravanzo, E., Pavan, M., Gibin, D., Di Piazza, G., Pasinato, L., Cappé, S., Verhagen, H., Robinson, T., Dorne, J.L., 2018. OpenFoodTox: EFSA's chemical hazards database, Zenodo (2018). doi:10.5281/zenodo.780543

Benfenati, E., Golbamaki, A., Raitano, G., Roncaglioni, A., Manganelli, S., Lemke, F., Norinder, U., Lo Piparo, E., Honma, M., Manganaro, A., Gini, G., 2018. A large comparison of integrated SAR/QSAR models of the Ames test for mutagenicity. SAR QSAR Environ. Res. 29, 591–611. https://doi.org/10.1080/1062936X.2018.1497702

Benfenati, E., Como, F., Manzo, M., Gadaleta, D., Toropov, A., Toropova, A., 2017. Developing innovative in silico models with EFSA's OpenFoodTox database. EFSA Support. Publ. https://doi.org/10.2903/sp.efsa.2017.en-1206

Benigni, R., Laura Battistelli, C., Bossa, C., Giuliani, A., Fioravanzo, E., Bassan, A., Fuart Gatnik, M., Rathman, J., Yang, C., Tcheremenskaia, O., 2019. Evaluation of the applicability of existing (Q)SAR models for predicting the genotoxicity of pesticides and similarity analysis related with genotoxicity of pesticides for facilitating of grouping and read across. EFSA Support. Publ. https://doi.org/10.2903/sp.efsa.2019.en-1598

Benigni, R., Bossa, C., Tcheremenskaia, O., 2013. Nongenotoxic carcinogenicity of chemicals: Mechanisms of action and early recognition through a new set of structural alerts. Chem. Rev. https://doi.org/10.1021/cr300206t

Benigni, R., Bossa, C., 2008. Structure alerts for carcinogenicity, and the Salmonella assay system: A novel insight through the chemical relational databases technology. Mutat. Res. - Rev. Mutat. Res. https://doi.org/10.1016/j.mrrev.2008.05.003

Carnesecchi, E., Svendsen, C., Lasagni, S., Grech, A., Quignot, N., Amzal, B., Toma, C., Tosi, S., Rortais, A., Cortinas-Abrahantes, J., Capri, E., Kramer, N., Benfenati, E., Spurgeon, D., Guillot, G., Dorne, J.L.C.M., 2019. Investigating combined toxicity of binary mixtures in bees: Meta-analysis of laboratory tests, modelling, mechanistic basis and implications for risk assessment. Environ. Int. https://doi.org/10.1016/j.envint.2019.105256

Como, F., Carnesecchi, E., Volani, S., Dorne, J.L., Richardson, J., Bassan, A., Pavan, M., Benfenati, E., 2017. Predicting acute contact toxicity of pesticides in honeybees (Apis mellifera) through a k-nearest neighbor model. Chemosphere. https://doi.org/10.1016/j.chemosphere.2016.09.092

Cooper, J.A., Saracci, R., Cole, P., 1979. Describing the validity of carcinogen screening tests. Br. J. Cancer. https://doi.org/10.1038/bjc.1979.10

Dao, P., Wang, K., Collins, C., Ester, M., Lapuk, A., Sahinalp, S.C., 2011. Optimally discriminative subnetwork markers predict response to chemotherapy. Bioinformatics. https://doi.org/10.1093/bioinformatics/btr245

Dorne, J. Lou, Richardson, J., Kass, G., Georgiadis, N., Monguidi, M., Pasinato, L., Cappe, S., Verhagen, H., Robinson, T., 2017. Editorial: OpenFoodTox: EFSA's open source toxicological database on chemical hazards in food and feed. EFSA J. https://doi.org/10.2903/j.efsa.2017.e15011

EFSA (European Food Safety Authority) Scientific Committee, S. More, V. Bampidis, D. Benford, J. Boesten, C. Bragard, T. Halldorsson, A. Hernandez-Jerez, S. Hougaard-Bennekou, K. Koutsoumanis, H. Naegeli, S.S. Nielsen, D. Schrenk, V. Silano, D. Turck, M. Younes, G. Aquilina, R. Crebelli, R. Gurtler, K.I. Hirsch-Ernst, P. Mosesso, E. Nielsen, R. Solecki, M. Carfì, C. Martino, D. Maurici, J. Parra Morte, J. Schlatter, Statement on the genotoxicity assessment of chemical mixtures, EFSA Journal 17(1):5519 (2019), 11 pp. doi:10.2903/j.efsa.2019.5519

EFSA (European Food Safety Authority), 2012. Compendium of botanicals reported to contain naturally occuring substances of possible concern for human health when used in food and food supplements. EFSA J. https://doi.org/10.2903/j.efsa.2012.2663

Ferrari, T., Gini, G., 2010. An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. Chem. Cent. J. https://doi.org/10.1186/1752-153X-4-S1-S2

Ferrari, T., Cattaneo, D., Gini, G., Golbamaki Bakhtyari, N., Manganaro, A., Benfenati, E., 2013. Automatic knowledge extraction from chemical structures: The case of mutagenicity prediction. SAR QSAR Environ. Res. https://doi.org/10.1080/1062936X.2013.773376

Fjodorova, N., Vračko, M., Novič, M., Roncaglioni, A., Benfenati, E., 2010. New public QSAR model for carcinogenicity. Chem. Cent. J. https://doi.org/10.1186/1752-153X-4-S1-S3

Golbamaki, A., Benfenati, E., Golbamaki, N., Manganaro, A., Merdivan, E., Roncaglioni, A., Gini, G., 2016. New clues on carcinogenicity-related substructures derived from mining two large datasets of chemical compounds. J. Environ. Sci. Heal. - Part C Environ. Carcinog. Ecotoxicol. Rev. https://doi.org/10.1080/10590501.2016.1166879

Honma, M., Kitazawa, A., Cayley, A., Williams, R. V., Barber, C., Hanser, T., Saiakhov, R., Chakravarti, S., Myatt, G.J., Cross, K.P., Benfenati, E., Raitano, G., Mekenyan, O., Petkov, P., Bossa, C., Benigni, R., Battistelli, C.L., Giuliani, A., Tcheremenskaia, O., DeMeo, C., Norinder, U., Koga, H., Jose, C., Jeliazkova, N., Kochev, N., Paskaleva, V., Yang, C., Daga, P.R., Clark, R.D., Rathman, J., 2019. Improvement of quantitative structure-activity relationship (QSAR) tools for predicting Ames mutagenicity: Outcomes of the Ames/QSAR International Challenge Project. Mutagenesis. https://doi.org/10.1093/mutage/gey031

Kirkland, D., Aardema, M., Henderson, L., Müller, L., 2005. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: I. Sensitivity, specificity and relative predictivity. Mutat. Res. - Genet. Toxicol. Environ. Mutagen. https://doi.org/10.1016/j.mrgentox.2005.02.004

Manganaro, A., Pizzo, F., Lombardo, A., Pogliaghi, A., Benfenati, E., 2016. Predicting persistence in the sediment compartment with a new automatic software based on the k-Nearest Neighbor (k-NN) algorithm. Chemosphere. https://doi.org/10.1016/j.chemosphere.2015.10.054

Metruccio, F., Castelli, I., Civitella, C., Galbusera, C., Galimberti, F., Tosti, L., Moretto, A., 2017. Compilation of a database, specific for the pesticide active substance and their metabolites, comprising the main genotoxicity endpoints. EFSA Support. Publ. https://doi.org/10.2903/sp.efsa.2017.en-1229

Milan, C., Schifanella, O., Roncaglioni, A., Benfenati, E., 2011. Comparison and possible use of in Silico tools for carcinogenicity within REACH legislation. J. Environ. Sci. Heal. - Part C Environ. Carcinog. Ecotoxicol. Rev. https://doi.org/10.1080/10590501.2011.629973

Morita, T., Hamada, S., Masumura, K., Wakata, A., Maniwa, J., Takasawa, H., Yasunaga, K., Hashizume, T., Honma, M., 2016. Evaluation of the sensitivity and specificity of in vivo erythrocyte micronucleus and transgenic rodent gene mutation tests to detect rodent carcinogens. Mutat. Res. - Genet. Toxicol. Environ. Mutagen. https://doi.org/10.1016/j.mrgentox.2016.03.008

Raitano, G., Roncaglioni, A., Manganaro, A., Honma, M., Sousselier, L., Do, Q.T., Paya, E., Benfenati, E., 2019. Integrating in silico models for the prediction of mutagenicity (Ames test) of botanical ingredients of cosmetics. Comput. Toxicol. https://doi.org/10.1016/j.comtox.2019.100108

Risk Assessment Information System (RAIS), *Toxicity values database*. (https://rais.ornl.gov/cgi-bin/tools/TOX_search?select=chem).

Silano, V., Bolognesi, C., Castle, L., Cravedi, J.P., Engel, K.H., Fowler, P., Franz, R., Grob, K., Gürtler, R., Husøy, T., Kärenlampi, S., Mennes, W., Milana, M.R., Penninks, A., de Fátima Tavares Poças, M., Smith, A., Tlustos, C., Wölfle, D., Zorn, H., Zugravu, C.A., 2008. Note for Guidance For the Preparation of an Application for the Safety Assessment of a Substance to be used in Plastic Food Contact Materials. EFSA J. https://doi.org/10.2903/j.efsa.2008.21r

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Sci. Sci. https://doi.org/10.1126/science.3287615

The Carcinogenic Potency Database (CPDB), https://toxnet.nlm.nih.gov/cpdb/

Toropova, A.P., Toropov, A.A., Marzo, M., Escher, S.E., Dorne, J. Lou, Georgiadis, N., Benfenati, E., 2018. The application of new HARD-descriptor available from the CORAL software to building up NOAEL models. Food Chem. Toxicol. https://doi.org/10.1016/j.fct.2017.03.060

Chapter 3 **Investigating combined toxicity of binary mixtures in bees: meta-analysis of laboratory tests, modelling, mechanistic basis and implications for risk assessment**

Edoardo Carnesecchi [a,b], Claus Svendsen [c], Stefano Lasagni [d], Audrey Grech [e], Nadia Quignot [f], Billy Amzal [f], Cosimo Toma [b], Simone Tosi [g], Agnes Rortais [h], Jose Cortinas-Abrahantes [h], Ettore Capri [i], Nynke Kramer [a], Emilio Benfenati [b], David Spurgeon [c], Gilles Guillot [j], Jean Lou Christian Michel Dorne [h,k,*]

[a] Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3584 Utrecht, the Netherlands

[b] Laboratory of Chemistry and Environmental Toxicology, Department of Environmental Health, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, via Mario Negri, 2, 20156 Milano, Italy

[c] Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford OX10 8BB, UK

[d] Arpae Emilia Romagna, Bologna, Italy

[e] INERIS, Paris, France

[f] CERTARA, Paris, France

[g] Epidemiology Unit, European Union Reference Laboratory (EURL) for Honeybee Health, University Paris Est, French Agency for Food, Environmental and Occupational Health and Safety, Paris, France

[h] European Food Safety Authority (EFSA), Scientific Committee and Emerging Risks Unit, Parma, Italy

[i] Università Cattolica del Sacro Cuore, Dipartimento di Scienze e Tecnologie Alimentari per una filiera agro-alimentare Sostenibile (DiSTAS), Piacenza, Italy

[j] International Prevention Research Institute, Lyon, France

[k] School of Biosciences and Phenome Centre Birmingham, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

[*] Corresponding author

Published in Environment International, 2019, 133: 105256

Abstract

Bees are exposed to a wide range of multiple chemicals "chemical mixtures" from anthropogenic (e.g. plant protection products or veterinary products) or natural origin (e.g. mycotoxins, plant toxins). Quantifying the relative impact of multiple chemicals on bee health compared with other environmental stressors (e.g. varroa, viruses, and nutrition) has been identified as a priority to support the development of holistic risk assessment methods. Here, extensive literature searches and data collection of available laboratory studies on combined toxicity data for binary mixtures of pesticides and non-chemical stressors has been performed for managed bees (honey bees (*Apis mellifera*), wild bees (*Bombus* spp.) and solitary bee species (*Osmia* spp.). From 957 screened publications, 14 publications provided 218 binary mixture toxicity data mostly for acute mortality (Lethal dose: LD_{50}) after contact exposure (61%), with fewer studies reporting chronic oral toxicity (20%) and acute oral LC_{50} values (19%). From the data collection, available dose response data for 92 binary mixtures were modelled using a Toxic Unit (TU) approach and the MIXTOX modelling tool to test assumptions of combined toxicity i.e. concentration addition (CA), and interactions (i.e. synergism, antagonism). The magnitude of interactions was quantified as the Model Deviation Ratio (MDRs). The CA model applied to 17% of cases while synergism and antagonism were observed for 72% (MDR > 1.25) and 11% (MDR < 0.83) respectively. Most synergistic effects (55%) were observed for the interaction between sterol-biosynthesis-inhibiting (SBI) fungicides and insecticide/acaricide. The mechanisms behind synergistic effects of binary mixtures in bees are known to involve direct cytochrome P450 (CYP) inhibition, resulting in an increase in internal dose and toxicity of the binary mixture. Moreover, bees are known to have the lowest number of CYP copies and other detoxification enzymes in the insect kingdom. In the light of these findings, occurrence of these binary mixtures in relevant crops (frequency and concentrations) would need to be investigated. Addressing this exposure dimension remains critical to characterise the likelihood and plausibility of such interactions to occur under field realistic conditions. Finally, data gaps and further work for the development of risk assessment methods to assess multiple stressors in bees including chemicals and non-chemical stressors in bees are discussed.

Keywords: Bees, pesticides, mixtures, interactions, laboratory toxicity, risk assessment.

1. Introduction

Worldwide, bee species such as the honey bee (*Apis mellifera*), bumble bees (*Bombus* spp.) and solitary bees (e.g. *Osmia* spp.) are essential organisms for the environment particularly for their critical roles in the pollination of crops, flowers and fruit trees and consequently their economic impact (Kennedy et al., 2013; Burkle et al., 2013; Potts et al., 2010). In the world, 75% of food crops (e.g. cacao, almond, apple etc.) relies on animal-mediated pollination (Klein et al., 2007) and the majority of human micronutrients (e.g. vitamins, minerals) derive from pollinator-dependent crop production (e.g. citrus fruits, walnuts tree, etc.) (Eilers et al., 2011). In contrast, crops such as wheat and rice providing mainly macronutrients (e.g. carbohydrate) are generally wind or self-pollinated (Culley et al., 2002). Moreover, it has been estimated that honey bees are responsible for providing pollination service to 96% of animal-pollinated crops and thus playing a key role in the maintenance and reproduction of 52 out of 115 leading global commodities (Vanengelsdorp and Meixner, 2010; Klein et al., 2007). In Europe, 84% of 264 cultivated crops are pollinated by insects and 4.000 vegetable varieties depend on bee pollination services as well as in the production of fruits (e.g. kiwi, raspberries, blueberries, etc.), seeds and vegetables (e.g. sunflower seeds, beetroot, carrots) through pollination services (Williams, 1994, Corbet et all 1991; Bommarco et al., 2012; Hoshide et al., 2018). Bees are also indirectly responsible for the reproduction and maintenance of wild plant communities and biodiversity (Aguilar et al., 2006; Ashman et al., 2004; De Groot et al., 2002). In addition, it is well known that managed honey bees provide honey, pollen, wax (e.g. for food processing), propolis (e.g. food technology), and royal jelly (used as a dietary supplement or as food ingredient) (Formato et al., 2011; Tinto et al., 2016). Overall, bees represent a very significant pollination service bridging agriculture, the food chain and the ecosystem thereby ensuring food production and security (Rose et al., 2015). In economic terms, the pollination services, from honey bees, bumble bees and wild bees contribute at least to 22 billion EUR each year of the European agriculture sector (Commission, E, 2016).

Over the last decade, important honey bee colony losses have been reported, particularly in North America and Western Europe (Jacques et al., 2016; Sanchez-Bayo et al., 2016; Steinhauer et al., 2014; Van der Zee et al., 2012). Scientific evidence shows that the weakening or death of bee colonies is mainly caused by the combined effects of multiple stressors rather than by one-off sudden attacks by a single factor (Goulson et al., 2015; EFSA, 2014a; Potts et al., 2010; Rortais et al., 2017). Such interactions can occur principally between i) biological factors (Nazzi et al., 2012; Nazzi and Pennacchio, 2014), ii) environmental factors (Di Pasquale et al., 2016; Goulson et al., 2015; Le Conte and Navajas, 2008), iii) chemical and nutritional stressors (Tosi et al., 2017; Tong et al., 2019), iv) chemical and biological factors (Williamson et al., 2013; Klein et al., 2017; Alaux et al., 2010; Vidau et al., 2011; Pettis et al., 2012) and v) multiple chemicals (EFSA, 2013a, b; Robinson et al., 2017;

Han et al., 2019; Sanchez-Bayo and Goka, 2016). In particular, the latter is raising concerns among scientists and regulatory bodies since bees can be exposed to a wide range of multiple chemicals, "chemical mixtures", including compounds from anthropogenic (e.g. plant protection products or veterinary drugs) or natural origin (e.g. mycotoxins, flavonoids, plant toxins) (Johnson, 2015; Tosi et al., 2019; EFSA PPR Panel, 2012; EFSA, 2014a). Hence, investigating the relative impact of multiple chemicals in comparison to non-chemical stressors (e.g. varroa, viruses) on bee health has been identified by the European Food Safety Authority (EFSA) as a priority to support the development of holistic risk assessment (RA) methods (EFSA AHAW Panel, 2016; EFSA, 2017a; Rortais et al., 2017). In this context, the Scientific Committee of EFSA has recently published a guidance document on "harmonised methodologies for human health, animal health and ecological RA of combined exposure to multiple chemicals" which provides a harmonised framework and step wise approaches for whole mixture and component-based approaches. The stepwise approaches are applied to very step of the RA process namely problem formulation, exposure assessment, hazard identification and characterisation, risk characterisation and uncertainty analysis (More et al., 2019). When dealing with component-based approaches, two main mathematical reference models are usually applied when predicting combined toxicity assuming non-interaction: dose/concentration addition (CA) (Loewe S., 1926) and independent action/response addition (IA) (Bliss C. I., 1939). When combined toxicity significantly deviates from the observed responses from CA or IA, predictions are usually referred to and modelled as interactions (Jonker et al., 2005; Kienzler et al., 2016; More et al., 2019). Interactions have been described as either antagonism (i.e. combined toxicity is below the sum of the components' toxicity) or synergism (i.e. toxicity of mixture greater than the sum of components' toxicity) (Kienzler et al., 2014). However, if only one of the chemicals in the binary mixture is expected to cause adverse effect (e.g. clothianidin + piperonyl butoxide), synergism is usually defined as potentiation (Heys et al., 2016; Robinson et al., 2017). In practice, mixtures of components with similar Modes of Action (MoA) are addressed using the CA model, whereas compounds with different MoAs are assessed using the IA model that mathematically combine probabilities of independent events (Jonker et al., 2005; Belden et al., 2007). Overall, evidence from the literature and scientific advisory bodies worldwide support the application of CA as a conservative approach compared to IA unless evidence for interactions can be demonstrated (Bopp et al., 2015; EFSA, 2013a, b; More et al., 2019).

The current manuscript provides the first quantitative review of the available laboratory toxicological studies of binary mixtures of chemicals (i.e. pesticides, veterinary drugs and environmental contaminants) in honey bees and wild bees. It aims to support hazard assessment by means of extensive literature searches, data collection, modelling and analysis of combined toxicity (dose addition, interactions (i.e. synergism, antagonism)) and their associated mechanisms.

First, extensive literature searches are perfomed to identify and collect combined toxicity endpoints (e.g. LD_{50} or LC_{50}) from acute and chronic laboratory studies on binary mixtures in honey bees and wild bees (solitary bees and bumble bees) together with available toxicity data and mode of action information from public databases. In addition, dose response from each individual binary mixture experiment are modelled to identify the nature and potency of the combined toxicity (dose addition, synergism, antagonism) and quantify its magnitude using a toxic unit approach and the MIXTOX model. Furthermore, new predictive hazard assessment tools applicable to large binary mixture datasets in bees are developed. The reader should note that exposure assessment (pesticides application rate, crop management, consumption patterns, etc.) and full risk characterisation are beyond the scope of this quantitative analysis. Implications for risk assessment and future directions concludes while considering mechanisms of interactions, data gaps, importance of exposure assessment scenarios and risk characterisation as well as the development of methods to assess multiple chemicals and multiple stressors in bees to support risk management.

2. Materials and Methods

2.1 Extensive Literature Searches

Extensive Literature Searches (ELS) were performed by two independent reviewers in January 2018 to critically appraise, collect and analyse data on toxicity of mixtures in bee species (EFSA, 2010), using structured search strategies (appendix S1). ELSs were carried out in PubMed (1975-2018), in Web of Science Core Collection (1975-2018), including Science Citation Index Expanded, CABI: CAB Abstracts®, Current Contents Connect®, Data Citation Index SM, FSTA® the food science resource, MEDLINE®, SciELO Citation Index, Zoological Record®, Conference Proceedings Citation Index-Science, Book Citation Index– Science, Current Chemical Reactions, Index Chemicus). All records were computed in the EndNote™ software. In addition, bibliographical sources from EFSA studies and database on mixture toxicity in bees were checked thoroughly for completeness (Quignot et al., 2015; Robinson et al., 2017). In addition, qualitative information on the Mode of Action (MoA) of the individual chemicals were collected from the literature and available databases (Sparks and Nauen, 2015; Hermann and Stenzel, 2019; Sanchez-bayo et al., 2012; Johnson et al., 2012, 2013; Leroux et al., 2008; De Castro et al., 2015; Huang et al., 2013).

Each individual publication retrieved in EndNote™ libraries was screened and assessed using inclusion and exclusion criteria reported in Table 3.1 in two steps i) screening of the titles and abstracts and ii) screening of the full-text of the publications. All included and excluded publications are available under individual EndNote™ libraries.

Table 3.1. Inclusion and exclusion criteria for the selection of relevant literature in the extensive literature search.

Inclusion criteria	
Review question	- Does the study provide toxicological outcomes for binary mixtures in bee species?
Population of interest	- Bees species (i.e. honey bee, bumble bees, *Osmia* spp.)
Study design	- *In vivo* experimental laboratory studies - *In vivo* field/semi-field studies; - Routes of exposure (i.e. contact, oral) - Study length (acute, chronic)
Outcome of interest	- Summary statistics or individual datasets on toxicity of mixtures and non-chemical stressors (e.g. LD_{50}, LC_{50} and related statistical descriptors) for single doses. - Summary statistics or individual datasets for multiple doses (dose response data) on toxicity of mixtures (e.g. LD_{50}, LC_{50} and related statistical descriptors)
Exclusion criteria	
Type of study	- *In vitro* studies - Studies reporting only qualitative data with no toxicological outcome; - Duplicated studies: studies reporting the same dataset in several publications, studies on non-chemical stressors or combined chemical and non-chemcial stressors, studies reporting results from systematic reviews, meta-analyses or predictive models.

2.2 Data collection and analysis

2.2.1 Data collection

Following the Extensive Literature Searches, individual toxicological endpoints (acute and chronic) from laboratory mixture experiments (e.g. LD_{50} or LC_{50}) were collected for the oral and contact exposure according to the inclusion criteria, including bee species, sample, size, summary statistics (mean, median, standard error of the mean, standard deviation, confidence intervals) and exposure patterns. Standardised templates were developed to structure the data into an excel database designed with relevant picklists. When papers reported only graphical information, quantitative data were extracted using "Plot Digitizer GNU" software (available at http://plotdigitizer.sourceforge.net/) or the R software (R Core Team, 2019).

In addition, reference points (e.g. LD_{50}, LC_{50}) for all individual chemicals i.e. mostly Plant Protection Products (PPPs) in honey bees were extracted from EFSA's Chemical Hazards database

"OpenFoodTox" (available at https://zenodo.org/record/1252752#.XLg-4Oj7SUm) (Dorne et al., 2017; EFSA, 2014b) and other publicly available databases were consulted including the US-EPA dashboard (https://cfpub.epa.gov/ecotox/), OECD e-ChemPortal (https://www.echemportal.org/echemportal/index.action), PPDB-Pesticide Properties Database (https://sitem.herts.ac.uk/aeru/ppdb/). All binary mixtures data were compiled in an excel database for further analysis (see chapter 2.2.2).

2.2.2 Quantification of magnitudes of interaction

Estimated Mean Ratios

A comprehensive analysis of magnitude of interactions (as potency or synergism ratios) was performed through the calculation of Estimated Mean Ratios (EMR) for each individual single compound and binary mixture toxicity dataset or for combined toxicity between a single chemical and a non-chemical stressor (biological or nutritional). EMR has been defined as the ratio between the estimated mean toxicity (e.g. LD_{50}, LC_{50}, EC_{50}) of a given single chemical (chemical A) for which the experimental dose is available (EM_A) and the estimated toxicity of the binary mixture chemical A + chemical B (EM_M) or chemical A + non-chemical stressor (Quignot et al., 2015):

$$\mathrm{EMR} = \frac{\mathrm{EM}_A}{\mathrm{EM}_M} \quad (1)$$

Each EMR for a given binary mixture (EM_M) is expressed on a harmonised scale starting at 1 to reflect changes in combined toxicity either as an increase (+) or a decrease (-) (Quignot et al., 2015).

It is noted that the EMR approach assumes that chemical B does not contribute to the mixture toxicity which does not fully comply with the principles of concentration addition (CA), which assumes that any amount of a chemical always contributes to the combined toxicity expressed in Toxic Units (Jonker et al., 2005). For each binary mixture, the statistical significance of the combined toxicity has been estimated using non-overlapping 95% confidence intervals (95% CI of the EM_A vs 95% CI of the EM_M for chemical A+B) as described in Johnson et al., (2012, 2013). All calculations were carried out in the R software (R Core Team, 2019).

In addition, risk of bias was assessed through the quantification of the variability across studies by calculating the Confidence Intervals (CIs) for each Estimated Mean Ratio (EMR). 95% CI were calculated according to the Fieller (1954) and Delta methods as described in the formulas 2, 3 and 4. Fieller's method (Fieller, 1954) is based on the assumption that $(\hat{\theta}_1, \hat{\theta}_2)$ follows a bivariate Normal distribution. For testing $\theta_1/\theta_2 = R_0$ (which amounts to testing $\theta_1 = R_0\theta_2$), the two-sided t-test is based on:

$$(\hat{\theta}_1 - R_0\hat{\theta}_2)/\mathrm{Var}[\hat{\theta}_1 - R_0\hat{\theta}_2]^{1/2} \quad (2)$$

The rejection region for this test is the set of values r satisfying:

$$(\hat{\theta}_1 - r\,\hat{\theta}_2) > t\mathrm{Var}[\hat{\theta}_1 - r\,\hat{\theta}_2]^{1/2} \quad (3)$$

Finding an explicit form for the confidence interval requires solving a quadratic equation in r. The confidence interval can be of the form (L, U), $(U, +\infty)$ $(0, U)$ or $(0, +\infty)$ depending on the number of solutions of the quadratic equation (Raftery and Schweder, 1993; Buonaccorsi and Iyer, 1984; Franz, 2007; Von Luxburg and Franz 2009; Hirschberg and Lye 2010).

The Delta method is based on a Taylor series approximation of:

$$\hat{R} = \hat{\theta}_1 / \hat{\theta}_2 \quad (4)$$

Around θ_1/θ_2 that is used to obtain estimates of the expectation and of the variance of $\hat{R}$ (Casella and Berger 2002; Faraggi et al., 2003; Franz, 2007; Hirschberg and Lye 2010). Assuming that $\hat{R}$ follows a normal distribution, the $1-\alpha$ confidence interval is obtained as $\hat{R} \pm z \times \mathrm{s.d.}$ where z is the upper $\alpha/2$ quantile of the standard normal distribution.

Standardised mortality ratios

For combined toxicity data reporting mean mortality or survival probability expressed in % of individual bees, the standardised mortality ratios (SMR) has been estimated as the ratio or percentage change in observed deaths compared to that occurring after exposure to the single compound. An SMR above 1 is simply interpreted as a higher number of observed deaths compared to the group exposed to the single compound (Everitt and Skrondal, 2010).

Toxic Unit Approach

Analysis of each experimental binary mixture for acute and chronic (contact or oral) toxicity was conducted using the Toxic Unit approach to standardise applied dose and critical endpoints (i.e. LD_{50}) using matching datasets for each chemical from OpenFoodTox and other databases. The toxic unit approach assumes that mixture toxicity predictions follow the Dose/Concentration Addition (DA/CA) model given the quantitative composition of each chemical within the mixture in relation to their relative potency (Jonker et al., 2005). Toxic Unit for chemical B (i.e. TU_B) is given as the ratio of the dose/concentration of chemical B applied in the binary mixture experiment relative to the selected critical endpoint (e.g. LD_{50}) used as reference as follows:

$$\mathrm{TU_B} = \frac{\text{Applied Dose}_{\,B}}{\text{Critical Endpoint}_B} \quad (5)$$

$TU_B = 0.1$ indicates that the dose of compound B applied in the mixture assay corresponds to 10% of the LD_{50} or LC_{50}. The expected combined potency of the mixture relative to a given acute (e.g.

LC_{50}, LD_{50}) or chronic (e.g. long-term NOEC) toxicological endpoint is also named "mixture strenght" or mixture potency symbolised as "TUm" (More et al., 2019) according to equation 6 (Jonker et al., 2005):

$$\mathrm{TUm} = \sum_{i=1}^{n} \frac{C_i}{ECx_i} \quad (6)$$

Effect Concentrations (ECx_i) relate to the critical endpoint selected as reference, and Concentration (Ci) refers to the concentration of the chemical (i) in the mixture. Consequently, while assuming CA as the default reference model, TUm is calculated by summing the individual TU_i values for each compound present in the mixture (binary, ternary or with more components) (SCCS, SCENHIR and SCHER, 2012; More et al., 2019) as follows:

$$\mathrm{TUm} = \sum_{i=1}^{n} \mathrm{TU}_i \quad (7)$$

A mixture with a TUm = 1 would be expected to produce the effect used as the critical endpoint in the TU calculations (e.g. EC_{10} => 10% effect, EC_{50} => 50% effect, LC_{50} => 50% lethality).

In addition, individual TU_B values were ranked into three classes in comparison with their corresponding EMR to plot and quantify the relative contribution of compound B (TU_B) to the overall combined mixture (TUm) (see results, 3.2.2):

- $TU_B \leq 0.10$
- $0.11 \leq TU_B \leq 0.30$
- $0.31 \leq TU_B \leq 0.60$

According to each TU_B class, the distribution of the EMR values against their "reverse cumulative frequency" has been plotted and fits were tested with Pearson product-moment correlation coefficient (R^2). This allowed quantifying the contribution of chemical B to the combined toxicity of the binary mixture.

2.2.3 Predictive models for combined toxicity and Model Deviation Ratios

For each individual binary mixture, predictive models of combined toxicity were compared to the experimental dose response data to assess deviation from DA/CA, i.e. interactions synergism, potentiation or antagonism. The DA/CA model assumes that the chemicals have a similar Mode of Action (MoA) in the mixture and they do not interact with each other, thus that they do not influence each other's uptake, distribution or metabolism at the site of the biological target (Faust et al., 2003; Jonker et al., 2005, Cedergreen et al., 2014, 2012; Backhaus et al., 2004, 2013). Therefore, if a mixture of *n* chemicals with TUm = 1 results in an *x%* (i.e. the selected critical endpoint reference value) effect compared to the control response, then the mixture is acting according to DA/CA as the following relationship holds:

$$\mathrm{TUm} = \sum_{i=1}^{n} \frac{C_i}{ECx_i} = 1 \quad (8)$$

Where *Ci* represent the concentration of chemical *i* in the mixture and ECx_i is the effect concentration of chemical *i* that results in the same effect ($x\%$) as observed in the mixture. However, as full dose response data are rarely reported in the literature it is difficult to derive all EC_x values to test mixtures yielding effects of different intensity (e.g. 10, 20, 50, and 80 %). Hence, because the most commonly reported critical endpoint values usually refer to 50% effects for both single chemicals and mixtures, the expected TUm for a mixture observed to give 50% mortality under CA would be TUm = 1, when the TU_i values are using the LC_{50} values of the individual chemicals as reference values. Based on the availability of critical endpoints from the data collection, EC_x in equation 6 and 8 were substituted with LC_{50} or LD_{50} values to quantify how well the observed effects fit the CA predictions for the binary mixture toxicity in bee species.

The magnitude of the deviation between the concentration addition-predicted model (predicted TUm) and the experimental data (observed TUm) was calculated as model deviation ratio (MDR) based on the TUm values according to Belden et al., (2007) using Observed TU values calculated as TUm in the mixture (50% mortality) compared to that from the expected TUm value of a mixture causing 50% lethality as TU of 1 as follows:

$$\mathrm{MDR} = \frac{\text{predicted TU}_m}{\text{observed TU}_m} \quad (9)$$

Here, MDR values (equation 9) represent the ratio between the expected or "predicted TUm" for a binary mixture causing 50% mortality (by definition a TUm = 1) (equation 8), and the "observed TUm" (equation 6) calculated as TUm causing 50% mortality (Belden et al., 2007; Coors and Frische 2011; Cedergreen et al., 2013, 2012). Thus, MDR values above 1 indicates toxicity above that expected from CA predictions, and MDR values below 1 indicates toxicity below that expected from CA predictions. According to the current scientific literature (Belden et al., 2007; Cedergreen, 2014), biologically significant synergism has been defined for a range of species as a deviation from CA superior to two-fold. As a consequence, mixtures are usually termed additive for $0.5 \leq \mathrm{MDR} \leq 2$, antagonistic for MDR values < 0.5 and synergistc for MDR values > 2 (Belden et al., 2007; Cedergreen, 2014). In our analysis, besides applying the MDR approach to characterise mixture effects, the statistical significance of the combined toxicity was assessed and calculated using non-overlapping 95% confidence intervals (i.e. 95% CI of the EM_A vs 95% CI of the EM_M for chemical A+B) as described by Johnson et al., (2012, 2013). From this analysis of statistical significance, MDR thresholds were refined as follows:

- MDR values between 0.83 and 1.25 indicate that combined toxicity follows **DA/CA** with observed TUm values deviating less than 1.5-fold from the expected TUm of 1.
- MDR values < 0.83 indicate that combined toxicity is below that predicted from CA and classified as **antagonism;**
- MDR > 1.25 indicates that combined toxicity is above that predicted from CA and classified as **synergism.**

2.2.4 Comparison of Estimated Mean Ratios and Model Deviation Ratios

A polynomial regression model (with formula y ~ x + I(x^2)+ I(x^3)) was fitted between the EMR from the individual dose response data and the corresponding individual MDR to assess the potential correlation between the two approaches by means of a Pearson product-moment correlation coefficient (R^2) (see result 3.2.4). R software has been used and R-script is provided in supplementary materials.

3. Results and Discussion

3.1 Extensive Literature searches

The results of the extensive literature search on combined toxicity of chemicals in honey bees and wild bees (solitary bees and bumble bees) are illustrated in figure 3.1 as a PRISMA flow diagram. 957 peer-reviewed articles were initially identified from the literature with total of 14 papers matching the inclusion criteria with relevant data providing a total of 218 binary mixtures (Moher et al., 2009) resulting from *in vivo* experimental laboratory studies. Overall, most publications (n= 10) reported mortality data in honey bees (*Apis mellifera*) for binary mixtures of pesticides with dose response data available for a total of 92 individual binary mixtures (Johnson et al., 2013, 2012, 2009, 2006; Zhu et al., 2017; Guseman et al., 2016; Rinkevich et al., 2015; Wilkinks et al., 2013; Iwasa et al., 2004; Ellis et al., 1997;). Similarly, four peer-reviewed articles provided relevant data for binary mixture toxicity in wild bees (*Bombus* spp.) and solitary bees (*Osmia* spp.) (Robinson et al., 2017; Sgolastra et al., 2017; Spurgeon et al., 2016; Biddinger et al., 2013). Finally, studies on chemical-non-chemical interactions were provided in two peer-reviewed articles and were excluded from the analysis (Tosi et al., 2017; Alaux et al., 2010).

Overall, toxicity data were mostly available for acute contact toxicity i.e. topical application (61%) with few studies reporting chronic oral effects (20%) or acute oral toxicity data (19%) as highlighted in a recent meta-analysis (Quignot et al., 2015). The rationale behind such findings lies in the fact that acute toxicity tests (24 and 48h) for honey bees are usually applied in the area of chemical risk assessment for regulated products such as pesticides. However, honey bees are exposed chronically to a range of chemicals (both alone and in combination), either by foraging on

contaminated areas, or through contaminated food, stored and consumed in the hive (EFSA, 2013a; EFSA AHAW Panel, 2016). Recently, the OECD (Organisation for Economic Co-operation and Development) proposed a new guideline (OECD, 2017) for chronic oral toxicity tests (10-days feeding test in the laboratory).

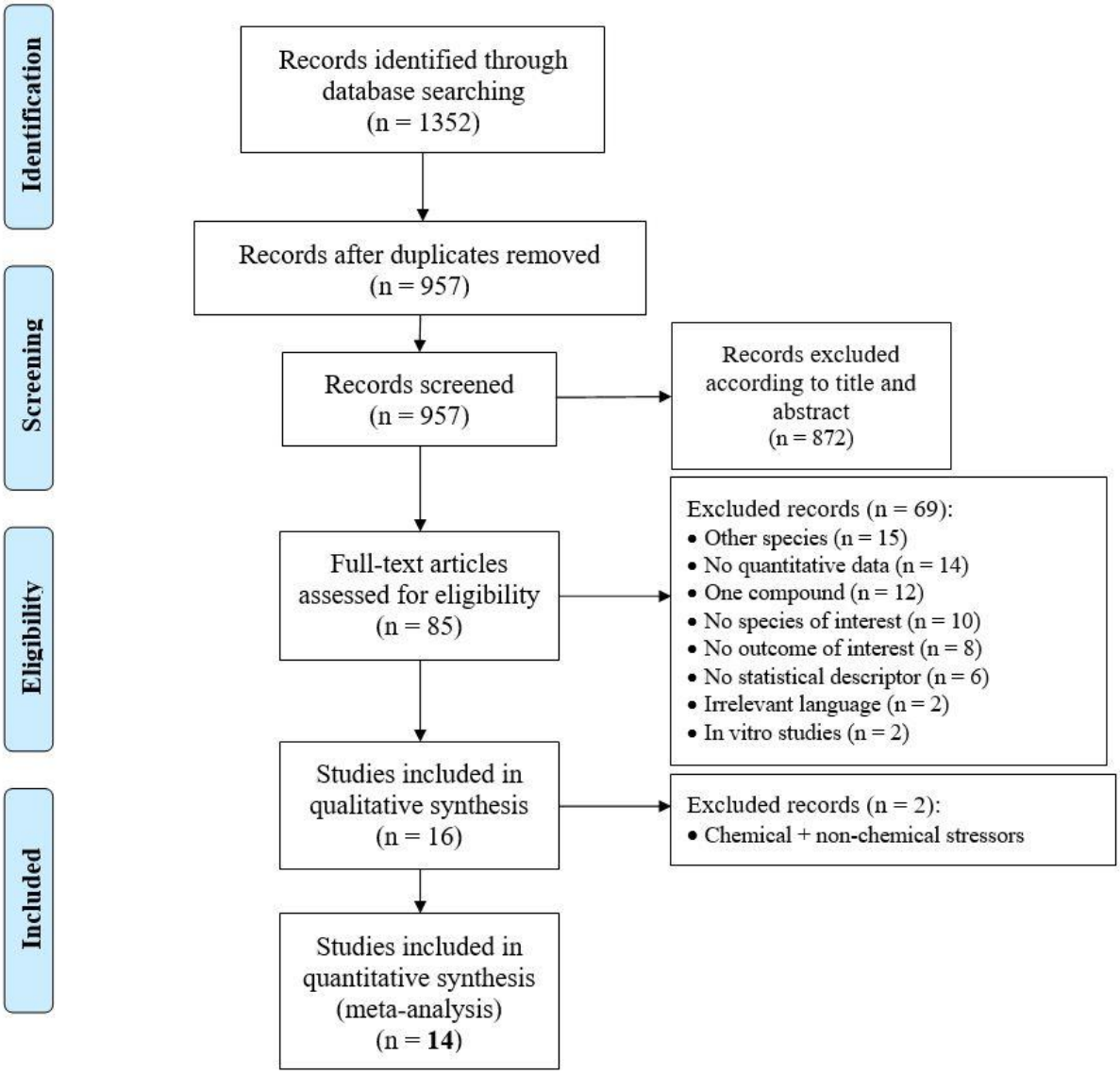


Figure 3.1. PRISMA 2009 Flow Diagram for the extensive literature searches on combined toxicity of binary mixtures in bee species.

From the extensive literature search, data for 51 chemicals, the vast majority as pesticides, were identified and their corresponding Modes of Action (MoA) were analysed for their pesticidal MoA for 23 insecticides and 16 fungicides respectively as well as MoA in honey bees as non-target species (Table 3.2). For the toxicological MoA in honey bees, classifications schemes from the Insecticide Resistance Action Committee (IRAC) and the Fungicide Resistance Action Committee's (FRAC) covering the specific target sites in target organisms for insecticides, acaricides and fungicides and

the scientific literature were reviewed (Leroux et al., 2008; Hermann and Stenzel, 2019; Sanchez-bayo, 2012; Johnson et al., 2012; 2013; Huang et al., 2013; de Castro et al., 2015; Sparks and Nauen, 2015). In this context, pyrethroids/pyrethrins insecticides and conazole fungicides were the most investigated pesticides (≈55%) belonging to the MoA groups of "sodium channel modulators" (≈25%) and "demethylation inhibitors" (≈30%) respectively, and amongst conazoles, triazole fungicides (Demethylation Inhibitors) provided the largest experimental datasets for binary mixtures in honey bees. Similarly, the combined exposure to neonicotinoid insecticides (Nicotinic acetylcholine receptor agonists) and conazole fungicides were the second most investigated mixtures (35%).

Table 3.2. Overview of xenobiotics with available binary mixture toxicity data in bees: class, chemical group and Mode of action (MoA) (Leroux et al., 2008; Hermann and Stenzel, 2019; Sanchez-bayo, 2012; Huang et al., 2013; Sparks and Nauen, 2015; Johnson et al., 2012, 2013; de Castro et al., 2015).

Compounds	**Class of compounds**	**Chemical group**		**MoA**
		Name	**Code [1]**	
Amitraz	Insecticide	Amitraz/formamidine	19	Octopamine receptor agonists (Nerve action)
Carbaryl	Insecticide	Carbamates	1A	Acetylcholinesterase (AChE) inhibitors (Nerve action)
Oxamyl				
Acephate	Insecticide	Organo-phosphates	1B	Acetylcholinesterase (AChE) inhibitors (Nerve action)
Coumaphos				
Dimethoate				
Fenpyroximate	Insecticide	METI acaricides and insecticides	21A	Mitochondrial complex I electron transport inhibitors (Energy metabolism)
Aldrin	Insecticide	Organo-chlorine	2A	GABA-gated chloride channel antagonists
Dieldrin				

Compounds	Class of compounds	Chemical group		MoA
		Name	Code [1]	
Bifenthrin **Cyfluthrin** **Fluvalinate** **Lambda-cyhalothrin** **Tau-fluvalinate** **Phenothrin**	Insecticide	Pyrethroids/Pyrethrins	3A	Sodium channel modulators (Nerve action)
Acetamiprid **Clothianidin** **Imidacloprid** **Thiacloprid** **Thiamethoxam**	Insecticide	Neonicotinoid	4A	Nicotinic acetylcholine receptor (nAChR) agonists (Nerve action)
Sulfoxaflor	Insecticide	Sulfoximines	4C	Nicotinic acetylcholine receptor (nAChR) agonists (Nerve action)
Oxalic acid	Insecticide	Natural insecticide	NA	NA
Azoxystrobin	Fungicide	Methoxy-acrylates (Strobilurin)	C3	Complex Ⅲ: cytochrome bc1 (ubiquinol oxidase) at Qo site (cyt b gene) (Respiration)
Boscalid	Fungicide	Pyridine-carboxamides	NA	Complex Ⅱ: succinate-dehydrogenase (Respiration)

Compounds	Class of compounds	Chemical group		MoA
		Name	Code [1]	
Metconazole **Epoxiconazole** **Fenbuconazole (Indar)** **Metconazole** **Myclobutanil** **Propiconazole** **Tebuconazole** **Tetraconazole** **Triadimefon** **Uniconazole-P**	NA	Fungicide Triazoles		C14- demethylase in sterol biosynthesis (erg11/cyp51) DMI-fungicides (DeMethylation Inhibitors)
Triflumizole **Prochloraz**	Fungicide	Imidazoles	NA	C14-demethylase in sterol biosynthesis (erg11/cyp51) DMI-fungicides (DeMethylation Inhibitors)
Chlorothalonil	Fungicide	Chloronitriles	NA	Multi-site contact activity

Compounds	Class of compounds	Chemical group		MoA
		Name	Code [1]	
Pyraclostrobin	Fungicide	methoxy-carbamates	C3	complex Ⅲ: cytochrome bc1 (ubiquinol oxidase) at Qo site (cyt b gene) QoI-fungicides (Quinone outside Inhibitors)
Glyphosate	Herbicide	Organophosphorus	NA	Enzyme inhibitor (it disrupts the shikimic acid pathway through inhibition of the enzyme 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase)
Diethyl maleate (DEM)	Chemical/ Synergist	NA	NA	Enzyme inhibitor Glutathione S-transferase (GST) inhibitor
Piperonyl butoxide (PBO)	Chemical/ Synergist	Cyclic aromatic	27A	Enzyme inhibitor Blocks pests natural detoxification system (P450-dependent monooxygenase inhibitor)
S,S,S-tributyl phosphorotrithioate (DEF)	Chemical/ Synergist	Organo-phosphorus	NA	Carboxylesterase inhibitor
Fumagillin	Veterinary products/drug	Antimicrobial agent	NA	Enzyme inhibitor (methionine aminopeptidase2 - MetAP2)
Ivermectin	Veterinary product/drug	Avermectins	NA	Receptor disrupter (γ-aminobutyric acid receptors, GABA-R)

Compounds	Class of compounds	Chemical group	MoA	Compounds
		Name	Code [1]	
Oxytetracycline	Veterinary products/drug	Antibiotic	NA	NA
Phenobarbital	Chemical	Barbituric acid derivate	NA	Receptor disrupter (γ-aminobutyric acid receptors, GABA-R)
Quercetin	Flavonoid	Flavonoid (polyphenol)	NA	Mammalian P-glycoprotein inhibitor
Salicylic acid	Acaricide (organic)	NA	NA	Cox inhibition Anti-inflammatory
Thymol	Veterinary product/drug	Monoterpenoid phenol	NA	Ergosterol biosynthesis disrupter
Tylosin	Veterinary product/drug	Antimicrobial agent	NA	Bacteriostatic
Verapamil	Drug	NA	NA	P-glycoprotein transport modulator
Xanthotoxin	Chemical	Furanocoumarin (produce by plants)		Enzyme inhibitor (xenobiotic-metabolizing P450s)

[1]= Code of chemical group name according to IRAC/FRAC classification shemes.

3.2 Data collection and analysis

3.2.1 Data Collection

218 individual binary mixtures were collected and included in the statistical analyses with the majority of toxicological endpoints reported as lethal doses or concentrations (e.g. LD_{50}, $LC_{50,}$) for pesticides or pesticides and veterinary drugs combinations with 133, 44 and 41 mixtures reporting acute contact toxicity (i.e. topical application), chronic oral toxicity and acute oral toxicity, respectively (tables S3.1, S3.2, S3.3, S3.7). Combined toxicity data for binary mixtures were

available as dose response data in honey bees species (Johnson et al., 2009, 2012, 2013) for acute contact toxicity (n= 92) and acute oral toxicity (n= 15) (tables S3.3, S3.7). All toxicity data are available as spreadsheets on EFSA knowledge junction under the DOI: 10.5281/zenodo.3383713 and as summary tables in supplementary materials (Tables S3.1 – S3.11) classified according to route and exposure patterns (i.e. oral, contact, acute and chronic) and toxicological endpoints (e.g. LD_{50}, LC_{50}) for the honey bees (*Apis mellifera*) and wild bee species *(Osmia bicornis, Bombus terrestris*). All the studies included in this meta-analysis were performed *in vivo* experimental laboratory tests according to standard toxicity tests as provided by the author(s). In honey bees, acute contact toxicity studies refer to "topical application" and toxicity tests were mainly conducted on group feeding tests (Johnson et al., 2013, 2009, 2006; Iwasa et al., 2004; Biddinger et al., 2013). Similarly, acute oral studies on honey bees and bumble bees were conducted on group feeding tests through consumption of contaminated food (e.g. nectar, pollen) (Robinson et al., 2017; Johnson et al., 2012). In contrast, toxicity studies on solitary bees such as *Osmia* spp. were conducted on individual feeding tests (Sgolastra et al., 2017; Biddinger et al., 2013).

3.2.2 Quantification of magnitudes of interaction

Estimated Mean Ratios

EMRs were calculated to characterise the magnitude of the combined toxicity for each individual binary mixture and expressed on a harmonised scale starting at 1 to reflect changes in the toxicological endpoint (EM_M) either as an increase (+) or a decrease (-) in combined toxicity (Quignot et al., 2015).

Acute contact toxicity

The acute contact toxicity database represented the largest database in honey bees with 133 LD_{50} for binary mixtures including dose response data (n = 92) (Tables S3.1-S3.3 and S3.15). A comprehensive analysis of the database provided an analysis of Toxic units below, prediction of combined toxicity and calculation of MDRs in section 3.2.2 and comparison of EMRs and MDRs in section 3.2.3. Overall, EMRs for binary mixtures reflecting statistically significant interactions (non-overlapping 95% CI) were highest (>100) for honey bees exposed to neonicotinoid insecticides (e.g. acetamiprid, thiacloprid) combined with cytochrome P450 (CYP) inhibitors (e.g. triazole fungicides such as propiconazole) and synergists (e.g. piperonyl butoxide (PBO) (table S3). Examples include EMRs of 1980 for pyrethroid tau-fluvalinate and prochloraz (TU_B= 0.07) and PBO (TU_B= 0.03) as well as EMRs of 235- and 101-fold for the neonicotinoid acetamiprid-triflumizole (TU_B= 0.50) and acetamiprid-propiconazole (TU_B= 0.10) (Table S3). In contrast, reduced combined toxicity through antagonistic interactions were also observed in a few instances (e.g. amitraz-oxalic acid, 4-fold) (Table S3).

Acute oral toxicity

EMR values were calculated for the 41 LC_{50} binary mixtures available for pesticides and veterinary drugs (Tables S3.4, S3.5, S3.6, S3.7 and S3.8). For honey bees, EMR values reflecting increase in combined toxicity were statistically significant for tau-fluvalinate (pyrethroid) with xanthotoxin (furanocoumarin produced by plants) with a value of ≈200 (Johnson et al., 2012), phenobarbital with lambda-cyhalothrin, aldrin and dieldrin with a 2.8-, 1.6- and 1.8-fold increase in combined toxicity respectively (Table 3.3). For veterinary products, EMRs also showed an increase in combined toxicity for ivermectin ($p<0.0001$) with verapamil (EMR= 4.1) > quercetin (EMR= 2.6) > fumagillin (EMR= 1.8) (Guseman et al., 2016) (table S5 and S6). In contrast, a slight decrease (1.5-1.7 fold) in combined toxicity of fenpyroximate (METI-acaricide) with oxytetracycline and fumagillin (veterinary products) was observed (tables S3.8).

Sgolastra et al., (2017) investigated combined toxicity, expressed as standardised mortality ratios (SMR), after exposure to binary mixtures of pesticides in three bee species (*A. mellifera, B. terrestris, O. bicornis*) at different time points (table S7) and found significant synergistic mortality in all species exposed to non-lethal doses of propiconazole (TU_B=0.07) and respective LD_{10} of the neonicotinoid insecticide clothianidin (TU_A=0.10). Such a significant increase in combined toxicity was measured for acute time points in *A. mellifera* (4h and 24h) and *B. terrestris* (4h), these persisted throughout the experiment (96h) in *O. bicornis*. Overall, SMR the magnitudes of synergism ranged from 4.4-fold in *A. mellifera* (at 24h) to 8.7 in *O. bicornis* (at 4h) (table S7).

Table 3.3. Ranking of combined toxicity for binary mixtures of pesticides and veterinary drugs in honey bees (expressed as LD_{50} µg/bee) following acute oral exposure (Johnson et al., 2013; 2012; Wilkins et al., 2013). Estimated Mean Ratio (EMR) for the binary mixture (chemical A + B) relative to chemical A alone as well as Confidence Interval (CI 95%) for EMR are provided.

Study_ID	**Chemical A**		**Binary Mixture (A+B)**								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95th)	CI_2 (95th)	EMR (+)	EMR (-)	CI_{EMR}	Slope (±SE)
Study_165	Tau-Fluvalinate[2]	8.1 (7.2-9.0)	Xanthotoxin	NA	0.04	0.001	0.13	201		na	0.3±0.09
Study_164	Tau-Fluvalinate[2]	8.1 (7.2-9.0)	Phenobarbital	NA	0.19	0.12	0.31	42*		20.7-64.0	1.5±0.12
Study_169	lambda-cyhalothrin[2]	0.048 (0.034-0.068)	Phenobarbital	NA	0.02	0.005	0.025	2.8*		0.67-4.1	2.9±0.4
Study_151	Tau-fluvalinate[1]	9.2 (7.9-10.8)	Fumagillin	NA	4.8	3.7	6.32	1.9a		1.3-2.5	2.0±0.22

Study_ID	Chemical A		Binary Mixture (A+B)								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95th)	CI_2 (95th)	EMR (+)	EMR (-)	CI_{EMR}	Slope (±SE)
Study_167	Tau-Fluvalinate[2]	8.1 (7.2-9.0)	Salicylic acid	NA	4.5	2.2	8.6	1.8		0.49-3.1	1.6±0.33
Study_171	Dieldrin[2]	0.037 (0.032-0.047)	Phenobarbital	NA	0.02	0.01	0.03	1.8*		0.86-2.9	3.5±0.30
Study_170	Aldrin[2]	0.061 (0.0527-0.071)	Phenobarbital	NA	0.04	0.03	0.05	1.6*		1.1-2.0	3.9±0.36
Study_163	Thymol[1]	38.1 (27.3-49.6)	Fumagillin	NA	25.3	21.3	29.7	1.5		0.99-2.0	4.0±0.41
Study_158	Amitraz[1]	5.47 (4.12-7.1)	Oxytetracycline	NA	3.7	3.0	4.7	1.5		0.96-2.0	4.2±0.54
Study_160	Amitraz[1]	5.5 (4.1-7.1)	Fumagillin	NA	3.9	2.9	5.2	1.4		0.85-2.0	3.8±0.52
Study_161	Thymol[1]	38.1 (27.3-49.6)	Oxytetracycline	NA	27.5	15.4	45.1	1.4		0.53-2.2	3.6±0.85
Study_152	Coumaphos[1]	26 (19.5-39.5)	Oxytetracycline	NA	20	15.1	27.6	1.3		0.65-1.9	2.5±0.36
Study_159	Amitraz[1]	5.5 (4.1-7.1)	Tylosin	NA	4.5	3.8	5.3	1.2		0.82-1.6	3.3±0.29
Study_162	Thymol[1]	38.1 (27.3-49.6)	Tylosin	NA	32.3	14.5	47.7	1.2		0.48-1.9	3.1±0.75
Study_153	Coumaphos[1]	27 (19.5-39.5)	Tylosin	NA	25.7	17.9	43.0	1.		0.38-1.6	3.5±0.72
Study_149	Tau-fluvalinate[1]	9.2 (7.95-10.8)	Oxytetracycline	NA	8.4	7.3	9.8	1.[b]		0.85-1.3	2.7±0.21
Study_155	Fenpyroximate[1]	3.2 (2.7-3.9)	Oxytetracycline	NA	4.7	3.9	5.7	0.69[a]	1.5	1.1-1.8	3.5±0.37
Study_157	Fenpyroximate[1]	3.24 (2.7-3.9)	Fumagillin	NA	5.5	4.4	6.9	0.59[a]	1.7	1.2-2.2	2.8±0.32
Study_168	Tau-Fluvalinate[2]	8.1 (7.2-9.0)	Indole-3-carbinol	NA	8.3	5.9	10.9	0.97	1.03	0.70-1.4	2.5±0.67
Study_150	Tau-fluvalinate[1]	9.2 (7.9-10.8)	Tylosin	NA	10.5	8.1	14.9	0.88	1.1	0.73-1.6	2.3±0.34
Study_156	Fenpyroximate[1]	3.24 (2.7-3.9)	Tylosin	NA	4.1	3.6	4.6	0.80	1.3	0.97-1.5	2.6±0.18
Study_154	Coumaphos[1]	28.0 (19.5-39.5)	Fumagillin	NA	33.3	25.5	49.2	0.78	1.3	0.60-1.9	2.1±0.28
Study_166	Tau-Fluvalinate[2]	8.1 (7.2-9.0)	Quercetin	NA	11.4	9.7	13.9	0.71	1.4	1.1-1.7	3.0±0.40

1= Johnson et al., 2013 (tau-fluvalinate + sucrose): significant differences compared to the respective treatment are indicated with a superscript letter "a" = significant pre-treatment effect, "b" = significant pre-treatment*acaricide dose effect. 2= Johnson et al., 2012: treatments with non-overlapping 95% confidence interval are considered significantly different. CI= 95% Confidence Interval.

EMR values for sub-chronic (LC_{50} 96h) and chronic (LC_{50} 240h) mortality (n= 44) after exposure to pesticide binary mixtures are provided in supplementary materials (table S9, S10, S11, S12 and S12;). Overall, EMRs for subchronic toxicity increased by a maximum of 1.5-fold in *A. meliferra* and *B. terrestris* whereas an EMR of 8.6-fold was reported for *O. bicornis* (Sgolastra et al., 2017) (Table S7). Chronic oral toxicity (LC_{50} 96h, 240h) of binary mixtures of pesticides (in *Apis mellifera*, *Bombus terrestris* and *Osmia bicornis*) showed an increase in toxicity with exposure time for all tested chemicals (n=6) (Table S9; Figure 3.2). In particular, effects on mortality increased from the 48h time interval until 240 h exposure time by 1.3- 1.6-fold in *B. terrestris* and *O. bicornis,* respectively (Robinson et al., 2017). Combined toxicity of tau-fluvalinate (pyrethroid) and propiconazole (SBI fungicide) showed potentiation via inhibition of metabolism by SBI fungicides (Berenbaum and Johnson 2015; Han et al., 2019). However, potentiation effects between SBI fungicides and clothianidin were not observed, and recent findings demonstrated that the expression of clothianidin induces the *CYP9q1* detoxification gene by (Yao et al., 2018). Zhu et al., (2017) show that none or very minor additive toxicity was for 5 binary mixture of imidacloprid and other pesticides (i.e. lambda-cyhalothrin, oxamyl, tetraconazole, glyphosate, sulfoxaflor) at concentrations similar to the residue levels detected in honey bee hives (i.e. field concentration) (Table S10). However, the author did not exclude that synergism may occur under other exposure situations particularly at higher concentrations or with different proportions of individual chemicals.

Combined chronic sub-lethal effects of coumaphos (organo-phosphate acaricide) and prochloraz (imidazole fungicide) in honey bee workers were investigated with regards to the molecular immune response at different developmental stages (prepupa, white-eyed pupa, adult) (Cizelj et al., 2016). Changes in mRNA level associated with upregulation of a range of genes (e.g. abaecin, defensin-1, cactus and basket) were reported for prochloraz and coumaphos. In addition, our results on mortality data suggest that an increased toxicity (EMR= 70) is observed when adult bees are exposed to coumaphos-prochloraz mixture, thus highlighting strong synergistic effects (MDR = 12.5) (Table S3; Figure 3.2).

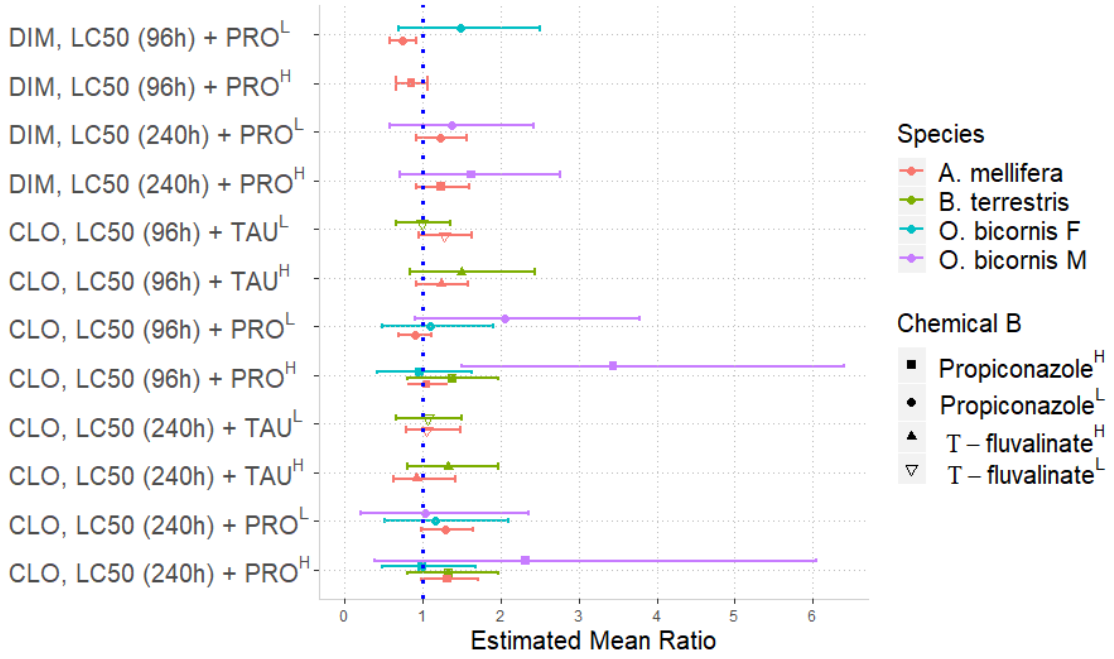


Figure 3.2. Estimated Mean Ratio (EMR) following chronic oral exposure to binary mixtures in different bee species i.e. A. mellifera, B. terrestris, and O. bicornis female (F) or male (M). EMRs (dots) and related 95% CI (lines) were reported with different shapes according to the chemical B (see legend) investigated in the assay (Spurgeon et al., 2016; Robinson et al., 2017). Chemicals are reported as follows: CLO = Clothianidin; DIM = Dimethoate; PRO = Propiconazole; TAU = Tau-fluvalinate. Dose of chemical B is reported according to the author "high" (H) or "low" (L).

Toxic Unit approach

Acute and chronic oral toxicity of multiple stressors

As described in the method section (2.2.2), the Toxic Unit (TU) approach has been applied to quantify potency of the binary mixture A+B (TUm) versus compound A (TU_A). In addition, the dose of compound B in each experiment (TU_B) has been estimated using matching potency information ($LC_{50\text{-}B}$) binary mixtures from available databases (EFSA PPR, 2012; SCCS, SCENHIR and SCHER, 2012; More et al., 2019). From the data available, this analysis could only be conducted for acute contact toxicity binary mixtures since no matching datasets for compound B were available for acute and chronic oral toxicity studies (Table S13, S14 and S15). OpenFoodTox and other databases (e.g. US-EPA, OECD e-chem portal, PPDB-Pesticide Properties Database, literature) provided 85% and 15% of values for compound B respectively. This approach is first described for available classes of chemicals namely a. insecticides-P-450 inhibitors and synergists, b. acaricides and insecticides, c. whole database. All individual binary mixtures data and summary statistics are available in supplementary material (tables S3.1 – S3.3, S3.13 and S3.15).

a) Insecticides-P-450 inhibitors (conazole fungicides, synergists)

Binary mixture experiments between insecticides and P450 inhibitors (e.g. conazole fungicides or synergists such as piperonyl butoxide - PBO) were the most investigated (55%) (Iwasa et al., 2004; Biddinger et al., 2013; Johnson et al., 2013; Spurgeon et al., 2016). For insecticides-conazole fungicides, the largest EMR values were observed for the pyrethroid insecticide tau-fluvalinate with prochloraz (≈1980 with TU_B=0.07), thiacloprid-triflumizole (EMR≈1460, TU_B=0.50), neonicotinoids acetamiprid and thiacloprid with propiconazole (EMR of 101 and 490, respectively with TU_B=0.07) (Johnson et al., 2013; 2012; Iwasa et al., 2004) (Table S3; Figure 3.3). Although prochloraz had the lowest concentration within the mixture, it showed the highest synergistic effects (MDR= 20) when combined with tau-fluvalinate (LD_{50}= 19.8 µg/bee) (figure 3.3). In contrast, very low doses of azole fungicides showed a slight antagonist effect of 1.5-fold on pyrethroids (tau-fluvalinate with propiconazole TU_B= 0.0003 or myclobutanil TU_B= 0.001) (Johnson et al., 2013) (Table S3; Figure 3.3). In addition, dose response data for the combined toxicity of tau-fluvalinate with myclobutanil (TU_B= 0.001, 0.01 and 0.07) and propiconazole (TU_B= 0.0003, 0.003 and 0.03) are illustrated following TU_B variation (figure 3.3). Insecticides and synergists such as tau-fluvalinate-PBO showed the highest EMR (≈ 1980) and MDR (≈32), thus demonstrating strong synergistic effects even at low doses of PBO (TU_B= 0.03) (Figure 3.4; Table S3 and S15). EMRs for tau-fluvalinate, lambda-cyhalothrin and cyfluthrin with PBO at higher dose (TU_B= 0.34) were also large and very significant (≈ 945, 78 and 30 respectively) (Table S3 and S15; Figure 3.4). It is interesting to note that when the three pyrethroids were tested without PBO, cyfluthrin shows the highest toxicity (LD_{50} 0.062 µg/bee) whereas tau-fluvalinate the least (LD_{50} 9.45 µg/bee) (table S3). Our results confirm that the differential synergistic effects observed amongst the three pyrethroids is likely to be due to esterases acting on the acid moiety (Johnson et al., 2006). Indeed, tau-fluvalinate has an aromatic acid group, so that it is not sequestered as readily as the other pyrethroids and shows the greatest magnitude of synergism (Moores et al., 2012; Gunning et al., 2007). Lower magnitude of interactions were shown for Cyfluthrin (EMR= 2.3) and S,S,S-tributyl phosphorotrithioate (DEF) (EMR= 30) with PBO (TU_B= 0.34) (Table S3 and S15). Similarly, combined toxicity of lambda-cyhalothrin with diethyl maleate (DEM) (EMR≈ 3) and PBO (TU_B= 0.34) (EMR ≈ 80) indicated greater synergism in the presence of PBO. The scientific basis for such interactions is of metabolic nature since PBO is a potent CYP inhibitor and DEF inhibits carboxylesterases (Johnson et al., 2013; 2015; Mao et al., 2017; Wu et al., 2007).

Overall, hymenoptera are known to have a specific metabolic profile with the lowest copy number of detoxification enzymes within the insect kingdom (Johnson et al., 2013; 2015; EFSA, 2013a). In particular, honey bees have one of the lowest numbers of CYP genes isoforms of any invertebrate sequenced (46 sequences). Therefore, our results confirm that sterol biosynthesis–

inhibiting (SBI) fungicides inhibit the CYP-mediated detoxification of some pyrethroids (e.g. tau-fluvalinate) and neonicotinoids (e.g. imidacloprid), thus increasing the acaricide and insecticide toxicity to bees, respectively (Wade et al., 2019; Pilling et al., 1995; Iwasa et al., 2004; Johnson et al., 2013).

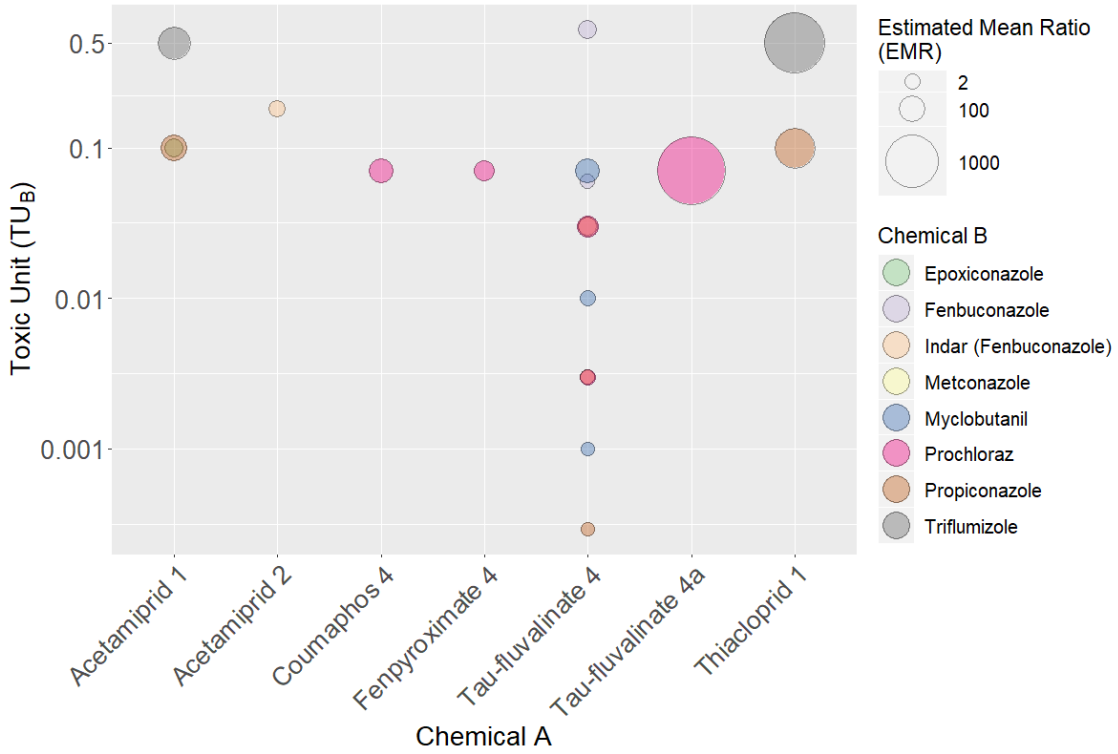


Figure 3.3. Bubble plot for acute contact toxicity of insecticides (chemical A) and conazole fungicides (chemical B) in honey bees: Estimated Mean Ratios (EMR) (A+B) and experimental potency-adjusted dose (chemical B: Toxic Unit - TU_B). Size of the bubble is proportional to the value of the EMR. Colours represent different chemicals as reported in the legend. 1= Iwasa et al., (2014). 2= Biddinger et al., (2013). 4, 4a= Johnson et al., (2013). All the studies were statistically significant according to non-overlapping 95% confidence intervals.

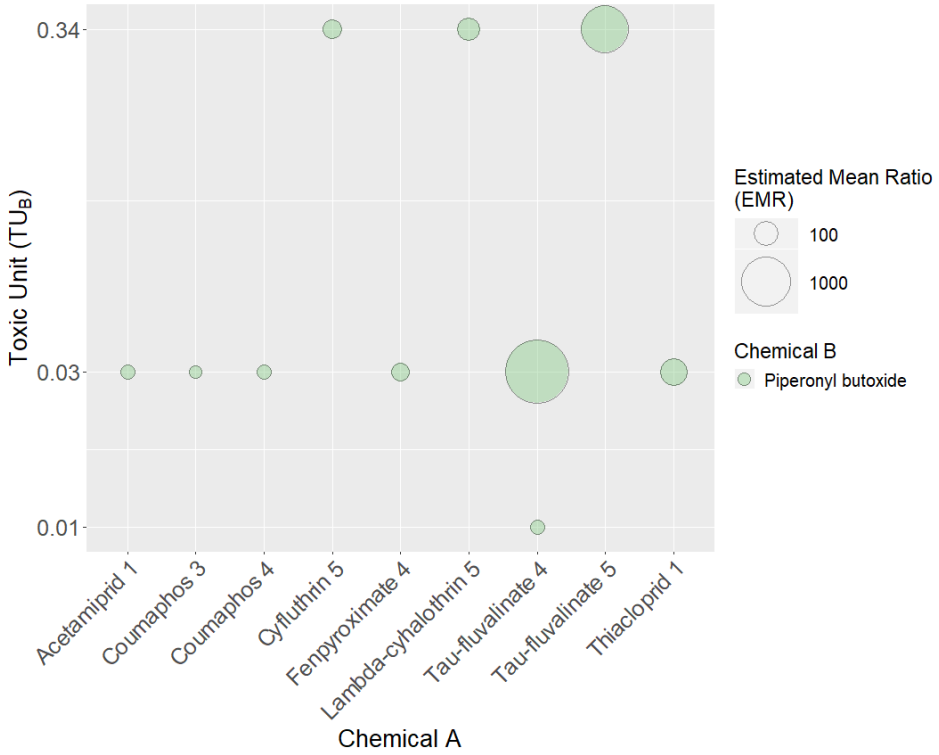


Figure 3.4. Bubble plot for acute contact toxicity of insecticides (chemical A) and synergists (PBO) (chemical B) in honey bees: Estimated Mean Ratio (EMR) (A+B) and experimental potency-adjusted dose (chemical B: Toxic Unit (TU_B). Size of the bubble is proportional to the value of the EMR. 1= Iwasa et al., (2014). 3= Johnson et al., (2009). 4= Johnson et al., (2013). 5= Johnson et al., (2006). All the studies were statistically significant according to non-overlapping 95% confidence intervals.

b) Acaricides-Insecticides

Combined toxicity was synergistic for tau-fluvalinate and coumaphos in a dose dependent fashion (TU_B= 0.005, 0.01, 0.05, 0.15, 0.49) with the highest EMR ≈ 30 (TU_B= 0.49). In contrast, the magnitude of synergism between coumaphos and tau-fluvalinate (TU_B= 0.01, 0.03, 0.08, 0.25) reached a maximum EMR of 3-fold at the highest doses (TU_B= 0.08, 0.25). Both compounds are known CYP inhibitors but based on the limited data available for these two binary mixtures further dose response data would be needed to better characterise the dose dependency of such interactions (Hesketh et al., 2016). In addition, both tau-fluvalinate and coumaphos are lipophilic and are absorbed by the wax component of the hive, thus persistent after repeated treatments and these aspects should be taken into account under field scenarios (EFSA PPR Panel, 2012). In addition, temporal transitivity (i.e. if the same effect occurs irrespective of the order of exposure) of the interactions should be taken into account when assessing acaricide-insecticide mixtures: fenpyroximate pre-treatment (TU_B= 0.06) increased tau-fluvalinate toxicity by 8 fold (MDR= 5.56), whereas the opposite is not observed (EMR= 1.2) thus showing additive effects (MDR= 1.09) (Figure 3.5; Table S15). Apparently, fenpyroximate can competitively inhibit CYP isoforms involved in tau-

fluvalinate detoxification while tau-fluvalinate does not interact with CYPs, thus allowing bees to tolerate fenpyroximate exposure (Mao et al., 2011; Johnson et al., 2013).

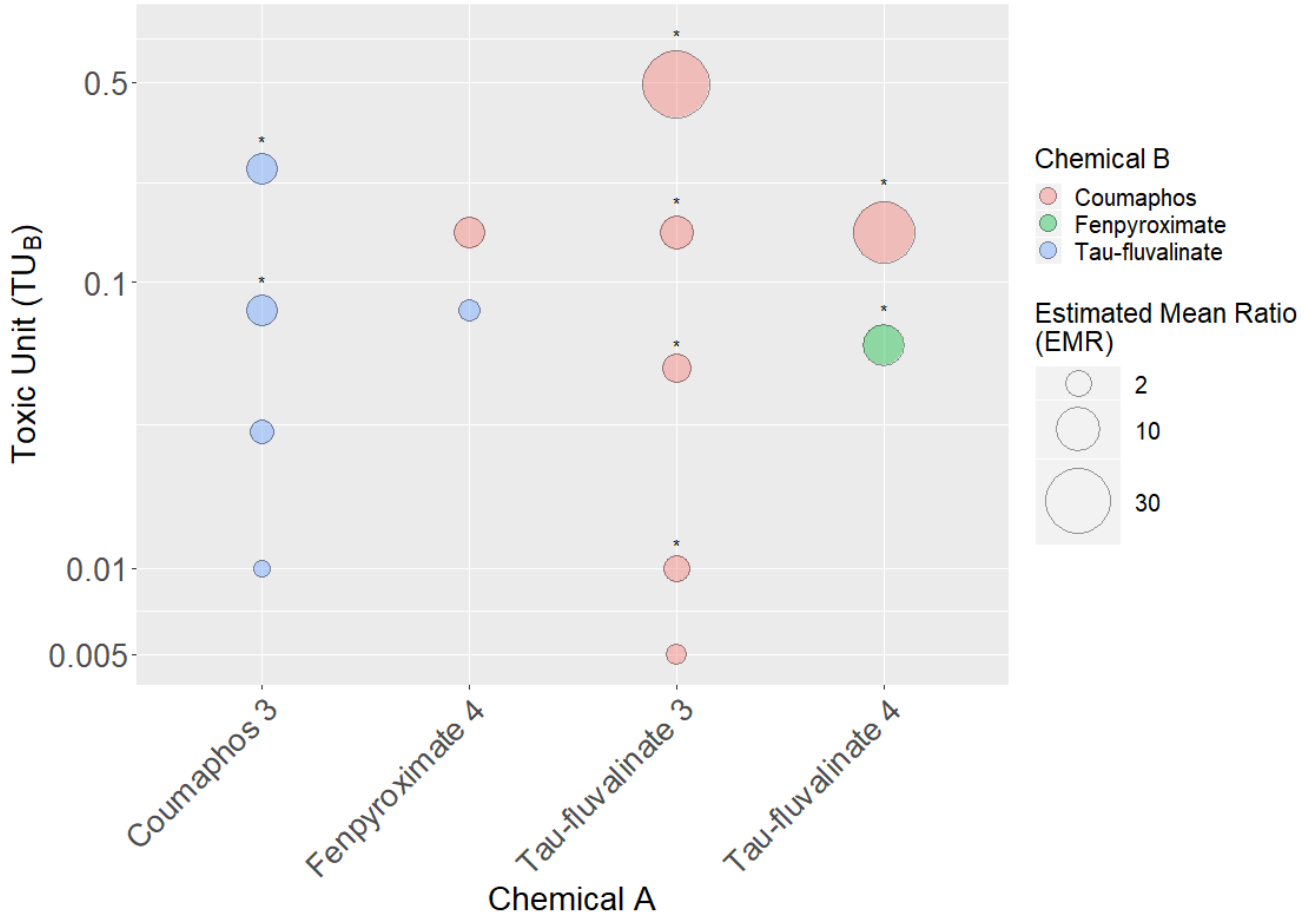


*Figure 3.5. Bubble plot for combined acute contact toxicity of acaricides (chemical A) and insecticides (chemical B) in honey bees. Estimated Mean Ratios (A+B) and experimental potency-adjusted dose (chemical B: Toxic Unit (TUB). Size of the bubble is proportional to the value of the EMR. References: 3= Johnson et al., (2009). 4= Johnson et al., (2013). * = for statistically significant studies.*

Experimental studies on combined toxicity (LD_{50}) following acute contact exposure to binary mixtures (PPPs - synergists) in different bee subspecies is presented in figure 3.6 (Rinkevich et al., 2015). Results show that bioassays using amitraz (acaricide), coumaphos (insecticide) and piperonyl butoxide (P450 inhibitor) increase phenothrin (insecticide) acute contact toxicity in all three different honey bee subspecies (i.e. Carniolan, Italian, and Russian bees). However, with regard to phenothrin sensitivity test (figure 3.6) between the three different honey bees subspecies, toxicity increased by a maximum of 7 fold in *A. mellifera primorski* down to 5 fold in *A. mellifera ligustica* following acute contact exposure to coumaphos (Rinkevich et al., 2015).

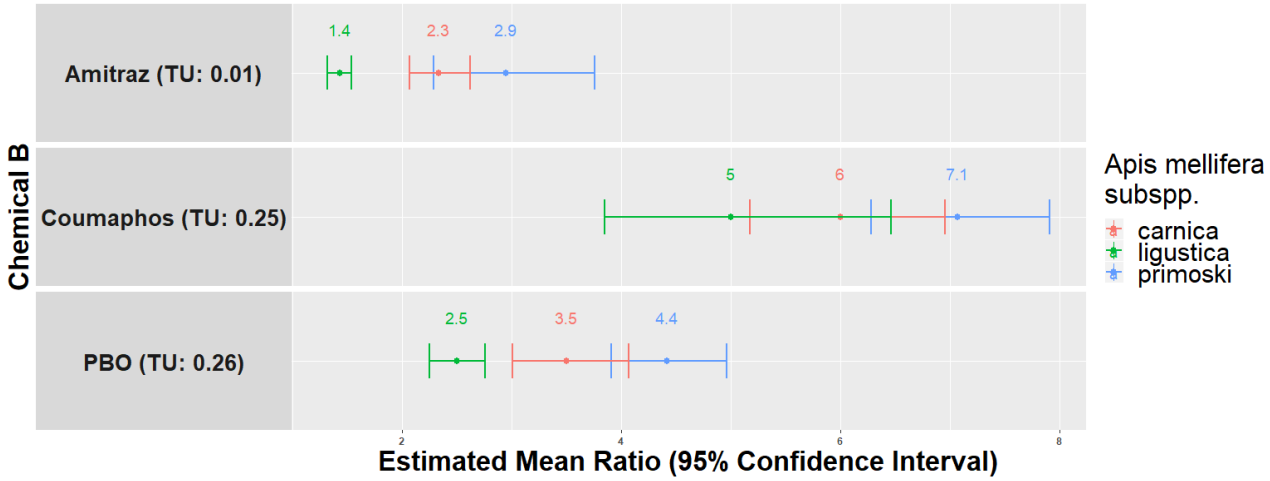


Figure 3.6. Fore stplot comparing honey-bee subspecies sensitivity to combined toxicity of binary mixtures (Rinkevich et al., 2015). Estimated Mean Ratio (EMR dots) and related 95% CI (lines) were reported in different honey bee subspecies (A. mellifera carnica, A. mellifera ligustica, A. mellifera primorski) following acute contact exposure to phenotrin with three different chemicals (amitraz, coumaphos and PBO).

c) Whole database

Figures 3.7, 3.8, and 3.9 compare EMRs for acute contact toxicity studies with their corresponding individual TU for compound B (TU_B) classified according to three different classes: $TU_B \leq 0.10$ (figure 3.7), $TU_B \leq 0.11$-0.30 and $TU_B \leq 0.31$-0.60 (figure 3.8). For each TU_B class, cumulative frequency distribution graphs are developed in order to quantify the sensitivity of the toxicological endpoint for chemical B contributing to the overall binary mixtures toxicity. The distribution of the EMR values against their "reverse cumulative frequency" is plotted and fits are tested with Pearson product-moment correlation coefficient (R^2). Results show that TU_B values range from 0.0001 to 0.61 (Table S15). Particularly, 63 (of out 133) binary mixtures experiments reported acute contact toxicity report TU_B values ≤ 0.1 (Figures 3.7 and 3.8). This indicates that most of the doses of chemical B applied in the binary mixtures assay correspond to less than 10% of their estimated relevant critical endpoint (e.g. LD_{50} or LC_{50}). Furthermore, if looking at EMRs, the highest binary mixture toxicity (EMR≈ 1980) is obtained when low doses of chemical B is applied in the binary mixture (i.e. $TU_B = 0.03$) (Figure 3.8; Table S15). Hence, our findings would raise a concern that mixtures of contaminants, although individually at low concentrations ($TU_B < 0.05$), frequently may enhance the whole binary mixture toxicity (Cedergreen, 2014; Belden et al., 2007). However, it should be noted that at very low dose of chemical B (i.e. $TU_B = 0.0003$) a decrease mixture toxicity i.e. EMR (-) = 1.55 was observed (3. 7, table S15). Overall, our results confirm that the observed synergism of binary mixtures in bees is, in most instances, explained as the result of toxicokinetic interactions at the level of metabolism either through the inhibition of a CYP or a transporter which then has toxicodynamic consequences

i.e. pyrethroids and CYP inhibitor piperonyl butoxide, insecticides with fungicides (Johnson et al., 2010; Moores et al., 2011). Generally speaking, toxicokinetic interactions of a mixture may cause deviations from additivity between components of the mixture either during absorption, distribution,metabolism or excretion.

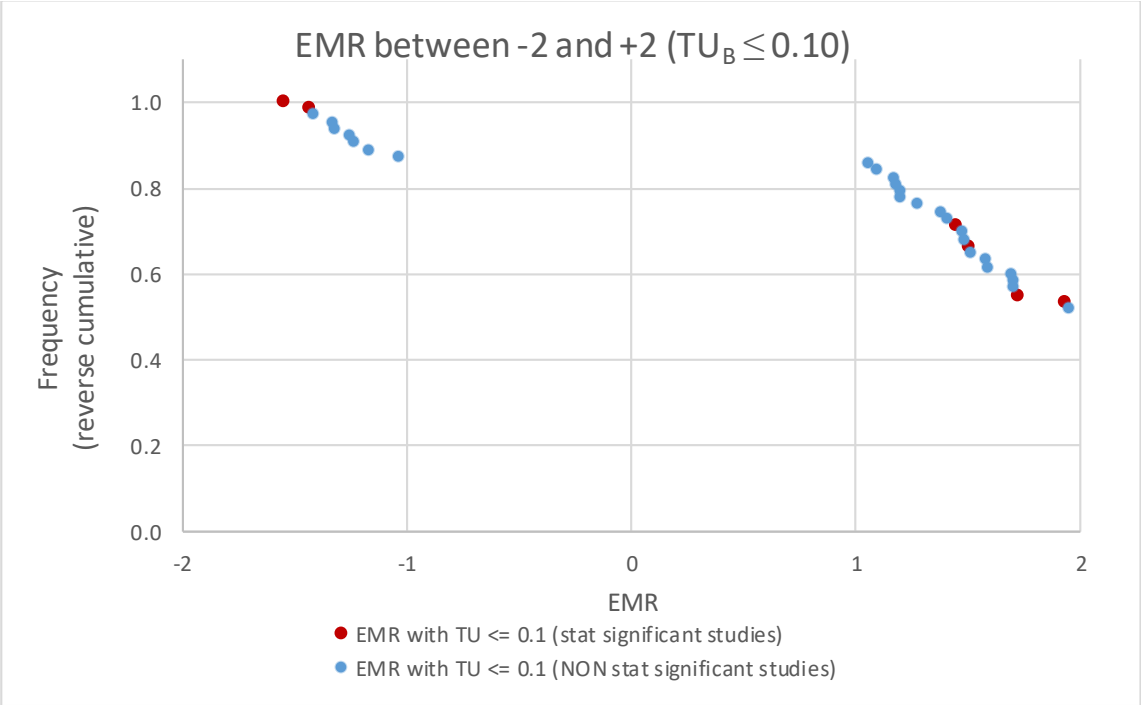


Figure 3.7. Cumulative frequency distribution for Estimated Mean Ratio (EMR) values (EMR -2 - +2), in acute contact toxicity studies in honey bees reporting Toxic Unit for chemical B (TU_B) ≤ 0.10.

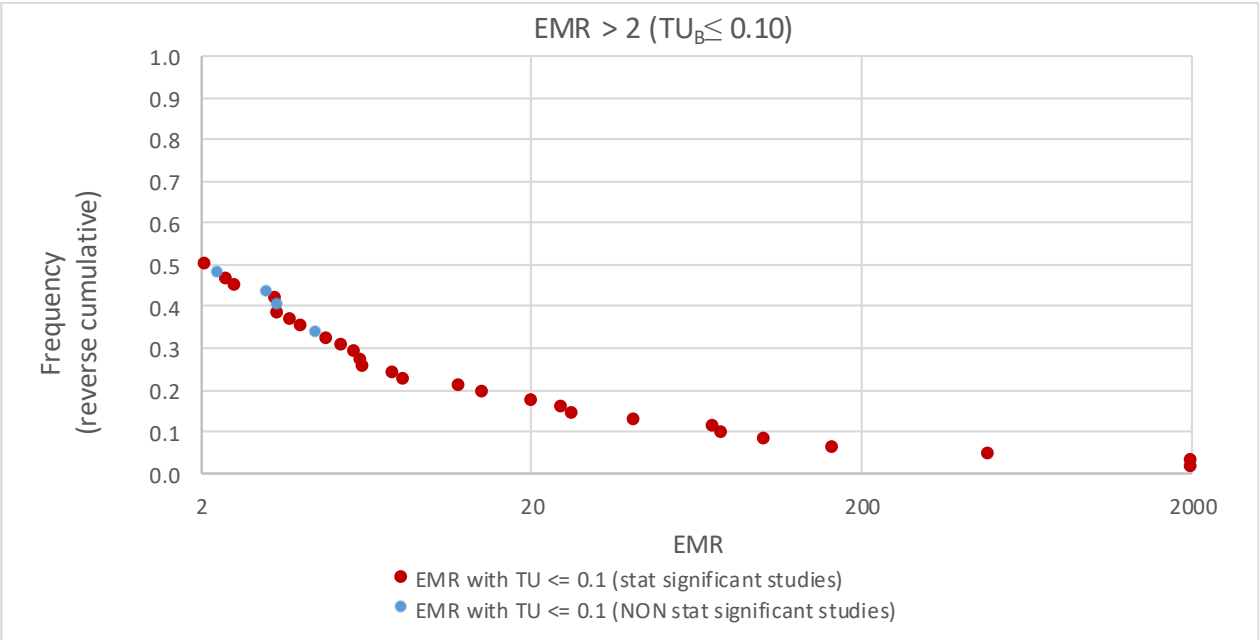


Figure 3.8. Cumulative frequency distribution for Estimated Mean Ratio (EMR) values > (+)2, in acute contact toxicity studies in honey bees showing Toxic Unit for the chemical B (TU_B) ≤ 0.10.

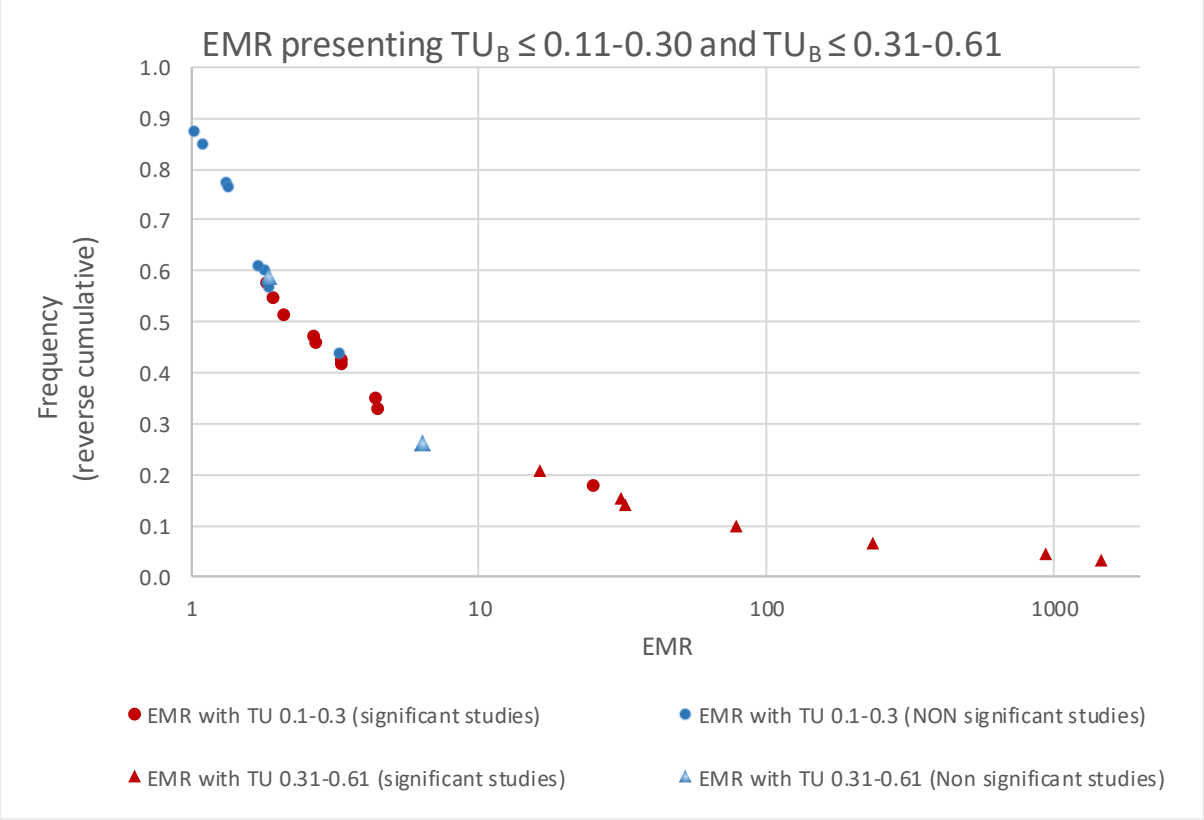


Figure 3.9. Cumulative frequency distribution for Estimated Mean Ratio (EMR) values > 2, in acute contact toxicity studies in honey bees showing Toxic Unit for the chemical B (TU_B) > 0.10. TU values are split into two classes as provided in the legend. Dots represent TU values ≤ 0.11 - 0.30. Triangles represent TU ≤ 0.31 - 0.61.

3.2.3 Predictive models for combined acute contact toxicity and Model Deviation Ratios

Comparison between predictive models of combined toxicity, to quantify the deviation from dose addition through the calculation of MDR values, are illustrated in figures 3.10 and 3.11 (see also Table S15) for the 92 acute contact toxicity binary mixtures (LD_{50} 24h) in honey bees with available experimental dose response data (Jonker et al., 2005). For the oral route, chronic binary mixtures and wild bee species, no data were available to conduct this analysis. Hence, for this analysis, individual TUs for each compound in the binary mixture experiment were added to calculate the observed TU of the mixture (TUm) assuming CA as default model.

As described in paragraph 2.2.4, MDR values were calculated according to Belden et al., (2007). However, in our analysis, we proposed refined MDR thresholds in order to provide more conservative predictions for quantifying deviations from the CA model (table 3.4). According to our MDR thresholds, from 92 binary mixtures of pesticides, combined toxicity of the binary mixtures was

synergistic in 72% (66 datasets with 48 statistically significant), 17% additive (16 datasets) and 11% antagonistic (10 datasets with 2 statistically significant) (Table S15 and Figures 3.10 and 3.11). Amongst synergies, the most commonly tested binary mixtures were conazole fungicide-insecticides combinations (Table S15). The statistical significance analysis for each binary mixture was performed using non-overlapping 95% CI of the experimental EM_A vs 95% CI of the experimental EM_M for chemical A+B. 16 out 66 mixtures were classified as statistical synergism according to our MDR thresholds (Table S15) although these were below the generic 2-fold deviation set as generic value by other authors regardless of target organism (e.g. *Daphnia* spp., honey bee), mixture (e.g. metals vs pesticides), exposure route (e.g. oral vs contact) and effects measured (e.g. lethal vs sublethal) in the experimental assays (Belden et al., 2007; Cedregreen 2014a). Here, we propose refined MDR thresholds to predict potential deviations from the DA model for the specific assessment of acute contact toxicity studies in honey bees.

Table 3.4. Comparison of Model Deviation Ratios (MDR) thresholds according to current scientific literature (Belden et al., 2007; Cedergreen, 2004) and refined MDR thresholds according to our analysis.

Mixture effect	**Thresholds for Model Deviation Ratio (MDR)** (according to Belden et al., 2007; Cedergreen, 2014)	**Refined thresholds for Model Deviation Ratio (MDR)**
Additive	0.5≤MDR≤2.0	0.83≤MDR≤1.25
Synergism	MDR> 2.0	MDR> 1.25
Antagonism	MDR< 0.5	MDR< 0.83

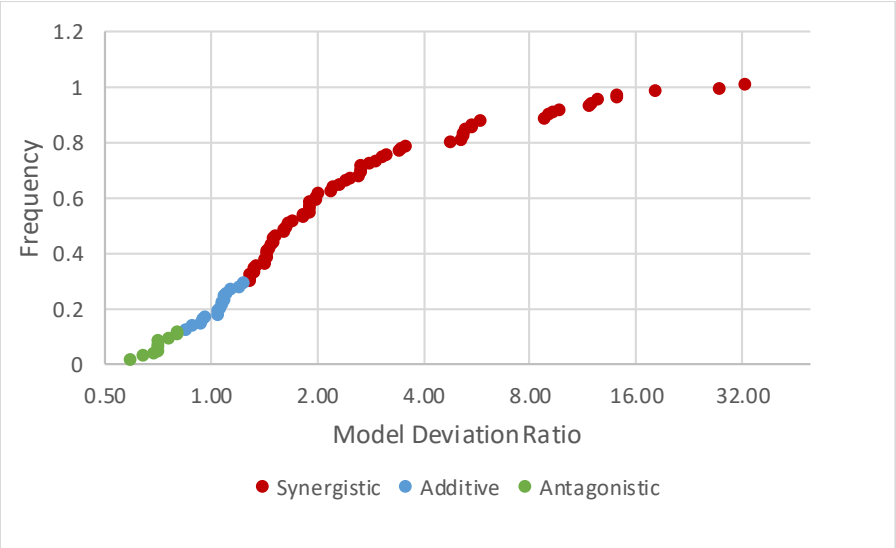


Figure 3.10. Cumulated frequency of Model Deviation Ratio. MDR for acute contact toxicity studies resulting from the meta-analysis of acute contact toxicity studies (Iwasa et al., 2004; Johnson et al., 2013; 2006; 2009; Ellis et al., 1997). MDR> 1.25 represents "synergistic" interactions, 0.83<MDR<1.25 represents "additive" effects; MDR < 0.83 represents "antagonistic" interactions.

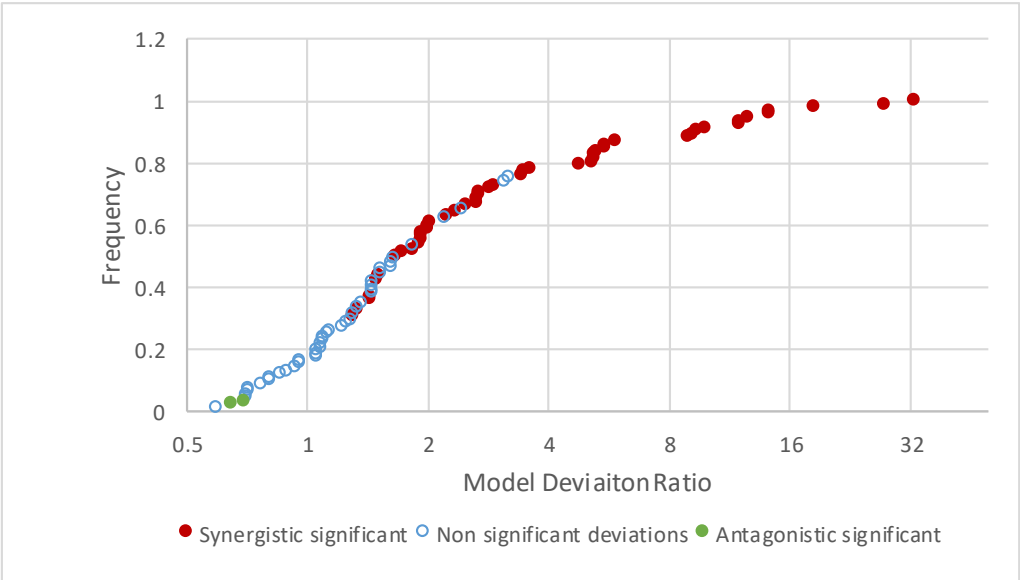


Figure 3.11. Cumulated frequency of Model Deviation Ratio. MDR for statistically significant studies resulting from the meta-analysis of acute contact toxicity studies (Iwasa et al., 2004; Johnson et al., 2013; 2006; 2009; Ellis et al., 1997). MDR > 1.25 represents "synergistic" interactions, 0.83<MDR<1.25 represents "additive" effects; MDR < 0.83 represents "antagonistic" interactions.

3.2.4 Comparison of Estimated Mean Ratios and Model Deviation Ratios

Correlations between the analyses of EMR (chapter 3.2.2) and MDR predictions (chapter 3.2.3) for acute contact toxicity of binary mixtures in honey bees (n=92) are presented on a scatterplot and Pearson product-moment correlation coefficient (R^2) in Figure 3.12 (see also Table S15). Correlations between EMR and MDR values showed different reliability according to the type of experiment. In fact, when considering binary mixtures for compounds used in potentiation experiments (i.e. synergists, thus presenting $TU_B < 0.05$), the correlation between the two variables was highly reliable ($R^2 = 1$) (Figure 3.12 - red dots). In contrast, for non-potentiation experiments ($TU_B > 0.05$) the correlation between the EMR and MDR slightly decreased ($R^2 = 0.72$). However, potentiation experiments of binary mixtures in honey bees are often reported as they reflect exposure to mixtures under field scenarios (Iwasa 2004; Johnson et al., 2012, 2013; Cedergreen, 2014; Spurgeon et al., 2016; Robinson et al., 2017). Hence, EMR analyses can provide a reliable tool to predict combined toxicity of binary mixtures, conducted as potentiation experiments, given the toxicity of chemical A (LD_{50} A) and the binary mixture (LD_{50} A +B). This tool can be potentially useful when dose response data are scarce and do not allow an MDR analysis to be performed, particularly for the identification of mixtures which cause synergistic interactions in honey bees. However, limitations of the EMR approach have to be acknowledged since it does not fully comply with the DA principles and does not assume any mathematical model for the prediction of combined toxicity (DA, Response Addition, etc.). The following thresholds for the EMR analysis are proposed:

- EMR < 0.95 indicates "antagonism" (i.e. corresponding to MDR < 0.83)
- 0.95 < EMR < 1.40 indicates " dose addition " (i.e. corresponding to 0.83<MDR< 1.25)
- EMR > 1.40 indicates "synergism" (i.e. corresponding to MDR > 1.25)

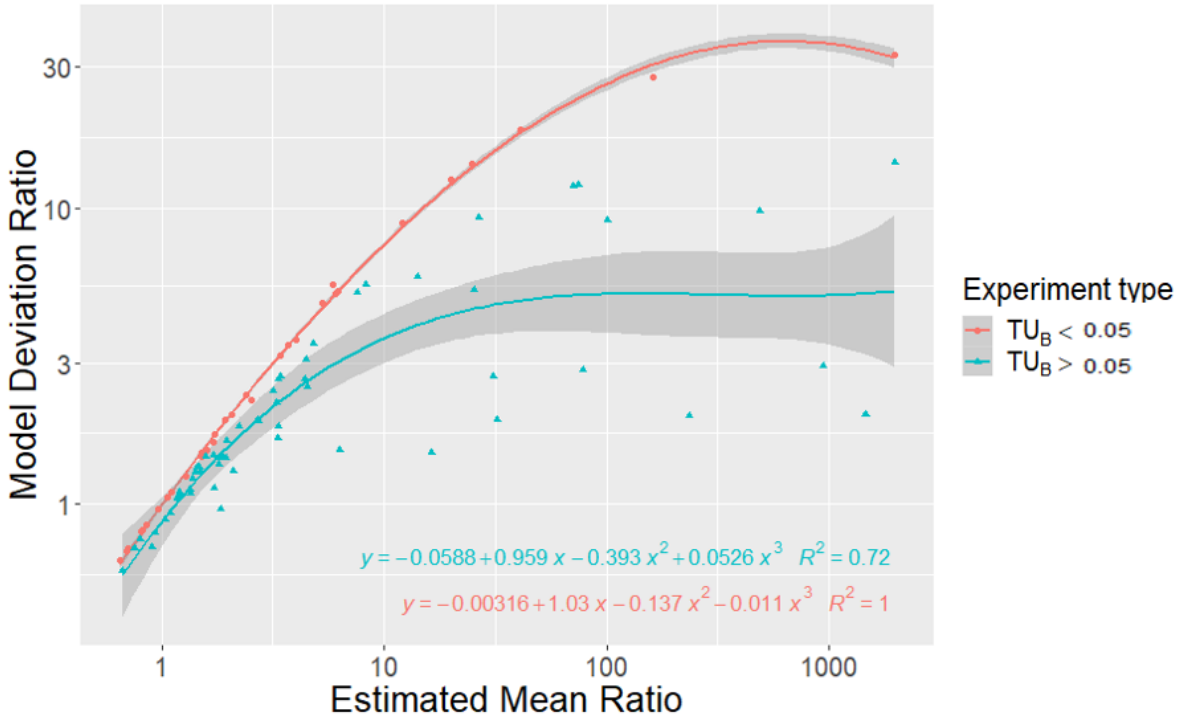


Figure 3.12. Scatter plot investigating the correlation between Estimated Mean Ratio (EMR) and Model Deviation Ratio (MDR) for acute contact toxicity of binary mixtures in honey bees. Red dots represent potentiation experiments (TU_B < 0.05). Blue triangles represent no-potentiation experiments (TU_B > 0.05).

4. Conclusions and implications for risk assessment

This manuscript constitutes the first consolidated **quantitative review of the available laboratory *in vivo* experiments on combined toxicity of binary mixtures in bee species** to support of hazard assessment. As noted in the introduction, exposure assessment and full risk characterisation are beyond the scope of the paper but their high relevance and implications for risk assessment are highlighted below together with future perspectives. Overall, 218 datasets were analysed with 61%, 20% and 19% reporting acute contact toxicity, chronic oral toxicity and acute oral toxicity, respectively. Magnitude of interactions were estimated using EMRs, from experimental studies lacking dose response data (133 acute contact, 54chronic oral and 41 acute oral datasets) and dose response data for 92 binary mixtures (acute contact data) allowed determining TU values, test deviation from dose addition and estimates of MDRs. Dose addition, synergism and antagonism were found in 17%, 72% and 11% respectively.

Strong correlations were found between EMRs and MDRs particularly for experimental studies involving potentiation experiments indicating toxicokinetic (TK) interactions as **key mechanisms** through mostly inhibition of metabolism (CYP, esterases, multiple drug resistancetransporters) as demonstrated with the potent CYP inhibitors piperonyl butoxide, triazole fungicides, tau-fluvalinate,

the carboxylesterase inhibitor DEF and the transporter inhibitors ivermectin (Johnson et al., 2013, 2015; Guseman et al., 2016; Mao et al., 2017; Wu et al., 2007). In addition, bees have also been shown are known to have the lowest copy number of detoxification enzymes within the insect kingdom, particularly for CYP isoforms, methyltransferases and glutathione-s-transferases and inhibition of such limited metabolic capacity also potentially leads to an increase in combined toxicity of chemicals (Johnson et al., 2013; 2015; EFSA, 2013a; Wade et al., 2019). Examples include fungicides CYP-mediated detoxification after exposure to the pyrethroid tau-fluvalinate or induction of imidacloprid metabolism leading to an increase in toxicity (Wade et al., 2019; Pilling et al., 1995; Iwasa et al., 2004; Johnson et al., 2013). A recent *in silico* docking study (Mao et al., 2017) with the active pocket of the honey bee, a broadly substrate-specific CYP with high quercetin-metabolising activity, and 121 pesticides, showed that six triazole fungicides inhibited CYP9Q1 though binding its catalytic site. In addition, five of six mitochondrion-related nuclear genes were down-regulated in adult honey bees fed binary mixtures of quercetin and the triazole myclobutanil and midgut metabolism of quercetin was reduced and was associated with reduced production of thoracic ATP, the energy source for flight muscles. Such findings have implications and authors conluded that, although fungicides have low acute toxicity, CYP inhibition interfering with quercetin detoxification may compromise mitochondrial regeneration, ATP production and bee health (Mao et al., 2017).

From this analysis, **key conclusions with regard to hazard assessment** of mixtures in bee species can be formulated:

1. Understanding the mechanistic basis of combined toxicity in bee species is critical for hazard assessment particularly because inhibition or induction of metabolism/transport may increase or decrease toxicity depending on the consequence of metabolism i.e. bioactivation to a toxic metabolite or detoxification (Spurgeon et al., 2016; Hesketh et al., 2016; Guseman et al., 2016; Mao et al., 2017).

2. Available data on binary mixtures were mostly generated with well-known inhibitors and may give a biased view of a more complex situation.

3. Applications of this analysis include: a) Use of the current open source database (DOI: 10.5281/zenodo.3383713) to provide scientific evidence for interactions and their magnitude for estimating mixture uncertainty factors for specific pesticide binary mixtures (see MIXTOX EFSA guidance (More et al., 2019)). b) Development of *in silico* tools such as Quantitative-structure activity relationship (QSAR) models to predict combined toxicity of mixtures in honey bees for acute contact toxicity and other endpoints (chronic, sub-lethal), bee species (solitary bees, bumble bees) and routes (oral) in the future. Such models have been

developed as classifiers to classify pesticides in different potency/threshold classes in bacteria (Toropova et al., 2012).

4. Key data gaps have been identified and include the need for: a) further laboratory testing and *in silico* docking studies in honey bees and wild bees to broaden our understanding of acute and chronic combined toxicity (contact and oral) and its dose dependency for different classes of pesticides and contaminants. This would support the characterisation of the synergistic potential of chemicals in bees TK interactions through either inhibition or induction of metabolism or through direct toxicodynamic (TD) interactions. It is noted that chemical adjuvants and additives applied in pesticide commercial formulations may have significant influence on combined toxicity and such formulations should be also tested either a components or as whole mixtures. b) Generation of basic TK (e.g. half-life) and bioaccumulation data for chemicals in bee species to allow for the development of use of Dynamic Energy Budget (DEB) models for hazard assessment of mixtures in bee species (EFSA, 2013b; EFSA 2014a; Hesketh et al., 2016; David et al., 2016; Rortais et al., 2017; EFSA, 2017a,b; Gradish et al., 2019).

Despite the above mentioned data gaps, availability of quantitative hazard metrics for combined toxicity will only provide a piece of the puzzle. Therefore, **addressing the exposure dimension** remains critical for (a) characterising the likelihood of co-occurrence of binary mixtures (or more complex mixtures) and (b) the potential magnitude of interactions at field relevant concentrations. Future directions to advance address exposure assessment science for honey bees and solitary bees include:

1. Data collection of realistic co-occurrence of multiple pesticide, veterinary drugs and contaminant residues in crops and plants visited by bees and bee matrices bearing in mind space and time.

2. Estimations of consumption data (e.g. contaminated sources such as nectar/pollen/water) for each bee species and life-stage (Tosi et al., 2018; EFSA AHAW Panel, 2016). This is particularly relevant for honey bees and wild bees (solitary and bumble bees) which can be exposed (via contact or oral routes) over a period of time, either directly through applications of multiple active ingredients in the field or indirectly through consumption of contaminated pollen or nectar (Tosi et al., 2018; Johnson, 2015; EFSA, 2013a; Simon-Delso et al., 2017; Prado et al., 2019).

3. Exposure assessment of multiple pesticide and contaminants for different routes (aerial, chemigation or ground application) and over different seasons in the same crop as tank mixes (Tosi et al., 2018).

For **risk characterisation**, a key recommendation for mixture assessment is the development of common risk metrics for honey bees and wild bee species which are then compared to protection goals defined by the risk managers. The choice of these methods is part of the iteration process of a mixture risk assessment to provide fit for purpose risk assessment and initiates in the problem formulation as part of the constant dialogue between risk assessors and risk managers (More et al., 2019). In principle, the risk metrics are selected using tiering principles depending on a) context of the risk assessment (regulated products, contaminants, bee species and level of biological organisation such as individual, hive, colony, population, landscape), b) data available on exposure (co-occurrence at field relevant concentrations, consumption patterns, routes of exposure), and hazard (evidence for combined toxicity (dose addition, toxicokinetic interactions (e.g. synergism), bioaccumulation, timelines and resources (More et al., 2019). In such contexts, harmonised risk metrics can be developed and will be dependent on data gaps identified here for the hazard and exposure dimensions ranging from low tier to high tier approaches. Low tier approaches include the application of the sum of TU i.e. individual TUs from laboratory LD_{50}s assuming dose addition and simple exposure estimates (e.g. rates of application of chemicals (e.g. pesticides) and default consumption in bees). High tier approaches can include probabilistic risk distributions for individuals, colony and population level based on the integration of model deviation ratios adjusted for internal dose (lethal or sub-lethal) using DEB models and probabilistic exposure assessment (co-occurrence, multiple routes, probabilistic consumption). For the population level and species, Species Sensitivity Distributions (SSDs) can also be used to identify hazard concentrations for the mixtures of concern according to the protection goal (HCx) and compared to population exposure estimates (More et al., 2019). Low or higher tier risk metrics are compared to a the given protection goal and the assessment may stop if no concerns are identified or may indicate a potential risk in which case either a risk management decision may be needed or higher tier risk metrics are applied to refine the risk characterisation (More et al., 2019).

Besides combined toxicity of multiple chemicals, a growing body of evidence has been published with regards to interactions between honey bee infectious agents (fungi, bacteria and viruses), predators, chemicals such as pesticides and contaminants (Collison et al., 2016; Hesketh et al., 2016), temperature and nutritional stressors (Tosi et al., 2017; Rortais et al., 2017). Examples provided in Table S10 include 1. combined exposure to clothianidin and imidacloprid and enhanced susceptibility of honey bees to deformed wing virus (DWV) (Di Prisco et al., 2013); 2. combined exposure to imidacloprid and Nosema ceranae (microsporidian parasite) in bees and increased sub-

lethal effects and individual mortality rates (Alaux et al., 2010; Vidau et al., 2011; Pettis et al., 2012); 3. combined exposure to clothianidin, thiamethoxam and nutritional stress reducing honey bee survival (Tosi et al., 2017).

In order to take into account such complex stressors on bee health, the scientific Committee of EFSA is currently developing holistic approaches for the **risk assessment of multiple stressors in honey bees** at the individual, hive level, colony, population and landscape level from a request of the European Parliament (EFSA Scientific Commitee, in preparation). Key challenges for implementing such harmonised methods into practice, need to be highlighted with particular reference to key data gaps in bees: combined toxicity (lethal and sub-lethal, TK data), occurrence and consumption patterns, the need to develop common risk metrics (e.g. toxic units, risk ratios, margin of exposure) while applying tiering principles depending on context of the assessment, data available, timelines and resources (More et al., 2019). Finally, data from OMICs technologies can provide inputs to the honey bee colony model (APISRAM - under development at EFSA) to develop biormarkers of sub-lethal effects at the individual, hive, colony and population level and to further quantify the impact of single and multiple stressors on bee health at the genome (transcriptomics), proteome (proteomics) and metabolome (metabolomics) level (EFSA, 2017a, b; Rortais et al., 2017; Aguilera et al., 2018).

Acknowledgements

The views expressed in this manuscript do not reflect the views of the European Food Safety Authority and are the authors only. This work has been financed by the European Food Safety Authority (EFSA) under the contract number NP/EFSA/AFSCO/2016/02 (Edoardo Carnesecchi), RC/EFSA/EMRISK/2013/02 (LASER Analytica Europe: Audrey Grech, Billy Amzal, Nadia Guignot) and OC/EFSA/SCER/2018/01 (Mario Negri Institute for Pharmacological Research, S-IN Soluzioni Informatiche, Molecular Networks GmbH and Altamira). Authors would like also to thank Lisa Hunston for the proof reading.

References

Aguilar, R., Ashworth, L., Galetto, L., Aizen, M.A., 2006. Plant reproductive susceptibility to habitat fragmentation: review and synthesis through a meta-analysis. Ecol. Lett. 9, 968–980. https://doi.org/10.1111/j.1461-0248.2006.00927.x.

Aguilera, J., Aguilera-Gomez, M., Barrucci, F., Cocconcelli, P.S., Davies, H., Denslow, N., Lou Dorne, J., Grohmann, L., Herman, L., Hogstrand, C., Kass, G.E.N., Kille, P., Kleter, G., Nogué, F., Plant, N.J., Ramon, M., Schoonjans, R., Waigmann, E., Wright, M.C., 2018. EFSA Scientific Colloquium 24 – 'omics in risk assessment: state of the art and next steps. EFSA Support. Publ. 15. https://doi.org/10.2903/sp.efsa.2018.EN-1512.

Alaux, C., Brunet, J.-L., Dussaubat, C., Mondet, F., Tchamitchan, S., Cousin, M., Brillard, J., Baldy, A., Belzunces, L.P., Le Conte, Y., 2010. Interactions between Nosema microspores and a neonicotinoid weaken honeybees (Apis mellifera). Environ. Microbiol. 12, 774–782. https://doi.org/10.1111/j.1462-2920.2009.02123.x.

Ashman, T.-L., Knight, T.M., Steets, J.A., Amarasekare, P., Burd, M., Campbell, D.R., Dudash, M.R., Johnston, M.O., Mazer, S.J., Mitchell, R.J., Morgan, M.T., Wilson, W.G., 2004. Pollen limitation of plant reproduction: ecological and evolutionary causes and consequences. Ecology 85, 2408–2421. https://doi.org/10.1890/03-8024.

Backhaus, T., Altenburger, R., Faust, M., Frein, D., Frische, T., Johansson, P., Kehrer, A., Porsbring, T., 2013. Proposal for environmental mixture risk assessment in the context of the biocidal product authorization in the EU. Environ. Sci. Eur. 25, 4. https://doi.org/10.1186/2190-4715-25-4.

Backhaus, T., Arrhenius, Å., Blanck, H., 2004. Toxicity of a mixture of dissimilarly acting substances to natural algal communities: predictive power and limitations of independent action and concentration addition. Environ. Sci. Technol. 38, 6363–6370. https://doi.org/10.1021/es0497678.

Belden, J.B., Gilliom, R.J., Lydy, M.J., 2007. How well can we predict the toxicity of pesticide mixtures to aquatic life? Integr. Environ. Assess. Manag. 3, 364–372. https://doi.org/10.1002/ieam.5630030307.

Berenbaum, M.R., Johnson, R.M., 2015. Xenobiotic detoxification pathways in honey bees. Curr. Opin. Insect Sci. 10, 51–58. https://doi.org/10.1016/j.cois.2015.03.005.

Biddinger, D.J., Robertson, J.L., Mullin, C., Frazier, J., Ashcraft, S.A., Rajotte, E.G., Joshi, N.K., Vaughn, M., 2013. Comparative toxicities and synergism of apple orchard pesticides to Apis mellifera (L.) and Osmia cornifrons (Radoszkowski). PLoS One 8, 1–6. https://doi.org/10.1371/journal.pone.0072587.

Bliss, C.I., 1939. The toxicity of poisons applied jointly. Ann. Appl. Biol. 26, 585–615. https://doi.org/10.1111/j.1744-7348.1939.tb06990.x.

Bommarco, R., Marini, L., Vaissière, B.E., 2012. Insect pollination enhances seed yield, quality, and market value in oilseed rape. Oecologia 169, 1025–1032. https://doi.org/10.1007/s00442-012-2271-6.

Bopp, S., Berggren, E., Kienzler, A., Van Der Linden, S., Worth, A., 2015. Scientific methodologies for the combined effects of chemicals – a survey and literature review. Use of novel and alternative methods in the assessment of effects from combined exposure to multiple chemicals. EUR 27471 EN. doi:10.2788/093511.

Buonaccorsi, J.P., Iyer, H.K., 1984. A comparison of confidence regions and designs in estimation of a ratio. Commun. Stat. - Simul. Comput. 13, 723–741. https://doi.org/10.1080/03610918408812411.

Burkle, L.A., Marlin, J.C., Knight, T.M., 2013. Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. Science 339, 1611–1615. https://doi.org/10.1126/science.1232728.

Casella, G., Berger, R., 2002. Statistical inference. vol. 2. Duxbury Pacific Grove.

Cedergreen, N., Svendsen, C., Backhaus, T., 2013. Toxicity prediction of chemical mixtures. Encycl. Environ. Manag. https://doi.org/10.1081/E-EEM-120046684.

Cedergreen, N., 2014. Quantifying synergy: a systematic review of mixture toxicity studies within environmental toxicology. PLoS One 9, e96580. https://doi.org/10.1371/journal.pone.0096580.

Cedergreen, N., Sørensen, H., Svendsen, C., 2012. Can the joint effect of ternary mixtures be predicted from binary mixture toxicity results? Sci. Total Environ. 427–428, 229–237. https://doi.org/10.1016/j.scitotenv.2012.03.086.

Cizelj, I., Glavan, G., Božič, J., Oven, I., Mrak, V., Narat, M., 2016. Prochloraz and coumaphos induce different gene expression patterns in three developmental stages of the Carniolan honey bee (Apis mellifera carnica Pollmann). Pestic. Biochem. Physiol. 128, 68–75. https://doi.org/10.1016/j.pestbp.2015.09.015.

Collison, E., Hird, H., Cresswell, J., Tyler, C., 2016. Interactive effects of pesticide exposure and pathogen infection on bee health – a critical analysis. Biol. Rev. https://doi.org/10.1111/brv.122066.

Commission, E., 2016. Bee health – EU Actions. Available at https://ec.europa.eu/food/animals/live_animals/bees_en (accessed 27 February 2019).

Coors, A., Frische, T., 2011. Predicting the aquatic toxicity of commercial pesticide mixtures. Environ. Sci. Eur. https://doi.org/10.1186/2190-4715-23-22.

Corbet, S.A., Williams, I.H., Osborne, J.L., 1991. Bees and the pollination of crops and wild flowers in the european community. Bee World 72, 47–59. https://doi.org/10.1080/0005772X.1991.11099079.

Culley, T.M., Weller, S.G., Sakai, A.K., 2002. The evolution of wind pollination in angiosperms. Trends Ecol. Evol. 17, 361–369. https://doi.org/10.1016/S0169-5347(02)02540-5.

David, A., Botías, C., Abdul-Sada, A., Nicholls, E., Rotheray, E.L., Hill, E.M., Goulson, D., 2016. Widespread contamination of wildflower and bee-collected pollen with complex mixtures of neonicotinoids and fungicides commonly applied to crops. Environ. Int. 88, 169–178. https://doi.org/10.1016/j.envint.2015.12.011.

de Castro, R.D., de Souza, T.M.P.A., Bezerra, L.M.D., Ferreira, G.L.S., de Brito Costa, E.M.M., Cavalcanti, A.L., 2015. Antifungal activity and mode of action of thymol and its synergism with nystatin against Candida species involved with infections in the oral cavity: an in vitro study. BMC Complement. Altern. Med. 15, 417. https://doi.org/10.1186/s12906-015-0947-2.

de Groot, R.S., Wilson, M.A., Boumans, R.M., 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. Ecol. Econ. 41, 393–408. https://doi.org/10.1016/S0921-8009(02)00089-7.

Di Pasquale, G., Alaux, C., Le Conte, Y., Odoux, J.-F., Pioz, M., Vaissière, B.E., Belzunces, L.P., Decourtye, A., 2016. Variations in the availability of pollen resources affect honey bee health. PLoS One 11, e0162818. https://doi.org/10.1371/journal.pone.0162818.

Di Prisco, G., Cavaliere, V., Annoscia, D., Varricchio, P., Caprio, E., Nazzi, F., Gargiulo, G., Pennacchio, F., 2013. Neonicotinoid clothianidin adversely affects insect immunity and promotes replication of a viral pathogen in honey bees. Proc. Natl. Acad. Sci. 110, 18466–18471. https://doi.org/10.1073/pnas.1314923110.

Dorne, J. Lou, Richardson, J., Kass, G., Georgiadis, N., Monguidi, M., Pasinato, L., Cappe, S., Verhagen, H., Robinson, T., 2017. Editorial: OpenFoodTox: EFSA's open source toxicological database on chemical hazards in food and feed. EFSA J. 15. https://doi.org/10.2903/j.efsa.2017.e15011.

EFSA (European Food Safety Authority), 2014a. Towards an integrated environmental risk assessment of multiple stressors on bees: review of research projects in Europe, knowledge gaps and recommendations. EFSA J. 12. https://doi.org/10.2903/j.efsa.2014.3594.

EFSA (European Food Safety Authority), 2013a. Towards holistic approaches to the risk assessment of multiple stressors in bees, EFSA Scientific Colloquium XVIII Summary Report. doi:10.2805/53269.

EFSA (European Food Safety Authority), 2013b. Guidance on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA J. 11. https://doi.org/10.2903/j.efsa.2013.3295.

EFSA Panel on Plant Protection Products and their Residues (PPR), 2012. Scientific Opinion on the science behind the development of a risk assessment of Plant Protection Products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA J.10, 2668. https://doi.org/10.2903/j.efsa.2012.2668.

EFSA (European Food Safety Authority), 2014b. Further development and update of EFSA's Chemical Hazards Database (NP/EFSA/EMRISK/2012/01). EFSA Support. Publ. 11. doi:10.2903/sp.efsa.2014.EN-654.

EFSA Panel on Animal Health and Welfare (AHAW), 2016. Assessing the health status of managed honeybee colonies (HEALTHY-B): a toolbox to facilitate harmonised data collection. EFSA J. 14, 04578. https://doi.org/10.2903/j.efsa.2016.4578.

EFSA (European Food Safety Authority), 2017a. Specifications for field data collection contributing to honey bee model corroboration and verification. EFSA Support. Publ.14. https://doi.org/10.2903/sp.efsa.2017.EN-1234.

EFSA (European Food Safety Authority), 2017b. Collecting and sharing data on bee health: towards a European Bee Partnership. EFSA Support. Publ. 14. https://doi.org/10.2903/sp.efsa.2017.en-1299.

EFSA Scientific Committee. A holistic approach for the risk assessment of multiple stressors in honey bees (Apis mellifera spp.). in preparation.

Eilers, E.J., Kremen, C., Smith Greenleaf, S., Garber, A.K., Klein, A.-M., 2011. Contribution of pollinator-mediated crops to nutrients in the human food supply. PLoS One 6, e21363. https://doi.org/10.1371/journal.pone.0021363.

Ellis, M.D., Baxendale, F.P., 1997. Toxicity of seven monoterpenoids to tracheal mites (Acari: Tarsonemidae) and their honey bee (Hymenoptera: Apidae) hosts when applied as fumigants. J. Econ. Entomol. 90, 1087–1091. https://doi.org/10.1093/jee/90.5.1087.

Everitt, Brian, Skrondal, Anders, 2010. Standardized mortality rate (SMR). The Cambridge dictionary of statistics. Cambridge University Press, New York, pp. 409 ISBN 9780521766999.

Faraggi, D., Izikson, P., Reiser, B., 2003. Confidence intervals for the 50 per cent response dose. Stat. Med. https://doi.org/10.1002/sim.1368.

Faust, M., Altenburger, R., Backhaus, T., Blanck, H., Boedeker, W., Gramatica, P., Hamer, V., Scholze, M., Vighi, M., Grimme, L., 2003. Joint algal toxicity of 16 dissimilarly acting chemicals is predictable by the concept of independent action. Aquat. Toxicol. 63, 43–63. https://doi.org/10.1016/S0166-445X(02)00133-9.

Fieller, E.C., 1954. Some Problems in Interval Estimation. J. R. Stat. Soc. Ser. B. https://doi.org/10.1111/j.2517-6161.1954.tb00159.x.

Formato, G., Zilli, R., Condoleo, R., Marozzi, S., Davis, I., Smulders, F.J.M., 2011. Risk management in primary apicultural production. Part 2: a Hazard Analysis Critical Control Point approach to assuring the safety of unprocessed honey. Vet. Q. 31, 87–97. https://doi.org/10.1080/01652176.2011.567755.

Franz, V.H., 2007. Ratios: A short guide to confidence limits and proper use. Cornell University. https://arxiv.org/abs/0710.2024.

Goulson, D., Nicholls, E., Botias, C., Rotheray, E.L., 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science 347, 1255957. https://doi.org/10.1126/science.1255957.

Gradish, A.E., van der Steen, J., Scott-Dupree, C.D., Cabrera, A.R., Cutler, G.C., Goulson, D., Klein, O., Lehmann, D.M., Lückmann, J., O'Neill, B., Raine, N.E., Sharma, B., Thompson, H., 2019. Comparison of pesticide exposure in honey bees (Hymenoptera: Apidae) and bumble bees (Hymenoptera: Apidae): implications for risk assessments. Environ. Entomol. 48, 12–21. https://doi.org/10.1093/ee/nvy168.

Gunning, R.V., Moores, G.D., Jewess, P., Boyes, A.L., Devonshire, A.L., Khambay, B.P., 2007. Use of pyrethroid analogues to identify key structural features for enhanced esterase resistance in Helicoverpa armigera (Hübner) (Lepidoptera: Noctuidae). Pest Manag. Sci. 63, 569–575. https://doi.org/10.1002/ps.1377.

Guseman, A.J., Miller, K., Kunkle, G., Dively, G.P., Pettis, J.S., Evans, J.D., VanEngelsdorp, D., Hawthorne, D.J., 2016. Multi-drug resistance transporters and a mechanism-based strategy for assessing risks of pesticide combinations to honey bees. PLoS One. https://doi.org/10.1371/journal.pone.0148242.

Han, W., Yang, Y., Gao, J., Zhao, D., Ren, C., Wang, S., Zhao, S., Zhong, Y., 2019. Chronic toxicity and biochemical response of Apis cerana cerana (Hymenoptera: Apidae) exposed to acetamiprid and propiconazole alone or combined. Ecotoxicology 28, 399–411. https://doi.org/10.1007/s10646-019-02030-4.

Hesketh, H., Lahive, E., Horton, A.A., Robinson, A.G., Svendsen, C., Rortais, A., Dorne, J.-L., Baas, J., Spurgeon, D.J., Heard, M.S., 2016. Extending standard testing period in honeybees to predict lifespan impacts of pesticides and heavy metals using dynamic energy budget modelling. Sci. Rep. 6, 37655. https://doi.org/10.1038/srep37655.

Hermann, D., Stenzel, K., 2019. FRAC Mode-of-action classification and resistance risk of fungicides. In: Modern Crop Protection Compounds. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp. 589–608. https://doi.org/10.1002/9783527699261ch14.

Heys, K.A., Shore, R.F., Pereira, M.G., Jones, K.C., Martin, F.L., 2016. Risk assessment of environmental mixture effects. RSC Adv. 6, 47844–47857. https://doi.org/10.1039/c6ra05406d.

Hirschberg, J., Lye, J., 2010. A geometric comparison of the delta and fieller confidence intervals. Am. Stat. 64, 234–241. https://doi.org/10.1198/tast.2010.08130.

Hoshide, A., Drummond, F., Stevens, T., Venturini, E., Hanes, S., Sylvia, M., Loftin, C., Yarborough, D., Averill, A., 2018. What is the value of wild bee pollination for wild blueberries and cranberries, and who values it? Environments 5, 98. https://doi.org/10.3390/environments5090098.

Huang, W.-F., Solter, L.F., Yau, P.M., Imai, B.S., 2013. Nosema ceranae escapes fumagillin control in honey bees. PLoS Pathog. 9, e1003185. https://doi.org/10.1371/journal.ppat.1003185.

Iwasa, T., Motoyama, N., Ambrose, J.T., Roe, R.M., 2004. Mechanism for the differential toxicity of neonicotinoid insecticides in the honey bee, Apis mellifera. Crop Prot. 23, 371–378. https://doi.org/10.1016/j.cropro.2003.08.018.

Jacques, A., Laurent, M., Ribiere-Chabert, M., Saussac, M., Bougeard, S., Hendrikx, P., Chauzat, M., 2016. Statistical analysis on the EPILOBEE dataset: explanatory variables related to honeybee colony mortality in EU during a 2 year survey. EFSA Support. Publ. 13. https://doi.org/10.2903/sp.efsa.2016.EN-883.

Johnson, R.M., 2015. Honey bee toxicology. Annu. Rev. Entomol. 60, 415–434. https://doi.org/10.1146/annurev-ento-011613-162005.

Johnson, R.M., Dahlgren, L., Siegfried, B.D., Ellis, M.D., 2013. Acaricide, fungicide and drug interactions in honey bees (Apis mellifera). PLoS One 8, e54092. https://doi.org/10.1371/journal.pone.0054092.

Johnson, R.M., Ellis, M.D., Mullin, C.A., Frazier, M., 2010. Pesticides and honey bee toxicity – USA. Apidologie 41, 312–331. https://doi.org/10.1051/apido/2010018.

Johnson, R.M., Mao, W., Pollock, H.S., Niu, G., Schuler, M.A., Berenbaum, M.R., 2012. Ecologically appropriate xenobiotics induce cytochrome P450s in Apis mellifera. PLoS One 7, 1–9. https://doi.org/10.1371/journal.pone.0031051.

Johnson, R.M., Pollock, H.S., Berenbaum, M.R., 2009. Synergistic interactions between in-hive miticides in Apis mellifera. J. Econ. Entomol. 102, 474–479. https://doi.org/10.1603/029.102.0202.

Johnson, R.M., Wen, Z., Schuler, M.A., Berenbaum, M.R., 2006. Mediation of pyrethroid insecticide toxicity to honey bees (Hymenoptera: Apidae) by cytochrome P450 monooxygenases. J. Econ. Entomol. 99, 1046–1050. https://doi.org/10.1093/jee/99.4.1046.

Jonker, M.J., Svendsen, C., Bedaux, J.J.M., Bongers, M., Kammenga, J.E., 2005. Significance testing of synergistic/antagonistic, dose level–dependent, or dose ratio-dependent effects in mixture dose–response analysis. Environ. Toxicol. Chem. 24, 2701. https://doi.org/10.1897/04-431R.1.

Kennedy, C.M., Lonsdorf, E., Neel, M.C., Williams, N.M., Ricketts, T.H., Winfree, R., Bommarco, R., Brittain, C., Burley, A.L., Cariveau, D., Carvalheiro, L.G., Chacoff, N.P., Cunningham, S.A., Danforth, B.N., Dudenhöffer, J.-H., Elle, E., Gaines, H.R., Garibaldi, L.A., Gratton, C., Holzschuh, A., Isaacs, R., Javorek, S.K., Jha, S., Klein, A.M., Krewenka, K., Mandelik, Y., Mayfield, M.M., Morandin, L., Neame, L.A., Otieno, M., Park, M., Potts, S.G., Rundlöf, M., Saez, A., Steffan-Dewenter, I., Taki, H., Viana, B.F., Westphal, C., Wilson, J.K., Greenleaf, S.S., Kremen, C., 2013. A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. Ecol. Lett. 16, 584–599. https://doi.org/10.1111/ele.12082.

Kienzler, A., Berggren, E., Bessems, J., Bopp, S., Van Der Linden, S., Worth, A., 2014. Assessment of mixtures - review of regulatory requirements and guidance. JRC Science and Policy Report. European Commission, Joint Research Center, Ispra, Italy. doi:10.2788/84264.

Kienzler, A., Bopp, S.K., van der Linden, S., Berggren, E., Worth, A., 2016. Regulatory assessment of chemical mixtures: Requirements, current approaches and future perspectives. Regul. Toxicol. Pharmacol. 80, 321–334. https://doi.org/10.1016/j.yrtph.2016.05.020.

Klein, A.-M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. Proc. R. Soc. B Biol. Sci. 274, 303–313. https://doi.org/10.1098/rspb.2006.3721.

Klein, S., Cabirol, A., Devaud, J.-M., Barron, A.B., Lihoreau, M., 2017. Why bees are so vulnerable to environmental stressors. Trends Ecol. Evol. 32, 268–278. https://doi.org/10.1016/j.tree.2016.12.009.

Le Conte, Y., Navajas, M., 2008. Climate change: impact on honey bee populations and diseases. Rev. Sci. Tech. l'OIE 27, 485–510. https://doi.org/10.20506/rst.27.2.1819.

Leroux, P., Bach, J., Debieu, D., Fillinger, S., Fritz, R., Walker, A.S., 2008. Mode of action of sterol biosynthesis inhibitors and resistance phenomena in fungi. Mod. Fungic. In: Antifung. Compd. V 15th Int. Reinhardsbrunn Symp. Friedrichroda, Ger. May 6–10, pp. 85–92.

Loewe, S., Muischnek, H., 1926. Über Kombinationswirkungen. Arch. für Exp. Pathol. Und Pharmakologie 114, 313–326. https://doi.org/10.1007/BF01952257.

Manning, P., Ramanaidu, K., Cutler, G.C., 2017. Honey bee survival is affected by interactions between field-relevant rates of fungicides and insecticides used in apple and blueberry production. FACETS 2, 910–918. https://doi.org/10.1139/facets-2017-0025.

Mao, W., Schuler, M.A., Berenbaum, M.R., 2011. CYP9Q-mediated detoxification of acaricides in the honey bee (Apis mellifera). Proc. Natl. Acad. Sci. 108, 12657–12662. https://doi.org/10.1073/pnas.1109535108.

Mao, W., Schuler, M.A., Berenbaum, M.R., 2017. Disruption of quercetin metabolism by fungicide affects energy production in honey bees (Apis mellifera). Proc. Natl. Acad. Sci. 114, 2538–2543. https://doi.org/10.1073/pnas.1614864114.

Moher, D., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann. Intern. Med. 151, 264. https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

Moores, G.D., Wegorek, P., Zamojska, J., Field, L., Philippou, D., 2012. The effect of a piperonyl butoxide/tau-fluvalinate mixture on pollen beetle (Meligethes aeneus) and honey bees (Apis mellifera). Pest Manag. Sci. 68, 795–800. https://doi.org/10.1002/ps.2328.

More, S.J., Bampidis, V., Benford, D., Bennekou, S.H., Bragard, C., Halldorsson, T.I., Hernández-Jerez, A.F., Koutsoumanis, K., Naegeli, H., Schlatter, J.R., Silano, V., Nielsen, S.S., Schrenk, D., Turck, D., Younes, M., Benfenati, E., Castle, L., Cedergreen, N., Hardy, A., Laskowski, R., Leblanc, J.C., Kortenkamp, A., Ragas, A., Posthuma, L., Svendsen, C., Solecki, R., Testai, E., Dujardin, B., Kass, G.E., Manini, P., Jeddi, M.Z., Dorne, J.C., Hogstrand, C., 2019. Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals. EFSA J. 17. https://doi.org/10.2903/j.efsa.2019.5634.

Nazzi, F., Brown, S.P., Annoscia, D., Del Piccolo, F., Di Prisco, G., Varricchio, P., Della Vedova, G., Cattonaro, F., Caprio, E., Pennacchio, F., 2012. Synergistic parasite-pathogen interactions mediated by host immunity can drive the collapse of honeybee colonies. PLoS Pathog. 8, e1002735. https://doi.org/10.1371/journal.ppat.1002735.

Nazzi, F., Pennacchio, F., 2014. Disentangling multiple interactions in the hive ecosystem. Trends Parasitol. 30, 556–561. https://doi.org/10.1016/j.pt.2014.09.006.

OECD (Organisation for Economic Co-operation and Development), 2017. Test No. 245: honey bee (Apis mellifera L.), chronic oral toxicity test (10-day feeding), OECD Guidelines for the Testing of Chemicals, Section 2. OECD. doi:10.1787/9789264284081-en.

Pettis, J.S., VanEngelsdorp, D., Johnson, J., Dively, G., 2012. Pesticide exposure in honey bees results in increased levels of the gut pathogen Nosema. Naturwissenschaften 99, 153–158. https://doi.org/10.1007/s00114-011-0881-1.

Pilling, E.D., Bromleychallenor, K.A.C., Walker, C.H., Jepson, P.C., 1995. Mechanism of synergism between the pyrethroid insecticide λ-cyhalothrin and the imidazole fungicide prochloraz, in the honeybee (Apis mellifera L.). Pestic. Biochem. Physiol. 51, 1–11. https://doi.org/10.1006/pest.1995.1001.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E., 2010. Global pollinator declines: trends, impacts and drivers. Trends Ecol. Evol. 25, 345–353. https://doi.org/10.1016/j.tree.2010.01.007.

Prado, A., Pioz, M., Vidau, C., Requier, F., Jury, M., Crauser, D., Brunet, J.-L., Le Conte, Y., Alaux, C., 2019. Exposure to pollen-bound pesticide mixtures induces longer-lived but less efficient honey bees. Sci. Total Environ. 650, 1250–1260. https://doi.org/10.1016/j.scitotenv.2018.09.102.

Quignot, N., Grech, A., Amzal, B., 2015. Data collection on combined toxicity of multiple chemicals for animal health and ecological risk assessment. EFSA Support. Publ. 12. https://doi.org/10.2903/sp.efsa.2015.EN-861.

Raftery, A.E., Schweder, T., 1993. Inference about the ratio of two parameters, with application to whale censusing. Am. Stat. 47, 259–264. https://doi.org/10.1080/00031305.1993.10475994.

Renzi, M.T., Amichot, M., Pauron, D., Tchamitchian, S., Brunet, J.L., Kretzschmar, A., Maini, S., Belzunces, L.P., 2016. Chronic toxicity and physiological changes induced in the honey bee by the exposure to fipronil and Bacillus thuringiensis spores alone or combined. Ecotoxicol. Environ. Saf. 127, 205–213. https://doi.org/10.1016/j.ecoenv.2016.01.028.

Rinkevich, F.D., Margotta, J.W., Pittman, J.M., Danka, R.G., Tarver, M.R., Ottea, J.A., Healy, K.B., 2015. Genetics, synergists, and age affect insecticide sensitivity of the honey bee, Apis mellifera. PLoS One 10, 2–13. https://doi.org/10.1371/journal.pone.0139841.

Robinson, A., Hesketh, H., Lahive, E., Horton, A.A., Svendsen, C., Rortais, A., Dorne, J. Lou, Baas, J., Heard, M.S., Spurgeon, D.J., 2017. Comparing bee species responses to chemical mixtures: Common response patterns? PLoS One 12, 1–21. https://doi.org/10.1371/journal.pone.0176289.

Rortais, A., Arnold, G., Dorne, J.-L., More, S.J., Sperandio, G., Streissl, F., Szentes, C., Verdonck, F., 2017. Risk assessment of pesticides and other stressors in bees: principles, data gaps and perspectives from the European Food Safety Authority. Sci. Total Environ. 587–588, 524–537. https://doi.org/10.1016/j.scitotenv.2016.09.127.

Rose, T., Kremen, C., Thrupp, A., Gemmill-Herren, B., Graub, B., Azzu, N., Antunes, V., Bruteig, I.E., Buchori, D., Donaldson, J., Dhyani, P.P., Garibaldi, L., Getz, A., Goss, M., Iqbal, J., Kasina, M., Kinuthia, W., Kwapong, P., Manetto, S., Martins, D., Nyamasio, G., Nyamongo, D.O., Odhiambo, C., América Suarez De Oliveira, D., Owusu, E.H., Pandey, B., Poole, C., Roubik, D.W., Roy, P., Waghchoure, E., Lusike, W., 2015. Policy analysis paper: mainstreaming of biodiversity and ecosystem services with a focus on pollination. United Nations Food and Agriculture Organization (FAO) with the contribution of participants at the "Policies for Pollination Management" Worksh. Available at: http://www.fao.org/3/a-i4242e.pdf.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. Vienna, Austria. Available at https://www.R-project.org/.

Sánchez-Bayo, F., 2012. Insecticides Mode of Action in relation to their toxicity to non-target organisms. J. Environ. Anal. Toxicol. s4. https://doi.org/10.4172/2161-0525S4-002.

Sanchez-Bayo, F., Goka, K., 2016. Impacts of pesticides on honey bees. In: Beekeeping and Bee Conservation - Advances in Research. InTech. https://doi.org/10.5772/62487.

SCCS, SCENHIR and SCHER (Scientific Committee on Health and Environmental Risks, S. C. on C.S., S.C. on E., N.I.H.R., 2012. Opinion on the toxicity and assessment of chemical mixtures. doi:10.2772/21444.

Sgolastra, F., Medrzycki, P., Bortolotti, L., Renzi, M.T., Tosi, S., Bogo, G., Teper, D., Porrini, C., Molowny-Horas, R., Bosch, J., 2017. Synergistic mortality between a neonicotinoid insecticide and an ergosterol-biosynthesis-inhibiting fungicide in three bee species. Pest Manag. Sci. 73, 1236–1243. https://doi.org/10.1002/ps.4449.

Simon-Delso, N., San Martin, G., Bruneau, E., Delcourt, C., Hautier, L., 2017. The challenges of predicting pesticide exposure of honey bees at landscape level. Sci. Rep. 7, 3801. https://doi.org/10.1038/s41598-017-03467-5.

Sparks, T.C., Nauen, R., 2015. IRAC: Mode of action classification and insecticide resistance management. Pestic. Biochem. Physiol. 121, 122–128. https://doi.org/101016/j.pestbp.2014.11.014.

Spurgeon, D., Hesketh, H., Lahive, E., Svendsen, C., Baas, J., Robinson, A., Horton, A., Heard, M., 2016. Chronic oral lethal and sub-lethal toxicities of different binary mixtures of pesticides and contaminants in bees (Apis mellifera, Osmia bicornis and Bombus terrestris). EFSA Support. Publ. 13. https://doi.org/10.2903/sp.efsa.2016.en-1076.

Steinhauer, N.A., Rennich, K., Wilson, M.E., Caron, D.M., Lengerich, E.J., Pettis, J.S., Rose, R., Skinner, J.A., Tarpy, D.R., Wilkes, J.T., VanEngelsdorp, D., 2014. A national survey of managed honey bee 2012–2013 annual colony losses in the USA: results from the Bee Informed Partnership. J. Apic. Res. 53, 1–18. https://doi.org/10.3896/IBRA.1.53.1.01.

Tinto, W.F., Elufioye, T.O., Roach, J., 2017. Waxes. In: Pharmacognosy. Elsevier, pp. 443–455.

Tong, L., Nieh, J.C., Tosi, S., 2019. Combined nutritional stress and a new systemic pesticide (flupyradifurone, Sivanto) reduce bee survival, food consumption, flight success, and thermoregulation. Chemosphere 237, 124408. https://doi.org/10.1016/j.chemosphere.2019.124408.

Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2012. CORAL: models of toxicity of binary mixtures. Chemom. Intell. Lab. Syst. 119, 39–43. https://doi.org/10.1016/j.chemolab.2012.10.001.

Tosi, S., Nieh, J.C., 2019. Lethal and sublethal synergistic effects of a new systemic pesticide, flupyradifurone (Sivanto), on honeybees. Proc. R. Soc. B Biol. Sci. 286, 20190433. https://doi.org/10.1098/rspb.2019.0433.

Tosi, S., Costa, C., Vesco, U., Quaglia, G., Guido, G., 2018. A 3-year survey of Italian honey bee-collected pollen reveals widespread contamination by agricultural pesticides. Sci. Total Environ. 615, 208–218. https://doi.org/10.1016/j.scitotenv.2017.09.226.

Tosi, S., Nieh, J.C., Sgolastra, F., Cabbri, R., Medrzycki, P., 2017. Neonicotinoid pesticides and nutritional stress synergistically reduce survival in honey bees. Proc. R. Soc. B Biol. Sci. 284, 20171711. https://doi.org/10.1098/rspb.2017.1711.

Van der Zee, R., Pisa, L., Andonov, S., Brodschneider, R., Charrière, J.-D., Chlebo, R., Coffey, M.F., Crailsheim, K., Dahle, B., Gajda, A., Gray, A., Drazic, M.M., Higes, M., Kauko, L., Kence, A.,

Kence, M., Kezic, N., Kiprijanovska, H., Kralj, J., Kristiansen, P., Martin Hernandez, R., Mutinelli, F., Kim Nguyen, B., Otten, C., Özkırım, A., Pernal, S.F., Peterson, M., Ramsay, G., Santrac, V., Soroker, V., Topolska, G., Uzunov, A., Vejsnaes, F., Wei, S., Wilkins, S., 2012. Managed honey bee colony losses in Canada, China, Europe, Israel and Turkey. J. Apic. Res. 51, 100–114. https://doi.org/10.3896/IBRA.1.51.1.12.

Vanengelsdorp, D., Meixner, M.D., 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. J. Invertebr. Pathol. 103, S80–S95. https://doi.org/10.1016/j.jip.2009.06.011.

Vidau, C., Diogon, M., Aufauvre, J., Fontbonne, R., Viguès, B., Brunet, J.-L., Texier, C., Biron, D.G., Blot, N., El Alaoui, H., Belzunces, L.P., Delbac, F., 2011. Exposure to sublethal doses of fipronil and thiacloprid highly increases mortality of honeybees previously infected by Nosema ceranae. PLoS One 6, e21550. https://doi.org/10.1371/journal.pone.0021550.

Von Luxburg, U., Franz, V.H., 2009. A geometric approach to confidence sets for ratios: Fieller's theorem, generalizations, and bootstrap. Stat. Sin. https://doi.org/10.2307/24308947.

Wade, A., Lin, C.-H., Kurkul, C., Regan, E.R., Johnson, R.M., 2019. Combined toxicity of insecticides and fungicides applied to California almond orchards to honey bee larvae and adults. Insects 10, 20. https://doi.org/10.3390/insects10010020.

Wilkins, S., Jarratt, N., Harkin, S., Thompson, H., Coulson, M., 2013. Effects of solvent on the toxicity of dimethoate in a honey bee in vitro larval study. Pest Manag. Sci. 69, 462–463. https://doi.org/10.1002/ps.3465.

Williams, I.H., 1994. The dependence of crop production within the European Union on pollination by honey bees. Agric. Zool. Rev. 6, 229–257.

Williamson, S.M., Wright, G.A., 2013. Exposure to multiple cholinergic pesticides impairs olfactory learning and memory in honeybees. J. Exp. Biol. 216, 1799–1807. https://doi.org/10.1242/jeb.083931.

Wu, G., Miyata, T., Kang, C.Y., Xie, L.H., 2007. Insecticide toxicity and synergism by enzyme inhibitors in 18 species of pest insect and natural enemies in crucifer vegetable crops. Pest Manag. Sci. 63, 500–510. https://doi.org/10.1002/ps.1361.

Yao, J., Zhu, Y.C., Adamczyk, J., 2018. Responses of honey bees to lethal and sublethal doses of formulated clothianidin alone and mixtures. J. Econ. Entomol. 111, 1517–1525. https://doi.org/10.1093/jee/toy140.

Zhu, Y.C., Yao, J., Adamczyk, J., Luttrell, R., 2017. Feeding toxicity and impact of imidacloprid formulation and mixtures with six representative pesticides at residue concentrations on honey bee physiology (Apis mellifera). PLoS One 12, e0178421. https://doi.org/10.1371/journal.pone.0178421.

Chapter 4 **Predicting acute contact toxicity of pesticides in honeybees (*Apis mellifera*) through a k-nearest neighbor model**

Francesca Como [a,*], Edoardo Carnesecchi [b], Stefania Volani [b], Jean Lou Dorne [b], Jane Richardson [b], Arianna Bassan [c], Manuela Pavan [c], Emilio Benfenati [a]

[a] Laboratory of Chemistry and Environmental Toxicology, Department of Environmental Health, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, via Mario Negri, 2, 20156 Milano, Italy

[b] European Food Safety Authority (EFSA), Scientific Committee and Emerging Risks Unit, Via Carlo Magno 1A, 43126 Parma, Italy

[c] S-IN Soluzioni Informatiche S.r.l., via G. Ferrari 14, 36100 Vicenza, Italy

[*] Corresponding author

Published in Chemosphere, 2017, 166: 438-444

Abstract

Ecological risk assessment of plant protection products (PPPs) requires an understanding of both the toxicity and the extent of exposure to assess risks for a range of taxa of ecological importance including target and non-target species. Non-target species such as honey bees (Apis mellifera), solitary bees and bumble bees are of utmost importance because of their vital ecological services as pollinators of wild plants and crops. To improve risk assessment of PPPs in bee species, computational models predicting the acute and chronic toxicity of a range of PPPs and contaminants can play a major role in providing structural and physico-chemical properties for the prioritisation of compounds of concern and future risk assessments. Over the last three decades, scientific advisory bodies and the research community have developed toxicological databases and quantitative structure-activity relationship (QSAR) models that are proving invaluable to predict toxicity using historical data and reduce animal testing. This paper describes the development and validation of a k-Nearest Neighbor (k-NN) model using in-house software for the prediction of acute contact toxicity of pesticides on honey bees. Acute contact toxicity data were collected from different sources for 256 pesticides, which were divided into training and test sets. The k-NN models were validated with good prediction, with an accuracy of 70% for all compounds and of 65% for highly toxic compounds, suggesting that they might reliably predict the toxicity of structurally diverse pesticides and could be used to screen and prioritise new pesticides.

Keywords: Pesticides, Honey bees, k-NN, In silico models, Acute contact toxicity.

1. Introduction

Pesticides as plant protection products (PPPs) have an important role in agriculture through the protection of crops from pests such as insects, plant diseases, fungus and weeds and consequently improve productivity and generally speaking food security. From an environmental perspective, scientific advisory bodies and governmental agencies have developed, over the last three decades, methods and frameworks to assess ecological risks of PPPs for a range of taxa including target and non-target species such as birds, bees and fish. The basic principles of these methods and frameworks follow the classic steps for risk assessment (hazard identification and characterization), exposure assessment combining the hazard (toxicological) dimension and the exposure dimension for risk characterization (EFSA, 2012; Dorne and Fink-Gremmels, 2013).

Non-target species of utmost importance in an environmental risk assessment include honey bees (Apis mellifera), solitary bees and bumble bees because of their vital ecological services as pollinators of wild plants and crops. The honey bee, Apis mellifera, is arguably the most important pollinator of agricultural crops and is especially exposed to chemicals when visiting melliferous plants (Klein et al., 2007). Special attention has therefore been paid to their protection, not only for their ecological importance by contributing to the maintenance of wild plant biodiversity, but also for their economic value as honey producers and crop-pollinating agents (EFSA, 2012), in fact, new methods of risk assessment specific to bees have been developed recently; see Sanchez-Bayo and Goka (2014). A large number of research studies have investigated pesticide toxicity on bees, new risk assessment methods for bees and the impact of single and multiple stressors (chemicals, diseases, nutrition, etc.) on honey bee colony losses influence honey bee colony losses, and poisoning incidents (Gallai et al., 2009; EFSA, 2012, 2013, 2014, 2015). During their foraging flights, while they collect nectar, pollen, plant resins, and ingest the nectar and water, honey bees inadvertently come into contact with a wide array of inorganic and organic xenobiotics, often taking them back to the colony where they may induce lethal and sub-lethal effects. Pesticide sensitivity in honey bees depends on the intrinsic toxicity of the chemical in relation to its structure and target as well as it toxicokinetics including its persistence in the organism and whether metabolism generate toxic metabolites or detoxifies the parent compound (EFSA, 2012). Many environmental factors can also modulate pesticide sensitivity in bees such as the season (Meled et al., 1998; Decourtye et al., 2003), age (Guez et al., 2001), brood rearing temperature (Medrzycki et al., 2010), and stressors such as mixtures of chemicals and pathogens (Alaux et al., 2010; Aufauvre et al., 2012). Another aspect of the sensitivity of bee species (including honey bees, solitary bees and bumble bees), has been demonstrated with the recent sequencing of the honey bee genome. Indeed, comparison with other insect genomes showed that the honey bee genome is markedly deficient in a number of genes encoding detoxification enzymes, including cytochrome P450 monooxygenases (P450s), glutathione-

S-transferases and carboxylesterases compared with other insects (Claudianos et al., 2006; Johnson, 2015).

Toxicity of PPPs in honey bees has been currently assessed using 24h or 48h acute toxicity laboratory bioassays to determine the median lethal concentration, LC_{50} (concentration that induces 50 percent death) or lethal dose, LD_{50} (dose that induces 50 percent death) as described in the OECD test guidelines (OECD,1998) and in European regulation (European Regulation (EU) No. 283/2013). To assess PPPs exposure honey bees for different routes (e.g. contact and oral routes) and exposure scenarios need to be addressed and these will depend on the product formulation and its intrinsic characteristics such as its physico-chemical properties. Such physico-chemical properties contribute to environmental persistence, spatial dispersal and potential contamination of feed sources for honey bees (e.g. nectar, bee bread). Overall, the risk assessment thus combines exposure data for possible routes of exposure such as the administered dose or concentration of a toxicant, the duration of the exposure and toxicity data for the active ingredient. In the last thirty years, scientific advisory bodies and the research community have developed toxicological databases and computational models such as quantitative structure-activity relationship (QSAR) models that are invaluable to predict hazard (toxicity) using historical data, reduce animal testing and serve as tools for regulatory purposes for both the prioritization of compounds of concern and future risk assessments. These QSAR models for species of ecological importance are often developed using the well-characterised relationships between the physico-chemical properties of chemicals, their persistence and toxicity as well as their global environmental fate (Domine et al., 1992; Devillers and Flatin, 2000). QSAR models have already been applied to the prediction of ecotoxicological endpoints in species of ecological importance including trout, daphnia, quail and bees within the project DEMETRA (Benfenati et al., 2011) and more recently for qualitative and quantitative toxicity prediction in bees using a global quantitative structure-toxicity relationship model (QSTR) (Singh et al., 2014). To address the needs for QSAR models predicting toxicity of PPPs in honey bees, we developed an in-house software using databases on acute contact data in honey bees from different sources and a k-Nearest Neighbor algorithm (k-NN).

2. Materials and Methods

2.1. Data

Toxicity data on honey bee acute contact toxicity were collected from different sources. The first one was the DEMETRA project (Benfenati et al., 2011). The second source was the Terrestrial USF EPA ECOTOX database present in the OECD QSAR Toolbox, vers. 3.3 (www.qsartoolbox.org). The third set of data was provided by the European Food Safety Authority (EFSA, http://www.efsa.europa.eu/it/). Criteria for data pruning were established following the official guideline (OECD, 1998) according to which pesticides are administered by contact routes to

represent the type of exposure under field conditions. *Apis mellifera* is used as a surrogate for assessing risks to bees. The mortality of honey bees is recorded after 48 h of exposure and results are presented in terms of mg active substance/bee as the median lethal dose (LD_{50} mg active substance/bee). The final dataset (in annex) was organized as a list of Simplified Molecular Input Line Entry System (SMILES) strings (Weininger, 1998) each one coupled with its experimental data and source. To obtain the chemical structure from each SMILES we used JChem for Office (Excel) (JChem for Office, version number: 15.3.2300.2514,2013, ChemAxon; http://www.chemaxon.com). Since in Europe no specific regulatory threshold has been set for 48 h ecotoxicological endpoints, reference thresholds were applied from the Pesticide Properties Database (PPDB) (http://sitem.herts.ac.uk/aeru/iupac/docs/Background_and_Support.pdf). The PPDB classifies pesticides as low toxicity for LD_{50} greater than 100 µg/bee, moderately toxic for LD_{50} between 1 and 100 µg /bee, and highly toxic for LD_{50} lower than 1 µg/bee. Based on these thresholds and criteria, three classes of pesticides were applied to both training and test sets (Table 1). Thus, we kept compounds associated with a well-defined numerical experimental value (precede by the symbol =) and with an experimental value up to 1 µg/bee or over 100 µg/bee. During the data pruning, acute contact toxicity data were not included in the database when any one of these four conditions was not fulfilled: (i) no specific structure available, (ii) inorganic compounds, (iii) no information on laboratory test duration or test duration less than 48 h (e.g. 24h), and (iv) laboratory toxicological test doses expressed as mg/kg or mg/m^3 (in all the other cases doses were normalized as µg/bee). For compounds with more than one experimental value, the associated variability was evaluated using the strategy employed by Benfenati and collaborators (Benfenati et al., 2011) as the ratio between duplicated experimental data (x/y) as follows:

a. If x/y was ≤5 the average of the experimental data was kept;

b. If x/y was >5 the compound was removed from the dataset.

Furthermore, chemicals with values preceded by the qualifier symbol, greater or lower than a value between 1 and 100 µg/bee with the extremes value excluded (e.g. >30 or <30), were not included. Indeed, it was not possible to classify them using the thresholds discussed above. After pruning, the database contained 256 pesticides with acute contact toxicity LC_{50}: 64 from EFSA, 86 from the DEMETRA project and 113 from the OECD toolbox. These 256 compounds were divided into a training set (80%) to develop the k-NN model and a test set (20%) to validate the model. For this reason, the compounds were sorted in ascending order relatively to the experimental value. Starting from the lowest to the highest experimental value, a compound every five of the database has been selected, and it has been included in the test set, finally composed by 51 compounds. The remaining number of compounds, 205, constituted the training set. The training set and test set

datasets (structures and experimental data) have been added as supplementary material (Appendix 4A).

2.2. Software

To build the k-NN we used in-house software istkNN, previously described (Manganaro et al., 2016). The algorithm of this software is a similarity-based approach that predicts the property of a substance in relation to the experimental data for the most similar compounds "nearest neighbors" on the training set. The algorithm of istKNN software is based on the first "k" compounds of the training set able to satisfy the similarity requirement related to the target compound, indicated by threshold "S1". If only one molecule satisfies this requirement then another, stricter threshold, indicated by threshold "S2", is applied. If no molecule satisfies this similarity requirement, no prediction is provided (missing value). The istkNN software takes as input a set of compounds expressed as SMILES and their related experimental values, each consisting of a set of vectors and class label for each vector. In the simplest case, it will be only a positive or negative value (+ or -). But the k-NN model can work equally well with an arbitrary number of classes (Zhang and Zhou, 2005). The istKNN software can provide predictions for a qualitative dataset or quantitative dataset. In the first case the final prediction will be a classification label, while in the second it will be a continuous value. The istKNN software also provides a batch mode, which allows the user to explore several possible parameter settings (K; S_1; S_2; E) to produce automatically a large number of k-NN models on the training set. Each k-NN model is defined by the settings chosen for the parameters, so users can analyse the output and choose the most suitable k-NN model for their purpose.

2.3. Development of the model

Since the dataset's experimental values consist of a classification label among three possible classes, as described in section 2.1, the qualitative approach was selected to develop the k-NN model on the training set.We notice that in most cases the regression in silico models use values in molar concentration. However, regulatory thresholds are given in weight. To be closer to the potential use for regulatory purposes, we simply adopted the threshold value given in weight. We used the batch mode to explore several possible parameter settings. The range of the number of substances (K), and of the thresholds (S1 and S2) were set as indicated:

- "K" = 2, 3, 4
- "S1" = 0.70, 0.75, 0.80, 0.85, 0.90
- "S2" = 0.70, 0.75, 0.80, 0.85, 0.90

After a close evaluation of the output of the batch mode, we selected the k-NN model with the best predictive power, with the following combination of parameters: K = 4; S1 = 0.70, S2 = 0.75. This means that four substances were used to provide the estimate, and that the similarity for these

substances was at least 0.70. However, if only one substance was found for modeling a certain target compound, the minimal similarity was 0.75. The selected model was developed and validated on the test set.

2.4. Statistical parameters

The k-NN models were evaluated using three standard parameters to depict the reliability of the classifiers: accuracy (Eq. (1)), sensitivity (Eq. (2)) and specificity (Eq. (3)). True positive (TP) and true negative (TN) compounds were considered as, respectively, toxic and non-toxic from the experimental data and the predictions. While false positive (FP) and false negative (FN) compounds were misclassified as toxic and non-toxic by the predictive model.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \quad \text{(Eq. 1)}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad \text{(Eq. 2)}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}} \quad \text{(Eq. 3)}$$

The Matthews Correlation Coefficient (MCC) (Eq. (4)) is generally regarded as a balanced measure for binary classifications and can be used for toxicity classes of different sample sizes. It was applied here as an indicator of the quality of binary classifications

$$\text{MCC} = \frac{\text{TP}*\text{TN}-\text{FP}*\text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \quad \text{(Eq. 4)}$$

MCC values range between -1 and +1: a value of +1 represents a perfect prediction, a value of 0 no better than random prediction, and a value of -1 indicates total disagreement between predicted and observed values (Dao et al., 2011). MCC values were calculated for both the training and test set since the three classes of compounds are not equally represented (Table 4.1). The most representative class, with 118 and 29 chemicals in the training and test sets, is the one with pesticides classified as low toxic, i.e. compounds with experimental LD_{50} greater than 100 µg/bee. The second class, with 51 and 14 chemicals in the training and test sets has pesticides classified as moderately toxic, i.e. with experimental LD_{50} between 1 and 100 µg/bee. The last class which is smaller in the training and test sets, has 36 and 8 chemicals, and the pesticides are classified as highly toxic (LD_{50} up to 1 µg/bee). Since the PPDB classification for contact acute toxicity in honey bees defined three classes of toxicity, these equations (1)-(4), which apply to a binary classification, were adapted using the method previously described (Diaza et al., 2015). Briefly, to bring back the

situation in a binary classification (toxic/non toxic) we analyzed the situation in two splits. In the first step, we considered exclusively the thresholds of 1 µg/bee, considering as positive (toxic) the chemicals with an experimental value equal or below to 1 µg/bee and negative (non toxic) all the others compounds. In a second step, we considered exclusively the threshold of 100 µg/bee, considering as positive (toxic) the chemicals with an experimental value equal or below to 100 µg/bee and negative (non toxic) all the others compounds.

Table 4.1. Number of compounds in the three classes, in the training and test sets.

Threshold	No. of compounds	Level of toxicity
Training set		
≤1 µg/bee	36	high
1 < X ≤ 100 µg/bee	51	moderate
>100 µg/bee	118	low
Test set		
≤1 µg/bee	8	high
1 < X ≤ 100 µg/bee	14	moderate
>100 µg/bee	29	low

3. Results

We developed a k-NN model using a training set, comprised by 205 compounds with 205 unique experimental values on honey bee acute contact toxicity. Toxicity data were collected from different sources: the DEMETRA project (Benfenati et al., 2011), the Terrestrial US-EPA ECOTOX database (www.qsartoolbox.org), and EFSA's openfoodtox database. The k-NN model developed is a classifier model under which the algorithm simply stores all its training examples, relating chemical structure and acute contact toxicity of PPPs in honey bees. To classify a new, unlabeled instance, the algorithm compares it to all stored instances with some defined similarity measure, and determines the k nearest neighbors. The class label assigned to the new instance can then be derived from the class labels of the nearest neighbors. The utility of the k-NN algorithm has been explored in other text classification applications (Sebastiani, 2002; Illa et al., 2004). For labelling compounds in the three classes described in the Materials and Methods sector, we referred to the two threshold values, 1 and 100, from PPDB. Based on these two thresholds we classified the 205 compounds of the training set as low, moderately and high toxic depending on their experimental LD_{50}. Then the model was validated on the 51 compounds of the test set, to check its predictive power.

3.1. Performance of the models

The k-NN model was unable to predict 13 compounds in the training set and only 1 in the test set, corresponding to 9% and 2% of the compounds tested, respectively. Table 4.2 shows the confusion matrix of the model on the training set (a) and test set (b). The diagonal lists the correctly classified compounds. The accuracy of the model was 0.70 in the training set and 0.76 in the test set. The performance of the k-NN models was tested with the statistical parameters usually used to test classification models. Statistical parameters were calculated for the training and test sets. Since for acute contact toxicity in honey bees the PPDB classification identifies three classes accuracy, sensitivity, specificity and MCC were calculated in two different splits. In the first we considered the threshold of 1 µg/bee; so compounds with experimental values < 1 µg/bee were considered positive and compounds with experimental values > 1 µg/bee were considered negative. In the second split we considered the threshold of 100 µg/bee, so compounds with experimental values < 100 µg/bee were considered positive and those with experimental values > 100 µg/bee were consider negative. The first split is the most stringent since only compounds that are "positive" for sure were considered.

Table 4.2. Confusion matrix on the training set (a) and test set (b). Diagonally and in bold is reported the number of compounds correctly predicted.

a)

Training set	**Predicted**			
	$LD_{50} \leq 1$ µg/bee	1 µg/bee < $LD_{50} \leq 100$ µg/bee	$LD_{50} > 100$ µg/bee	Total
Experimental				
$LD_{50} \leq 1$ µg/bee	22	5	7	34
1 µg/bee < $LD_{50} \leq 100$ µg/bee	8	14	25	47
$LD_{50} > 100$ µg/bee	2	11	98	111
Total	32	30	130	192

b)

Test Test	Predicted			
	$LD_{50} \leq 1$ µg/bee	1 µg/bee < $LD_{50} \leq 100$ µg/bee	$LD_{50} > 100$ µg/bee	Total
Experimental				
$LD_{50} \leq 1$ µg/bee	6	1	1	8
1 µg/bee < $LD_{50} \leq 100$ µg/bee	3	7	3	13
$LD_{50} > 100$ µg/bee	2	2	25	29
Total	11	10	29	50

3.1.1. Split 1 (threshold 1 µg/bee)

In split 1 we only considered the threshold of 1 µg/bee. Compounds with experimental values up to mg/bee were considered positive and compounds with experimental values greater than 1 µg/bee were considered negative. Table 4.3 (a) shows the confusion matrix and statistical parameters of this model on the training set. The accuracy is 0.88 since in this case only compounds that are "positive" for sure were considered. Specificity, sensitivity and MCC were good too. Table 4.3 (b) shows the validation of this model on the test set, which was comparable to the training set. We have not found values for other studies addressing classification of substances in insects. Cappelli et al., (2015) reported an evaluation of in silico models for fish acute toxicity. Accuracy ranged from 0.68 to 0.86 depending on the software. In the case of the highly toxic compounds accuracy values range from 0.80 to 0.49. In particular if we consider the prediction on the highly toxic compounds (< 1 µg/bee) there are 28 chemicals correctly predicted out of 43 (Table 4.3 a) this corresponds to an accuracy of 0.65.

Table 4.3. Split 1 (threshold 1 µg/bee) results of the training set (a) and validation of the model on the test set (b). The numbers in bold correspond to the sum of the compounds of each subset.

a)			
Training set	**Predicted**		
	$LD_{50} \leq 1$ µg/bee	$LD_{50} > 1$ µg/bee	Total
Experimental			
$LD_{50} \leq 1$ µg/bee	**22 (TP)**	5+7= **12 (FN)**	34
$LD_{50} > 1$ µg/bee	8+2= **10 (FP)**	14+25+11+98= **148 (TN)**	158
Total	32	160	192
Statistical parameters	**Statistical indicators**		
Accuracy 0.88	TP True positive		
Sensitivity 0.68	FN False negative		
Specificity 0.93	TN True negative		
MCC 0.59	FP False positive		
b)			
Test Test	Predicted		
	$LD_{50} \leq 1$ µg/bee	$LD_{50} > 1$ µg/bee	Total
Experimental			
$LD_{50} \leq 1$ µg/bee	**6 (TP)**	1+1= **2 (FN)**	8
$LD_{50} > 1$ µg/bee	3+2= **5 (FP)**	7+3+2+25= **37 (TN)**	42
Total	11	39	50
Statistical parameters	**Statistical indicators**		
Accuracy 0.86	TP True positive		
Sensitivity 0.75	FN False negative		
Specificity 0.88	TN True negative		
MCC 0.55	FP False positive		

3.1.2. Split 2 (threshold 100 µg/bee)

In split 2 we considered the threshold of 100 mg/bee. Compounds with experimental values up to 100 mg/bee were considered positive and compounds with experimental values greater than 100 mg/bee negative. Table 4.4 (a) shows the confusion matrix and statistical parameters of this model on the training set. Accuracy equal to 0.76, in this case was lower than with the threshold of 1

mg/bee, since compounds with a moderate level of toxicity here were considered "positive". Sensitivity, specificity and MCC, however, were higher than with the lower threshold.

Table 4.4. Split 2 (threshold 100 µg/bee) results of the training set (a) and validation of the model on the test set (b). The numbers in bold correspond to the sum of the compounds of each subset.

a)			
Training set	**Predicted**		
	$LD_{50} \leq 100$ µg/bee	$LD_{50} > 100$ µg/bee	Total
Experimental			
$LD_{50} \leq 1$ µg/bee	22+5+8+14= **49 (TP)**	7+25= **32 (FN)**	81
$LD_{50} > 1$ µg/bee	2+11= **13 (FP)**	**98 (TN)**	111
Total	62	130	192
Statistical parameters	**Statistical indicators**		
Accuracy 0.76	TP True positive		
Sensitivity 0.60	FN False negative		
Specificity 0.88	TN True negative		
MCC 0.51	FP False positive		
b)			
Test Test	Predicted		
	$LD_{50} \leq 100$ µg/bee	$LD_{50} > 100$ µg/bee	Total
Experimental			
$LD_{50} \leq 1$ µg/bee	6+1+3+7= **17 (TP)**	1+3= **4 (FN)**	21
$LD_{50} > 1$ µg/bee	**4 (FP)**	**25 (TN)**	29
Total	21	29	50
Statistical parameters	**Statistical indicators**		
Accuracy 0.84	TP True positive		
Sensitivity 0.80	FN False negative		
Specificity 0.86	TN True negative		
MCC 0.67	FP False positive		

4. Discussion

This K-NN model to predict acute contact toxicity in honey bees is a classifier for screening pesticides of concern, and has been validated using training and tests sets. In the future, it is foreseen that

such a model may serve as a screening for pesticides for the purpose of prioritisation or further development of in silico models to reduce in vivo testing. There are still only very few QSAR models for the honey bee and few studies of adverse effects of pesticides on insect species (Lo Piparo et al., 2006; Devillers et al., 2003, 2015). Within the DEMETRA project, a quantitative QSAR model has been developed for honey bees (Benfenati et al., 2011). The previous model is quantitative, while the present one is a classifier. Another major difference is that the present model starts from a larger datasets combining the DEMETRA database, the terrestrial US-EPA ECOTOX database and the EFSA's toxicological database, and may have broader applicability. The present model is much simpler compared with the DEMETRA model, since it does not require the calculation of a number of descriptors or values using the model algorithm. The present model does not require any descriptor. Thus, the present model may find a broader spectrum of application, particularly because it is implemented within the open source platform VEGA (www.vega-qsar.eu) (Benfenati et al., 2013).

5. Conclusion

Pesticides are biologically active compounds, which include many functional groups and chemical structures, and building computational models for this heterogeneous situation remains a challenge. The validation of the k-NN models in this work suggests their suitability to reliably predict the toxicity of structurally diverse pesticides so they have the potential to be applied to the screening and prioritisation of new pesticides in the future. To make the model available to the scientific community, the KNN has been integrated into the VEGA platform (http://www.vegaqsar.eu/) ensuring free and easy access to users. Implementation within the VEGA platform will also give a clear quantitative value for the measurement of the applicability domain, according to requirements from scientific advisory bodies (ECHA, 2016). Further research to develop QSAR models for bees will be important for risk assessment of PPPs and will provide a way to reduce testing when sufficient data are available. In the future, with the availability of new data, a model could be developed to predict acute toxicity in bees for the oral route, using the current K-NN QSAR models presented in this paper. The method developed in this paper was tested on the currently available data on oral toxicity, from EFSA, DEMETRA and OECD toolbox, predictions did not perform compared with the data for contact toxicity. A recent OECD guideline has proposed revisions of laboratory toxicity testing of PPPs in adult bees to move to 10-day chronic feeding studies to replace 48 h tests (OECD, 2016); when such data become available, another predictive model can be built for the 10-day toxicity datasets using the K-NN model as a starting point.

Acknowledgements

We acknowledge from the LIFE project PROSIL (www.prosil-life.eu/) LIFE12 ENV/IT/000154 financial support and from the EFSA framework awarded to S-IN (OC/EFSA/SCER/2014/01 'Further development and update of EFSA's Chemical Hazards Database') the data collection of the acute contact toxicity data in honey bees.

References

Alaux, C., Brunet, J.L., Dussaubat, C., Mondet, F., Tchamitc Han, S., Cousin, M., Brillard, J., Baldy, A., Belzunces, L.P., Le Conte, Y., 2010. Interactions between Nosema microspores and a neonicotinoid weaken honeybees (Apis mellifera). Environ. Microbiol. l12 (3), 774-782.

Aufauvre, J., Biron, D.G., Vidau, C., Fontbonne, R., Roudel, M., Diogon, M., Vigues, B., Belzunces, L.P., Delbac, F., Blot, N., 2012. Parasite-insecticide interactions: a case study of Nosema ceranae and fipronil synergy on honeybee. Sci. Rep. 2012 (2), 326. Epub 2012 Mar 22.

Benfenati, E., Boriani, E., Craciun, M., Malazizi, L., Neagu, D., Roncagioni, A., 2011. Databases for pesticide ecotoxicity. In: Benfenati, E. (Ed.), Quantitative Structure-activity Relationships (QSAR) for Pesticide Regulatory Purposes. Elsevier, pp. 59-81.

Benfenati, E., Pardoe, S., Martin, T., Gonella Diaza, R., Lombardo, A., Manganaro, A., Gissi, A., 2013. Using toxicological evidence from QSAR models in practice. ALTEX 30 (1), 19-40.

Cappelli, C.I., Cassano, A., Golbamaki, A., Moggio, Y., Lombardo, A., Colafranceschi, M., Benfenati, E., 2015. Assessment of in silico models for acute aquatic toxicity towards fish under REACH regulation. SAR QSAR Environ. Res. 26 (12), 977-999.

Claudianos, C., Ranson, H., Johnson, R.M., Biswas, S., Schuler, M.A., Berenbaum, M.R., Feyereisen, R., Oakeshott, J.G., 2006. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Mol. Biol. 15, 615-636.

Dao, P., Wang, K., Collins, C., Ester, M., Lapuk, A., Sahinalp, S.C., 2011. Optimally discriminative subnetwork markers predict response to chemotherapy. Bioinformatics 27, 205-213.

Decourtye, A., Lacassie, E., Pham-Delegue, M.H., 2003. Learning performances of honeybees (Apis mellifera) are differentially affected by imidacloprid according to the season. Pest Manag. Sci. 59, 269-278.

Devillers, J., Flatin, J., 2000. A general QSAR model for predicting the acute toxicity of pesticides to Oncorhynchus mykiss. SAR QSAR Environ. Res. 11, 25-43.

Devillers, J., Decourtye, A., Budzinski, H., Pham-Delègue, M.H., Cluzeau, S., Maurin, G., 2003. Comparative toxicity and hazards of pesticides to Apis and non-Apis bees. A chemometrical study. SAR QSAR Environ. Res. 14 (5e6), 389-403.

Devillers, J., Doucet-Panaye, A., Doucet, J.P., 2015. Structureeactivity relationship (SAR) modelling of mosquito larvicides. SAR QSAR Environ. Res. 26 (4), 263-278.

Diaza, R.G., Manganelli, S., Esposito, A., Roncaglioni, A., Manganaro, A., Benfenati, E., 2015. Comparison of in silico tools for evaluating rat oral acute toxicity. SAR QSAR Environ. Res. 26 (1), 1-27.

Domine, D., Devillers, J., Chastrette, M., Karcher, W., 1992. Multivariate structure-property relationships (MSPR) of pesticides. Pestic. Sci. 35, 73-82.

Dorne, J.L., Fink-Gremmels, J., 2013. Human and animal health risk assessments of chemicals in the food chain: comparative aspects and future perspectives. Toxicol. Appl. Pharmacol. 270 (3), 187e195. http://dx.doi.org/10.1016/j.taap.2012.03.013, 1.

ECHA (European Chemicals Agency), 2016. Practical Guide 5 e How to Use and Report (Q)SARs, ECHA-16-B-09-EN.

EFSA (European Food Safety Authority), 2012. Scientific Opinion on the science behind the development of a risk assessment of Plant Protection Products on bees (Apis mellifera Bombus spp. and solitary bees). EFSA J. 10 (5), 266.

EFSA (European Food Safety Authority), 2013. Towards Holistic Approaches to the Risk Assessment of Multiple Stressors in Bees. EFSA supporting publication, p. 76, 2013:EN- 509.

EFSA (European Food Safety Authority), 2014. Modern methodologies and tools for human hazard assessment of chemicals. EFSA J. 12 (4), 13. http://dx.doi.org/10.2903/j.efsa.2014.3638, 3638.

EFSA (European Food Safety Authority), 2015. Harmonisation of Human and Ecological Risk Assessment of Combined Exposure to Multiple Chemicals. EFSA supporting publication, p. 39, 2015:EN-784.

European regulation (EU) No 283/2013, 2013. Setting Out the Data Requirements for Active Substances, in Accordance with Regulation (EC) No. 1107/2009 of the European Parliament and of the Council Concerning the Placing of Plant Protection Products on the Market.

Gallai, N., Salles, J.M., Settele, J., Vaissiere, B.E., 2009. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. Ecol. Econ. 68, 810e821.

Guez, D., Suchail, S., Gauthier, M., Maleszka, R., Belzunces, L.P., 2001. Contrasting effects of imidacloprid on habituation in 7-and 8-day-old honeybees (Apis mellifera). Neurobiol. Learn. Mem. 76 (2), 183-191.

Illa, J.M.G., Alonso, J.B., Marrè, M.S., 2004. Nearest neighbours for time series: special issue: machine learning policies. J. Appl. Intell. 20, 21-35.

Johnson, R.M., 2015 Jan 7. Honey bee toxicology. Annu. Rev. Entomol. 60, 415-434.

Klein, A.M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., et al., 2007. Importance of pollinators in changing landscapes for world crops. Proc. R. Soc. Lond B 274, 303-313.

Lo Piparo, E., Smiesko, M., Mazzatorta, P., Benfenati, E., Idinger, J., Blümel, S., 2006. Preliminary analysis of toxicity of benzoxazinones and their metabolites for Folsomia candida. J. Agric. Food Chem. 54 (4), 1099-1104.

Manganaro, A., Pizzo, F., Lombardo, A., Pogliaghi, A., Benfenati, E., 2016. Predicting persistence in the sediment compartment with a new automatic software based on the K-Nearest Neighbor (k-NN) algorithm. Chemosphere 144, 1624-1630.

Medrzycki, P., Sgolastra, F., Bortolotti, L., Bogo, G., Tosi, S., Padovani, E., Sabatini, A.G., 2010. Influence of brood rearing temperature on honey bee development and susceptibility to poisoning by pesticides. J. Apic. Res. 49 (1), 52-59.

Meled, M., Thrasyvoulou, A., Belzunces, L.P., 1998. Seasonal variations in susceptibility of Apis mellifera to the synergistic action of prochloraz and deltamethrin. Environ. Toxicol. Chem. 17, 2517e-2520. http://dx.doi.org/10.1002/etc.5620171220.

OECD, 1998. Test No. 214: Honeybees, Acute Contact Toxicity Test, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris.

OECD, 2016. Honey Bee (Apis mellifera l.), Chronic Oral Toxicity Test 10 Day Feeding Test in the Laboratory Draft Guideline February 2016. In: Side-effects on Honeybees, vol. 40. EPPO Bulletin, pp. 313e319. http://dx.doi.org/10.1111/j.1365-2338.2010.02418.x. PP 1/170(4) (2010). https://www.oecd.org/env/ehs/testing/Draft%20TG%2010d%20Honeybee%20feeding_Feb%202016.pdf.

Sanchez-Bayo, F., Goka, K., 2014. Pesticide residues and bees e a risk assessment. PLoS one 9 (4), e94482. http://dx.doi.org/10.1371/journal.pone.0094482.

Sebastiani, F., 2002. Machine learning in automated text categorization. ACM Comput. Surv. 34, 1-47.

Singh, K.P., Gupta, S., Basant, N., Mohan, 2014. QSTR modeling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes.D. Chem. Res. Toxicol. 27 (9), 1504e1515, 2014 Sep. 15.

Weininger, D., 1998. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28(1), 31-26.

Zhang, M.L., Zhou, Z.H., 2005. A k-nearest neighbor based algorithm for multi-label classification. In: 2005 IEEE International Conference on Granular Computing, 2, pp. 718-721. http://dx.doi.org/10.1109/GRC.2005.1547385.

Chapter 5 **Predicting acute contact toxicity of organic binary mixtures in honey bees (*A. mellifera*) through innovative QSAR models**

Edoardo Carnesecchi [a,b,*], Andrey A. Toropov [b], Alla P. Toropova [b], Nynke Kramer [a], Claus Svendsen [c], Jean Lou Dorne [d], Emilio Benfenati [b]

[a] Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3584 Utrecht, the Netherlands
[b] Laboratory of Chemistry and Environmental Toxicology, Department of Environmental Health, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, via Mario Negri, 2, 20156 Milano, Italy
[c] Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford OX10 8BB, UK
[d] European Food Safety Authority (EFSA), Scientific Committee and Emerging Risks Unit, Parma, Italy
[*] Corresponding author

Published in Science of the Total Environment, 2020, 704: 135302

Abstract

Pollinators such as honey bees are of considerable importance, because of the crucial pollination services they provide for food crops and wild plants. Since bees are exposed to a wide range of multiple chemicals "mixtures" both of anthropogenic (e.g. plant protection products) and natural origin (e.g. plant toxins), understanding their combined toxicity is critical. Although honey bees are employed worldwide as surrogate species for Apis and non-Apis bees in toxicity tests, it is practically unfeasible to perform in vivo tests for all mixtures of chemicals. Therefore, Quantitative Structure-Activity Relationships (QSAR) models can be developed using available data and can provide useful tools to predict such combined toxicity. Here, three different QSAR models within the CORAL software have been calibrated and validated for honey bees (A. mellifera) to predict the acute contact mixtures potency ($LD_{50\text{-}mix}$), in two regression based-models, and the nature of combined toxicity (synergism / non-synergism) in a classification-based model. Experimental data on binary mixtures (n = 123) ($LD_{50\text{-}mix}$) including dose response data (n = 97) and corresponding Toxic Unit values were retrieved from EFSA databases. The models were built using the principle of extraction of attributes from SMILES (or quasi-SMILES) while calculating so-called correlation weights for these attributes using Monte Carlo techniques. The two regression models were validated for their reliability and robustness (R^2 = 0.89, CCC = 0.92, Q^2 = 0.81; R^2 = 0.87, CCC = 0.89, Q^2 = 0.75). The classification model was validated using sensitivity (=0.86), specificity (=1), accuracy (=0.96), and Matthews correlation coefficient (MCC = 0.90) as qualitative statistical validation parameters. Results indicate that these QSAR models successfully predict acute contact toxicity of binary mixtures in honey bees and can support prioritisation of multiple chemicals of concerns. Data gaps and further development of QSAR models for honey bees are highlighted particularly for chronic and sublethal effects.

Keywords: Mixtures toxicity, Honey bees, Quantitative structure-activity relationship, CORAL software, Monte Carlo method

1. Introduction

Honey bees (Apis mellifera), solitary bees and bumble bees represent important environmental non-target species particularly because of their crucial pollination services for food crops and their contribution to the maintenance and reproduction of wild plant communities and biodiversity (Klein et al., 2007; Vanengelsdorp and Meixner, 2010; Lambert et al., 2012). Honey bees are employed worldwide as surrogate species for Apis and non-Apis bees to perform toxicity tests on single pesticides (EFSA, 2013; U.S. EPA, 2014). In addition, they also represent sentinel species together with their hive products as bioindicators (i.e. honey, propolis, pollen) to investigate environmental contamination by regulated products (e.g. Plant Protection Products (PPPs), veterinary drugs) or anthropogenic (polycyclic aromatic hydrocarbons, heavy metals) and natural contaminants (mycotoxins, plant alkaloids and flavonoids) (Lambert et al., 2012; Johnson et al., 2012, 2013; Barganska et al., 2016; Tosi et al., 2018; Skorbiłowicz et al., 2018). As a matter of fact, bees are exposed to these as multiple substances "mixtures", either by foraging on contaminated areas or through contaminated food stored and consumed in the hive.

Over the last decade, scientific advisory bodies and governmental agencies have developed methods and frameworks to assess such mixtures issues (U.S. EPA, 2003; Kemi, 2015; EFSA, 2009; EFSA PPR Panel, 2012, 2013; EFSA, 2014; Backhaus and Faust, 2012; Kienzler et al., 2016; Nys et al., 2018). In this context, the recent EFSA MIXTOX guidance document (More et al., 2019) has illustrated methods and case studies in bees providing opportunities to investigate their contribution to bee health compared with other stressors (e.g. varroa, viruses) and to develop holistic risk assessment methodologies (EFSA, 2013, 2016; EFSA, 2017; Rortais et al., 2017). In order to further understand combined toxicity in honey bees, a recent meta-analysis (Carnesecchi et al., 2019) of acute contact laboratory toxicity assays on PPPs and veterinary drugs highlighted synergisms and antagonisms in 72% and 11% of datasets, respectively. For most observed synergisms, cytochrome P450 (CYP) inhibition was the major mechanism resulting in a decrease in elimination and an increase in the toxicity of the binary mixture (Carnesecchi et al., 2019; Johnson et al., 2013; Wade et al., 2019). Although the authors identified numerous data gaps, such combined toxicity databases potentially allow developing predictive Quantitative Structure-Activity Relationships (QSAR) tools particularly because it is rather impossible to test all possible mixtures in bees for their acute or chronic effects. Such QSAR tools are only available for single chemicals but to date not for the prediction of combined toxicity (Venko et al., 2017; Singh et al., 2014; Hamadache et al., 2018; Como et al., 2017). Hence, this manuscript describes the development and application of three innovative predictive QSAR models for honey bees within the CORAL software namely (i) two regression-based QSAR models predicting acute (contact) mixtures potency ($pLD_{50\text{-}mix}$) in a quantitative manner, and (ii) a classification-based model predicting the nature of combined toxicity

for organic binary mixtures (i.e. synergism / non-synergism). Calibration and validation of the models are described using available experimental data, simplified molecular input-line entry system (SMILES) and attributes. Validation is assessed using independent datasets and associated statistics (i.e. correlation weights) (Toropova et al., 2012; Toropov et al., 2012a, 2019). Finally, conclusions highlight the potential application of such in silico tools for the hazard assessment and prioritisation of organic binary mixtures in ecological risk assessment.

2. Materials and methods

2.1. Experimental data

Experimental data from laboratory studies on honey bees measuring the combined toxicity ($LD_{50\text{-}mix}$) and Toxic Units (TUs) following acute contact exposure to organic binary mixtures were retrieved from an EFSA database described in our recent meta-analysis (Carnesecchi et al., 2019). The database provides quantitative information (e.g. Toxic Unit, LD_{50}= 24 h) on 123 mixtures studies. First, a simple regression model (Approach A) to predict mixture potency (LD50-mix) was developed from the $LD_{50\text{-}mix}$ dataset (n = 123) while considering as input two chemical structures represented by simplified molecular input-line entry system (SMILES) (Table S1). As second step, a dataset including only dose response data (n = 97) on binary mixtures (Approach B) was created to develop (i) a regression-based model to predict the potency of the binary mixtures ($pLD_{50\text{-}mix}$), and (ii) a classification-based model to predict the nature of the combined toxicity (synergism/non-synergism), taking into account Toxic Units (TUs) for each chemical in the mixture (Table 5.1). Hence, in Approach B, quantitative data on TUs were used as additional features and were represented as quasi-SMILES (Toropov et al., 2018; Toropova et al., 2019a). The TU approach assumes that predictions for combined toxicity in the binary mixture follow the Concentration Addition (CA) model (Figure 5.1) given the quantitative composition of each chemical within the binary mixture in relation to their relative potency (Jonker et al., 2005). A detailed account of the methodologies for TU calculation is provided elsewhere (Carnesecchi et al., 2019). With regard to data pruning, no OECD guidelines are available for designing binary mixtures toxicity experiments in honey bees. Hence, it was not possible to follow any harmonised criteria in the data pruning steps, in contrast to what is available in the OECD guideline for testing single chemicals (OECD, 1998). Since the scientific literature most often reports in vivo $LD_{50\text{-}mix}$ as µg active substance/bee as a median lethal dose after 24 h, model results were expressed as $pLD_{50\text{-}mix}$ (i.e. negative decimal logarithm log [$1/LD_{50}$]), logarithm of the inverse of the lethal dose to kill 50% of honey bees in the tested sample (Toropova et al., 2012; Iwasa et al., 2004; Johnson et al., 2012, 2013).

Table 5.1. Overview of the QSAR models developed here. Input data (e.g. number and class of substances) and endpoints are reported for both models.

Input data (*n*)	Features	Statistical method	Classes and number (*n*) of substances	Endpoint/Model
		Approach A		
123	• $SMILES_A$ - $SMILES_B$	• Method Monte Carlo (MCC)	• Fungicide (13) • Insecticide/acaricide (12) • Synergist (3)	• Regression-based model ($pLD_{50\text{-}mix}$ 24h)
Input data (n)	**Features**	**Statistical method**	**Classes and number (n) of substances**	**Endpoint/Model**
		Approach B		
97	• $SMILES_A$ - $SMILES_B$ • Toxic Unit	• Method Monte Carlo (MCC) • Quasi-SMILES	• Insecticide/acaricide (11) • Fungicide (11) • Synergist (1)	• Regression based model ($pLD_{50\text{-}mix}$ 24h) • Classification-based model (synergism / non-synergism)

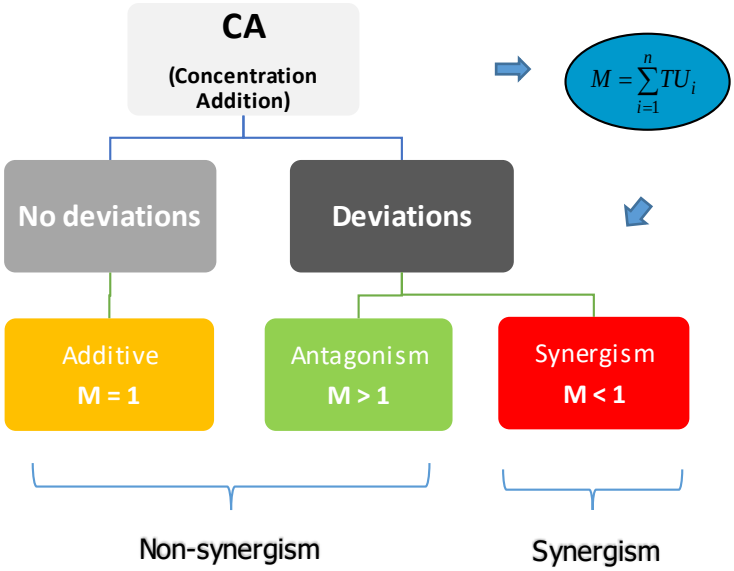


Figure 5.1. Decision tree for predicting combined toxicity of binary mixtures (M) (i.e. dose addition, synergism and antagonism) according to Concentration addition (CA) model and Toxic Unit (TU)

approach. Two classes (synergism and non-synergism) are identified as outputs of the classification model (approach B).

2.2. Development of the models

2.2.1. Approach A – Regression model ($LD_{50\text{-}mix}$)

The chemical structure of the two-component mixtures is represented using the disconnected simplified molecular input-line entry system (SMILES) (Table S1), where the "." (period or "dot") is used to represent disconnections (Hunter et al., 1987). The total number of data (n = 123) are randomly split into four sets: training set (≈25%; TRN), invisible training set (≈25%; iTRN), calibration set (≈25%; CLB), and validation set (≈25%; VLD) sets, each of which contains independent datasets and has a specific purpose:

1. The TRN set is the core dataset of the QSAR model. Compounds from this set are used to generate correlation weights giving maximal value of the target function using Monte Carlo optimisation;

2. The iTRN set inspects whether the model predictions are satisfactory using data for compounds that are independent from the TRN set;

3. The CLB set detects the start of the overtraining of the model.

4. The VLD set is used for the validation of the prediction model as a final step.

2.2.1.1. Optimal descriptor

The optimal descriptor used to develop the QSAR model for the combined toxicity of binary mixtures in honey bees (Approach A) is the following:

$$DCW(T^*,N^*) = \sum_{k=1} CW(S_k) + \sum_{k=1} CW(SS_k) + \sum_{k=1} CW(SSS_k) \quad (1)$$

The S_k is the "SMILES-atom" i.e. one or two symbols (e.g. 'C', 'N', 'O', etc.) and cannot be examined separately (e.g. 'Cl', 'Si', etc.). The SS_k is a combination of two SMILES-atoms. The CW(S_k), CW(SS_k), and CW(SSS_k) are so-called correlation weights of the above-mentioned attributes of SMILES. The numerical data on the CW(S_k), CW(SS_k), and CW(SSS_k) are calculated using the Monte Carlo method, i.e. the optimisation procedure which gives maximal value of target function (*TF*):

$$TF = r_{TRN} + r_{iTRN} - |r_{TRN} - r_{iTRN}| * 0.1 \quad (2)$$

Where the r_{TRN} and r_{iTRN} are correlation coefficients between observed and predicted endpoints for the training and invisible training sets, respectively.

2.2.1.2. Statistical criteria

In order to evaluate a regression model on combined toxicity of binary mixtures ($LD_{50\text{-}mix}$) in honey bees, the following statistical criteria are used: determination coefficient (R^2), cross-validated determination coefficient (Q^2) which measures prediction power, root mean squared error (RMSE), mean absolute error (MAE), Fischer F-ratio (F) and concordance correlation coefficient (CCC) (Roy et al., 2012; Chirico and Gramatica, 2012). The latter is defined as a complementary or alternative statistical criterion for external validation measures, particularly when other statistical criteria are in conflict. Results of the analyses are provided in Table 5.2.

2.2.2. Approach B – Classification and regression models based on Toxic Unit

Approach B aims at developing two QSAR models on experimental dose response data (n = 97) while using the CORAL software. TUs for each chemical in the binary mixture are used as additional features to develop i) a regression model to predict combined toxicity of the binary mixture ($LD_{50\text{-}mix}$) and ii) a classification model to predict the nature of the combined toxicity (synergism / non-synergism). In both models, toxicity data for binary mixtures are represented by the so-called quasi–simplified molecular input-line entry system (quasi-SMILES) which is analogue of the traditional SMILES applied in QSPR/QSAR analyses but makes use of all available data (not only information about the molecular structure) (Toropov et al., 2018). The total number of data (n = 97) were randomly split into the training (≈ 25%), invisible training (≈ 25%), calibration (≈25%), and validation (≈25%) sets.

2.2.2.1. Optimal descriptors

Two kinds of optimal descriptors are calculated for i) the regression models and ii) the classification models, respectively:

$$DCW(T^*,N^*) = \sum_{k=1}^{NA} CW(S_k) + \sum_{k=1}^{NA-1} CW(SS_k) + \sum_{k=1}^{NA-2} CW(SSS_k) \quad (3)$$

$$DCW(T^*,N^*) = \sum_{k=1}^{NA} CW(S_k) + \sum_{k=1}^{NA-1} CW(SS_k) \quad (4)$$

All parameters are already described in Eqs. (1). However, the number of attributes in SMILES (NA) is added for including TU values (TU_A, TU_B) as additional features. As consequence, QSAR-models

are calculated with the Monte Carlo optimisation based on two kinds of target functions TF_1 (see Eqs. 2) and TF_2:

$$TF_2 = TF_1 + IIC_{CLB} * 0.1 \quad (5)$$

Where, the IIC_{CLB} is calculated with data on the calibration (*CLB*) set as follows:

$$IIC_{CLB} = r_{CLB} \frac{\min(^{-}MAE_{CLB}, ^{+}MAE_{CLB})}{max(^{-}MAE_{CLB}, ^{+}MAE_{CLB})} \quad (6)$$

$$^{-}MAE_{CLB} = \frac{1}{^{-}N} \sum_{k=1}^{^{-}N} |\Delta_k|, \qquad \Delta_k < 0; ^{-}N \text{ is the number of } \Delta_k < 0 \quad (7)$$

$$^{+}MAE_{CLB} = \frac{1}{^{+}N} \sum_{k=1}^{^{+}N} |\Delta_k|, \qquad \Delta_k \geqslant 0; ^{+}N \text{ is the number of } \Delta_k \geqslant 0 \quad (8)$$

$$\Delta_k = observed_k - calculated_k \quad (9)$$

The observed and calculated are corresponding values of the endpoint. Having the numerical data on the $CW(S_k)$, $CW(SS_k)$ and $CW(SSS_k)$, the predictive model is calculated using the Least Squares method and data for compounds within the training set:

$$pEC_{50} = C_0 + C_1 * DCW(T^*, N^*) \quad (10)$$

2.2.2.2. Statistical criteria

Statistical criteria for the regression based QSAR model (Approach B) have been applied as described in section 2.2.1 (Approach A). In addition, TUs were used here as an additional feature of the models to improve the performance of the regression and classification of the models developed according to Approach B. In addition, other statistical criteria were used namely: sensitivity, specificity, accuracy, and Mattews correlation coefficient (MCC) to in order to build up the classification model for two classes i) synergism (1) and ii) non-synergism (0) (Toropova and Toropov, 2017; Toropov et al., 2012b). Generally, the MCC coefficient is applied in machine learning to measure the quality of binary classifications and it can be used when the classes present very different sizes (Dao et al., 2011).

$$Sensitivity = \frac{TP}{TP + FN} \quad (11)$$

$$Specificity = \frac{TN}{TN + FP} \quad (12)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (13)$$

$$MCC = \frac{TP \text{ x } TN - FP \text{ x } FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (14)$$

TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively, in a confusion matrix. MCC values range between -1 and +1 while the latter indicates a perfect prediction, a value of 0 indicate a prediction no better than a random one, and a value of +1 show total disagreement between predicted and observed values (Dao et al., 2011).

2.2.2.3. Model(s) validation and mechanistic interpretation (Approach A and B)

Model validation for both approaches (A and B) used statistical parameters for modelling internal (i.e. calibration) and external validations in order to estimate the predictive capability and the goodness of fitness of the QSAR models (Qin et al., 2018). With regard to regression based QSAR models, the statistical quality of the fitted equations was evaluated using the determination coefficient (R^2), concordance correlation coefficient (CCC), cross-validated determination coefficient (Q^2), root mean squared error (RMSE), mean absolute error (MAE) and Fischer F-ratio (F) (Roy et al., 2012). Results for the assessment of statistical quality of the regression models following approaches A and B are presented in Table 5.2 and 5.3, respectively (sections 3.1 and 3.2.1). Similarly, the classification model (synergism / non-synergism) has been validated according to the following statistical parameters: sensitivity, specificity, accuracy, and Mattews' correlation coefficient and results for the assessment of the statistical quality are presented in Table 5.4 (section 3.2.2).
The "Mechanistic interpretation" is defined as the causality between a substance and its toxicity (or not-toxicity). It is required when developing a QSAR model as described in the OECD guidelines (OCED, 2007), *"a mechanistic interpretation, if possible"* so that establishing a correlation and a causal relationship between the chemical structure of the compound and its toxicity (OECD principle 5) (Thoreau, 2016). The CORAL models provide the mechanistic interpretation in the form of promoters either increasing or decreasing potency of a chemical (Toropova and Toropov 2017, 2018; Toropov et al., 2019). Here, the mechanistic interpretation is obtained by means of the results of several runs of the Monte Carlo optimisation. In particular, molecular features extracted from SMILES, providing stable positive correlation weights in several runs of the Monte Carlo optimisation,

can be recognised as promoters of increase in the toxic potency of the mixture ($LD_{50\text{-}mix}$). In contrast, molecular features presenting only negative correlation weights in several runs of the optimisation are promoters of a decrease in the toxic potency of the mixture.

2.2.2.4. Applicability Domain (Approach A and B)

The applicability domain is an important component of QSAR analyses (OECD, 2007). According to OECD principle 3, a QSAR model should have a well-defined applicability domain. Applicability domain is defined as the area or chemical space represented by the molecular properties or structural information of the chemicals used for the model development. A collection of conceptions of applicability domains for different QSAR approaches is available and include: (i) physico-chemical domain, (ii) structural domain, (iii) response domain and (iv) integrated methods (Gadaleta et al., 2016). However, for models developed in the CORAL software, the statistical defects of SMILES calculated according to the distribution of available data into the training (TRN), invisible training (iTRN), calibration (CLB), and validation (VLD) sets are the basis to define the applicability domain. Here for both approaches (A and B) the same methods (Toropova et al., 2018) were applied for the assessment of the applicability domain defined according to distribution of SMILES attributes in the training and calibration sets following two steps:

Step 1: the definition of statistical defect (d_k) for each SMILES attribute (A_k = S_k, SS_k, SSS_k) involved (non-blocked) to construct the model;

$$d_k = \frac{|P(A_k) - P'(A_k)|}{N(A_k) + N'(A_k)} \tag{15}$$

where $P(A_k)$ and $P'(A_k)$ are the probabilities of A_k in the training and calibration sets, respectively; $N(A_k)$ and $N'(A_k)$ are the frequencies of A_k in the training and calibration sets, respectively.

Step 2: The calculation for all substances of the statistical SMILES-defect (D_j):

$$D_j = \sum_{k=1}^{NA} d_k \tag{16}$$

where NA is the number of non-blocked SMILES attributes in the SMILES. A substance falls in the domain of applicability if

$$Dj < 2 * \overline{D} \tag{17}$$

where $\overline{D}$ is average of the statistical SMILES-defect for the training set.

3. Results and discussions

3.1. Approach A – Regression model (pLD50-mix)

3.1.1. Model validation and mechanistic interpreation

The regression-based QSAR model developed here aimed at predicting the combined toxicity ($pLD_{50\text{-}mix}$) of binary mixtures of organic compounds in honey bees (n= 123) while considering chemical structures of the two components represented by disconnected SMILES as features. The regression-based QSAR model has been built using three equations according to three random splits, which include the TRN, iTRN, CLB, and VLD sets. Results with the following equations for each random splits are:

$$LD_{50\text{-}mix} = -3.6089(\pm 0.0708) + 0.1126(\pm 0.0020) * DCW(1,15) \quad (18)$$

$$LD_{50\text{-}mix} = -3.2651(\pm 0.0668) + 0.1498(\pm 0.0038) * DCW(1,15) \quad (19)$$

$$LD_{50\text{-}mix} = -3.9621(\pm 0.0855) + 0.1338(\pm 0.0030) * DCW(1,15) \quad (20)$$

Table 5.2 provides the results of the applications of statistical criteria and the characteristics of the regression-based models with the corresponding equations (18, 19 and 20). The predictability of the models has been assessed using (i) determination coefficient (R^2) (a model has desired predictability if $R^2 > 0.65$) (Roy et al., 2012); (ii) concordance correlation coefficient (CCC) which indicates good predictability of the model if CCC > 0.85 (Chirico and Gramatica, 2012); and (iii) cross-validated determination coefficient (Q^2) requiring a value larger than 0.70 to be interpreted as reliable models (Chirico and Gramatica, 2012). According to our statistics (Table 5.2), results can be considered satisfactory. The most reliable model "best split" is represented by Eq. (19) showing $R^2 = 0.87$, CCC = 0.89 and $Q^2 = 0.75$ (Table 5.2 and figure 5.2), respectively. Similarly, Eq. (20) provides good statistics for split 3 ($R^2 = 0.83$; CCC = 0.84; $Q^2 = 0.72$). To date, no QSAR models for predicting acute toxicity of organic binary mixtures in insects have been published in the scientific literature. Toropova et al. (2012) developed a QSAR model using CORAL software for toxicity of binary mixtures (expressed as pEC_{50} - decrease in light emission in *Photobacterium phosphoreum*) presenting $R^2 > 0.86$ across six different splits.

Table 5.2. Statistical quality of QSAR models for the prediction of acute contact toxicity of binary mixtures developed using approach A. Statistical values highlighted in bold indicate the most reliable model across 3 different splits.

Split	Set	n*	R^2	CCC	Q^2	RMSE	MAE	F
1	TRN*	31	0.72	0.84	0.68	0.70	0.51	75
	iTRN	31	0.55	0.66	0.50	0.81	0.63	35
	CLB	31	0.74	0.82	0.71	0.47	0.40	82
	VLD	30	0.77			0.49	0.40	
2	TRN*	30	0.62	0.76	0.57	0.77	0.61	45
	iTRN	31	0.62	0.75	0.58	0.68	0.51	48
	CLB	31	0.79	**0.89**	**0.75**	0.51	0.39	111
	VLD	31	**0.87**			0.45	0.34	
3	TRN*	31	0.60	0.75	0.55	0.75	0.56	43
	iTRN	30	0.60	0.77	0.55	0.66	0.50	42
	CLB	31	0.75	0.84	0.72	0.57	0.48	88
	VLD	31	0.83			0.48	0.37	

*) n = number of pairs of SMILES in a set; R^2 = determination coefficient; CCC = concordance correlation coefficient; Q^2 = cross-validated determination coefficient; RMSE = root mean squared error; MAE = mean absolute error; F = Fischer F-ratio.

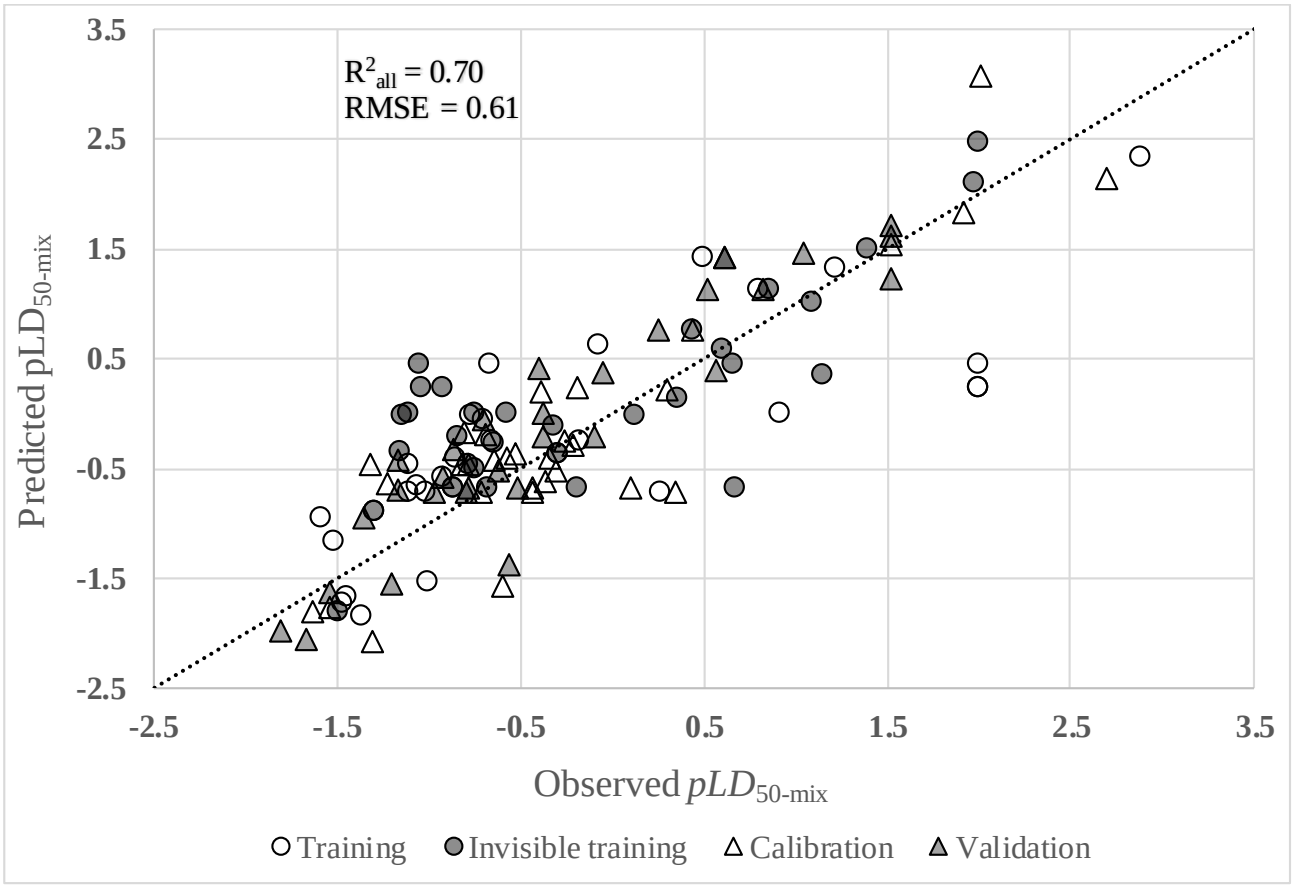


Figure 5.2. Observed versus predicted log[1/$LD_{50\text{-}mix}$] of binary mixtures for the regression-based model (Approach A), split 2 "best split". R^2_{all} (determination coefficient) and RMSE (root mean squared error) are provided for all compounds (i.e. compounds from training, invisible training, calibration, and validation sets).

The mechanistic interpretation has been obtained through three runs of the Monte Carlo optimisation. Molecular features providing stable positive correlation weights have been recognised as promoters of increase in the toxic potency of the mixture ($pLD_{50\text{-}mix}$). In contrast, molecular features presenting only negative correlation weights in the three runs of the Monte Carlo optimisation are considered as promoters of decrease in the toxic potency of the mixture ($pLD_{50\text{-}mix}$). Hence, according to our results oxygen atoms connected with double bonds, presence of rings, as well as presence of atoms of nitrogen are promoters of $pLD_{50\text{-}mix}$ increase. Similarly, carbon atoms connected with double bonds, triple bonds, presence of fluorine as well as nitrogen involved in a ring are promoters of $pLD_{50\text{-}mix}$ decrease (Table S2). However, it is necessary to take into account the prevalence of corresponding features in the training set and validation set so that rare attributes are not considered as source of reliable heuristic hypotheses.

3.2. Approach B- Classification and regression models based on Toxic Units

According to materials and methods, dose response data (n= 97) were used to develop two QSAR models taking into account mixtures ratios expressed as TUs. Results for each model to provide i) quantitative predictions of acute contact toxicity of binary mixtures ($pLD_{50\text{-}mix}$) and ii) classification of combined toxicity (qualitative) (synergism / non-synergism) are illustrated below in sections 3.2.1 and 3.2.2, respectively.

3.2.1. Regression model ($pLD_{50\text{-}mix}$; Toxic Unit)

3.2.1.1. Model validation and mechanistic interpretation

The statistical quality of the regression-based QSAR model ($pLD_{50\text{-}mix}$) including TUs as additional features is shown in Table 5.3. The QSAR model is built on four statistical models according to four random splits, which include the TRN, iTRN, CLB, and VLD sets. Results with the following equations for each random splits are:

$$pLD_{50\text{-}mix} = -3.1822116\,(\pm 0.0231229) + 0.0791914\,(\pm 0.0006750) * DCW(1,15) \quad (21)$$

$$pLD_{50\text{-}mix} = -6.0814987\,(\pm 0.0642473) + 0.1193315\,(\pm 0.0013066) * DCW(1,15) \quad (22)$$

$$pLD_{50\text{-}mix} = -3.1281529\,(\pm 0.0260826) + 0.0720372\,(\pm 0.0007152) * DCW(1,15) \quad (23)$$

$$pLD_{50\text{-}mix} = -6.2876839\,(\pm 0.0806048) + 0.0469456\,(\pm 0.0006665) * DCW(1,15) \quad (24)$$

Results of statistical characteristics for the four models calculated with corresponding equations (21-24) are presented in Table 5.3. Similarly to Approach A for the regression model (section 3.1), the predictability of the models has been assessed according to: (i) determination of the R^2 coefficient so that the model has a desired predictability with $R^2 > 0.65$) (Roy et al., 2012); (ii) concordance correlation coefficient (CCC) for which good predictability is represented by CCC > 0.85) (Chirico and Gramatica, 2012); and (iii) cross-validated determination coefficient (Q^2) which is supposed to be larger than 0.70 for reliable models (Chirico and Gramatica, 2012). Hence, according to our statistics (Table 5.3 and figure 5.3), results can be considered satisfactory. In particular, the most reliable model is represented by Eq. (21) showing $R^2 = 0.89$, CCC = 0.92 and $Q^2 = 0.81$. Similarly, Eq. (24) provides good statistics ($R^2 = 0.78$; CCC = 0.92; $Q^2 = 0.84$), being R^2 slightly smaller than the one resulting from Eq. (21). Recently, Qin et al. (2018) developed a regression model for predicting mixture toxicities (additive and non-additive) of antibiotics and pesticide in *Allivibrio fischeri* showing $R^2_m = 0.68$. However, results are not comparable due to the diversity of chemical classes and statistical approach used.

Mechanistic interpretation and statistical robustness of the CORAL model has been investigated using quasi-SMILES attributes and one run of Monte Carlo optimisation for the calculation of correlation weights for characterising either an overestimation or underestimation in predictions of the acute contact toxicity values for honey bees. According to our results (Table S3-S4), (i) positive correlation weights with SMILES attributes are interpreted as an increase in the acute contact toxicity (synergy) of the binary mixture; and (ii) negative correlation weights with SMILES attributes are associated with a decrease in the acute contact toxicity of the binary mixture (antagonism).

Table 5.3. Mechanistic interpretation and statistical robustness of the QSAR model for the prediction of acute contact toxicity of binary mixtures in honey bees based on Monte Carlo calculation with target functions 2 (TF_2). Statistical values highlighted in bold indicate the most reliable model "best split" across three different splits.

Split	Set	n	R^2	CCC	Q2	RMSE	F
1	TRN*	25	0.97	0.99	0.97	0.19	863
	iTRN	24	0.98	0.96	0.98	0.29	1140
	CLB	24	0.85	**0.92**	**0.81**	0.30	122
	VLD	24	**0.89**			0.53	
2	TRN	25	0.96	0.98	0.95	0.25	512
	iTRN	24	0.96	0.96	0.95	0.31	492
	CLB	24	0.79	0.88	0.73	0.35	81
	VLD	24	0.81			0.62	
3	TRN	24	0.96	0.98	0.95	0.21	522
	iTRN	24	0.95	0.95	0.94	0.33	462
	CLB	24	0.74	0.84	0.69	0.55	63
	VLD	25	0.75			0.61	
4	TRN	24	0.94	0.97	0.92	0.29	324
	iTRN	24	0.94	0.96	0.92	0.29	325
	CLB	24	0.86	**0.92**	**0.84**	0.45	133
	VLD	25	**0.78**			0.51	

*) TRN, iTRN, CLB, and VLD are the training, invisible training, calibration, and validation sets, respectively; n is the number of mixtures in a set; R^2 is determination coefficient, CCC is concordance correlation coefficient; Q^2 is leave-one-out cross-validated correlation coefficient; RMSE is root means squared error; F is Fischer F-ratio.

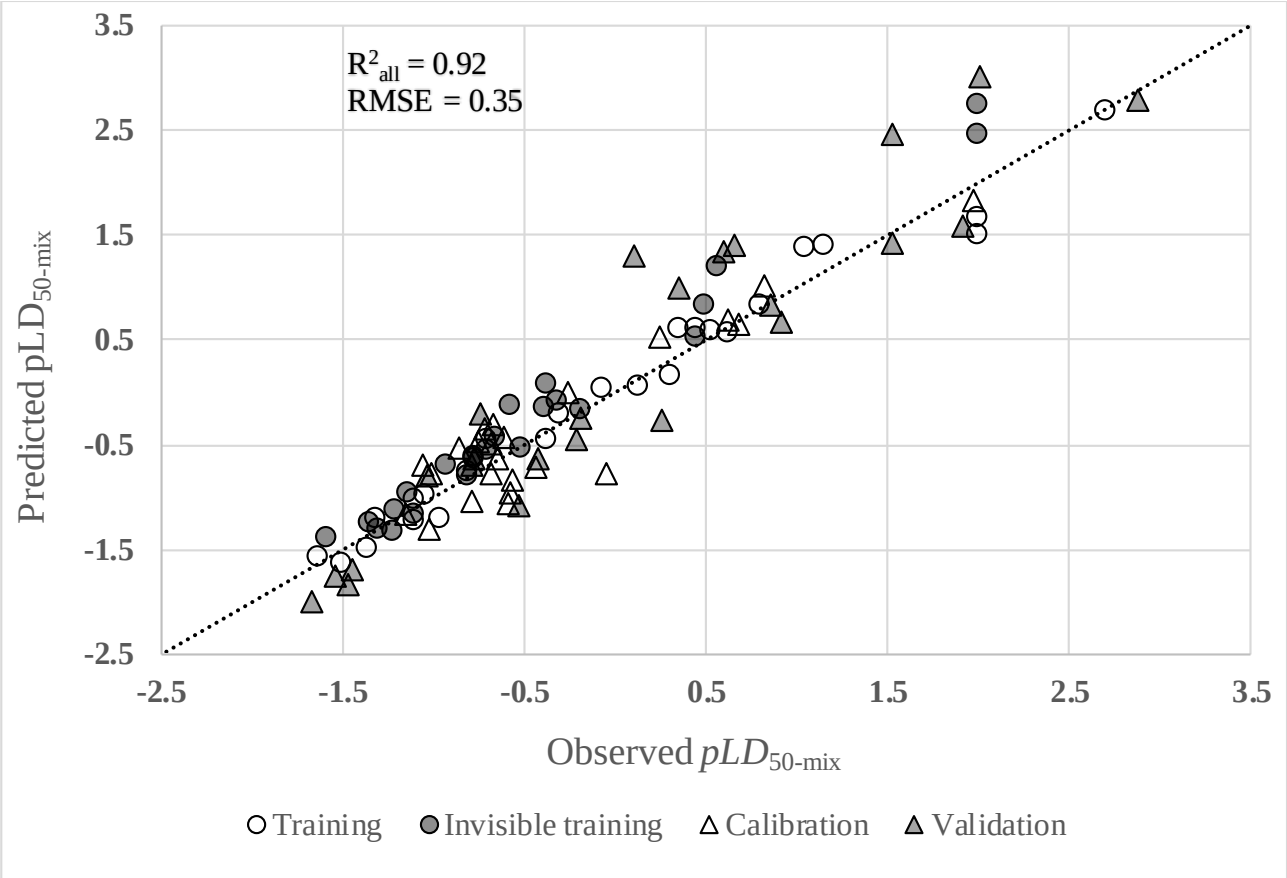


Figure 5.3. Observed versus predicted log [1/$LD_{50\text{-mix}}$] of binary mixtures for the regression-based model (Approach B), split 1 "best split". R^2_{all} (determination coefficient) and RMSE (root mean squared error) are provided for all compounds (i.e. compounds from training, invisible training, calibration, and validation sets).

3.2.2. Classification model (synergism / non-synergism and Toxic Unit)

3.2.2.1. Model validation and mechanistic interpretation

Statistical robustness of the classification model on organic binary mixtures effect (synergism / non-synergism) has been tested and presented in Table 5.4. Table S5 contains the observed versus predicted synergism data. According to our results, the accuracy of the model in validation set is 0.96 and 0.87 in the calibration set. Similarly, specificity (= 1.00), sensitivity (= 0.86) and MCC (= 0.90) for validation demonstrated robust results. To date, only classification models for single substances in honey bees are available in the literature such as Venko et al. (2018) and Como et al. (2017) presenting an accuracy = 0.84, sensitivity = 0.80, specificity = 0.86 and MCC = 0.67 (test set). The logic behind the mechanistic interpretation of the classification model is based on three runs of Monte Carlo optimisation (see section 3.2.1). Results on the classification model built up by means of the semi-correlation (Toropova et al., 2019b) are shown in Table 5.5. Similarly, the full list of correlation weights (CW) used for calculations of the classification model is provided in Table S6.

Table 5.4. Statistical quality of the classification model for binary mixtures effect (i.e. synergism / non-synergism).

Set	Statistical quality of model
Training	TP *= 5; TN= 19; FP= 0; FN= 0; N= 24
	Sensitivity = 1.00
	Specificity = 1.00
	Accuracy = 1.00
	MCC = 1.00
Invisible Training	TP = 5; TN = 18; FP = 0; FN = 1; N = 24
	Sensitivity = 0.83
	Specificity = 1.00
	Accuracy = 0.96
	MCC = 0.8885
Calibration	TP = 3; TN = 18; FP = 0; FN = 3; N = 24
	Sensitivity = 0.50
	Specificity = 1.00
	Accuracy = 0.88
	MCC = 0.66
Validation	TP = 6; TN = 18; FP = 0; FN = 1; N = 25
	Sensitivity = 0.86
	Specificity = 1.00
	Accuracy = 0.96
	MCC = 0.90

*) TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively, in a confusion matrix.

Table 5.5. Mechanistic interpretation for categorical model "synergism / non-synergism" according to Approach B. Full list is provided in supplementary materials.

No.*	SAk	CWs Run 1	CWs Run 2	CWs Run 3	N1	N2	N3	Dj
1	%.....	1.67	1.07	1.58	24	24	24	0.000
2	=...1.....	1.34	1.46	1.80	24	23	24	0.000
3	C...(.....	0.40	0.25	0.29	24	24	24	0.000
4	C...2.....	0.39	0.63	0.55	24	24	24	0.000
5	C...=.....	0.11	0.03	0.40	24	24	24	0.000
6	C...C.....	0.24	0.37	0.43	24	24	24	0.000
7	^...%.....	1.28	1.52	0.79	24	24	24	0.000
8	^.....	0.89	0.85	1.59	24	24	24	0.000
9	^...C.....	1.64	0.53	1.03	24	24	24	0.000
10	1...(.....	0.52	0.58	0.93	23	22	22	0.001
11	5.....	0.73	0.88	0.62	23	22	22	0.001
12	N...C.....	0.58	0.37	0.39	21	15	18	0.003
13	3.....	0.61	0.46	0.78	20	18	17	0.003
14	C...3.....	0.46	0.06	0.59	20	18	17	0.003
15	1...%.....	1.95	1.46	1.95	12	13	10	0.004
1	1.....	-0.73	-0.30	-0.55	24	24	24	0.000
2	=.....	-0.18	-0.08	-0.07	24	24	24	0.000
3	C.....	-0.02	-0.30	-0.27	24	24	24	0.000
4	C...1.....	-0.39	-0.42	-0.72	24	24	24	0.000
5	N...(.....	-0.15	-0.29	-0.12	23	20	21	0.002
6	(...(.....	-0.51	-0.31	-0.58	22	18	18	0.004
7	O...=.....	-0.54	-0.09	-1.05	22	22	20	0.002
8	3...(.....	-0.12	-0.17	-0.46	20	18	17	0.003
9	C...#.....	-0.39	-0.52	-0.47	16	10	14	0.003
10	N...#.....	-0.41	-0.38	-0.42	16	10	14	0.003
11	F...(.....	-0.06	-0.11	-0.32	15	9	12	0.005
12	Cl..1.....	-0.54	-0.70	-0.50	10	7	8	0.005
13	4...%.....	-0.42	-0.21	-0.23	5	5	2	0.018
14	Cl..2.....	-0.39	-0.37	-0.37	5	3	4	0.005
15	2...%.....	-0.28	-0.55	-0.41	4	3	7	0.011

*) N1, N2, and N3 are number of SAk in the training, invisible training, and calibration sets, respectively. CW = Correleation Weights as in Eq. (1); Dj= SMILES-defect as in Eq. (16).

3.3. Comparative assessment

3.3.1. Applicability of Approach A and Approach B

This manuscript presents three innovative QSAR models (two regression- and one classification-based) developed for the prediction of combined toxicity of binary mixtures in honey bees to date, notwithstanding that other models for predicting binary mixtures toxicity already exist in the literature (Qin et al., 2018; Wang D. et al., 2018; Toropova et al., 2012; Muratov et al., 2012; Kim et al., 2018; Tian et al., 2013).

Two different approaches (A and B) have been applied and validated using the CORAL software for predicting both acute mixtures potency ($pLD_{50\text{-}mix}$) and the nature of combined toxicity (synergism / non-synergism). In particular, we demonstrated how depending on the availability of quantitative data on binary mixtures (e.g. dose-response, toxic unit, etc.), two different notation methods (traditional SMILES or quasi-SMILES) can be applied for developing robust regression- and classification-based QSAR models (Table 5.1).

Approach A (section 3.1) allowed building up one regression-based QSAR model using as input the chemical structure of two-component mixtures codified as disconnected SMILES. Indeed, this is a simplistic representation of combined toxicity of binary mixture understanding, since the model does not take into account mixture ratio (e.g. toxic unit), thus ignoring the relative potency of each chemical contributing to the overall mixture toxicity (Bopp et al., 2015). Similar models have been already published in the literature for predicting binary mixtures toxicity in bacteria (Toropova et al., 2012; Wang T. et al., 2018) and flammability of binary liquid mixtures (Toropova et al., 2019a). As a consequence, Approach A can be considered as valuable tool for the preliminary screening of chemical mixtures of concern in the case of lack of information on the potency of each component, by integrating available mechanistic and biological data on the specific target species (More et al., 2019; EFSA, 2013).

In contrast, Approach B considered additional quantitative information such as TUs for each chemical codified in quasi-SMILES (Toropov et al., 2018; Toropova et al., 2018). Hence, Approach B results of more interpretable than Approach A for predicting (i) the potency of organic binary mixtures ($pLD_{50\text{-}mix}$) and (ii) the nature of combined toxicity (i.e. synergism / non-synergism). In addition, our results confirm that the statistical quality of regression-based model (Approach B) improved when including Toxic Unit as additional feature (figure 5.4). Regarding the classification-based model (Approach B), the scientific literature describes the effects of a given chemical mixture either as interactive (e.g. synergism, antagonism) or non-interactive (i.e. additive) when assuming concentration addition (CA) as default model (More et al., 2019). However, due to lack of experimental data, our classification model is based on two classes synergism (1) and non-synergism (0) (Toropova and Toropov, 2017). Nevertheless, this approach can result as highly conservative for

predicting synergistic effects of binary mixture interactions. Indeed, a recent meta-analysis (Carnesecchi et al., 2019) confirmed that in honey bees, interactions were observed for 72% of cases as synergism and 28% as non-synergism (i.e. 17% additive, 11% antagonism). Similarly, such QSAR models can be applied to further refine current thresholds used in Model Deviation Ratio (MDR) and Estimated Mean Ratio (EMR) calculations (Belden et al., 2007; Carnesecchi et al. 2019).

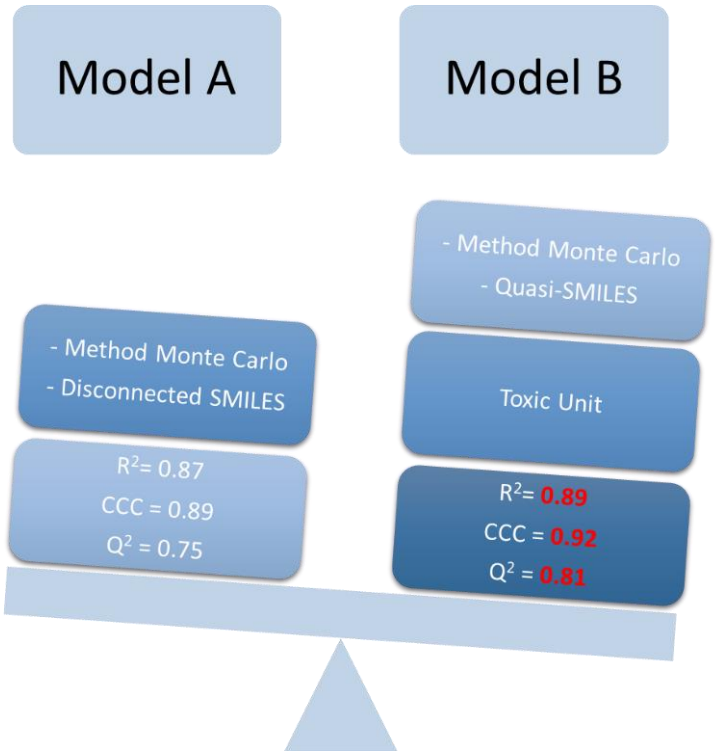


Figure 5.4. Comparison of two QSAR regression models following Approaches A and B. Results of the statistical quality and statistical methods are presented. Results of the statistical quality are highlighted in red for the best model here developed. R^2 = determination coefficient; CCC = concordance correlation coefficient; Q^2 = cross-validated determination coefficient.

3.3.2. SMILES attributes as drivers of binary mixture toxicity

Currently, most QSAR models investigating binary mixture toxicity provide qualitative or quantitative predictions of the toxicological endpoint without consideration of molecular feature(s) that may be responsible for an increase or a decrease in toxicity (Kim and Kim, 2015; Wang et al., 2019). Here, this study demonstrates how the CORAL software can be applied to identify molecular features as drivers of binary mixture toxicity ($pLD_{50\text{-}mix}$) i.e. SMILES attributes (SAk) using Monte Carlo optimisation. According to our results (Tables S5.2-S5.4 and S5.6), the regression- and classification-based QSAR models (Approach A and B) showed that SMILES attributes with stable positive correlation weights (CWs) can be interpreted as promoters of an increase (synergism) in acute

contact toxicity ($pLD_{50\text{-}mix}$) of the binary mixture in bees. In contrast, molecular features presenting negative CWs, from the Monte Carlo optimisation, can be interpreted as promoters of a decrease (antagonism) in the acute toxicity of the binary mixture. Generally speaking, it is recommended to avoid generating rules which are supported by few experimental data, and these can be identified by the software and qualified as "rare attributes". The attributes which are used by the QSAR models can be interpreted as associated to specific chemical features related to the acute contact toxicity of the binary mixture. From the statistical basis of the CORAL models, Figure 5.5 illustrates examples of molecular features associated with either an increase or a decrease in the acute contact toxicity of the binary mixtures in bees. Overall, these innovative bee QSAR models developed within the CORAL software identify the most frequent molecular features that are associated with the binary mixture and its acute contact toxicity including the co-presence of functional groups associated with reactivity (e.g. double or triple bonds), the branching level of the molecule and steric components or the presence of certain atoms associated with polarity (e.g. nitrogen).

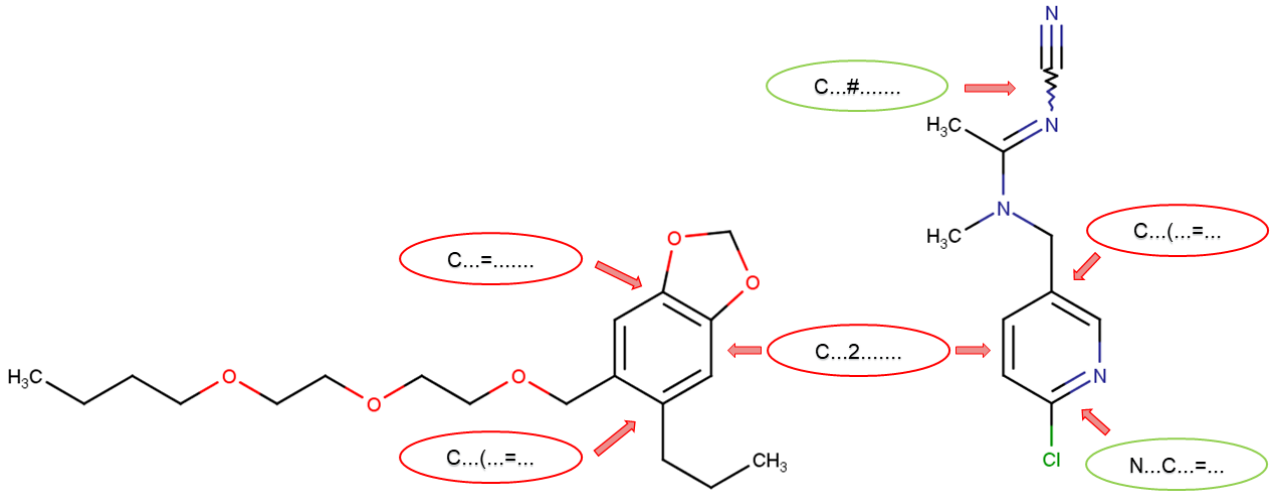


Figure 5.5. Example of SMILES attributes associated with positive or negative correlation weights (CW) in binary mixture toxicity (piperonyl butoxide + acetamiprid). CW were demonstrated in two out of three the QSAR models following three runs of Monte Carlo optimisation. Red circles indicate SMILES attributes with positive correlation coefficient (i.e. increase combined toxicity). Green circles indicate SMILES attributes with negative correlation coefficient (i.e. decrease combined toxicity). [C...#.....] indicates the presence of carbon atom(s) connected with triple bond; [C...=.....] indicates the presence of carbon atom(s) connected with double bonds. [C...2.....] indicates the presence of two rings in the binary mixtures. [C...(...=...] indicates the presence of carbon atom(s) with double bonds and branched chain.

4. Conclusions and further perspectives

This manuscript has explored the development of innovative QSAR models for the prediction of acute contact toxicity of binary mixtures in bees through their calibration, validation and mechanistic interpretation:

- Two regression-based models predicting acute (contact) mixtures potency ($pLD_{50\text{-}mix}$) in a quantitative manner validated with statistical tests and were demonstrated to have be reliable and robust (Approach A: $R^2 = 0.87$, CCC = 0.89, $Q^2 = 0.75$; Approach B: $R^2 = 0.89$, CCC = 0.92, $Q^2 = 0.81$).

- A classification-based model predicting the nature of the combined toxicity (synergism, non-synergism) validated using qualitative statistical validation parameters i.e. sensitivity (= 0.86), specificity (= 1.00), accuracy (= 0.96), and Mattews correlation coefficient (MCC = 0.90). These models are currently being implemented within the VEGA-HUB platform (https://www.vegahub.eu/) as well as on CORAL software/databases (http://www.insilico.eu/coral/).

To date, such QSAR models were not previously available and this manuscript shows the potential use of *in silico* tools as part of New Approaches Methodologies (NAMs), particularly in the context of prioritisation and hazard assessment of potentially hazardous mixtures in honey bees (More et al., 2019). With regard to the chemical space of the models here developed, it is worth noting that the training sets are mostly constructed with available binary mixture data of limited toxicological and structural diversity (mostly conazole fungicides, pyrethroids and neonicotinoid insecticides) which reflects the limited applicability domain of these models. As a consequence, it is recommended to apply these QSAR models for predicting the combined toxicity ($pLD_{50\text{-}mix}$) of similar binary mixtures such as plant protection products (PPPs), veterinary drugs and their (potential) formulations in honey bees.

Finally, data gaps remain and still limit the development and broader applications of such QSAR models in honey bees:

- Acute contact toxicity data (e.g. mortality) are the major available datasets for a significant number of compounds and their mixtures but chronic toxicity oral data and toxicity data for sub-lethal effects in honey bees and wild bees are still lacking (Carnesecchi et al., 2019).

- Experimental toxicokinetic data (e.g. half-life, bioaccumulation) are also lacking particularly to develop ad-hoc QSAR models for further characterisation of the impact of persistence on combined toxicity of multiple chemicals in bees (including interactions). With such dataset, QSAR models can be integrated with Dynamic Energy Budget models to provide a refined

understanding of combined toxicity at the honey bee population level (Spurgeon et al., 2016; Hesketh et al., 2016).

- Although QSAR models for the prediction of the Mode of Action (MoA) of chemicals have been developed for aquatic species (Kienzler et al., 2019), the integration of such (qualitative) information into QSAR models for predicting mixture toxicity results challenging. In this context, the ongoing EFSA's OpenFoodTox 2.0 "OptiTox" project (figure 5.6) aims to develop new predictive tools and integrate results from multiple *in silico* methods ((Q)SAR, MoA predictors, read-across and Activity-Activity relationship within a Weight of Evidence (WoE) strategy (Hardy et al., 2017; Benfenati et al., 2019).

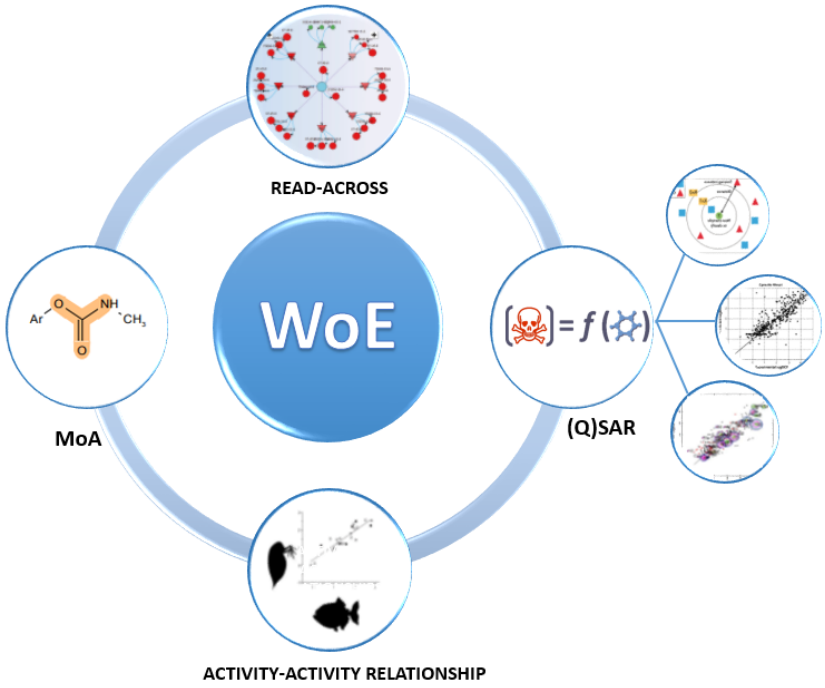


Figure 5.6. Conceptual framework for integrating results from multiple in silico tools (Mode of Action, Read-across, (Q)SAR and Activity-Activity relationship) within a Weight of Evidence (WoE) approach (Hardy et al., 2017; Benfenati et al., 2019).

Acknowledgements

The views expressed in this manuscript do not reflect the views of the European Food Safety Authority and are the authors only. This work was supported by the European Food Safety Authority (EFSA) [contract number: OC/EFSA/SCER/2018/01 and NP/EFSA/AFSCO/2016/02 (Edoardo Carnesecchi)]; and LIFE-VERMEER (LIFE16 ENV/IT/000167). Authors would like to thank Cosimo Toma (European Chemicals Agency / Institute for Pharmacological Research Mario Negri) for constructive criticism of the manuscript and Matteo Sironi (Istituto di Ricerche Farmacologiche Mario Negri IRCCS) for the graphic design support.

References

Backhaus, T., Faust, M., 2012. Predictive environmental risk assessment of chemical mixtures: A conceptual framework. Environ. Sci. Technol. https://doi.org/10.1021/es2034125

Bargańska, Ż., Ślebioda, M., Namieśnik, J., 2016. Honey bees and their products: Bioindicators of environmental contamination. Crit. Rev. Environ. Sci. Technol. 46, 235–248. https://doi.org/10.1080/10643389.2015.1078220

Belden, J.B., Gilliom, R.J., Lydy, M.J., 2007. How well can we predict the toxicity of pesticide mixtures to aquatic life? Integr. Environ. Assess. Manag. 3, 364–372. https://doi.org/10.1002/ieam.5630030307

Benfenati, E., Chaudhry, Q., Gini, G., Dorne, J. Lou, 2019. Integrating in silico models and read-across methods for predicting toxicity of chemicals: A step-wise strategy. Environ. Int. https://doi.org/10.1016/j.envint.2019.105060

Bopp, S., Berggren, E., Kienzler, A., van der Linden, S., Worth, A., 2015. Scientific methodologies for the assessment of combined effects of chemicals a survey and literature review, Publications Office of the European Union. https://doi.org/doi:10.2788/093511

Carnesecchi E., Svendsen C., Lasagni S., Grech A., Quignot N., Amzal B., Toma C., Tosi S., Rortais A., Cortinas-Abrahantes J., Capri E., Kramer N., Benfenati B., Spurgeon D., Guillot G., Dorne J.L., *in press*. Investigating combined toxicity of binary mixtures in bees: Meta-analysis of laboratory tests, modelling, mechanistic basis and implications for risk assessment. Environ. Int

Cedergreen, N., 2014. Quantifying synergy: a systematic review of mixture toxicity studies within environmental toxicology. PLoS One 9, e96580. doi:10.1371/journal.pone.0096580.

Chirico, N., Gramatica, P., 2012. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. J. Chem. Inf. Model. https://doi.org/10.1021/ci300084j

Como, F., Carnesecchi, E., Volani, S., Dorne, J.L., Richardson, J., Bassan, A., Pavan, M., Benfenati, E., 2017. Predicting acute contact toxicity of pesticides in honeybees (Apis mellifera) through a k-nearest neighbor model. Chemosphere. https://doi.org/10.1016/j.chemosphere.2016.09.092

Dao, P., Wang, K., Collins, C., Ester, M., Lapuk, A., Sahinalp, S.C., 2011. Optimally discriminative subnetwork markers predict response to chemotherapy. Bioinformatics. https://doi.org/10.1093/bioinformatics/btr245

Dorne, J. Lou, Richardson, J., Kass, G., Georgiadis, N., Monguidi, M., Pasinato, L., Cappe, S., Verhagen, H., Robinson, T., 2017. Editorial: OpenFoodTox: EFSA's open source toxicological database on chemical hazards in food and feed. EFSA J. https://doi.org/10.2903/j.efsa.2017.e15011

EFSA (European Food Safety Authority), 2009. Risk Assessment for Birds and Mammals. EFSA J. https://doi.org/10.2903/j.efsa.2009.1438

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Guidance on tiered risk assessment for Plant Protection Products for aquatic organisms in edge-of-field surface waters. EFSA Journal 2013;11(7):3290, 268 pp. doi: 10.2903/j.efsa.2013.3290

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2012. Scientific Opinion on the science behind the development of a risk assessment of Plant Protection Products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA Journal 2012; 10(5) 2668, 275 pp. doi: 10.2903/j.efsa.2012.2668.

EFSA (European Food Safety Authority), 2013. EFSA Guidance Document on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA J. 11 (7):3295. http://dx.doi.org/10.2903/j.efsa.2013.3295 (268 pp.).

EFSA (European Food Safety Authority), 2014. Towards an integrated environmental risk assessment of multiple stressors on bees: review of research projects in Europe, knowledge gaps and recommendations. EFSA Journal 2014;12(3):3594, 102 pp. doi: 10.2903/j.efsa.2014.3594

EFSA Panel on Animal Health and Welfare (AHAW), 2016. EFSA Panel on Animal Health and Welfare (AHAW) (2016). Assessing the health status of managed honeybee colonies (HEALTHY-B): a toolbox to facilitate harmonised data collection. https://doi.org/10.2903/j.efsa.2016.4578

EFSA (European Food Safety Authority), 2017. Specifications for field data collection contributing to honey bee model corroboration and verification. EFSA supporting publication 2017:EN-234.54pp.doi:10.2903/sp.efsa.2017.EN-1234.

Gadaleta, D., Catto, M., 2016. Applicability Domain for QSAR Models: Where Theory Meets Reality. Int. J. Quant. Struct. Relationships 1. https://doi.org/10.4018/IJQSPR.2016010102

Hamadache, M., Benkortbi, O., Hanini, S., Amrane, A., 2018. QSAR modeling in ecotoxicological risk assessment: application to the prediction of acute contact toxicity of pesticides on bees (Apis mellifera L.). Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-017-0498-9

Hardy, A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, H.K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J.R., Silano, V., Solecki, R., Turck, D., Benfenati, E., Chaudhry, Q.M., Craig, P., Frampton, G., Greiner, M., Hart, A., Hogstrand, C., Lambre, C., Luttik, R., Makowski, D., Siani, A., Wahlstroem, H., Aguilera, J., Dorne, J., Fernandez Dumont, A., Hempen, M., Valtueña Martínez, S., Martino, L., Smeraldi, C., Terron, A., Georgiadis, N., Younes, M., 2017. Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J. https://doi.org/10.2903/j.efsa.2017.4971

Hesketh, H., Lahive, E., Horton, A.A., Robinson, A.G., Svendsen, C., Rortais, A., Dorne, J.L., Baas, J., Spurgeon, D.J., Heard, M.S., 2016. Extending standard testing period in honeybees to predict lifespan impacts of pesticides and heavy metals using dynamic energy budget modelling. Sci. Rep. https://doi.org/10.1038/srep37655

Hunter, R.S., F.D. Culver, and A. Fitzgerald. 1987. SMILES User Manual. A Simplified Molecular Input Line Entry System. Includes extended SMILES for defining fragments. Review Draft, Internal Report, Montana State University, Institute for Biological and Chemical Process Control (IPA), Bozeman, MT.

Iwasa, T., Motoyama, N., Ambrose, J.T., Roe, R.M., 2004. Mechanism for the differential toxicity of neonicotinoid insecticides in the honey bee, Apis mellifera. Crop Prot. 23, 371–378. https://doi.org/10.1016/j.cropro.2003.08.018

Johnson, R.M., Mao, W., Pollock, H.S., Niu, G., Schuler, M.A., Berenbaum, M.R., 2012. Ecologically appropriate xenobiotics induce cytochrome P450s in Apis Mellifera. PLoS One 7, 1–9. https://doi.org/10.1371/journal.pone.0031051

Johnson, R.M., Dahlgren, L., Siegfried, B.D., Ellis, M.D., 2013. Acaricide, Fungicide and Drug Interactions in Honey Bees (Apis mellifera). PLoS One 8. https://doi.org/10.1371/journal.pone.0054092

Jonker, M.J., Svendsen, C., Bedaux, J.J.M., Bongers, M., Kammenga, J.E., 2005. Significance testing of synergistic/antagonistic, dose level-dependent, or dose ratio-dependent effects in mixture dose-response analysis. Environ. Toxicol. Chem. https://doi.org/10.1897/04-431R.1

KEMI (The Swedish Chemicals Agency), 2015. An additional assessment factor (MAF) – A suitable approach for improving the regulatory risk assessment of chemical mixtures?, Swedish Chemicals Agency. Print: Arkitektkopia, Stockholm 2015. ISSN 0284-1185. Article number: 361 159.

Kienzler, A., Bopp, S.K., van der Linden, S., Berggren, E., Worth, A., 2016. Regulatory assessment of chemical mixtures: Requirements, current approaches and future perspectives. Regul. Toxicol. Pharmacol. https://doi.org/10.1016/j.yrtph.2016.05.020

Kienzler, A., Connors, K.A., Bonnell, M., Barron, M.G., Beasley, A., Inglis, C.G., Norberg-King, T.J., Martin, T., Sanderson, H., Vallotton, N., Wilson, P., Embry, M.R., 2019. Mode of action (MOA) classifications in the EnviroTox database: Development and implementation of a consensus MOA classification. Environ. Toxicol. Chem. https://doi.org/10.1002/etc.4531

Kim, J., Kim, S., 2015. State of the art in the application of QSAR techniques for predicting mixture toxicity in environmental risk assessment. SAR QSAR Environ. Res. 26, 41–59. https://doi.org/10.1080/1062936X.2014.984627

Kim, J., Fischer, M., Helms, V., 2018. Prediction of Synergistic Toxicity of Binary Mixtures to Vibrio fischeri Based on Biomolecular Interaction Networks. Chem. Res. Toxicol. https://doi.org/10.1021/acs.chemrestox.8b00164

Klein, A.-M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. Proc. R. Soc. B Biol. Sci. 274, 303–313. https://doi.org/10.1098/rspb.2006.3721

Lambert, O., Veyrand, B., Durand, S., Marchand, P., Bizec, B. Le, Piroux, M., Puyo, S., Thorin, C., Delbac, F., Pouliquen, H., 2012. Polycyclic aromatic hydrocarbons: Bees, honey and pollen as sentinels for environmental chemical contaminants. Chemosphere 86, 98–104. https://doi.org/10.1016/j.chemosphere.2011.09.025

More, S.J., Hardy, A., Bampidis, V., Benford, D., Hougaard Bennekou, S., Bragard, C., Boesten, J., Halldorsson, T.I., Hernández-Jerez, A.F., Jeger, M.J., Knutsen, H.K., Koutsoumanis, K.P., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J.R., Silano, V., Nielsen, S.S., Schrenk, D., Solecki, R., Turck, D., Younes, M., Benfenati, E., Castle, L., Cedergreen, N., Laskowski, R., Leblanc, J.C., Kortenkamp, A., Ragas, A., Posthuma, L., Svendsen, C., Testai, E., Dujardin, B., Kass, G.E.N., Manini, P., Zare Jeddi, M., Dorne, J.L.C., Hogstrand, C., 2019. Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals. EFSA J. https://doi.org/10.2903/j.efsa.2019.5634

Muratov, E.N., Varlamova, E. V., Artemenko, A.G., Polishchuk, P.G., Kuz'Min, V.E., 2012. Existing and developing approaches for QSAR analysis of mixtures. Mol. Inform. https://doi.org/10.1002/minf.201100129

Nys, C., Van Regenmortel, T., Janssen, C.R., Oorts, K., Smolders, E., De Schamphelaere, K.A.C., 2018. A framework for ecological risk assessment of metal mixtures in aquatic systems. Environ. Toxicol. Chem. https://doi.org/10.1002/etc.4039

OECD, 1998. Test No. 213: Honeybees, Acute Oral Toxicity Test. OECD Guidel. Test. Chem. https://doi.org/10.1787/9789264070165-en

OECD, 2007. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, ENV/JM/MONO(2007)2. Paris: OECD publishing.

Qin, L.T., Chen, Y.H., Zhang, X., Mo, L.Y., Zeng, H.H., Liang, Y.P., 2018. QSAR prediction of additive and non-additive mixture toxicities of antibiotics and pesticide. Chemosphere. https://doi.org/10.1016/j.chemosphere.2018.01.142

Rortais, A., Arnold, G., Dorne, J. Lou, More, S.J., Sperandio, G., Streissl, F., Szentes, C., Verdonck, F., 2017. Risk assessment of pesticides and other stressors in bees: Principles, data gaps and perspectives from the European Food Safety Authority. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2016.09.127

Roy, K., Mitra, I., Kar, S., Ojha, P.K., Das, R.N., Kabir, H., 2012. Comparative studies on some metrics for external validation of QSPR models, in: Journal of Chemical Information and Modeling. https://doi.org/10.1021/ci200520g

Singh, K.P., Gupta, S., Basant, N., Mohan, D., 2014. QSTR modeling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes. Chem. Res. Toxicol. https://doi.org/10.1021/tx500100m Sobati, M. A., Abooali, D., Maghbooli, B., & Najafi, H. (2016). A new structure-based model for estimation of true critical volume of multi-component mixtures. Chemometrics and Intelligent Laboratory Systems, 155, 109-119.

Skorbiłowicz, M., Skorbiłowicz, E., Cieśluk, I., 2018. Bees as Bioindicators of Environmental Pollution with Metals in an Urban Area. J. Ecol. Eng. 19, 229–234. https://doi.org/10.12911/22998993/85738

Spurgeon, D., Hesketh, H., Lahive, E., Svendsen, C., Baas, J., Robinson, A., Horton, A., Heard, M., 2017. Chronic oral lethal and sub-lethal toxicities of different binary mixtures of pesticides and contaminants in bees (Apis mellifera, Osmia bicornis and Bombus terrestris). EFSA Support. Publ. 13. https://doi.org/10.2903/sp.efsa.2016.en-1076

Thoreau, F., 2016. 'A mechanistic interpretation, if possible': How does predictive modelling causality affect the regulation of chemicals? Big Data Soc. https://doi.org/10.1177/2053951716670189

Tian, D., Lin, Z., Yin, D., 2013. Quantitative structure activity relationships (QSAR) for binary mixtures at non-equitoxic ratios based on toxic ratios-effects curves. Dose-Response. https://doi.org/10.2203/dose-response.11-042.Lin

Toropov, A.A., Toropova, A.P., Lombardo, A., Roncaglioni, A., De Brita, N., Stella, G., Benfenati, E., 2012a. CORAL: The prediction of biodegradation of organic compounds with optimal SMILES-based descriptors. Cent. Eur. J. Chem. https://doi.org/10.2478/s11532-012-0031-4

Toropov, A.A., Raška, I., Toropova, A.P., Raškova, M., Veselinović, A.M., Veselinović, J.B., 2019. The study of the index of ideality of correlation as a new criterion of predictive potential of QSPR/QSAR-models. Sci. Total Environ. 659, 1387–1394. https://doi.org/10.1016/j.scitotenv.2018.12.439

Toropov, A.A., Toropova, A.P., Benfenati, E., Diomede, L., Salmona, M., 2018. Use of quasi-SMILES to model biological activity of "micelle–polymer" samples. Struct. Chem. https://doi.org/10.1007/s11224-018-1115-3

Toropov, A.A., Toropova, A.P., Rasulev, B.F., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2012b. CORAL: binary classifications (active/inactive) for Liver-Related Adverse Effects of Drugs. Curr. Drug Saf. 7, 257–61.

Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2012. CORAL: Models of toxicity of binary mixtures. Chemom. Intell. Lab. Syst. https://doi.org/10.1016/j.chemolab.2012.10.001

Toropova, A.P., Toropov, A.A., Carnesecchi, E., Benfenati, E., Dorne, J. Lou, 2019a. The index of ideality of correlation: models for flammability of binary liquid mixtures. Chem. Pap. https://doi.org/10.1007/s11696-019-00903-w

Toropova, A.P., Toropov, A.A., 2017. CORAL: Binary classifications (active/inactive) for drug-induced liver injury. Toxicol. Lett. 268, 51–57. https://doi.org/10.1016/j.toxlet.2017.01.011

Toropova, A.P., Toropov, A.A., 2018. Use of the index of ideality of correlation to improve models of eco-toxicity. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-018-3291-5

Toropova, A.P., Toropov, A.A., Benfenati, E., Castiglioni, S., Bagnati, R., Passoni, A., Zuccato, E., Fanelli, R., 2018. Quasi-SMILES as a tool to predict removal rates of pharmaceuticals and dyes in sewage. Process Saf. Environ. Prot. https://doi.org/10.1016/j.psep.2018.07.003

Toropova, A.P., Toropov, A.A., Benfenati, E., 2019b. Semi-correlations as a tool to build up categorical (active/inactive) model of GABA A receptor modulator activity. Struct. Chem. 30, 853–861. https://doi.org/10.1007/s11224-018-1226-x

Tosi, S., Costa, C., Vesco, U., Quaglia, G., Guido, G., 2018. A 3-year survey of Italian honey bee-collected pollen reveals widespread contamination by agricultural pesticides. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2017.09.226

U.S. EPA (Environmental Protection Agency), 2014. Guidance for assessing pesticide risks to bees. Office of Research and Development, National Center for Environmental Assessment, Washington, DC 20460. Available at: https://www.epa.gov/sites/production/files/2014-06/documents/pollinator_risk_assessment_guidance_06_19_14.pdf

U.S. EPA (Environmental Protection Agency), 2003. Framework for Cumulative Risk Assessment. Office of Research and Development, National Center for Environmental Assessment, Washington, DC 20460. EPA/630/P-02/001F. Available at: https://www.epa.gov/sites/production/files/2014-11/documents/frmwrk_cum_risk_assmnt.pdf

VanEngelsdorp, D., Meixner, M.D., 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. J. Invertebr. Pathol. 103, S80–S95. https://doi.org/10.1016/j.jip.2009.06.011

Venko, K., Drgan, V., Župerl, Š., Vračko, M., Novič, M., 2017. In silico evaluation of toxicity towards honey bees with QSAR models. Toxicol. Lett. https://doi.org/10.1016/j.toxlet.2017.07.787

Wade, A., Lin, C.-H., Kurkul, C., Regan, E.R., Johnson, R.M., 2019. Combined Toxicity of Insecticides and Fungicides Applied to California Almond Orchards to Honey Bee Larvae and Adults. Insects 10, 20. https://doi.org/10.3390/insects10010020

Wang, D., Wu, X., Lin, Z., Ding, Y., 2018. A comparative study on the binary and ternary mixture toxicity of antibiotics towards three bacteria based on QSAR investigation. Environ. Res. https://doi.org/10.1016/j.envres.2017.12.015

Wang, T., Tang, L., Luan, F., Cordeiro, M.N.D.S., 2018. Prediction of the toxicity of binary mixtures by qsar approach using the hypothetical descriptors. Int. J. Mol. Sci. https://doi.org/10.3390/ijms19113423

Chapter 6 **Integrating QSAR models predicting acute contact toxicity and mode of action profiling in honey bees (*A. mellifera*): data curation using open source databases, performance testing and validation**

Edoardo Carnesecchi [a,b,*], Cosimo Toma [a,b], Alessandra Roncaglioni [b], Nynke Kramer [a], Emilio Benfenati [b], Jean Lou CM Dorne [c]

[a] Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3584 Utrecht, the Netherlands
[b] Laboratory of Chemistry and Environmental Toxicology, Department of Environmental Health, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, via Mario Negri, 2, 20156 Milano, Italy
[c] European Food Safety Authority (EFSA), Scientific Committee and Emerging Risks Unit, Parma, Italy
[*] Corresponding author

Published in Science of the Total Environment, 2020, 735:139243

Abstract

Honey bees (*Apis mellifera*) provide key ecosystem services as pollinators bridging agriculture, the food chain and ecological communities, thereby ensuring food production and security. Ecological risk assessment of single Plant Protection Products (PPPs) requires an understanding of the exposure and toxicity. *In silico* tools such as QSAR models can play a major role for the prediction of structural, physico-chemical and pharmacokinetic properties of chemicals as well as toxicity of single and multiple chemicals. Here, the first integrative honey bee QSAR model has been developed for PPPs using EFSA's OpenFoodTox, US-EPA ECOTOX and Pesticide Properties DataBase i) to predict acute contact toxicity (LD_{50}) and ii) to profile the Mode of Action (MoA) of pesticides active substances. Three different classification-based and four regression-based models were developed and tested for their performance, thus identifying two models providing the most reliable predictions based on k-NN algorithm. The two-category QSAR model (toxic / non-toxic; n= 411) was validated using sensitivity (=0.93), specificity (=0.85), balanced accuracy (=0.90), and Matthews correlation coefficient (MCC = 0.78) as statistical parameters. The regression-based model (n= 113) was validated for its reliability and robustness (R^2= 0.74; MAE= 0.52). Current study proposes the MoA profiling for 113 pesticides active substances and the first harmonised MoA classification scheme for acute contact toxicity in honey bees, including LD_{50s} data points from three different databases. The classification allows to further define MoAs and the target site of PPPs active substances, thus enabling regulators and scientists to refine chemical grouping and toxicity extrapolations for single chemicals and component-based mixture risk assessment of multiple chemicals. Relevant future perspectives are briefly addressed to integrate MoA, adverse outcome pathways (AOPs) and toxicokinetic information for the refinement of single-chemical/combined toxicity predictions and risk estimates at different levels of biological organisation in the bee health context.

Keywords: QSAR models, honey bees, Mode of Action, ecological risk assessment, chemical mixtures

1. Introduction

The importance of pesticides as plant protection products (PPPs) in agriculture forestry, urban gardens, parks has been recognised worldwide, particularly to protect crops against pests (e.g. insects, weeds), diseases or pathogens (e.g. fungi), which may affect plants health and potentially reduce crop yield, thus potentially threatening food security (FAO, ITPS, 2017; Frische et al. 2018). However, concerns due to the potential harmful effects of PPPs including insecticides on ecosystems, particularly towards non-target species such as pollinators, have raised (Douglas et al. 2020; EFSA, 2018; Tosi and Nieh, 2019; Sanchez-Bayo and Goka, 2016; Simon-Delso et al. 2015; Tosi et al. 2018).

Indeed, pollinators such as honey bees (*Apis mellifera*), bumble bees (*Bombus* spp.) and solitary bees (e.g. *Osmia* spp.) play a key-role as ecosystems service providers (ESP) contributing to the maintenance, reproduction of wild plant communities and biodiversity as well as bridging agriculture, the food chain and the ecological communities, thereby ensuring food production and security (Breeze et al. 2011; Schulp et al. 2014; Rose et al., 2015). Honey bees also represent sentinel species together with their hive products as bioindicators (i.e. honey, pollen, beebread) to monitor environmental contamination by regulated products (e.g. PPPs, veterinary residues), anthropogenic (e.g. persistent organic pollutants, heavy metals, particulate matter) and natural contaminants (mycotoxins, plant alkaloids) (Negri et al. 2015; Bargańska et al. 2016; Tosi et al. 2018).

Moreover, honey bees are employed worldwide as surrogate species for *Apis* and non-*Apis* bees to perform toxicity tests on single pesticides (EFSA, 2013; USEPA, PMRA, CALDPR, 2014). In the EU, pesticide risk assessment (RA) requires the settings of protection goals and the evaluation of the environmental impact associated with the exposure and toxicity of PPPs use (Regulation EC, 1107/2009). For honey bees, the Regulation lays downs that "*an active substance should only be approved if it results in negligible exposure or has no unacceptable acute or chronic effects on colony survival and development, taking into account effects on honeybee larvae and honeybee behaviour*". In this context, the European Food Safety Authority (EFSA) performs the RA of single active substances based on toxicity data (e.g. LD_{50}) provided with the pre-market registration dossiers (e.g. Draft Assessment Report) submitted by applicants (EFSA PPR Panel, 2012). Although there is growing evidence that bees are exposed to a wide range of multiple chemicals "mixtures" (David et al. 2016; Tosi et al. 2018; Prado et al. 2019) which, in some instances, potentially trigger interactions such as synergistic effects (Carnesecchi et. al 2019a,b; Spurgeon et al. 2016), further work is needed to integrate information on such combined toxicity in RA practice (Rortais et al. 2017; Bopp et al. 2019; Topping et al. 2020). EFSA has recently published a MIXTOX guidance document to support harmonised methodologies for ecological RA of combined exposure to multiple chemicals while

specifically illustrating the integration of information on combined toxicity for bees (More et al. 2019). In this context, key recommendations include the need to further develop and implement generic *in silico* models such as quantitative structure–activity relationship (QSAR) to predict combined toxicity for bees and a broader range of species of ecological relevance. These models can support the integration of toxicity and mechanistic data for hazard assessment of single chemicals as well as for component-based approaches for mixture risk assessment (MRA). As a consequence, innovative QSAR models to predict combined toxicity of pesticides active substances in honey bees have been developed recently allowing the identification of structural features that may drive an increase or decrease in combined toxicity (Carnesecchi et al. 2020). In addition, authors have also highlighted that current data gaps regarding information on the mode of action (MoA) of single chemicals is still limiting the development and broader applications of such innovative QSAR models.

In the human-health and animal health areas, MoA refers to the major steps leading to an adverse health effect following interaction of the chemical with biological targets at the sub-cellular level, whilst not necessarily implying the full understanding of the mechanism of action at the molecular level (WHO, 2009; Boobis et al. 2006; OECD, 2017; EFSA PPR, 2013). Similarly, in ecological RA, MoA has been defined as a functional change at the cellular level triggered by the substance entering the organism which then involves levels of biological organisation from organisms, multiple species, to populations all the way to ecosystems (Kienzler et al. 2017, 2019; Segner, 2011). Several different MoA frameworks exist for classifying chemicals (Verhaar et al. 1992; Russom et al. 1997; Kienzler et al. 2017), which allowed developing robust predictive tools for MoA classification such as EnviroTox database (Kienzler et al. 2019; Connors et al. 2019) and TEST software (Martin et al. 2013, 2015). However, such tools are mostly based only on vertebrate information specifically fish toxicity data, thus limiting their application to aquatic environmental RA. Similarly, different QSAR models predicting toxicity for single chemicals in honey bees are available but to date these do not address the challenge of the integration of toxicity prediction together with MoA profiling (Venko et al., 2018; Singh et al., 2014; Hamadache et al., 2018; Como et al., 2017; Toropov and Benfenati, 2007; Devillers et al. 2002). MoA information together with information on adverse outcome pathways (AOPs) in honey bees can provide a sound understanding of the link between molecular targets , as molecular initiating event and key events leading to adverse effects at individual and colony level as recently for neonicotinoids targeting nicotinic acetylcholine receptors in honey bees (LaLone et al. 2017).

The present manuscript aims to address the challenge of integrating MoA information in QSAR models with the first integrative honey bee QSAR models for PPPs using open source databases i) to predict acute contact toxicity (LD_{50}) and ii) to profile the MoA of active substances. In addition, the current study explores the development of harmonised MoA classification schemes

to relate the structure toxicological information with the target sites of PPPs active substances for a range of applications, including toxicity predictions and refining the grouping of chemicals for component-based RA of multiple chemicals (More et al. 2019).

2. Materials and Methods

2.1 QSAR model development

2.1.1 Data curation

Pesticide toxicity data for honey bees (*Apis mellifera*) expressed as LD_{50} µg/bee (acute contact, 48h) were retrieved in June 2018 from three publicly available databases (1) EFSA's chemical hazards database "OpenFoodTox" (Benfenati et al. 2020; DOI: 10.5281/zenodo.3693783), (2) US-EPA ECOTOXicology knowledgebase (ECOTOX; available at https://cfpub.epa.gov/ecotox/) and (3) Pesticide Properties DataBase (PPDB; available at https://sitem.herts.ac.uk/aeru/ppdb/en/index.htm).

Criteria for data pruning were applied following to the official guideline (OECD, 1998) according to which pesticides are administered by contact routes to represent the type of exposure under field conditions. Overall, information on the specific criteria applied for the data pruning within each database are presented in Table 6.1.

Table 6.1. Criteria applied for the data pruning of toxicity data (expressed as LD_{50}) as reported in three different databases (OpenFoodTox, US-EPA ECOTOX, PPPDB).

Database	Species	Organism life stage	Exposure duration	Route	Dose unit (LD_{50})	Qualifier (tested chemical)	Chemical purity
OpenFoodTox	Honey bee	Adult	48h	Dermal	µg/bee µg/piece ng/bee	"as such"	NA
US-EPA ECOTOX	*Apis mellifera* (with all subspecies)	Adult	48h	Topical, Dermal	µg/piece µg/org µg/bee µg/g org ng/µl ng/org AI ng/org ppm ppb AI mg/org	NA	> 80
PPDB	Honey bees (*Apis* spp.)	NA	48h	Contact	µg/bee	NA	NA

NA= Not Available

After the creation of a list of unique CAS numbers and names, all the SMILES have been retrieved with a semi-automated workflow (Gadaleta et al. 2018). SMILES provided with the original databases have been used for manual check and no differences have been found. After an analysis on stereoisomers, it has been found that a high percentage of molecule had chiral points without any specification of chirality. This finding led us to the decision of stripping all the stereoisomer information from SMILES in order to have a more homogenous dataset.

SMILES have then been associated with original values and two different procedures have been adopted whether it was a classification or regression dataset building:

- **Classification-based models;** the threshold used for toxicity classification was 100 µg/bee, which corresponds to the limit test (OECD, 1998). If the values associated to the

same SMILES fell under and up this threshold, the relative compound have been excluded for classification modeling. The final dataset is constituted by 413 compounds.

- **Regression-based model;** all compounds presenting the qualifier (>) were excluded. All values were converted in µmol/bee and grouped by SMILES; geometric means for each value associated with the same SMILES were calculated. When values associated with the same SMILES showed a 3-fold difference between the maximum and the minimum, the relative compound was excluded from the regression modelling. When continuous data were available, these were transformed on the logarithmic scale. In addition, compounds excluded from the classification modelling were also excluded from the regression modelling. This is of particular relevance to abamectin and avermectin B1 which were both manually excluded because they are used as a mixture. The final dataset was built from 113 compounds.

2.1.2 Data splitting

Both datasets for classification-based and regression-based models were divided into a Training (TS) and Validation Set (VS) in a ratio of 80:20. The number of compounds in each set is shown in Table 6.2. In order to ensure a uniform distribution of the endpoint values and to have the widest possible chemical space in the two subsets, we applied an activity/structure sampling method. Pubchem fingerprints are calculated starting from SMILES. In case of regression-based models five equal-sized bins were created based on fixed ranges of experimental values. For classification-based models, only two groups have been considered (high toxicity – low toxicity). For each bin, a deterministic algorithm selected the 80% of compounds starting from a selected group of compounds (the first 5 compounds of the dataset) and looking for the most diverse molecules using as metric the Tanimoto similarity coefficient (Tanimoto, 1958) calculated on fingerprint. The picking algorithm is called MaxMin (Ashton et al. 2002). The resulting 80% of each bin was regrouped in TS and the remaining compounds constituted the VS (Worth et al. 2005; Golbraikh and Tropsha, 2002; Golbraikh et al. 2003).

Table 6.2. Datasets splitting for classification- and regression-based models. The complete datasets are reported in Table S6.1 and table S6.2.

Set	Classification-based models	Regression-based models
Train	328	88
Test	83	25
Tot	411	113

2.1.3 Calculation of molecular descriptors

Dragon 7.0 was used for the calculation of 2 D molecular descriptors while stereoisomer information was removed. Moreover, descriptors with constant values (standard deviation 0) or correlated over 95% (Pearson correlation coefficient) with another descriptor (stronger correlation with the endpoint) were rejected. Centering and scaling as well as a range of methods of variable selection to fit the algorithm used for the model derivation were applied to all descriptors. Genetic algorithm has been used for for Decision Trees (DT), k-nearest neighbors (k-NN), Multiple linear regression (MLR) and Partial least squares regression (PLS) (OECD, 2007), while VSURF (Genuer et al. 2015) has been used for random forest (RF) (Breiman, 2001). Genetic algorithm (OECD, 2007) has been applied with gaselect (Kepplinger et al. 2017) R package implementation, using a custom fitness function. The same user function (Underlying Algorithm for the derivation of the fitness function) is based on the same package of the algorithm used after the descriptor selection and it is the same of the following model derivation. In particular, a custom function has been implemented using cross-validation error as given in the output of DT R implementation and Cohen's Kappa (Cohen, 1960) between experimental and Cross validated predictions for k-NN.

2.1.4 Learning algorithms

In order to build classification-based models, DT, RF and k-NN were employed. With regard to regression models MLR and PLS were used. All the parameters are reported in table S6.3-S6.4 for classification- and regression-based models, respectively.

2.1.4.1 Multi Linear regression

MLR is the most popular algorithm for QSAR development since it produces a transparent and an algorithm that is easily reproducible (OECD, 2007). MLR describes how a single response variable "Y" depends linearly on a number of predictor variables. A MLR can lead easily to overfitting, especially when dealing with a high number of predictors. In order to avoid overfitting, genetic algorithm has been applied using as fitness function Q^2. The algorithm used is "lm" as implemented in the package caret (Kuhn, 2008).

2.1.4.2 Partial Least Squares

PLS is a combination of MLR and principal component analysis (PCA). It performs a MLR using as predictors the principal components of the original data matrix. The algorithm used is "pls" as implemented in the package caret. In order to select the best parameters "Caret" hyper parameter tuning grid has been used. For pls implementation only the number of components is tuned.

2.1.4.3 K-nearest neighbor

The k-NN identifies a k number of neighbors for the target compound that will be used to provide a prediction of the endpoint. It is a transparent algorithm widely used for QSAR datasets with different similarity metrics to select neighbors (Manganaro et al. 2016). The algorithm uses a metric to measure distances between molecules, after pruning of molecular descriptors using Genetic Algorithm. The algorithm used is "KKNN" (Samworth, 2012) implemented in the package caret. In order to select the best parameters "Caret" hyper parameter tuning grid has been used. For k-NN implementation these parameters are tuned: the maximum number of neighbors (kmax), Parameter of Minkowski distance (distance) and the type of kernel estimate of the densities, used to weight the mean (kernel).

2.1.4.4 Decision Trees

DTs (Quinlan, 1986, 1987) are flowchart-like structures formed by a series of nodes that generates a set of rules that follow a IF Variable A is X THEN…" pattern. All the rules are hierarchically connected until a terminal node is reached, which assigns the class to the compound. The used algorithm is "RPART" as implemented in the package caret. In order to select the best parameters "Caret" hyper parameter tuning grid has been used. For RPART implementation only the complexity parameter (CP) is optimised, leading to a minimum improvement in the model needed at each node.

2.1.4.5 Random forest

After VSURF variable selection, a RF variant as implemented in the package "Ranger" (Wright et al. 2017) has been used for model derivation. In order to select the best parameters "Caret" (Kuhn, 2008), hyper parameter tuning grid has been used. Three parameters were tuned by grid search: mtry (number of randomly selected descriptors used in each tree of the RF), splitrule (the rule used to choose descriptors for a single tree, i.e. "gini" or "extratrees" for classification; "variance" or "extra trees" for regression), and min.node.size (minimal node size of trees). The number of trees was left as default value (500).

2.1.5 Statistical criteria

2.1.5.1 Classification-based models

In order to evaluate the performance of the classification-based models for two classes $LD_{50} < 100$ µg/bee (1) and $LD_{50} \geq 100$ (0), the following statistical criteria were: sensitivity, specificity, accuracy (BA) and Matthews correlation coefficient (MCC). Generally, the MCC coefficient is applied in machine learning to measure the quality of binary classifications, particularly when the classes present very different sizes (Dao et al., 2011).

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \qquad \text{(Eq. 1)}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad \text{(Eq. 2)}$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad \text{(Eq. 3)}$$

$$\text{MCC} = \frac{\text{TP}*\text{TN} - \text{FP}*\text{FN}}{\sqrt{(\text{TP+FP})(\text{TP+FN})(\text{TN+FP})(\text{TN+FN})}} \quad \text{(Eq. 4)}$$

TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively. MCC values range between -1 and +1, while the latter indicates a perfect prediction, a value of 0 indicate a prediction no better than a random one, and a value of -1 show total disagreement between predicted and observed values (Dao et al. 2011).

2.1.5.2 Regression-based models

The determination coefficient (R^2) is the fitness function used to evaluate the goodness of fit and is calculated as shown in Equation 5.

$$R^2 = 1 - \frac{\sum(yi - \hat{y}i)^2}{\sum(yi - \bar{y}i)^2} \quad \text{(Eq. 5)}$$

where yi is the experimental value of the i-th chemical in the dataset; $\hat{y}i$ is the calculated value of the i-th query compound in the dataset for the determination of R^2; $\bar{y}i$ is the mean of the experimental values of the compounds in the dataset, for all the N compounds. Similarly, RMSE (root mean square error) is an additional parameter used in the evaluation (Eq.6) which is calculated as follows:

$$RMSE = \sqrt{\sum \frac{(\hat{y}i - yi)^2}{N}} \quad \text{(Eq. 6)}$$

The Cross-validated determination coefficient (Q^2) has been used for the calculation of statistics in cross-validation (Eq. 7):

$$Q^2 = 1 - \frac{\sum(y_k - \acute{y}_k)^2}{\sum(y_k - \bar{y}_k)^2} \quad \text{(Eq. 7)}$$

$y_k, \acute{y}_k, \bar{y}$ are observed, cross validated prediction and average values of the dependent variable, respectively (Golbraikh and Tropsha, 2002). For all the models the reported Q^2 is the mean value of the Q^2 of a 5-fold cross-validation repeated 3 times. Similarly, additional statistical parameters such as Q2-F1, Q2-F2, Q2-F3, CCC, r^2_0, r^2_m, $\overline{r^2}$, Δr^2_m, k and k' are calculated according to Gramatica and Sangion (2016).

2.1.6 Applicability Domain

The applicability domain (AD) of a QSAR model is defined as "the physico-chemical, structural, or biological space, knowledge or information on which the TS of the model has been developed, and for which it is applicable to make predictions for new compounds[…]. Ideally, the QSAR should only be used to make predictions within that domain by interpolation not extrapolation" (Eriksson et al. 2003). The models are specifically designed to deal with pesticides. The model performance is taken into account without considering AD. However, since the set used for modeling contains all the molecules under 800 of molecular weight, this value should be considered as the upper limit to predict compounds in a reliable way. Moreover, the QSAR models here developed will be implemented in the open source platform VEGA-HUB (https://www.vegahub.eu/; Benfenati et al. 2017), therefore the reliability of the prediction will be evaluated using the Applicability Domain Index (ADI), which is an aggregated result taking into account several aspects:

1) Similar molecules with known experimental value and their accuracy (or average error) in their prediction,
2) Concordance among the target and similar molecules for the experimental data,
3) Atom Centered Fragments similarity check,
4) Descriptors noise sensitivity analysis,
5) Model descriptors range check.

As additional value, the information of the MoA of the closest neighbors will be provided in order to better assess the reliability of the prediction.

2.2 Mode of Action

Several definitions of MoA are available from the literature, and scientific advisory bodies (WHO, 2009; Boobis et al. 2006; OECD, 2017) although the common goal is facilitating classification of chemicals according to mechanistic information (e.g. chemical class, molecular target) and through robust schemes (Verhaar et al. 1992; Enoch et al. 2008; Carriger et al. 2016; Kienzler et al. 2019). However, since here the authors refer exclusively to pesticides active substances used in PPPs, focus is given to classification schemes which are principally based on MoA as the target site of the substance (Sparks and Nauen, 2015; Casida, 2009; Wing et al. 2005). Hence, the main criterion for the refined classification of the MoA relies on the availability of the information describing the interaction with the receptor of the target species (e.g. sodium channel modulator, Acetylcholinesterase (AChE) inhibitors).

Qualitative information on the MoA of substances (n= 113) present in the regression-based QSAR model (see chapter 2.1.5.2) were collected from different sources such as publicly available databases and the peer-reviewed scientific literature. Priority was given to well-defined schemes used for the classification of pesticides such as the one proposed by the Insecticide Resistance Action

Committee (IRAC; Sparks and Nauen, 2015), Fungicide Resistance Action Committee (FRAC; Hermann and Stenzel, 2019) and Herbicide Resistance Action Committee (HRAC; Beffa et al. 2019). Similarly, MoA information was retrieved from the PPDB (available at https://sitem.herts.ac.uk/aeru/ppdb/en/index.htm). When no data were available on the MoA from the above mentioned sources, the publicly available scientific literature were investigated (Sanchez-Bayo, 2012; Simon-Delso et al. 2015; Johnson et al. 2012, 2013; Leroux et al. 2008; De Castro et al. 2015). In addition, a comparison of MoA nomenclatures (i.e. site of action) reported in the different databases/schemes was carried out taking as main reference the Resistance Action Committee classifications (e.g. IRAC, FRAC, HRAC) in order to provide users with a harmonised classification scheme. Pesticides active substances were classified according to their i) function (e.g. insecticide, fungicide, acaricides, herbicides, etc.), ii) chemical class (e.g. carbamates, pyrethroids, etc.) and iii) site of action (e.g. AChE inhibitors, sodium channel modulators). Finally, chemicals were grouped according to the harmonised MoA (i.e. site of action) to allow an assessment of potential variability in acute contact toxicity (potency expressed as $\log LD_{50}$ µmol/bee) in honey bees across and within MoA groups. This exercise provides means for a refined prioritisation and grouping of chemicals for RA of combined exposure to multiple chemicals as recommended by EFSA MIXTOX Guidance, as well as supports to move towards an understanding of mechanisms of toxicity of active substances in PPPs (More et al. 2019).

3. Results and discussions

3.1 QSAR model development

3.1.1 Data collection and analysis

PCA has been performed in order to check if the TS covers the space of the test set (Figure 6.1). An ellipsoid has been drawn for both sets in order to check the overlap. The test resulted structurally covered by the TS. As reported in the Materials and Methods, 132 substances belong to the minor class (toxic) while 279 substances to the major one (non-toxic), having a ratio higher than 1:2. The dataset is slightly unbalanced towards the low toxicity substances (LD50 ≥ 100 µg/bee), nonetheless there is enough coverage of the minor class, making the dataset suitable for modelling. Similarly, he negative logarithm of LD_{50} (pLD_{50} µmol/bee) have been used for the quantitative modelling. The data distribution curve of pLD_{50} is plotted in Figure 6.2.

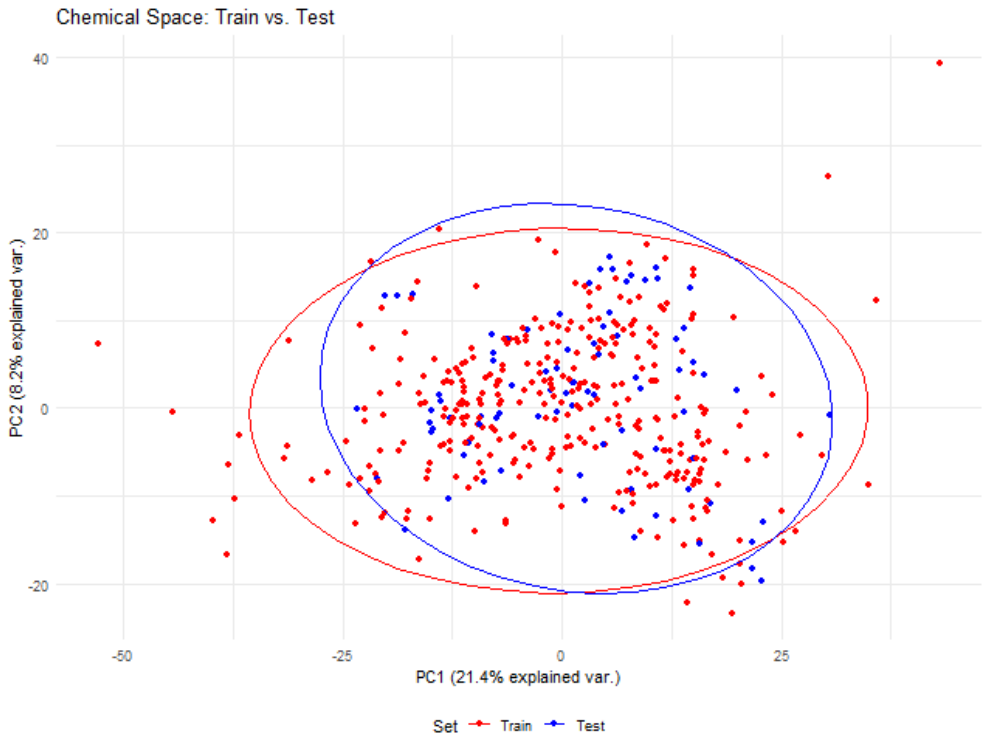


Figure 6.1. Principal Component Analysis (PCA) on DRAGON descriptors for classification-based models (ellipsoid calculated at 0.95 probability).

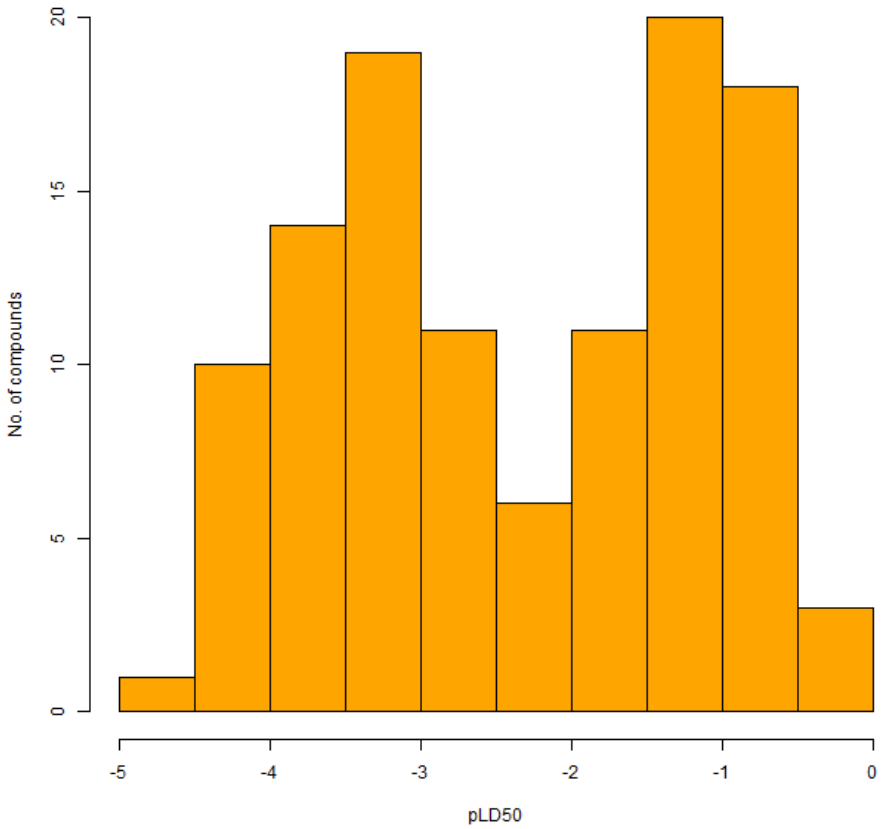


Figure 6.2. Histogram of pLD50 (pLD50 µmol/bee) data used for regression-based model.

3.1.2 Chemical descriptors selection

VSURF was used to select best descriptors as input for both classification and regression RF. Each RF used for the feature selection was constituted by 100 trees. Genetic Algorithm used for the other models (MLR, PLS, DT, k-NN) was set up to find the best number of descriptors between 5 and 12. Tables 6.3 and 6.4 report the selected descriptors selected for classification and regression models, respectively. Additionally, plots showing the most important variables for each model are reported in the supplementary material (Figures S6.1-S6.2).

Table 6.3. Selected descriptors for random forest (RF), decision tree (DT) and k-nearest neighbors (k-NN) classification-based models.

RF	DT	KNN
X0Av	B09[C-O]	ICR
F02[C-P]	F02[C-P]	LOC
ATSC1e	SpMax_A	VE1_B(p)
MATS3v	ChiA_B(m)	MATS8e
SpMax1_Bh(s)	ATS1m	GATS3m
SpMAD_X	SpMax2_Bh(p)	Eta_sh_x
RBF	P_VSA_s_1	CATS2D_04_DA
B05[C-P]	-	CATS2D_00_LL
MATS1s	-	T(Cl..Cl)
-	-	SAdon

Table 6.4. Selected descriptors for random forest (RF), multi linear regression (MLR), partial least squares (PLS), decision tree (DT) and k-nearest neighbors (k-NN) regression-based models.

RF	MLR	PLS	DT	KNN
CATS2D_07_LL	CATS2D_07_LL	CATS2D_07_LL	piPC06	X4v
MATS8v	MATS4m	F06[N-S]	ChiA_B(p)	GGI10
X3v	GATS8s	MATS4m	MATS8v	SpMin2_Bh(s)
CATS2D_03_DL	GGI8	GATS4e	SpMax2_Bh(p)	Eig02_AEA(dm)
TI2_L	JGI4	GATS8s	P_VSA_m_2	SsOH
IVDE	CATS2D_00_DD	JGI8	SpMaxA_EA(ed)	NssO
SpMax2_Bh(p)	CATS2D_03_DA	JGT	Eig03_EA(bo)	CATS2D_02_DD
CATS2D_09_LL	CATS2D_03_DL	CATS2D_03_DL	Eig05_EA(dm)	CATS2D_07_LL
SaaN	F03[C-N]	F02[N-S]	F04[N-P]	F01[N-O]
SpPosA_B(i)	Psychotic-50	F04[O-O]	F04[O-S]	F05[C-S]
GATS5m	-	-	-	F06[N-S]
MATS4v	-	-	-	F07[C-N]
N-074	-	-	-	-

3.1.3 Classification-based models

In this study, three classification-based models were developed using different approaches as RF, DT and k-NN. For each model, the performance in TS, 5-Fold cross validation (CV) and VS was

evaluated in order to identify the best model (Figure S6.3). Results for statistical quality are shown in Table 5. All the models showed an acceptable sensitivity (>0.89), while the performance for RF and DT decreased when evaluating the specificity in CV (0.53 and 0.54, respectively). MCC for VS across the three models is above 0.75, thus showing that the models are able to identify both classes when predicting an external set of compounds (Table 6.5 and Figure S6.3). However, our results show that k-NN is the most robust model when identifying minority class (<100 µg/bee) compounds. Additional Radar plots of the models are reported in supplementary materials (Figure S6.3), while the predictions and the descriptors are reported in Table S6.5-S6.7.

Table 6.5. Results of the statistical quality for random forest (RF), decision tree (DT) and k-nearest neighbor (k-NN) classification-based models. Test set (TS), cross validation set (CV) and validation set (VS) are reported.

Algorithm	***RF***			***DT***			***K-NN***		
Set	***TS***	***CV***	***VS***	***TS***	***CV***	***VS***	***TS***	***CV***	***VS***
Sensitivity	1.00	0.94	0.96	0.95	0.92	0.89	0.77	0.90	0.93
Specificity	1.00	0.54	0.74	0.58	0.53	0.89	0.96	0.67	0.85
Accuracy	1.00	0.81	0.89	0.83	0.80	0.89	0.90	0.83	0.90
MCC	1.00	0.54	0.75	0.59	0.50	0.76	0.76	0.59	0.78

In the literature, other classification-based QSAR models to predict pesticides toxicity in honey bees are available, and thus a comparison of their predictive performance is illustrated here (Table 6.6). Overall, the k-NN model reported here shows higher performance with regards to statistical quality of the results compared with those from existing k-NN models (Como et al. 2017; Venko et al. 2018). Similarly, the PNN-QSTR model (Sing et al. 2018) demonstrated a comparable performance compared with the current k-NN model, although based on a probabilistic neural network approach. Here, it is important to highlight that the two-category model presented here is the first classifier built based on the largest available dataset encompassing data for 411 pesticides active substances from three different open access sources (i.e. EFSA's OpenFoodTox, US-EPA ECOTOX and PPDB) and thus it has the advantage to increase the application domain of the model for a large range of pesticides.

Table 6.6. Comparison of the classification-based model (k-NN; best model) developed here with publicly available two-category models for acute toxicity towards honey bees (threshold 100 µg/bee). Results of the statistical quality are reported for training (TS) and validation sets (VS) in each model.

Reference	Set	Compounds (n)	Sensitivity	Specificity	Accuracy
Venko et al., 2018	TS	205	0.88	0.90	0.89
	VS	49	0.75	0.79	0.78
Como et al., 2017	TS	192	0.60	0.88	0.76
	VS	50	0.80	0.86	0.84
Singh et al., 2014	TS	175	1.00	1.00	1.00
	VS	62	0.86	1.00	0.87
Present study	TS	328	0.77	0.96	0.90
(k-NN model)	VS	83	0.93	0.85	0.90

3.1.4 Regression-based model

As for the classification-based models, we evaluated the performance in TS, CV and VS. Overall, all models provided good predictivity with regards to goodness of fit (TS) except for PLS and DT, which were associated with significantly R^2 lower than those from other models (Table 6.7). Statistical quality for DT model was not satisfactory for CV and VS, while RF had an R^2 close to 0.5 in CV. In terms of reproducibility and overall performance, our results show that the best model is represented by k-NN (R^2= 0.74) followed by RF (R^2= 0.72) (Table 6.7). Additional Scatter plots (predicted vs experimental) of all models here developed are reported in Figure S6.4, while the predictions and the descriptors are reported in Table S6.8-S6.12.

Table 6.7. Statistical robustness and performance for random forest (RF), multi linear regression (MLR), partial least squares (PLS), decision tree (DT) and k-nearest neighbors (k-NN) regression-based models. Test Set (TS), Cross Validation set (CV) and Validation Set (VS) are reported. Statistical parameters are reported and defined according to Gramatica and Sangion (2016).

Parameter	RF	MLR	PLS	DT	KNN	Acceptability criteria
RMSE -TS	0.41	0.66	0.76	0.71	0.39	
R^2 - TS	0.88	0.7	0.61	0.66	0.9	0.6
MAE - TS	0.34	0.56	0.61	0.53	0.3	
CCC - TS	0.93	0.82	0.76	0.79	0.94	
$Q^2_{5\text{-FOLD CV}}$	0.49	0.59	0.53	0.35	0.63	0.5
RMSE - VS	0.8	0.93	0.9	1.01	0.71	
R^2 - VS	0.72	0.55	0.59	0.46	0.74	0.6
MAE - VS	0.63	0.76	0.66	0.79	0.52	
Q2-F1	0.65	0.53	0.56	0.44	0.72	
Q2-F2	0.65	0.52	0.56	0.44	0.72	
Q2-F3	0.57	0.41	0.45	0.31	0.65	
CCC - VS	0.75	0.73	0.71	0.66	0.83	
R^2_0 - VS	0.65	0.53	0.58	0.46	0.73	
R^2_M - VS	0.53	0.48	0.52	0.43	0.66	0.5
$\overline{R^2}$ - VS	0.35	0.42	0.35	0.31	0.54	0.5
ΔR^2_M - VS	0.36	0.11	0.34	0.23	0.23	<0.3
K - VS	0.98	0.96	0.93	0.94	0.96	0.85 < k < 1.15
K' - VS	0.93	0.91	0.95	0.91	0.97	0.85 <k' < 1.15

The best model developed here (k-NN) was used for comparative purposes with the published QSAR regression models predicting honey bee toxicity. Results are provided in Table 6.8 and show that is the partial within the current model is related to the VS and the characterisation of the data points. In contrast to Hamadache et al. (2018), our k-NN model is highly curated, and all the data above 100 µg/bee (limit test) were filtered out (OECD, 1998). In fact, these data points, when included in a model, behave as attractors of the regression giving an over-estimation of optimistic performances. Hence, after filtering those compounds, the predictivity of the refined QSAR model increased significantly, thus providing a tool to predict honey bee toxicity for substances of concern (i.e. moderate/high toxicity) with high precision. For this reason, the classification- and regression-based QSAR models developed here are not meant to be used as two distinct models but rather as an integrative tool following a specific hierarchical workflow (Figure 6.3).

Table 6.8. Comparison of regression-based model here developed with publicly available models for acute toxicity towards honey bees. Determination coefficient (R2) and root-mean-square error (RMSE) are reported.

Reference	Set	Compounds (n)	R^2	RMSE
Devillers et al., 2002	TS	86	0.82	0.430
	VS	11	0.94	0.39
Toropov and Benfenati, 2007	TS	85	0.68	0.82
	VS	20	0.72	0.68
Dulin et al., 2012	TS	39	0.81	0.350
	VS	6	0.85	0.218
Singh et al., 2014	TS	190	0.85	0.50
	VS	47	0.86	0.33
Hamadache et al., 2018	TS	95	0.98	0.36
	VS	16	0.96	0.71
Present study	TS	88	0.90	0.39
	VS	25	0.74	0.71

3.1.5 Integrative honey bee QSAR model

As illustrated in the above paragraph, the current study aimed to develop the first integrative honey bee QSAR model to allow predicting the toxicity of unknown compounds following a specific hierarchical workflow (Figure 6.3). Here, 12 compounds were tested as independent datasets compared to the training sets of both classification- and regression-based models for the evaluation of the predictivity of the models and the potential application of the integrative tool. Results show that for the majority of compounds (n= 10), the quantitative prediction provides satisfactory predictions compared to those from the corresponding experimental data (Table 6.9) with the exception of two compounds (momfluorothrin, CAS n. 609346-29-4; chloroxuron, CAS n. 1982-47-4) for which the predictions from the classification-based model underestimated the toxicity resulting in an incorrect classification (Table 6.9).

An example of output from the integrative honey bee QSAR model is shown in Table 6.9 which reports CAS number, experimental MoA, predicted toxicity class (1= toxic / 0= non-toxic; threshold 100 µg/bee) and quantitative prediction (experimental LD_{50} µg/org) for the target chemical as well as the five most similar compounds identified by the k-NN model for each target. Hence, one of the novelty of the tool is to provide users with toxicity predictions (LD_{50} µg/org) as well as experimental MoA and LD_{50} for the five most similar compounds identified by the k-NN model.

Table 6.9. Results of the validation of the integrative honey bee QSAR model for 12 target chemicals that are independent from the training sets of both classification- and regression-based models. The output illustrates the CAS number, experimental MoA, predicted toxicity class (1= toxic / 0= non-toxic; threshold 100 µg/bee), experimental LD50 (µg/org), and quantitative prediction (pred LD50 µg/org) of the target substance. In addition, the table provides the five most similar compounds identified through the k-NN model for each target chemical.

Target	*Chemical#1*	*Chemical#2*	*Chemical#3*	*Chemical#4*	*Chemical#5*
momfluorothrin **CAS: 609346-29-4** **MoA: Sodium channel modulators_Na channel(+)** **Pred Class: 0** **Exp (µg/org): 0.2** **Pred (µg/org): 0.22**	4-(2,4-Dichlorophenoxy)butanoic acid CAS: 94-82-6 MoA: Synthetic auxins (action like indole acetic acid) Exp (µg/org): 14.5	chlorpropham CAS: 101-21-3 MoA: Inhibition of mitosis/microtubule organization Exp (µg/org): 96.1	2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate) CAS: 22781-23-3 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.43	methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.16	alachlor CAS: 15972-60-8 MoA: Inhibition of very-long-chain fatty acid synthesis (VLCFAs) Exp (µg/org): 16.0
terbufos **CAS: 13071-79-9** **MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)** **Pred Class: 1** **Exp (µg/org): 4.1** **Pred (µg/org): 2.79**	tefluthrin CAS: 79538-32-2 MoA: Sodium channel modulators_Na channel(+) Exp (µg/org): 0.28	resmethrin CAS: 10453-86-8 MoA: Sodium channel modulators_Na channel(+) Exp (µg/org): 0.06	imiprothrin CAS: 72963-72-5 MoA: Sodium channel modulators_Na channel(+) Exp (µg/org): 0.4	5-Amino-1-[2,6-dichloro-4-(trifluoromethyl)phenyl]-4-[(trifluoromethyl)sulfinyl]-1H-pyrazole-3-carbonitrile CAS: 120068-37-3 MoA: GABA-gated chloride channel blockers_GABA-R(-) Exp (µg/org): 0.01	(1,3,4,5,6,7-Hexahydro-1,3-dioxo-2H-isoindol-2-yl)methyl ester 2,2-dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid CAS: 7696-12-0 MoA: Sodium channel modulators_Na channel(+) Exp (µg/org): 0.16

Target	Chemical#1	Chemical#2	Chemical#3	Chemical#4	Chemical#5
chloroxuron **CAS: 1982-47-4** **MoA: Inhibition of photosynthesis at PS II** **Pred Class: 0** **Exp (µg/org): 16.0** **Pred (µg/org): 19.88**	alachlor CAS: 15972-60-8 MoA: Inhibition of very-long-chain fatty acid synthesis (VLCFAs) Exp (µg/org): 16.0	methiocarb CAS: 2032-65-7 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.29	chlorpropham CAS: 101-21-3 MoA: Inhibition of mitosis/microtubule organization Exp (µg/org): 96.1	4-Bromo-2-(4-chlorophenyl)-1-(ethoxymethyl)-5-(trifluoromethyl)-1H-pyrrole-3-carbonitrile CAS: 122453-73-0 MoA: Uncoupler of oxidative phosphorylation Exp (µg/org): 0.12	2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate) CAS: 22781-23-3 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.43
pyrethrins (cinerin II) **CAS: 121-20-0** **MoA: Sodium channel modulators_Na channel(+)** **Pred Class: 1** **Exp (µg/org): 0.01** **Pred (µg/org): 0.02**	methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.16	thiodicarb CAS: 59669-26-0 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 3.1	chlorbromuron CAS: 13360-45-7 MoA: Inhibition of photosynthesis at PS II Exp (µg/org): 16.0	chlorpropham CAS: 101-21-3 MoA: Inhibition of mitosis/microtubule organization Exp (µg/org): 96.1	2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate) CAS: 22781-23-3 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.43
pyrethrins (jasmolin II) **CAS: 1172-63-0** **MoA: Sodium channel modulators_Na channel(+)** **Pred Class: 1** **Exp (µg/org): 0.01** **Pred (µg/org): 0.02**	methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.16	thiodicarb CAS: 59669-26-0 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 3.1	chlorbromuron CAS: 13360-45-7 MoA: Inhibition of photosynthesis at PS II Exp (µg/org): 16.0	chlorpropham CAS: 101-21-3 MoA: Inhibition of mitosis/microtubule organization Exp (µg/org): 96.1	2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate) CAS: 22781-23-3 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.43

Target	Chemical#1	Chemical#2	Chemical#3	Chemical#4	Chemical#5
pyrethrins (cinerin I) **CAS: 25402-06-6** **MoA: Sodium channel modulators_Na channel(+)** **Pred Class: 1** **Exp (µg/org): 0.01** **Pred (µg/org): 0.07**	methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.16	thiodicarb CAS: 59669-26-0 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 3.1	chlorbromuron CAS: 13360-45-7 MoA: Inhibition of photosynthesis at PS II Exp (µg/org): 16.0	chlorpropham CAS: 101-21-3 MoA: Inhibition of mitosis/microtubule organization Exp (µg/org): 96.1	2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate) CAS: 22781-23-3 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.43
pyrethrins (pyrethrin I) **CAS: 121-21-1** **MoA: Sodium channel modulators_Na channel(+)** **Pred Class: 1** **Exp (µg/org): 0.01** **Pred (µg/org): 0.06**	methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.16	thiodicarb CAS: 59669-26-0 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 3.1	chlorbromuron CAS: 13360-45-7 MoA: Inhibition of photosynthesis at PS II Exp (µg/org): 16.0	chlorpropham CAS: 101-21-3 MoA: Inhibition of mitosis/microtubule organization Exp (µg/org): 96.1	2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate) CAS: 22781-23-3 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.43
2,2-Dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid (3-phenoxyphenyl)methyl ester **CAS: 26002-80-2** **MoA: Sodium channel modulators_Na channel(+)** **Pred Class: 1** **Exp (µg/org): 0.07** **Pred (µg/org): 0.02**	methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.16	thiodicarb CAS: 59669-26-0 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 3.1	chlorbromuron CAS: 13360-45-7 MoA: Inhibition of photosynthesis at PS II Exp (µg/org): 16.0	oryzalin CAS: 19044-88-3 MoA: Inhibition of microtubule assembly Exp (µg/org): 40.8	chlorpropham CAS: 101-21-3 MoA: Inhibition of mitosis/microtubule organization Exp (µg/org): 96.1
carbosulfan| Carbosulfan **CAS: 55285-14-8** **MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)** **Pred Class: 1**	2-Methyl-2-(methylthio)propanol O-[(methylamino)carbonyl]oxime CAS: 116-06-3 MoA:	thiofanox CAS: 39196-18-4 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)	methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)	2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate) CAS: 22781-23-3	imiprothrin CAS: 72963-72-5 MoA: Sodium channel modulators_Na channel(+) Exp (µg/org): 0.4

Exp (µg/org): 0.18 **Pred (µg/org): 0.26**	Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.29	Exp (µg/org): 0.06	Exp (µg/org): 0.16	MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.43	
ethoprophos **CAS: 13194-48-4** **MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)** **Pred Class: 1** **Exp (µg/org): 5.56** **Pred (µg/org): 0.56**	tefluthrin CAS: 79538-32-2 MoA: Sodium channel modulators_Na channel(+) Exp (µg/org): 0.28	imiprothrin CAS: 72963-72-5 MoA: Sodium channel modulators_Na channel(+) Exp (µg/org): 0.4	resmethrin CAS: 10453-86-8 MoA: Sodium channel modulators_Na channel(+) Exp (µg/org): 0.06	5-Amino-1-[2,6-dichloro-4-(trifluoromethyl)phenyl]-4-[(trifluoromethyl)sulfinyl]-1H-pyrazole-3-carbonitrile CAS: 120068-37-3 MoA: GABA-gated chloride channel blockers_GABA-R(-) Exp (µg/org): 0.01	(1,3,4,5,6,7-Hexahydro-1,3-dioxo-2H-isoindol-2-yl)methyl ester 2,2-dimethyl-3-(2-methyl-1-propenyl)cyclopropane carboxylic acid CAS: 7696-12-0 MoA: Sodium channel modulators_Na channel(+) Exp (µg/org): 0.16
fenitrothion| Phosphorothioic acid O,O-dimethyl O-(3-methyl-4-nitrophenyl)ester ***CAS: 122-14-5*** ***MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)*** ***Pred Class: 1*** ***Exp (µg/org): 0.25*** ***Pred (µg/org): 3.0***	N,N-Bis(2-methylpropyl)carbamothioic acid S-ethyl ester CAS: 2008-41-5 MoA: Inhibition of lipid synthesis – not ACCase Exp (µg/org): 29.0	triazamate CAS: 112143-82-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 27.0	chlorpyrifos-methyl CAS: 5598-13-0 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.15	6-Methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one CAS: 2439-01-2 MoA: N/A Exp (µg/org): 66.47	Phosphoramidothioic acid, O,S-Dimethyl ester CAS: 10265-92-6 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 1.37

Target	Chemical#1	Chemical#2	Chemical#3	Chemical#4	Chemical#5
Phosphorothioic acid, O,O-Diethyl-O-(4-nitrophenyl)ester CAS: 56-38-2 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Pred Class: 1 Exp (µg/org): 0.18 Pred (µg/org): 0.19	triazamate CAS: 112143-82-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 27.0	N,N-Bis(2-methylpropyl)carbamothioic acid S-ethyl ester CAS: 2008-41-5 MoA: Inhibition of lipid synthesis – not ACCase Exp (µg/org): 29.0	6-Methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one CAS: 2439-01-2 MoA: N/A Exp (µg/org): 66.47	methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.16	chlorpyrifos-methyl CAS: 5598-13-0 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (µg/org): 0.15

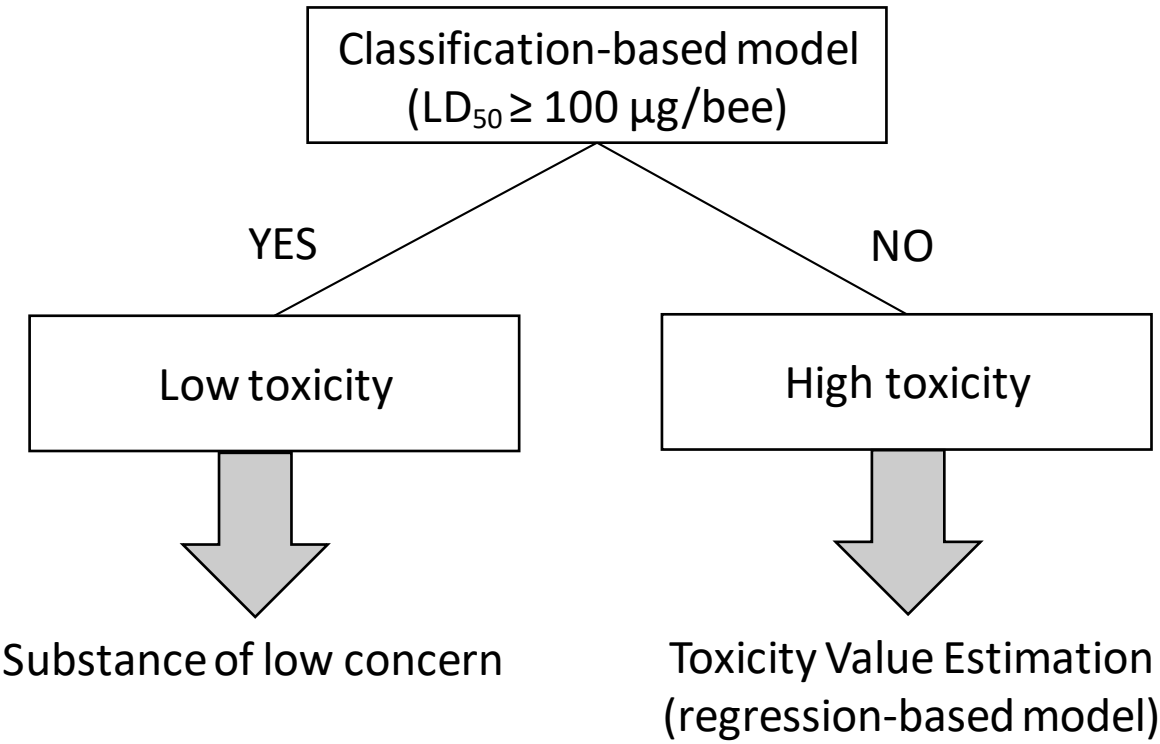


Figure 6.3. Hierarchical workflow to apply classification- and regression-based models for hazard assessment of pesticides active substances in honey bees.

3.2 MoA assessment

3.2.1 Harmonised MoA classification scheme

A total of 113 substances were included in the analysis of the MoA nomenclatures in order to provide an harmonised classification of pesticides active substances (used in PPPs) according to i) function

(e.g. insecticide, fungicide, etc.), ii) chemical class (e.g. carbamates, organophosphate, etc.), and iii) site of action (e.g. sodium channel modulators) (Table S6.13).

According to the above-mentioned criteria, an example of different nomenclatures applied in MoA classification for insecticides and fungicides is provided in Table 6.10 and Table 6.11, respectively. Qualitative information on MoA were extracted from three different sources: PPDB, Resistance Action Committee classifications (i.e. IRAC, FRAC, HRAC), and the peer reviewed scientific literature. Our results suggest that PPDB provides users with additional general and toxicological information on a number of substances (e.g. systemic/non-systemic, authorisation status), although the harmonisation of MoA has not been fully explored yet. Similarly, studies from the peer reviewed scientific literature have often applied a range of criteria (chemical class vs target site) when classifying chemicals, and did not prove univocal MoA nomenclature and classification schemes (Sanchez-Bayo, 2012; Johnson et al. 2012, 2013; Simon-Delso et al. 2015). In contrast, IRAC, FRAC and HRAC provide structured schemes, classifying pesticide active substances according to their target site (and cross-resistance), thus encompassing 32 insecticides, 56 fungicides and 26 herbicides MoAs, respectively (Sparks and Nauen, 2015; Hermann and Stenzel, 2019; Beffa et al. 2019). Therefore, although the different classification schemes analysed here have not been developed for the same purposes, the authors acknowledge that Resistance Action Committee schemes (IRAC, FRAC, HRAC) provide users with a more systematic and sound classification of MoAs (i.e. target site) for pesticide active substances used in PPPs.

A similar research effort to harmonise MoA classification schemes has been proposed by Kienzler et al. (2017, 2019). However, authors focused only on aquatic toxicity (e.g. fish) while taking into account several different definitions of MoA having degree of specificity based on fish behavioural responses, toxicological responses or weight of evidence classification. Therefore, this manuscript provides the first harmonised MoA classification for terrestrial non-target species and has been applied to honey bees, taking into account existing knowledge on the specific target site of pesticide active substances.

Overall, such MoA harmonised classification is valuable for the development and the testing of the validity of QSAR model and other *in silico* tools, thereby contributing to a sound mechanistic interpretation of the model (OECD, 2007). A specific application of this MoA analysis is in the refinement of applicability domain index (ADI) definitions for k-NN models in the open source VEGA platform (see chapter 2.1.6). In a similar fashion, k-NN QSAR models can be used to identify similar compounds "neighbors" and their structural features responsible for the toxicological mechanism(s) of the active substance (as illustrated in chapters 3.1.3 and 3.1.4), thus allowing prediction of potential target sites in terrestrial organisms such as earthworms and honey bees (Roy et al. 2020; Ghosh et al. 2020).

It should be noted that the harmonised classification proposed here (Table S6.13) has been carried out for 113 chemicals for which data were applied for the development of the regression-based QSAR model (chapter 3.1.4), thereby the full list of substances provided by the Resistance Action Committee classifications was not included (i.e. IRAC, FRAC, HRAC). However, although a smaller number of chemicals was included in the MoA analysis, the substances underwent data curation through a structured workflow in order to avoid ambiguous structures, and thus providing high-quality and curated datasets (Gadaleta et al. 2018).

Table 6.10. Example of different MoA nomenclatures used for the classification of substances according to different databases/schemes (PPDB; IRAC (Sparks and Nauen, 2015); Sanchez-Bayo, 2012; Johnson et al., 2012, 2013).

Substance name	**CAS n.**	**MoA/Site of action**			
		PPDB	***IRAC***	***Others (Sanchez-Bayo, 2012; Johnson et al., 2012, 2013)***	***Harmonised classification***
Triazamate	112143-82-5	Systemic with contact and stomach action.	Acetylcholinesterase (AChE) inhibitors Nerve action (Strong evidence that action at this protein is responsible for insecticidal effects)	Neurotoxic AChE(-)	Acetylcholinesterase inhibitors [AChE(-)]
Thiodicarb	59669-26-0	Mainly stomach action but some contact effects. Cholinesterase inhibitor.			
Chlorpyrifos	2921-88-2	Non-systemic with contact, inhalation and stomach action. Acetylcholinesterase (AChE) inhibitor.			

Table 6.11. Example of different MoA nomenclatures used for the classification of fungicides according to different databases/schemes (PPDB; FRAC (Hermann and Stenzel, 2019)).

Substance name	CAS n.	MoA/Site of action		
		PPDB	***FRAC***	***Harmonised classification***
Tetraconazole	112281-77-3	Systemic with protectant, eradicant and curative properties. Sterol biosynthesis inhibitor, acts mainly on the vegetative stages of fungi by blocking the mycelial growth either inside or on the surface of the host plant.	DMI-fungicides (DeMethylation Inhibitors) (SBI: Class I). TARGET: C14-demethylase in sterol biosynthesis (erg11/cyp51)	SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)
Spiroxamine	118134-30-8	Systemic with protective, curative and eradicative action. Disrupts membrane function. Inhibits sterol biosynthesis in membranes.	Amines ("morpholines") (SBI: Class II). TARGET: Δ14-reductase and Δ8 to Δ7-isomerase in sterol biosynthesis (erg24, erg2)	SBI: Class II_Δ14-reductase and Δ8 to Δ7-isomerase in sterol biosynthesis (erg24, erg2)

3.2.2 Pesticide active substances toxicity in relation to their MoA

While focus is given to the application of MoA for the grouping of multiple chemicals for combined exposure RA using component-based approaches in the ecological area, the honey bee example

provides an application for non-target terrestrial organisms as illustrated in the EFSA MIXTOX guidance (More et al. 2019).

Following the results of the critical analysis of MoA nomenclatures (chapter 3.2.1), 17 classes of MoAs have been defined for pesticide active substances (n= 113) according their specific target site (Sparks and Nauen, 2015; Hermann and Stenzel, 2019; Beffa et al. 2019) (Figure 6.4 and Table S6.13). Results show that 38% of chemicals were classified as insecticide/acaricides "*Acetylcholinesterase (AChE) inhibitors*", 18% as insecticide/acaricide "*Sodium channel modulators*", 6% as herbicides "*Inhibitor of photosynthesis at PS II*", 4% as insecticides "*Nicotinic acetylcholine receptor (nAChR) competitive modulators*", 4% as acaricide "*Mitochondrial complex I-II electron transport inhibitors*" and 4% as fungicides "*Sterol Biosynthesis Inhibiting (SBI) class I-II (erg11/cyp51)*" (Figure 6.4).

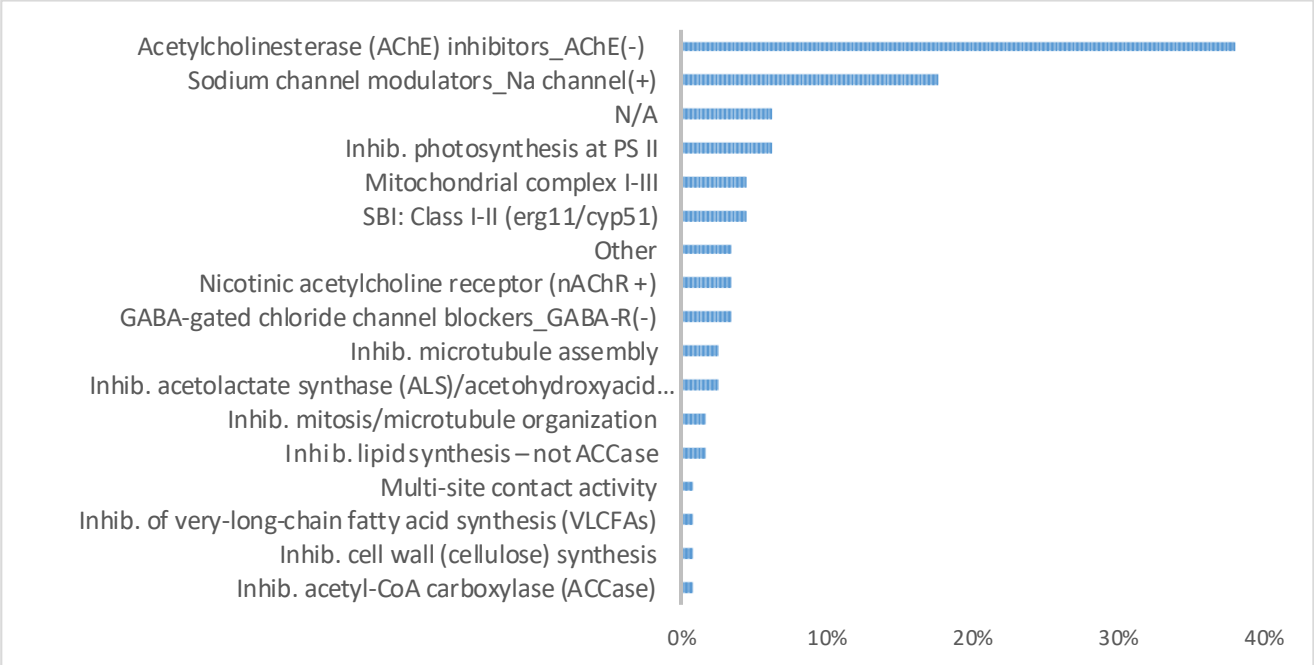


Figure 6.4. Percentage of pesticides active substances (n= 113) that are classified into each harmonised MoA group.

Overall, the method for grouping chemicals into assessment groups (AGs) proposed here has been carried out for chemicals (n= 113) used in the development of the regression-based QSAR model, thus including substances with moderate/high acute (contact) toxicity in honey bees (LD_{50}< 100 µg/bee) (Figure 6.5 and Table S6.13).

Insecticides/acaricides

According to our results, pyrethroid/pyrethrin insecticides and/or acaricides belonging to the MoA group of "sodium channel modulators", showed the highest acute contact toxicity in honey bees

(LD_{50} = 0.013-23.57 µg/bee), among which 66% with LD_{50} = 0.013-0.10 µg/bee (Figure 5 and Table S6.13). However, the most potent toxic insecticide was by far fipronil (CAS. 120068-37-3) which reported LD_{50}= 0.00389-0.00593 µg/bee (EFSA, 2006) and it is classified as a GABA-gated chloride channel blocker (Sanchez-Bayo, 2012). Similarly, insecticides/acaricides within the MoA group "Acetylcholinesterase (AChE) inhibitors", showed high (contact) toxicity in honey bees (LD_{50} = 0.0049-59.8 µg/bee), 69% of which presenting LD_{50} <1 µg/bee (Figure 5 and Table S6.13). Interestingly, all AChE inhibitors insecticides/acaricides were classified as carbamate or organophosphate substances (Table S6.13). By definition, insecticides and acaricides are generally harmful to non-target terrestrial organisms such as bees (Douglas et al. 2020; Sanchez-Bayo and Goka, 2016). However, differences in insecticide toxicity within the same MoA appear to be driven by their structural features and reactivity, as demonstrated for cyano-substituted neonicotinoids (e.g. thiacloprid and acetamiprid) which are three orders of magnitude less toxic to honey bees compared with other compounds of the same MoA group (Iwasa et al. 2004; Sanchez-Bayo, 2012; Carnesecchi et al. 2019a,b).

Herbicides and fungicides

The majority of triazine- and urea-derived herbicides (acting as "inhibitors of photosynthesis at PS II") presented $LD_{50s} \leq 20$ µg/bee (Figure 5 and Table S6.13). However, such active substances have nor to date been authorised on the EU market (EU Pesticide Database, 2019). Interestingly, oryzalin (dinitroaniline herbicide acting as inhibitor of microtubule assembly in weeds; CAS n. 19044-88-3), is currently authorised in some EU countries (ES, FR, IT, PT) and has its high toxicity in honey bees with LD_{50} = 40.8 µg/bee (Table S6.13). Oryzalin also has LD_{50}= 32 µg/bee following oral exposure to honey bees (EFSA, 2010). Similarly, sulfonylurea herbicides such as nicosulfuron (CAS n. 111991-09-4) acting as inhibitor of acetolactate synthase (currently authorised in several EU member states) exhibits moderate contact toxicity towards honey bees (LD_{50} = 76 µg/bee) (EFSA, 2008). In contrast, Sterol Biosynthesis Inhibiting (SBI class I) fungicides such as triazoles and pyrimidine have acute contact LD_{50} in honey bees ranging from 20-69 µg/bee with spiroxamine (SBI class II; CAS n. 118134-30-8) as the most potent active substance (LD_{50} = 4.22 µg/bee).

Overall, our results suggest that, although the majority of active substances (60%) were classified as insecticides and/or acaricides with MoA groups including AChE inhibitors, sodium channel modulators and nAChR competitive modulators, some herbicides and fungicides (authorised in the EU) exhibit moderate to high toxicity (LD_{50}= 0.8-69 µg/bee; Figure 5) in honey bees . Similarly, high variability in the acute contact toxicity of PPPs in honey bees is often reported within the same MoA as shown for insecticides classified under the MoA AChE inhibitors, GABA-gated chloride channel blockers", acaricides as mitochondrial complex I electron transport inhibitors (METI) (Figure 5).

However, looking at the size of the database, more toxicity data points would be needed to allow a more robust statistical analysis of LD_{50} variability in honey bees.

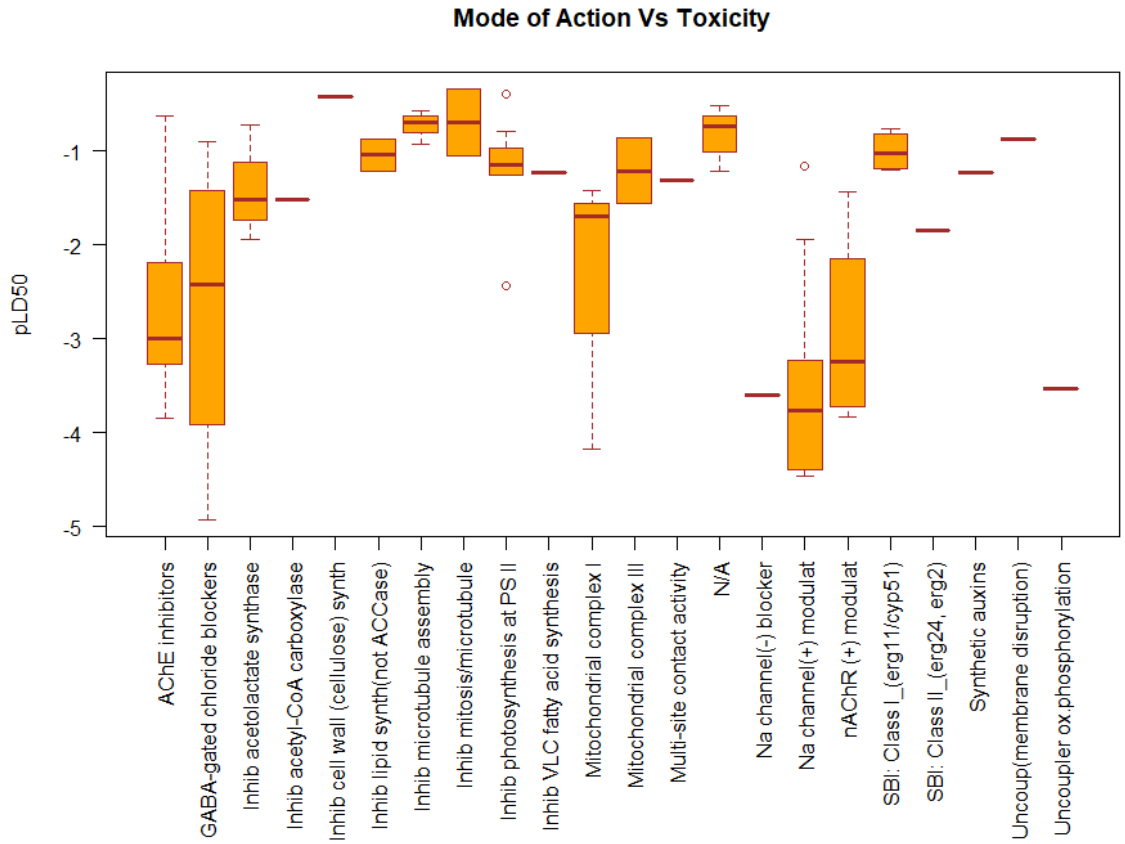


Figure 6.5. Distribution of honey bee acute effects (pLD50 µmol/bee) for 22 different classes of MoA i.e. target site for pesticide active substances (n = 113).

3.2.3 MoA as tool for grouping chemicals into assessment groups for ecological risk assessment of multiple chemicals

Notwithstanding that several definitions of MoA have been proposed, a range of applications of MoA schemes have also been reported for chemical RA. In the human health area, scientific advisory bodies such as the US-EPA, the WHO, the OECD and EFSA have proposed the application of MoA when defining AGs in component-based approaches for the identification and characterisation of combined toxicity of pesticide active substances as for example potential neurotoxicity in the thyroid to set cumulative AGs (Meek et al., 2011; EFSA PPR, 2013; OECD, 2019; More et al. 2019). Similarly, MoA can be applied as a grouping tool for ecological and human-health RA to group chemicals when applying mathematical models such as concentration addition or independent action for predicting mixtures toxicity (Kienzler et al. 2016; More et al. 2019; Carnesecchi et al. 2019a). Furthermore, MoA has been used to investigate potential pesticide resistance in target organisms (i.e. field

population), thus representing a key-aspect of pest management worldwide (Sparks and Nauen, 2015; Hermann and Stenzel, 2019).

In the present manuscript, we provided the first MoA harmonised classification for pesticide active substances used in PPPs. However, a broad question remains: can one use MoA information to group chemicals in the broad context of ecological RA of multiple chemicals? In order to answer this question, we should clarify that two fundamental aspects intrinsically characterise a given pesticide active substance: its toxicity and MoA at target site (i.e. site of action). As in the case of insecticides, their toxicity and specificity are a consequence of the MoA at the cellular or physiological level in the organism (Simon-Delso et al. 2015). Similarly, whilst toxicity is triggered by the internal dose causing the adverse effect (e.g. death of the organism), the specificity depends on the key events leading to the adverse outcome as biochemical or physiological mechanisms targeted by the insecticide in the specie(s) of interest (Sanchez-Bayo, 2012; Meek et al. 2011). Examples of how pesticide specificity vary among taxa (as in the case of selective insecticides), or are conserved across taxa (e.g. broad-spectrum insecticides) are available in the literature (Sanchez-Bayo, 2012). The author collected publicly available LD_{50s} of insecticides in honey bees while reporting avermectins as the most toxic insecticide class (LD_{50} 0.04 µg/bee) > neonicotinoids (typical LD_{50} 0.03-3.6 µg/bee) > pyrethroids (typical LD_{50} 0.07-1.3 µg/bee). Although our results are in line with Sanchez-Bayo (2012), we further demonstrated that high variability of pesticide toxicity (LD_{50}) among the same MoA group is often reported (figure 5). Therefore, in order to reply to the question above, we suggest that further develop and test the proposed approach for grouping using the MoA harmonised classification in bees, while an additional analysis of the variability of single pesticide toxicity within the same MoA group (e.g. Sodium channel modulator) to integrate the potency aspect when prioritising chemicals of concern in component-based approaches. Nonetheless, besides the intrinsic characteristics of each pesticide active substance (i.e. toxicity and MoA), it is noteworthy that TK-TD variability and inter-colonies variability also play a key role when assessing pesticide (hazard) toxicity (Medrzycki et al. 2013; Wang et al. 2020; Heard et al. 2017; Chmiel et al. 2020).

4. Conclusions

This manuscript has explored the application of EFSA's OpenFoodTox, the US-EPA ECOTOX database and PPDB to develop the first integrative honey bee QSAR models i) to predict acute contact toxicity (LD50) and ii) to profile the MoA of pesticides active substances. Here, seven different QSAR models have been developed, tested and validated for their performance. Two models which can be applied as integrative tools according to a specific hierarchical workflow provided the best predictions:

- A k-NN classification-based model (LD_{50} ≥100 µg/bee (non-toxic) vs. LD_{50} <100 µg/bee (toxic)) which has been validated using statistical parameters i.e. sensitivity (=0.93), specificity (=0.85), accuracy (=0.90), and Matthews correlation coefficient (MCC = 0.78);
- A k-NN Regression-based model (LD_{50} <100 µg/bee; only continuous data) validated with statistical parameters, which were demonstrated to be reliable and robust (R^2 = 0.72; MAE = 0.52; in validation test).

These models are currently being implemented within the VEGA-HUB platform (https://www.vegahub.eu/) and all supplementary materials will be available open source on EFSA's Knowledge Junction platform (doi: 10.5281/zenodo.3755675). The authors acknowledge that k-NN based models present advantages for their implementation, such as qualitative/quantitative predictions based on the most similar compounds i.e. "neighbour" (Mansouri et al. 2018; ECHA, 2016). Similarly, k-NN models also offer further advantages for the implementation of MoA profiling in VEGA models such as visualisation of the neighbors used to predict the target compound as well as reliability of the prediction while assessing their known MoA.

Similarly, the current study also proposes the first harmonised MoA classification scheme for 113 pesticides active substances (used as insecticides, acaricides, herbicides, fungicides and plant growth regulator), including potency of the chemicals as acute contact toxicity as LD_{50s} values from three different sources (EFSA's OpenFoodTox, US-EPA ECOTOX and PPDB). Hence, this exercise allowed further defining toxicological MoAs and the target site of PPPs active substances in honey bees, thus enabling regulators and scientists to refine chemical grouping and toxicity extrapolations for component-based RA of multiple chemicals (More et al. 2019; Carnesecchi et al. 2020). In addition, this approach can be of value for research and development to design new active substances for which potency in non-target species such as honey bees is known and controlled, by relying more on the "*a priori*" knowledge of the pesticide chemical structure and its potential target site (Commission Regulation (EU) 283/2013). In a broader context, this manuscript also highlights how New Approach Methodologies (NAMs) such as *in silico* tools can shift RA to a more mechanistically based understanding of the MoA and mechanism of action of chemicals, thus reducing traditional *in vivo* experiments and providing alternative testing methods.

Finally, many data gaps remain and still limit the development and broader applications of such QSAR models in honey bees:

- Acute contact toxicity data (e.g. mortality) are the major available datasets for a significant number of compounds and their mixtures but acute/chronic toxicity oral data and toxicity data for sub-lethal effects of pesticides and contaminants in honey bees and wild bees are still lacking (Carnesecchi et al. 2019a,b, 2020). Hence, further work is needed to generate acute and chronic toxicity oral data and sub-lethal toxicity data as the basis to develop QSAR

models to predict these effects and integrate such quantitative metrics in RA of single and multiple chemicals (Carnesecchi et al. 2019a; Toma et al. in preparation).

- Statistical variability for single toxicity tests (e.g. LD_{50}) as well as sample size are not usually reported in toxicological studies on bees, thus increasing uncertainty for hazard and RA of chemicals towards non-target organisms (Denton et al. 2003). Such data would be of value when predicting the toxicity of unknown compounds, thereby enhancing their reliability.
- Experimental toxicokinetic data (e.g. absorption, distribution, metabolism and excretion including elimination rate, half-life and bioaccumulation) are also lacking and are an entire part of the MoA of chemicals. This type of information would further support the development of ad-hoc QSAR models to further characterise the impact of fast elimination or persistence of chemicals on the toxicity of single substances as well as on combined toxicity of multiple chemicals in bees (including interactions). Availability of such datasets will allow the further development of QSAR models combined with biometric and life cycle information to generate the next generation of Dynamic Energy Budget (DEB) models for honey bees and wild bees.

Finally, since the understanding of single AOPs and their networks is unfolding and, new tools are being developed to monitor species of ecological relevance at the landscape level; integration of mechanistic understanding at different levels of biological organisation is foreseen as a mid-term perspective. Ultimately, it will allow risk assessors to tackle toxicity of single chemicals, combined toxicity for multiple chemicals and the resulting risk for honey bees and wild bees at the individual, colony, population and landscape level (Topping et al. 2020; Carnesecchi et al. 2019a, 2020; Baas et al. 2018; LaLone et al. 2017; Spurgeon et al. 2017; Hesketh et al. 2016;).

Acknowledgments

The views expressed in this manuscript do not reflect the views of the European Food Safety Authority and are the authors only. This work was supported by the European Food Safety Authority (EFSA) [contract number: OC/EFSA/SCER/2018/01 and NP/EFSA/AFSCO/2016/02 (Edoardo Carnesecchi)]; and LIFE-VERMEER (LIFE16 ENV/IT/000167).

References

Ashton, M., Barnard, J., Casset, F., Charlton, M., Downs, G., Gorse, D., Holliday, J., Lahana, R., Willett, P., 2002. Identification of diverse database subsets using property-based and fragment-based molecular descriptions. Quant. Struct. Relationships. https://doi.org/10.1002/qsar.200290002

Baas, J., Augustine, S., Marques, G.M., Dorne, J. Lou, 2018. Dynamic energy budget models in ecological risk assessment: From principles to applications. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2018.02.058

Bargańska, Ż., Ślebioda, M., Namieśnik, J., 2016. Honey bees and their products: Bioindicators of environmental contamination. Crit. Rev. Environ. Sci. Technol. 46, 235–248. https://doi.org/10.1080/10643389.2015.1078220

Beffa, R., Menne, H., Köcher, H., 2019. Herbicide Resistance Action Committee (HRAC): Herbicide Classification, Resistance Evolution, Survey, and Resistance Mitigation Activities, in: Modern Crop Protection Compounds. https://doi.org/10.1002/9783527699261.ch1

Benfenati, E., Como, F., Manzo, M., Gadaleta, D., Toropov, A., Toropova, A., 2017. Developing innovative in silico models with EFSA's OpenFoodTox database. EFSA Support. Publ. https://doi.org/10.2903/sp.efsa.2017.en-1206

Benfenati, E., Carnesecchi, E., Roncaglioni, A., Baldin, R., Ceriani, L., Ciacci, A., Kovarich, S., Sartori, L., Mostrag, A., Magdziarz, T., Yang, C., 2020. Maintenance, update and further development of EFSA's Chemical Hazards: OpenFoodTox 2.0. EFSA supporting publication 2020:EN-1822. 36 pp. doi:10.2903/sp.efsa.2020.EN-1822.

Boobis, A.R., Cohen, S.M., Dellarco, V., McGregor, D., Meek, M.E., Vickers, C., Willcocks, D., Farland, W., 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit. Rev. Toxicol. https://doi.org/10.1080/10408440600977677

Bopp, S.K., Kienzler, A., Richarz, A.N., van der Linden, S.C., Paini, A., Parissis, N., Worth, A.P., 2019. Regulatory assessment and risk management of chemical mixtures: challenges and ways forward. Crit. Rev. Toxicol. https://doi.org/10.1080/10408444.2019.1579169

Breeze, T.D., Bailey, A.P., Balcombe, K.G., Potts, S.G., 2011. Pollination services in the UK: How important are honeybees? Agric. Ecosyst. Environ. https://doi.org/10.1016/j.agee.2011.03.020

Breiman L., 2001. Machine Learning, 45(1), 5–32. Stat. Dep. Univ. California, Berkeley, CA 94720. https://doi.org/10.1023/A:1010933404324

Carnesecchi, E., Svendsen, C., Lasagni, S., Grech, A., Quignot, N., Amzal, B., Toma, C., Tosi, S., Rortais, A., Cortinas-Abrahantes, J., Capri, E., Kramer, N., Benfenati, E., Spurgeon, D., Guillot, G., Dorne, J.L.C.M., 2019a. Investigating combined toxicity of binary mixtures in bees: Meta-analysis of laboratory tests, modelling, mechanistic basis and implications for risk assessment. Environ. Int. 133. https://doi.org/10.1016/j.envint.2019.105256.

Carnesecchi, E., Svendsen, C., Lasagni, S., Grech, A., Quignot, N., Amzal, B., Toma, C., Benfenati, E., Guillot, G., Dorne, J.L.C.M., 2019b. Individual datasets investigating combined toxicity of binary mixtures in bees from laboratory tests [Data set]. Zenodo. Available at http://doi.org/10.5281/zenodo.3383713

Carnesecchi, E., Toropov, A.A., Toropova, A.P., Kramer, N., Svendsen, C., Dorne, J. Lou, Benfenati, E., 2020. Predicting acute contact toxicity of organic binary mixtures in honey bees (A. mellifera) through innovative QSAR models. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2019.135302

Carriger, J.F., Martin, T.M., Barron, M.G., 2016. A Bayesian network model for predicting aquatic toxicity mode of action using two dimensional theoretical molecular descriptors. Aquat. Toxicol. https://doi.org/10.1016/j.aquatox.2016.09.006

Casida, J.E., 2009. Pest toxicology: The primary mechanisms of pesticide action. Chem. Res. Toxicol. https://doi.org/10.1021/tx8004949

Chmiel, J.A., Daisley, B.A., Pitek, A.P., Thompson, G.J., Reid, G., 2020. Understanding the Effects of Sublethal Pesticide Exposure on Honey Bees: A Role for Probiotics as Mediators of Environmental Stress. Front. Ecol. Evol. https://doi.org/10.3389/fevo.2020.00022

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. Educ. Psychol. Meas. https://doi.org/10.1177/001316446002000104

Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market.

Como, F., Carnesecchi, E., Volani, S., Dorne, J.L., Richardson, J., Bassan, A., Pavan, M., Benfenati, E., 2017. Predicting acute contact toxicity of pesticides in honeybees (Apis mellifera) through a k-nearest neighbor model. Chemosphere. https://doi.org/10.1016/j.chemosphere.2016.09.092.

Connors, K.A., Beasley, A., Barron, M.G., Belanger, S.E., Bonnell, M., Brill, J.L., de Zwart, D., Kienzler, A., Krailler, J., Otter, R., Phillips, J.L., Embry, M.R., 2019. Creation of a Curated Aquatic Toxicology Database: EnviroTox. Environ. Toxicol. Chem. https://doi.org/10.1002/etc.4382

Dao, P., Wang, K., Collins, C., Ester, M., Lapuk, A., Sahinalp, S.C., 2011. Optimally discriminative subnetwork markers predict response to chemotherapy. Bioinformatics. https://doi.org/10.1093/bioinformatics/btr245.

David, A., Botías, C., Abdul-Sada, A., Nicholls, E., Rotheray, E.L., Hill, E.M., Goulson, D., 2016. Widespread contamination of wildflower and bee-collected pollen with complex mixtures of neonicotinoids and fungicides commonly applied to crops. Environ. Int. 88, 169–178. https://doi.org/10.1016/j.envint.2015.12.011.

Denton, D.L., Fox, J.F., Fulk, F.A., 2003. Enhancing toxicity test performance by using a statistical criterion. Environ. Toxicol. Chem. https://doi.org/10.1897/02-384

De Castro, R.D., de Souza, T.M.P.A., Bezerra, L.M.D., Ferreira, G.L.S., de Brito Costa, E.M.M., Cavalcanti, A.L., 2015. Antifungal activity and mode of action of thymol and its synergism with nystatin against Candida species involved with infections in the oral cavity: an in vitro study. BMC Complement. Altern. Med. 15, 417. https://doi.org/10.1186/s12906-015-0947-2.

Devillers, J., Pham-Delègue, M.H., Decourtye, A., Budzinski, H., Cluzeau, S., Maurin, G., 2002. Structure-toxicity modeling of pesticides to honey bees. SAR QSAR Environ. Res. https://doi.org/10.1080/1062936021000043391

Douglas, M.R., Sponsler, D.B., Lonsdorf, E. V., Grozinger, C.M., 2020. County-level analysis reveals a rapidly shifting landscape of insecticide hazard to honey bees (Apis mellifera) on US farmland. Sci. Rep. https://doi.org/10.1038/s41598-019-57225-w

ECHA (European Chemicals Agency), 2016. Practical Guide - How to use and report (Q)SARs. ECHA-16-B-09-EN. Available at https://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf

EFSA (European Food Safety Authority), 2018. Evaluation of the data on clothianidin, imidacloprid and thiamethoxam for the updated risk assessment to bees for seed treatments and granules in the EU. EFSA supporting publication 2018:EN-1378. 31pp.doi:10.2903/sp.efsa.2018.EN-1378

EFSA (European Food Safety Authority), 2013. Guidance on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA J. 11. https://doi.org/10.2903/j.efsa.2013.3295.

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2012. Scientific Opinion on the science behind the development of a risk assessment of Plant Protection Products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA J. 10, 2668. https://doi.org/10.2903/j.efsa.2012.2668.

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Scientific Opinion on the identification of pesticides to be included in cumulative assessment groups on the basis of their toxicological profile (2014 update). EFSA Journal 2013;11(7):3293, 131 pp. https://doi.org/10.2903/j.efsa.2013.3293

EFSA (European Food Safety Authority), Crivellente, F., Hart, A., Hernandez-Jerez, A.F., Hougaard Bennekou, S., Pedersen, R., Terron, A., Wolterink, G., Mohimont, L., 2019. Establishment of cumulative assessment groups of pesticides for their effects on the thyroid. EFSA J. https://doi.org/10.2903/j.efsa.2019.5801

EFSA (European Food Safety Authority), 2006. Conclusion regarding the peer review of the pesticide risk assessment of the active substance fipronil, 2006. EFSA J. https://doi.org/10.2903/j.efsa.2006.65r

EFSA (European Food Safety Authority), 2010. Conclusion on the peer review of the pesticide risk assessment of the active substance oryzalin. EFSA Journal 2010; 8(9):1707. [59 pp.]. doi:10.2903/j.efsa.2010.1707. Available online: www.efsa.europa.eu

Enoch, S.J., Hewitt, M., Cronin, M.T.D., Azam, S., Madden, J.C., 2008. Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree. Chemosphere. https://doi.org/10.1016/j.chemosphere.2008.06.052

Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., Gramatica, P., 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ. Health Perspect. https://doi.org/10.1289/ehp.5758

European Pesticide Database, 2019. Available at https://ec.europa.eu/food/plant/pesticides/eu-pesticides-atabase/public/?event=activesubstance.selection&language=EN (accessed 23 March 2020)

FAO and ITPS, 2017. Global assessment of the impact of plant protection products on soil functions and soil ecosystems, Rome, FAO. 40 pp.

Frische, T., Egerer, S., Matezki, S., Pickl, C., Wogram, J., 2018. 5-Point programme for sustainable plant protection. Environ. Sci. Eur. https://doi.org/10.1186/s12302-018-0136-2

Gadaleta, D., Lombardo, A., Toma, C., Benfenati, E., 2018. A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications. J. Cheminform. https://doi.org/10.1186/s13321-018-0315-6

Genuer, R., Poggi, J.M., Tuleau-Malot, C., 2015. VSURF: An R package for variable selection using random forests. R J. https://doi.org/10.32614/rj-2015-018

Golbraikh, A., Tropsha, A., 2002. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J. Comput. Aided. Mol. Des. https://doi.org/10.1023/A:1020869118689

Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y. De, Lee, K.H., Tropsha, A., 2003. Rational selection of training and test sets for the development of validated QSAR models. J. Comput. Aided. Mol. Des. https://doi.org/10.1023/A:1025386326946

Ghosh, S., Ojha, P.K., Carnesecchi, E., Lombardo, A., Roy, K., Benfenati, E., 2020. Exploring QSAR modeling of toxicity of chemicals on earthworm. Ecotoxicol. Environ. Saf. https://doi.org/10.1016/j.ecoenv.2019.110067

Gramatica, P., Sangion, A., 2016. A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology. J. Chem. Inf. Model. https://doi.org/10.1021/acs.jcim.6b00088

Hamadache, M., Benkortbi, O., Hanini, S., Amrane, A., 2018. QSAR modeling in ecotoxicological risk assessment: application to the prediction of acute contact toxicity of pesticides on bees (Apis mellifera L.). Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-017-0498-9.

Heard, M.S., Baas, J., Dorne, J. Lou, Lahive, E., Robinson, A.G., Rortais, A., Spurgeon, D.J., Svendsen, C., Hesketh, H., 2017. Comparative toxicity of pesticides and environmental contaminants in bees: Are honey bees a useful proxy for wild bee species? Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2016.10.180

Hermann, D., Stenzel, K., 2019. FRAC Mode-of-action Classification and Resistance Risk of Fungicides, in: Modern Crop Protection Compounds. https://doi.org/10.1002/9783527699261.ch14

Hesketh, H., Lahive, E., Horton, A.A., Robinson, A.G., Svendsen, C., Rortais, A., Dorne, J.L., Baas, J., Spurgeon, D.J., Heard, M.S., 2016. Extending standard testing period in honeybees to predict lifespan impacts of pesticides and heavy metals using dynamic energy budget modelling. Sci. Rep. https://doi.org/10.1038/srep37655

Johnson, R.M., Mao, W., Pollock, H.S., Niu, G., Schuler, M.A., Berenbaum, M.R., 2012. Ecologically appropriate xenobiotics induce cytochrome P450s in Apis mellifera. PLoS One 7, 1–9. https://doi.org/10.1371/journal.pone.0031051

Johnson, R.M., Dahlgren, L., Siegfried, B.D., Ellis, M.D., 2013. Acaricide, fungicide and drug interactions in honey bees (Apis mellifera). PLoS One 8, e54092. https://doi.org/10.1371/journal.pone.0054092.

Kepplinger, D., Filzmoser, P., Varmuza, K., 2017. Variable selection with genetic algorithms using repeated cross-validation of PLS regression models as fitness measure. arXiv preprint arXiv:1711.06695.

Kienzler, A., Barron, M.G., Belanger, S.E., Beasley, A., Embry, M.R., 2017. Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.7b02337

Kienzler, A., Connors, K.A., Bonnell, M., Barron, M.G., Beasley, A., Inglis, C.G., Norberg-King, T.J., Martin, T., Sanderson, H., Vallotton, N., Wilson, P., Embry, M.R., 2019. Mode of Action Classifications in the EnviroTox Database: Development and Implementation of a Consensus MOA Classification. Environ. Toxicol. Chem. https://doi.org/10.1002/etc.4531

Kienzler, A., Bopp, S.K., van der Linden, S., Berggren, E., Worth, A., 2016. Regulatory assessment of chemical mixtures: Requirements, current approaches and future perspectives. Regul. Toxicol. Pharmacol. 80, 321–334. https://doi.org/10.1016/j.yrtph.2016.05.020

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. https://doi.org/10.18637/jss.v028.i05

LaLone, C.A., Villeneuve, D.L., Wu-Smart, J., Milsk, R.Y., Sappington, K., Garber, K. V., Housenger, J., Ankley, G.T., 2017. Weight of evidence evaluation of a network of adverse outcome pathways linking activation of the nicotinic acetylcholine receptor in honey bees to colony death. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2017.01.113

Leroux, P., Bach, J., Debieu, D., Fillinger, S., Fritz, R., Walker, A.S., 2008. Mode of action of sterol biosynthesis inhibitors and resistance phenomena in fungi. Mod. Fungic. In: Antifung. Compd. V 15th Int. Reinhardsbrunn Symp. Friedrichroda, Ger. May 6–10, pp. 85–92.

Manganaro, A., Pizzo, F., Lombardo, A., Pogliaghi, A., Benfenati, E., 2016. Predicting persistence in the sediment compartment with a new automatic software based on the k-Nearest Neighbor (k-NN) algorithm. Chemosphere. https://doi.org/10.1016/j.chemosphere.2015.10.054

Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J., 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. J. Cheminform. https://doi.org/10.1186/s13321-018-0263-1

Martin, T.M., Grulke, C.M., Young, D.M., Russom, C.L., Wang, N.Y., Jackson, C.R., Barron, M.G., 2013. Prediction of aquatic toxicity mode of action using linear discriminant and random forest models. J. Chem. Inf. Model. https://doi.org/10.1021/ci400267h

Martin, T.M., Young, D.M., Lilavois, C.R., Barron, M.G., 2015. Comparison of global and mode of action-based models for aquatic toxicity. SAR QSAR Environ. Res. https://doi.org/10.1080/1062936X.2015.1018939

Medrzycki, P., Giffard, H., Aupinel, P., Belzunces, L.P., Chauzat, M.P., Claßen, C., Colin, M.E., Dupont, T., Girolami, V., Johnson, R., Conte, Y. Le, Lückmann, J., Marzaro, M., Pistorius, J., Porrini, C., Schur, A., Sgolastra, F., Delso, N.S., Van Der Steen, J.J.M., Wallner, K., Alaux, C., Biron, D.G.,

Blot, N., Bogo, G., Brunet, J.L., Delbac, F., Diogon, M., El Alaoui, H., Provost, B., Tosi, S., Vidau, C., 2013. Standard methods for toxicology research in Apis mellifera. J. Apic. Res. https://doi.org/10.3896/IBRA.1.52.4.14

More, S.J., Bampidis, V., Benford, D., Bennekou, S.H., Bragard, C., Halldorsson, T.I., Hernández-Jerez, A.F., Koutsoumanis, K., Naegeli, H., Schlatter, J.R., Silano, V., Nielsen, S.S., Schrenk, D., Turck, D., Younes, M., Benfenati, E., Castle, L., Cedergreen, N., Hardy, A., Laskowski, R., Leblanc, J.C., Kortenkamp, A., Ragas, A., Posthuma, L., Svendsen, C., Solecki, R., Testai, E., Dujardin, B., Kass, G.E., Manini, P., Jeddi, M.Z., Dorne, J.C., Hogstrand, C., 2019. Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals. EFSA J. 17. https://doi.org/10.2903/j.efsa.2019.5634.

Negri, I., Mavris, C., Di Prisco, G., Caprio, E., Pellecchia, M., 2015. Honey bees (Apis mellifera, L.) as active samplers of airborne particulate matter. PLoS One. https://doi.org/10.1371/journal.pone.0132491

OECD (Organisation for Economic Co-operation and Development), 2017. Revised guidance document on developing and assessing Adverse Outcome Pathways. Ser. Test. Assessment, OECD Publ. Off. Paris-France. Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 1998. Test No. 214: Honeybees, Acute Contact Toxicity Test, OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris. Available at https://doi.org/10.1787/9789264070189-en.

OECD (Organisation for Economic Co-operation and Development), 2007. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, ENV/JM/MONO(2007)2. OECD Publishing, Paris. Available at https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm

OECD (Organisation for Economic Co-operation and Development), 2018. Considerations for Assessing the Risks of Combined Exposure to Multiple Chemicals, Series on Testing and Assessment No. 296, Environment, Health and Safety Division, Environment Directorate. Available at https://www.oecd.org/chemicalsafety/risk-assessment/considerations-for-assessing-the-risks-of-combined-exposure-to-multiple-chemicals.pdf

Prado, A., Pioz, M., Vidau, C., Requier, F., Jury, M., Crauser, D., Brunet, J.-L., Le Conte, Y., Alaux, C., 2019. Exposure to pollen-bound pesticide mixtures induces longer-lived but less efficient honey bees. Sci. Total Environ. 650, 1250–1260. https://doi.org/10.1016/j.scitotenv.2018.09.102.

Quinlan, J.R., 1986. Induction of Decision Trees. Mach. Learn. https://doi.org/10.1023/A:1022643204877

Quinlan, J.R., 1987. Simplifying decision trees. Int. J. Man. Mach. Stud. https://doi.org/10.1016/S0020-7373(87)80053-6

Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 Concerning the Placing of Plant Protection Products on the Market and Repealing Council Directives 79/117/EEC and 91/414/EEC. OJ L 309/1, 24.11.2009(2009), pp. 1-50

Rose, T., Kremen, C., Thrupp, A., Gemmill-Herren, B., Graub, B., Azzu, N., Antunes, V., Bruteig, I.E., Buchori, D., Donaldson, J., Dhyani, P.P., Garibaldi, L., Getz, A., Goss, M., Iqbal, J., Kasina, M., Kinuthia, W., Kwapong, P., Manetto, S., Martins, D., Nyamasio, G., Nyamongo, D.O., Odhiambo, C., América Suarez De Oliveira, D., Owusu, E.H., Pandey, B., Poole, C., Roubik, D.W., Roy, P., Waghchoure, E., Lusike, W., 2015. Policy analysis paper: mainstreaming of biodiversity and ecosystem services with a focus on pollination. United Nations Food and Agriculture Organization (FAO) with the contribution of participants at the "Policies for Pollination Management" Worksh. Available at: http://www.fao.org/3/a-i4242e.pdf.

Rortais, A., Arnold, G., Dorne, J.-L., More, S.J., Sperandio, G., Streissl, F., Szentes, C., Verdonck, F., 2017. Risk assessment of pesticides and other stressors in bees: principles, data gaps and perspectives from the European Food Safety Authority. Sci. Total Environ. 587–588, 524–537. https://doi.org/10.1016/j.scitotenv.2016.09.127.

Rovida, C., Hartung, T., 2009. Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals - A report by the transatlantic think tank for toxicology (t4). ALTEX. https://doi.org/10.14573/altex.2009.3.187

Roy, J., Kumar Ojha, P., Carnesecchi, E., Lombardo, A., Roy, K., Benfenati, E., 2020. First report on a classification-based QSAR model for chemical toxicity to earthworm. J. Hazard. Mater. https://doi.org/10.1016/j.jhazmat.2019.121660

Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A., 1997. Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (pimephales promelas). Environ. Toxicol. Chem. https://doi.org/10.1897/1551-5028(1997)016<0948:pmotaf>2.3.co;2

Samworth, R.J., 2012. Optimal weighted nearest neighbour classifiers. Ann. Stat. https://doi.org/10.1214/12-AOS1049

Sanchez-Bayo, F., Goka, K., 2016. Impacts of Pesticides on Honey Bees, in: Beekeeping and Bee Conservation - Advances in Research. https://doi.org/10.5772/62487

Sanchez-Bayo, F., 2012. Insecticides Mode of Action in Relation to Their Toxicity to Non-Target Organisms. J. Environ. Anal. Toxicol. https://doi.org/10.4172/2161-0525.s4-002

Simon-Delso, N., Amaral-Rogers, V., Belzunces, L.P., Bonmatin, J.M., Chagnon, M., Downs, C., Furlan, L., Gibbons, D.W., Giorio, C., Girolami, V., Goulson, D., Kreutzweiser, D.P., Krupke, C.H., Liess, M., Long, E., Mcfield, M., Mineau, P., Mitchell, E.A., Morrissey, C.A., Noome, D.A., Pisa, L., Settele, J., Stark, J.D., Tapparo, A., Van Dyck, H., Van Praagh, J., Van Der Sluijs, J.P., Whitehorn, P.R., Wiemers, M., 2015. Systemic insecticides (Neonicotinoids and fipronil): Trends, uses, mode of action and metabolites. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-014-3470-y

Schulp, C.J.E., Lautenbach, S., Verburg, P.H., 2014. Quantifying and mapping ecosystem services: Demand and supply of pollination in the European Union. Ecol. Indic. https://doi.org/10.1016/j.ecolind.2013.07.014

Segner H, 2011. Moving beyond a descriptive aquatic toxicology: The value of biological process and trait information. Aquatic Toxicology, 105, 50–55. Available online: http://ec.europa.eu/health/scientific_committee s/docs/challenges_mandate_en.pdf

Singh, K.P., Gupta, S., Basant, N., Mohan, D., 2014. QSTR modeling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes. Chem. Res. Toxicol. https://doi.org/10.1021/tx500100m

Sparks, T.C., Nauen, R., 2015. IRAC: Mode of action classification and insecticide resistance management. Pestic. Biochem. Physiol. https://doi.org/10.1016/j.pestbp.2014.11.014

Spurgeon, D., Hesketh, H., Lahive, E., Svendsen, C., Baas, J., Robinson, A., Horton, A., Heard, M., 2016. Chronic oral lethal and sub-lethal toxicities of different binary mixtures of pesticides and contaminants in bees (Apis mellifera, Osmia bicornis and Bombus terrestris). EFSA Support. Publ. 13. https://doi.org/10.2903/sp.efsa.2016.en-1076.

Tanimoto T.T., 1958. An elementary mathematical theory of classification and prediction in IBM Internal Report.

Toma, C., Carnesecchi E., Benfenati E., Dorne J.L., in preparation. First report on regression based- and classification-based QSAR models for predicting oral acute toxicity of pesticides active substances towards honey bee (Apis mellifera). Toxicology Letters journal.

Topping, C.J., Aldrich, A., Berny, P., 2020. Overhaul environmental risk assessment for pesticides. Science (80-.). https://doi.org/10.1126/science.aay1144

Toropov, A.A., Benfenati, E., 2007. SMILES as an alternative to the graph in QSAR modelling of bee toxicity. Comput. Biol. Chem. https://doi.org/10.1016/j.compbiolchem.2007.01.003

Tosi, S., Nieh, J.C., 2019. Lethal and sublethal synergistic effects of a new systemic pesticide, flupyradifurone (Sivanto®), on honeybees. Proc. R. Soc. B Biol. Sci. 286, 20190433. https://doi.org/10.1098/rspb.2019.0433

Tosi, S., Costa, C., Vesco, U., Quaglia, G., Guido, G., 2018. A 3-year survey of Italian honey bee collected pollen reveals widespread contamination by agricultural pesticides. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2017.09.226

USEPA, PMRA, CALDPR, 2014. Guidance for Assessing Pesticide Risks to Bees. U.S. Environmental Protection Agency (USEPA), Health Canada Pest Management Regulatory Agency (PMRA) and California Departament of Pesticide Regulation, Office of Pesticide Programs USEPA, Health Canada PMRA, CDPR. Available at: https://www.epa.gov/pollinator-protection/pollinator-risk-assessment-guidance

Venko, K., Drgan, V., Novič, M., 2018. Classification models for identifying substances exhibiting acute contact toxicity in honeybees (Apis mellifera)$. SAR QSAR Environ. Res. https://doi.org/10.1080/1062936X.2018.1513953

Verhaar, H.J.M., Van Leeuwen, C.J., Hermens, J.L.M., 1992. Classifying environmental-pollutants. 1: Structure-activity-relationships for prediction of aquatic toxicity. Chemosphere. https://doi.org/10.1016/0045-6535(92)90280-5

Wang, M., Braasch, T., Dietrich, C., 2020. Reduction of variability for the assessment of side effects of toxicants on honeybees and understanding drivers for colony development. PLoS One. https://doi.org/10.1371/journal.pone.0229295

WHO (World Health Organization), 2009. Environmental Health Criteria 240: Principles and Methods for the Risk Assessment of Chemicals in Food. WHO, Geneva, (Definitions page A-25). Available at http://www.who.int/foodsafety/publications/chemical-food/en

WHO (World Health Organization), 2020. WHO recommended classification of pesticides by hazard and guidelines to classification, 2019 edition. Geneva. Licence: CC BY-NC-SA 3.0 IGO. Available at https://apps.who.int/iris/bitstream/handle/10665/332193/9789240005662-eng.pdf?ua=1

Wing, K.D., Andaloro, J.T., McCann, S.F., Salgado, V.L., 2005. Indoxacarb and the Sodium Channel Blocker Insecticides: Chemistry, Physiology, and Biology in Insects, in: Comprehensive Molecular Insect Science. https://doi.org/10.1016/B0-44-451924-6/00070-3

Worth, A. P., Bassan, A., Gallegos, A., Netzeva, T. I., Patlewicz, G., Pavan, M., Vračko, M., 2005. The characterisation of (quantitative) structure-activity relationships: preliminary guidance. Institute for Health and Consumer Protection, Toxicology and Chemical Substances Unit, European Chemical Bureau. Available at https://core.ac.uk/download/pdf/93511332.pdf

Wright, M.N., Ziegler, A., 2017. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. https://doi.org/10.18637/jss.v077.i01

Chapter 7 **Summary, general discussion and conclusions**

Summary

Multiple stressors, including diseases and chemicals, affect the health of ecosystems, animals and humans (EEA, 2018; Topping et al., 2020). The risks posed by chemical products such as biocides, medicines, veterinary products and plant protection products (PPPs) are assessed via Human Health Risk Assessment (HHRA), Animal Health Risk Assessment (AHRA), and Ecological Risk Assessment (ERA).

Regulatory authorities such as EFSA perform ERA and HHRA for single chemical stressors while supporting scientific, industrial and societal progress (EFSA Scientific Committee, 2016; UNEP, 2019). In order to assess the risk of chemicals, toxicity tests on laboratory animals are traditionally used to determine safe levels of exposure to single chemicals (Rovida and Hartung, 2009). Since 2007, the so-called Tox21 strategy in the United States has provided a basis to shift toxicological assessments away from traditional animal test to *in vitro* and *in silico* studies providing a mechanistic basis for chemical toxicity often considered as new approach methodologies (NAMs) (Thomas et al., 2018). In the EU, the Registration, Evaluation, Authorization and restriction of Chemicals (REACH) Regulation (EC No, 1907/2006) promotes the use of NAMs where suitable, for the assessment of industrial chemicals toxicity (ECHA, 2016, 2020). This thesis demonstrated how NAMs, particularly quantitative structure-activity relationship (QSAR) models, can provide a reliable means in next generation RA of chemicals, particularly supporting traditional *in vivo* experiments with non-animal alternatives for HHRA and ERA of regulated chemicals, emerging contaminants and their mixtures.

In **chapter 1** of this thesis, the basis of traditional HHRA and ERA has been introduced and the ethical, economical and scientific issues associated with traditional RA approaches have been highlighted. The RA of chemicals such as PPP active substances in the EU is mainly performed for the pre-market authorisation (*a priori*) of single compounds (EC, 2009, 2013). In particular, the hazard assessment of PPP active substances is based on animal testing to derive dose-response relationships for individual chemicals. Uncertainty factors are applied to no observed adverse effect levels (NOAEL) from (sub)-chronic studies in animals to account for the uncertainty and variability in species differences and human variability in toxicokinetic (TK) and toxicodynamic (TD) processes (EFSA, 2014). This approach is costly from a time, budget and laboratory animal perspective (EC, 2020). NAMs such as QSARs potentially help to overcome these shortcomings (Rajes and Bajic, 2016). QSAR models allow the prediction of physicochemical properties (e.g. lipophilicity, flammability), biological activities (e.g. toxicity, clearance) and environmental fate of compounds from the knowledge of their chemical structure (OECD, 2007). Key applications of QSARs include prioritisation and screening of chemicals of concern and data gap filling for substances with non available *in vivo* toxicity data (Pradeep et al., 2016). Overall, QSARs can potentially refine, reduce and replace animal testing in current RA approaches (referred to as the 3Rs principle), and neatly

fit in the Tox21 paradigm (Thomas et al., 2018; Hamadache et al., 2020; Villaverde et al., 2020). However, QSAR models predicting chemical mixture toxicity and the nature of the combine effects (e.g. synergism, antagonism) are still under development for the majority of non-target organisms (Bopp et al., 2019; Carnesecchi et al., 2020a).

Chapters 2 and **4** describe QSAR models developed for human and environmental hazard assessment of single chemicals, which have not been fully characterised with regards to their toxicological profile. In **chapter 2**, the applicability of freely available QSAR models (i.e. VEGA-HUB platform) is proposed for HHRA of chemicals in food. The chapter presents the capability of such *in silico* tools to predict genotoxicity and carcinogenicity endpoints of PPP active substances and metabolites, mandated by Regulation (EU) 1107/2009 (EC, 2009). It also evaluates QSAR predictions of natural complex compounds present in novel food, such as botanicals, according to Regulation (EU) 2015/2283 (European Parliament and Cuncil, 2015). Overall, an accuracy of 90% and 70% is reported for genotoxicity and carcinogenicity QSARs predictions, respectively. **Chapter 4** provides a case study on the application of QSARs in ERA, showing the development and validation of a classification-based model (k-nearest neighbor - k-NN) for predicting acute contact toxicity of PPP active substances on honey bees. With an accuracy of 70%, the k-NN model allows classifying PPPs into three toxicity groups: high (LD_{50} <100 µg), moderate (1< LD_{50} <100 µg /bee), and low (LD_{50} >100 µg) toxic compounds.

As illustrated in **chapters 2** and **4,** the vast majority of QSAR models available in the literature is designed for predicting properties of single chemicals (e.g. VEGA-HUB, OECD QSAR Toolbox) and are not designed to predict combined effects of multiple chemicals so-called "mixtures" (Khan et al., 2020; Bopp et al., 2019; Carnesecchi et al., 2020a). In the real world, combined exposure to multiple chemicals is more frequent than exposure to single substances (Drakvik et al., 2020; Escher et al., 2020; Crépet et al., 2013; Mitchell et al., 2017; Tosi et al., 2018). The assessment of the effects of "intentional mixtures" in commercial products is often performed (ECHA, 2017), but RA of combined exposure to multiple chemicals in the environment i.e. "unintentional mixtures", is most often overlooked (Kienzler et al., 2014; Bopp et al., 2019). In order to address this issue, **chapter 3** provides the first quantitative review on combined effects of chemical mixtures in non-target organisms such as bees relevant for ERA of PPP active substances. In particular, **chapter 3** investigated and reviewed the methodologies available for mixture risk assessment (MRA), with the view of collecting and generating harmonised data for developing ad hoc *in silico* tools for predicting mixtures toxicity in honey bees. This systematic review demonstrated that bees as beneficial pollinators can be exposed to complex combinations of multiple chemicals which may cause synergistic and sublethal (indirect) effects in a spatial and temporal manner (Prado et al., 2019; Tosi et al., 2018; Carnesecchi et al., 2019a). It is impossible to experimentally test all possible

mixtures in species of ecological relevance, such as bees, birds, fish and amphibians (Carnesecchi et al., 2019a; Bopp et al., 2018). Therefore, innovative and sound scientific methodologies supporting efficient and animal-free methods for MRA are needed (Tickner et al., 2019; OECD, 2019a).

In line with the EFSA Guidance Document "MixTox" on methodologies for RA of combined exposure to multiple chemicals (More et al., 2019), **chapters 5 and 6** of this thesis tackled the current challenges in MRA by developing and applying non-animal test methods to predict the combined effects of multiple chemicals. **Chapter 5** presents the first open source QSAR model written in CORAL software which predicts acute contact toxicity of binary mixtures of PPPs to honeybees (i.e. the median lethal dose of mixtures, $LD_{50\text{-}mix}$). Using a weight of evidence (WoE) approach, the model also provides the user with an estimate of the nature of the combined toxicity (i.e. whether the mixture acts synergistically or non-synergistically). However, while the model in **chapter 5** is demonstratively useful for MRA of PPPs, it does not integrate mechanistic data such as mode of action (MoA) of chemicals (Carnesecchi et al., 2020a,b; Kienzler et al., 2019). In ERA, MoA is defined as a functional change at a cellular level triggered by the substance entering the organism (Kienzler et al. 2017, 2019; Segner, 2011). Consequently, **chapter 6** addressed the challenge of integrating MoA information into QSARs. The chapter presents the first integrative honey bee QSAR models using open source databases to i) predict acute contact toxicity (LD_{50}) and ii) profile the MoA of PPPs active substances. Similarly, a harmonised MoA classification scheme has been developed for classifying PPPs active substances according to their MoA. Such a refined MoA approach allowed the establishment of the relationship between structural information and target sites of PPPs active substances for a range of properties, including toxicity predictions and refinement of the grouping of chemicals using component-based approaches for RA of multiple chemicals (More et al. 2019; Carnesecchi et al., 2020b). The specific software to predict MoA of PPPs was implemented within the VEGA-HUB platform, and thus it can be used to group substances acting through a common MoA while applying the component-based approach and the dose addition default assumption.

The current chapter, **chapter 7,** provides a discussion on the scientific principles to support the application of QSAR models in HHRA and ERA of chemicals. In line with the Tox21 strategy, it illustrates that animal-free methods such as QSARs can play a key role in the RA of emerging chemicals and their mixtures, thus providing a better mechanistic understanding of toxicity and "a priori" knowledge of chemical structures. The methods described herein can be applied to datasets beyond those listed in this thesis and relevant to HHRA, AHRA and ERA of chemical mixtures. In addition, this chapter also presents current data gaps in the development and application of QSARs in MRA. It is foreseen that for chemical RA, multiple NAMs will need to be used to integrate different

lines of evidence (i.e. *ex vivo*, *in vitro*, *in silico* data) in an iterative manner. Therefore, the reporting of results from multiple NAM applications and the overall uncertainty is highlighted, thus advocating the use of integrated approaches to testing and assessment (IATA) and harmonised WoE methods (OECD, 2020a; OECD, 2019b; Hardy et al., 2017).

General discussion

QSARs in the new toxicity testing paradigm

NAMs such as QSARs play a vital role in next generation RA of chemicals. They provide practical solutions to the urgent needs raised by the EU Chemical Strategy for Sustainability, which is a key commitment of the European Green Deal. The EU Chemical Strategy for Sustainability aims to boost innovation for the safe and sustainable use of chemicals. It specifically calls for multidisciplinary research to move away from animal testing (EC, 2020). Similarly, the goal of the US Tox21 programme is to develop, evaluate, and use NAMs for better predicting how substances may affect humans and the environment (Thomas et al., 2018). To contribute to the EU Green Deal and the US Tox21 strategy, the overall aim of this thesis is to promote the use of animal-free methods such QSARs in order to innovate current chemicals safety testing. In particular, this thesis demonstrates how NAMs can improve the quality, efficiency and speed of Ecological Risk Assessment (ERA) and Human Health Risk Assessment (HHRA) of regulated chemicals, emerging contaminants and their mixtures. These concepts have been widely illustrated and discussed through the analysis of five different case studies applied to single chemicals RA (**chapters 2** and **4**) and mixture risk assessment (MRA) (**chapter 3**, **5** and **6**) in line with the recent EFSA Guidance Document "MIXTOX" on harmonised methodologies for risk assessment of combined exposure to multiple chemicals (More et al., 2019).

QSARs as screening tools for predicting genotoxicity and carcinogenicity endpoints

Results from **chapter 2** demonstrate that current *in silico* applications such as regression-based and classification-based QSARs are efficient screening tools for HHRA of chemicals. They allow the prioritisation of PPP active substances, botanicals and their metabolites present in feed and food. In addition, these QSARs allow the identification of PPPs and botanicals which require further assessment for their potential genotoxicity and carcinogenicity properties in humans (ECHA, 2016). In particular, this chapter illustrates the capability of freely available QSARs (implemented on VEGA-HUB *in silico* platform) to predict genotoxicity endpoints such as "bacterial reverse mutation assay" (i.e. Ames test) with an accuracy close to 90% and carcinogenicity with an accuracy of 70%. Finally, **chapter 2** also stresses that more predictive QSAR models are still needed to be developed for experimental/clinical endpoints of mutagenicity other than the Ames test readout. Worldwide, the

Ames test is the most accepted *in vitro* method for assessing gene mutation potential, but the *in vitro* micronucleus test represents the preferred method to detect numerical and structural chromosomal aberrations, which allows for the identification of aneugens and clastogens (OECD,1997; Kirkland et al., 2011). To date, most of the available QSAR models are designed to predict the *in vitro* Ames test rather than the *in vivo/in vitro* micronucleus test (Van Bossuyt et al., 2020; Benfenati et al., 2018), which regulatory bodies require for CMR assessment (ECHA, 2016). Further work is required to better assess aneugenic and clastogenic substances specifically and to improve QSAR models for predicting *in vivo/in vitro* genotoxicity test (Van Bossuyt et al., 2020). To do so, larger and curated collections of experimental data on *in vivo* and *in vitro* micronucleus assays are needed (Kirkland et al., 2019).

A 3Rs-based toolbox has been recently proposed by Luijten et al. (2020) to assess non-genotoxic carcinogens (NGTxC) using *in silico*, *in vitro*, and (*ad hoc*) short-term *in vivo* models with the aim to identify the different MoAs involved in carcinogenesis. This exercise describes an overarching integrated approach to the testing and assessment (IATA) of NGTxC, which enables regulators and scientist to integrate and weigh all available data for hazard assessment (e.g. Cell Transformation Assay, and other relevant assays) (Jacobs et al., 2020; OECD, 2018a; OECD, 2020a). Overall, the NGTxC IATA including the application of QSAR models as developed in **chapter 2** will allow for a qualitative approach to predict both the presence and absence of carcinogenic potential of chemicals across industrial sectors in line with the relevant worldwide regulatory requirements (Jacobs et al., 2020).

QSARs as prioritisation tools for ecological risk assessment of regulated products

Chapter 4 illustrates how QSARs can play a key role in future ERA of chemicals, particularly when assessing the effects of potential harmful PPP active substances towards non-target species such as honey bees. According to OECD QSAR principles (OECD, 2007), a k-Nearest Neighbor (k-NN) model is developed and validated using in-house software for the prediction of acute contact toxicity (i.e. the median lethal dose, LD_{50}) of PPPs towards honey bee (*Apis mellifera*). The classification-based model (k-NN) shows good predictions capacity with an accuracy of 70% for all compounds (n= 255) and of 65% for highly toxic compounds (n= 43). The k-NN model identifies a k number of similar compounds i.e. "neighbours" to the target chemical that will be used to provide the prediction of the endpoint. It is a transparent algorithm, easy to implement on *in silico* platforms and widely used for QSAR datasets presenting different similarity metrics to select neighbours (Manganaro et al., 2016; Mansouri et al., 2018). Hence, the validation of the k-NN models in this work suggests their suitability to reliably predict the toxicity of structurally diverse compounds such as PPPs. Such QSARs show their potential use as screening and prioritisation tools for new PPPs active substances in future ERA, while encouraging the scientific community to further develop *in silico* models to reduce *in vivo*

testing. However, it is noteworthy that PPPs are biologically active compounds, which are characterised by many functional groups (WHO, 2020). Therefore, additional mechanistic information such as MoA of PPP active substances might be of value when developing computational models (Klüver et al., 2019; Zhang et al., 2020; Carnesecchi et al., 2020b).

Although **chapters 2** and **4** provide practical examples for the application of *in silico* tools in ERA and HHRA such as CM(R) assessment, to date the majority of QSARs are mainly available for predicting hazard properties of single chemicals rather than mixtures of chemicals (Moon et al., 2020; Carnesecchi et al., 2020b). A few models have been published in the literature for predicting binary mixtures toxicity in bacteria (Toropova et al., 2012; Wang et al., 2018) and flammability of binary liquid mixtures (Toropova et al., 2019). No QSAR models have been published that predict mixtures toxicity in keystone species such as honey bees, which are particularly relevant for ERA of regulated chemicals such as PPPs. Consequently, **chapter 3** aimed to collect available laboratory *in vivo* experiments on combined toxicity of binary mixtures in bee species, thus providing the first consolidated quantitative review of mixtures toxicity data for the development of QSAR models. This exercise demonstrates that systematic review methodologies are reliable tools to generate robust datasets and scientific evidence to support the hazard assessment of combined exposure to multiple chemicals (Carnesecchi et al., 2019a). The database provides scientific evidence for toxic interactions (e.g. synergism, additive and antagonism) and support the derivation of science-based uncertainty factors for specific PPP binary mixtures in line with EFSA's MIXTOX guidance (Carnesecchi et al., 2019b; More et al., 2019). However, beyond QSAR models to address combined toxicity, the generation of basic TK data (e.g. half-life, bioaccumulation) for chemicals in bee species is needed to allow for the development of *ad hoc* QSAR and Dynamic Energy Budget (DEB) models for hazard assessment of mixtures in bee species (Hesketh et al., 2016). The use of DEB models, in combination with QSARs, might better capture both the toxic dose responses (TD) and the metabolic/elimination processes (TK) of a chemical, thus quantifying time-related effects in non-target organisms (Baas et al., 2018; Marques et al., 2018).

QSAR models to predict chemical mixture potency and the nature of combined toxicity

The database developed in **chapter 3** was employed to develop the first innovative QSAR models to predict the acute contact toxicity of PPPs binary mixtures ($LD_{50\text{-}mix}$) in honey bee, and the nature of combined toxicity (synergism / non-synergism) within a WoE approach. As detailed in **chapter 5,** three different QSAR models are calibrated and validated within CORAL software according to OECD principles (OECD, 2007). Two different notation methods namely "traditional SMILES" and "quasi-SMILES" are used. The "traditional SMILES" approach allows building up one regression-based QSAR model using as input the chemical structure of two-component mixtures codified as disconnected SMILES (i.e. "*chemicalA.chemicalB*") (Toropova et al., 2012). In contrast, the "quasi-

SMILES" method considers additional quantitative information such as the toxicity of each chemical in the mixture (expressed as toxic unit - TU) codified in quasi-SMILES (i.e. "*chemicalA* TU_i*.chemicalB* TU_j"). Consequently, the latter results provide a more interpretable means compared to that relying on traditional SMILES, particularly for predicting (i) the potency of organic binary mixtures ($pLD_{50\text{-}mix}$) and (ii) the nature of combined toxicity (i.e. synergism / non-synergism) in future MRA. Similar models are available in the literature (Qin et al., 2018; Wang et al., 2018; Toropova et al., 2012; Muratov et al., 2012; Kim et al., 2018; Tian et al., 2013). A caveat is the fact that they do not identify the molecular features (i.e. SMILES attributes) responsible for the increase or decrease in the binary mixture toxicity ($pLD_{50\text{-}mix}$). The models validated in **chapter 5**, show that SMILES attributes with stable positive correlation weights (CWs) can be interpreted as promoters of an increase (synergism) in acute contact toxicity ($pLD_{50\text{-}mix}$) of the binary mixture in bees. In contrast, molecular features presenting negative CWs, from the Monte Carlo optimisation, can be interpreted as promoters of a decrease (antagonism) in the acute toxicity of the binary mixtures. Overall, **chapters 3** and **5** also highlight the current data gaps in component-based approaches for MRA in bees, and consequently in QSAR modelling. Such key data gaps include the need for:

a) Laboratory testing and *in silico* docking studies in honey bees and wild bees to broaden our understanding of acute and chronic combined toxicity (contact and oral) and dose dependency for different classes of PPP active substances and contaminants. This would support the characterisation of the synergistic potential of chemicals in bees and TK interactions through either inhibition or induction of metabolism or through direct toxicodynamic (TD) interactions (Cedergreen et al., 2017; More et al., 2019; Gradish et al., 2019).
b) Experimental TK data (e.g. half-life, bioaccumulation) and MoA information would be essential to develop ad-hoc QSAR models for further characterisation of the impact of persistence on combined toxicity of multiple chemicals in bees (including interactions). With such dataset, QSAR models can be integrated DEB models for the hazard assessment of mixtures in bee species while providing a refined mechanistic understanding of the combined toxicity at the honey bee population level (Baas et al., 2018; Hesketh et al., 2016; EFSA Scientific Committee, 2021).

Integrating NAMs for component-based mixture risk assessment

Chapter 6 aimed to address the challenge of integrating MoA information into *in silico* predictions for component-based MRA, thus validating the first integrative QSAR models i) to predict acute contact toxicity (LD_{50}) and ii) to profile the MoA of PPP active substances using open source databases (i.e. EFSA's OpenFoodTox, US-EPA ECOTOX and Pesticide Properties DataBase). This

chapter provided the methodology and the computational tools to derive the first harmonised MoA classification scheme for 113 PPPs active substances, including potency of the chemicals (as LD_{50s} values) towards honey bees. However, this exercise raised a broad question, applicable to both ERA and HHRA: whether MoA can be used to group chemicals in the context of RA of multiple chemicals?.

To address the above-mentioned question, there is a need to clarify two fundamental aspects that intrinsically characterise a given PPP active substance: its toxicity and MoA at the target site (i.e. site of action). The prediction and evaluation of toxicity can benefit from *in silico* models, particularly those predicting the effect and those aiming to predict or anyhow characterise the mechanism; it is noted that the outputs from these two specific models do not necessarily overlap (Carnesecchi et al., 2020b). For instance, the toxicity of insecticides and their specificity towards animals are a consequence of the MoA at the cellular or physiological level in the organism (Simon-Delso et al., 2015; Legradi et al., 2018). Similarly, whilst toxicity is triggered by the internal dose causing the adverse effect (e.g. death of the organism), the specificity depends on the key events leading to the adverse outcome as biochemical events or physiological targets hit by the insecticide in the specie(s) of interest (Sanchez-Bayo, 2012; Meek et al. 2014). Examples are available in the literature with regards to variability in the specificity of PPPs amongst taxa (as in the case of selective insecticides), or are conservation across taxa (e.g. broad-spectrum insecticides) (Sanchez-Bayo, 2012; Legradi et al., 2018). Hence, in **chapter 6**, it is further demonstrated that high variability of PPP active substances toxicity (i.e. LD_{50}) among the same MoA group is often reported in honey bees. Overall, in order to reply to the question above, it is suggested to further develop and test the proposed approach for grouping chemicals using the MoA harmonised classification in other species of ecological relevance. Similarly, an additional analysis of the variability of single PPP active substance toxicity within the same MoA group (e.g. Sodium channel modulator) is recommended in order to integrate the potency aspect when prioritising chemicals of concern in component-based approaches. Nonetheless, besides the intrinsic characteristics of each PPP active substance (i.e. toxicity and MoA), it is noteworthy that TK-TD variability (intra and inter species) plays a key role in the assessment of (hazard) toxicity of PPPs (Medrzycki et al., 2013; Wang et al., 2020; Heard et al., 2017; Chmiel et al., 2020).

Although **chapter 6** focused on bee toxicity, it provides further evidence for the development of harmonised MoA classification schemes to relate the structural and toxicological information with target sites of PPPs active substances for a range of applications: a) toxicity extrapolation and predictions, b) MoA profiling and c) refined grouping of chemicals for component-based MRA in human health and ecological areas (OECD, 2018b; Bopp et al., 2019). In future applications, such methodologies can be used in combination with adverse outcome pathway (AOP) and AOP networks, which then provide the mechanistic understanding needed to integrate data from e.g. *in silico* and

in vitro methods as well as to support of (quantitative) *in vitro* to *in vivo* extrapolation (IVIVE) (Beronius et al., 2020; OECD, 2016).

Key findings and future perspectives in a nutshell

This thesis demonstrated how NAMs such as QSARs can provide regulatory bodies and industry with robust animal-free methods to perform future ERA, AHRA and HHRA of regulated chemicals, emerging contaminants and their mixtures. In line with the Tox21 strategy, this work presents practical examples of QSAR model applications in order to i) shift RA to give us a more mechanistic understanding of toxicity and ii) innovate current chemicals safety testing by improving the quality, efficiency and speed of ERA, AHRA and HHRA of single and multiple chemicals. In particular, this thesis provides risk assessors and scientists with sound methodologies and smart strategies such as *in silico* tools, and responds to the recent call of the EU Chemical Strategy for Sustainability (EC, 2020) aimed to replace animal testing through multidisciplinary research, methods and models, and data analysis capacities.

Below, key findings and recommendations of this thesis are summarised:

- Current QSAR models available on VEGA-HUB *in silico* platform (https://www.vegahub.eu/) are reliable animal-free methods for HHRA purposes, particularly when i) prioritising hazardous chemicals present in the food chain such as PPP active susbtances, metabolites and botanicals; ii) screening/characterising each structurally defined substance for component-based approach in MRA; iii) predicting human health related endpoints such as carcinogenicity and genotoxicity (e.g. *in vitro* Ames test, *in vivo* micronucleus) relevant for CMR (carcinogenic, mutagenic or toxic for reproduction) assessment according to REACH Regulation (Van Bossuyt et al., 2020; Carnesecchi et al., 2020c; Chinen and Malloy, 2020). To date, VEGA provides users with more than 70 QSAR models able to predict (eco)toxicological, environmental and physico-chemical properties of a chemical using its structure (Benfenati et al., 2020). The VEGA software has been recently included in the OECD QSAR ToolBox (OECD, 2020b; Schultz et al., 2018). It will be further integrated into EFSA's OpenFoodTox database, thus fostering the use of NAMs in chemical risk assessment to allow integrating different lines of evidence in an iterative manner (i.e. *in vivo*, *in vitro*, *in silico* data) and reporting of the overall uncertainty using harmonised WoE approaches (Dorne et al., 2021; Hardy et al., 2017).
- K-nearest neighbors (k-NN) based models present advantages in both development and applications compared to other computational models (Mansouri et al., 2018). K-NN is a transparent algorithm which can be applied for developing regression- and classification-based models, thus offering either quantitative (e.g. LD_{50} value) or qualitative/categorial

predictions (e.g. toxic, non-toxic) based on on the most similar compounds i.e. "neighbour" (Carnesecchi et al., 2020b). Similarly, k-NN allows identifying structural alerts or molecular features responsible for the biological activity and environmental fate properties (Manganaro et al., 2016). It is easy to implement on *in silico* platforms (e.g. VEGA-HUB, OECD QSAR ToolBox, OPERA), thus resulting widely used for regulatory purposes such as REACH regulation (Mansouri et al., 2018; Pizzo et al., 2016). In particular, K-NN models offer further advantages for the implementation of MoA profiling in VEGA platform such as visualisation of the neighbours used to predict the target compound as well as reliability of the prediction while assessing their known MoA (Carnesecchi et al., 2020b). Different QSARs methods are available to date (Triebe et al., 2017). However, it is unrealistic to identify "a priori" the best QSAR algorithm to use for chemical RA purposes. Rather, the choice of the QSAR method is case-specific and it depends on several aspects such as i) the nature of available experimental data (e.g. categorial or continuos), ii) the quality of (raw) datasets, iii) the application purpose of the model, and iv) the feasibility of its implementation and reproducibility (Gómez-Jiménez et al., 2018; Cherkasov et al., 2014; OECD, 2007). From a regulatory point of view, international authorities recommend applying different QSAR models, and thus combining multiple predictions within a WoE approach to draw safe conclusions on potential hazardous chemicals (Benfenati et al., 2019; ECHA, 2016).

- QSARs have been historically used for human health and ecological hazard assessment of single chemicals, particularly for predicting phys-chem and toxicological properties of one chemical at a time (Khan et al., 2020; Luechtefeld and Hartung, 2017). Here, the first innovative QSAR models i) to predict binary mixtures toxicity and the nature of the combined toxicity (i.e. synergism, non-synergism) in bees (Carnesecchi et al., 2020a), and iii) to profile mechanistic information such as MoA of chemicals (Carnesecchi et al., 2020b) were developed and validated according to OECD Principles (OECD, 2007). The mixtures models are freely available on CORAL software/databases (http://www.insilico.eu/coral/) and, together with the MoA profiling model are currently being implemented within VEGA-HUB *in silico* platform. Although being primarily designed for predicting mixtures toxicity in species of ecological relevance (i.e. honey bee), such algorithms and methodologies can be applied to datasets comprising phys-chem properties (Toropova et al., 2019), ecological and human-health related endpoints (e.g. mixtures short-term toxicity, carcinogenicity, mutagenicity), thus offering reliable non-animal testing tools for HHRA and ERA of chemicals and their mixtures (Benfenati et al., 2018; Bopp et al., 2019; Carnesecchi et al., 2020a,b). However, the application of such innovative *in silico* tools should be considered as a stepwise approach, thus integrating results from different available NAMs (e.g. high-throughput *in vitro* models)

according to IATA and WoE strategies (OECD, 2020a; OECD, 2019b; Hardy et al., 2017). Similarly, besides integrating NAMs to predict complex toxicological responses, combining NAMs in tiered approaches may also enable more efficient testing of a large number of chemicals and thus reducing in vivo toxicity testing (USEPA, 2018; Thomas et al., 2018).

- Since the understanding of single Adverse Outcome Pathways (AOPs) and their networks is unfolding in both HHRA and ERA (Perkins et al., 2019; LaLone et al., 2017), integration of mechanistic information on the molecular initiating event (MIE), key events (KE) and key event relationships (KER) is foreseen as a mid-term perspective as part of quantitative Adverse Outcome Pathways (qAOPs) approach (OECD, 2020a; Spinu et al., 2020). AOPs can provide the mechanistic understanding needed to integrate data from NAMs such as *in silico* and *in vitro* methods (OECD, 2020a). Similarly, computational models such as QSARs can be employed to predict key MIEs relevant to AOPs such as hepatic steatosis (Cotterill et al., 2020). Additional *in vivo* TK data (e.g. bioavailability, half-life) on tested chemicals would help scientists to better capture the internal (effective) dose triggering either the MIE or the KE (Spinu et al., 2020). TK data can be derived from (high-throughput) *in vitro* methods as well as *in silico* ones (e.g. QSAR, PBK models), while integrating different lines of evidence in read-across or grouping approaches, before being assessed in an overall WoE (OECD 2020b; Benfenati et al., 2019). This will allow risk assessors to tackle the mechanistic understanding of chemicals toxicity as well as the interaction effects of the combined exposure to multiple chemicals (Beronius et al., 2020; More et al., 2019).
- Last, but not the least: **data sharing = progress in science**!

References

Baas, J., Augustine, S., Marques, G.M., Dorne, J. Lou, 2018. Dynamic energy budget models in ecological risk assessment: From principles to applications. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2018.02.058

Benfenati, E., Golbamaki, A., Raitano, G., Roncaglioni, A., Manganelli, S., Lemke, F., Norinder, U., Lo Piparo, E., Honma, M., Manganaro, A., Gini, G., 2018. A large comparison of integrated SAR/QSAR models of the Ames test for mutagenicity $. SAR QSAR Environ. Res. 29, 591–611. https://doi.org/10.1080/1062936X.2018.1497702

Benfenati, E., Chaudhry, Q., Gini, G., Dorne, J. Lou, 2019. Integrating in silico models and read-across methods for predicting toxicity of chemicals: A step-wise strategy. Environ. Int. https://doi.org/10.1016/j.envint.2019.105060

Benfenati, E., Carnesecchi, E., Roncaglioni, A., Baldin, R., Ceriani, L., Ciacci, A., Kovarich, S., Sartori, L., Mostrag, A., Magdziarz, T., Yang, C., 2020. Maintenance, update and further development of EFSA's Chemical Hazards: OpenFoodTox 2.0. EFSA supporting publication 2020:EN-1822. 36 pp. doi:10.2903/sp.efsa.2020.EN-1822.

Beronius, A., Zilliacus, J., Hanberg, A., Luijten, M., van der Voet, H., van Klaveren, J., 2020. Methodology for health risk assessment of combined exposures to multiple chemicals. Food Chem. Toxicol. 143, 111520. https://doi.org/10.1016/j.fct.2020.111520

Boobis, A.R., Cohen, S.M., Dellarco, V., McGregor, D., Meek, M.E., Vickers, C., Willcocks, D., Farland, W., 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit. Rev. Toxicol. https://doi.org/10.1080/10408440600977677

Bopp, S.K., Kienzler, A., Richarz, A.N., van der Linden, S.C., Paini, A., Parissis, N., Worth, A.P., 2019. Regulatory assessment and risk management of chemical mixtures: challenges and ways forward. Crit. Rev. Toxicol. https://doi.org/10.1080/10408444.2019.1579169

Bopp, S.K., Barouki, R., Brack, W., Dalla Costa, S., Dorne, J.L.C.M., Drakvik, P.E., Faust, M., Karjalainen, T.K., Kephalopoulos, S., van Klaveren, J., Kolossa-Gehring, M., Kortenkamp, A., Lebret, E., Lettieri, T., Nørager, S., Rüegg, J., Tarazona, J. V., Trier, X., van de Water, B., van Gils, J., Bergman, Å., 2018. Current EU research activities on combined exposure to multiple chemicals. Environ. Int. https://doi.org/10.1016/j.envint.2018.07.037

Carnesecchi, E., Toropov, A.A., Toropova, A.P., Kramer, N., Svendsen, C., Dorne, J. Lou, Benfenati, E., 2020a. Predicting acute contact toxicity of organic binary mixtures in honey bees (A. mellifera) through innovative QSAR models. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2019.135302

Carnesecchi, E., Svendsen, C., Lasagni, S., Grech, A., Quignot, N., Amzal, B., Toma, C., Tosi, S., Rortais, A., Cortinas-Abrahantes, J., Capri, E., Kramer, N., Benfenati, E., Spurgeon, D., Guillot, G., Dorne, J.L.C.M., 2019a. Investigating combined toxicity of binary mixtures in bees: Meta-analysis of laboratory tests, modelling, mechanistic basis and implications for risk assessment. Environ. Int. 133. https://doi.org/10.1016/j.envint.2019.105256

Carnesecchi, E., Svendsen, C., Lasagni, S., Grech, A., Quignot, N., Amzal, B., Toma, C., Benfenati, E., Guillot, G., Dorne, J.L.C.M., 2019b. Individual datasets investigating combined toxicity of

binary mixtures in bees from laboratory tests [Data set]. Zenodo. Available at http://doi.org/10.5281/zenodo.3383713

Carnesecchi, E., Toma, C., Roncaglioni, A., Kramer, N., Benfenati, E., Dorne, J.L.C.M., 2020b. Integrating QSAR models predicting acute contact toxicity and mode of action profiling in honey bees (A. mellifera): Data curation using open source databases, performance testing and validation. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.139243

Carnesecchi, E., Raitano, G., Gamba, A., Benfenati, E., Roncaglioni, A., 2020c. Evaluation of non-commercial models for genotoxicity and carcinogenicity in the assessment of EFSA's databases. SAR QSAR Environ. Res. https://doi.org/10.1080/1062936X.2019.1690045

Carpenter, D.O., Arcaro, K., Spink, D.C., 2002. Understanding the human health effects of chemical mixtures. Environ. Health Perspect. https://doi.org/10.1289/ehp.02110s125

Cedergreen, N., Dalhoff, K., Li, D., Gottardi, M., Kretschmann, A.C., 2017. Can Toxicokinetic and Toxicodynamic Modeling Be Used to Understand and Predict Synergistic Interactions between Chemicals? Environ. Sci. Technol. https://doi.org/10.1021/acs.est.7b02723

Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'Min, V.E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A., 2014. QSAR modeling: Where have you been? Where are you going to? J. Med. Chem. https://doi.org/10.1021/jm4004285

Chinen, K., Malloy, T., 2020. QSAR Use in REACH Analyses of Alternatives to Predict Human Health and Environmental Toxicity of Alternative Chemical Substances. Integr. Environ. Assess. Manag. https://doi.org/10.1002/ieam.4264

Chmiel, J.A., Daisley, B.A., Pitek, A.P., Thompson, G.J., Reid, G., 2020. Understanding the Effects of Sublethal Pesticide Exposure on Honey Bees: A Role for Probiotics as Mediators of Environmental Stress. Front. Ecol. Evol. https://doi.org/10.3389/fevo.2020.00022

Cotterill, J., Price, N., Rorije, E., Peijnenburg, A., 2020. Development of a QSAR model to predict hepatic steatosis using freely available machine learning tools. Food Chem. Toxicol. https://doi.org/10.1016/j.fct.2020.111494

Crépet, A., Tressou, J., Graillot, V., Béchaux, C., Pierlot, S., Héraud, F., Leblanc, J.C., 2013. Identification of the main pesticide residue mixtures to which the French population is exposed. Environ. Res. 126, 125–133. https://doi.org/10.1016/j.envres.2013.03.008

Drakvik, E., Altenburger, R., Aoki, Y., Backhaus, T., Bahadori, T., Barouki, R., Brack, W., Cronin, M.T.D., Demeneix, B., Hougaard Bennekou, S., van Klaveren, J., Kneuer, C., Kolossa-Gehring, M., Lebret, E., Posthuma, L., Reiber, L., Rider, C., Rüegg, J., Testa, G., van der Burg, B., van der Voet, H., Warhurst, A.M., van de Water, B., Yamazaki, K., Öberg, M., Bergman, Å., 2020. Statement on advancing the assessment of chemical mixtures and their risks for human health and the environment. Environ. Int. https://doi.org/10.1016/j.envint.2019.105267

Dorne, J.L.C.M., Richardson, J., Livaniou, A., Carnesecchi, E., Ceriani, L., Baldin, R., Kovarich, S., Pavan, M., Saouter, E., Biganzoli, F., Pasinato, L., Zare Jeddi, M., Robinson, T.P., Kass, G.E.N., Liem, A.K.D., Toropov, A.A., Toropova, A.P., Yang, C., Tarkhov, A., Georgiadis, N., Di Nicola, M.R., Mostrag, A., Verhagen, H., Roncaglioni, A., Benfenati, E., Bassan, A., 2021. EFSA's OpenFoodTox: An open source toxicological database on chemicals in food and feed and its future developments. Environ. Int. 146, 106293. https://doi.org/10.1016/j.envint.2020.106293

ECHA (European Chemicals Agency), 2020. The use of alternatives to testing on animals for the REACH Regulation. Fourth report under Article 117(3) of the REACH Regulation. ECHA-20-R-08-EN. Available at https://echa.europa.eu/documents/10162/0/alternatives_test_animals_2020_en.pdf/db66b8a3-00af-6856-ef96-5ccc5ae11026

ECHA (European Chemicals Agency), 2017. Guidance on the biocidal products regulation. Volume III Human health - assessment & evaluation, Helsinki, Finland: European Chemicals Agency. (Parts B + C), Version 3.0, November 2017. Reference: ECHA-17-G-25-EN. Available at https://echa.europa.eu/documents/10162/23036412/biocides_guidance_human_health_ra_iii_part_bc_en.pdf/30d53d7d-9723-7db4-357a-ca68739f5094

ECHA (European Chemicals Agency), 2016. Preparation of an inventory of substances suspected to meet REACH Annex III criteria. Preparation of the Annex III inventory – Technical Documentation. Helsinki, Finland: European Chemicals Agency. Available at: https://echa.europa.eu/documents/10162/22332820/annex_iii_preparation_inventory_en.pdf/e42ea5b1-28f0-4390-8d34-6d09e2875ebd

EEA (European Environment Agency), 2018. Chemicals for a sustainable future: report of the EEA Scientific Committee Seminar. Copenhagen, Denmark. Available at https://www.eea.europa.eu/about-us/governance/scientific-committee/reports/chemicals-for-a-sustainable-future

EFSA (European Food Safety Authority), 2009. Guidance of the Scientific Committee on Transparency in the Scientific Aspects of Risk Assessments carried out by EFSA. Part 2: General Principles, 2009. EFSA J. https://doi.org/10.2903/j.efsa.2009.1051

EFSA (European Food Safety Authority), 2014. Modern methodologies and tools for human hazard assessment of chemicals. EFSA Journal 2014;12(4):3638, 87 pp. doi:10.2903/j.efsa.2014.3638.

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Scientific Opinion on the identification of pesticides to be included in cumulative assessment groups on the basis of their toxicological profile (2014 update). EFSA Journal 2013;11(7):3293, 131 pp. https://doi.org/10.2903/j.efsa.2013.3293

EFSA Scientific Committee, More S, Bampidis V, Benford D, Bragard C, Halldorsson T, Hernández-Jerez A, Hougaard Bennekou S, Koutsoumanis, Machera K, Hanspeter N, Nielsen SS, Schlatter J, Dieter Schrenk D, Silano V, Turck D, Younes M, Arnold G, Dorne JL, Maggiore A, Pagani S, Szentes C, Terry S, Tosi S, Vrbos D, Zamariola G and Rortais A, *in press*. A systems-based approach to the environmental risk assessment of multiple stressors in honey bees. EFSA Journal.

Escher, B.I., Stapleton, H.M., Schymanski, E.L., 2020. Tracking complex mixtures of chemicals in our changing environment. Science (80-.). https://doi.org/10.1126/science.aay6636

European Commission (EC), 2020a. Report from the Commission to the European Parliament and the Council: 2019 report on the statistics on the use of animals for scientific purposes in the Member States of the European Union in 2015-2017. COM/2020/16 final. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0016&from=EN

European Commission (EC), 2020b. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: The European Green Deal. COM/2019/640 final. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2019%3A640%3AFIN

European Commission (EC), 2009. Regulation (EC) No. 1107/2009 of the European Parliament and of the council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. Off. J. Eur. Union 309, 1–50.

European Commission (EC), 2013. Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. Available at https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32013R0283

European Parliament and Council, 2015. Regulation (EU) 2015/2283 of the European Parliament and of the Council of 25 November 2015 on novel foods, amending Regulation (EU) No 1169/2011 of the European Parliament and of the Council and repealing Regulation (EC) No 258/97 of the European Parliament and of the Council and Commission Regulation (EC) No 1852/2001. Available at https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32015R2283

Gómez-Jiménez, G., Gonzalez-Ponce, K., Castillo-Pazos, D.J., Madariaga-Mazon, A., Barroso-Flores, J., Cortes-Guzman, F., Martinez-Mayorga, K., 2018. The OECD Principles for (Q)SAR Models in the Context of Knowledge Discovery in Databases (KDD), in: Advances in Protein Chemistry and Structural Biology. https://doi.org/10.1016/bs.apcsb.2018.04.001

Gradish, A.E., van der Steen, J., Scott-Dupree, C.D., Cabrera, A.R., Cutler, G.C., Goulson, D., Klein, O., Lehmann, D.M., Lückmann, J., O'Neill, B., Raine, N.E., Sharma, B., Thompson, H., 2019. Comparison of pesticide exposure in honey bees (Hymenoptera: Apidae) and bumble bees (Hymenoptera: Apidae): implications for risk assessments. Environ. Entomol. 48, 12–21. https://doi.org/10.1093/ee/nvy168.

Hamadache M., Benkortbi O., Amrane A., Hanini S. (2020) QSAR Approaches and Ecotoxicological Risk Assessment. In: Roy K. (eds) Ecotoxicological QSARs. Methods in Pharmacology and Toxicology. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-0150-1_25

Hardy, A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, H.K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J.R., Silano, V., Solecki, R., Turck, D., Benfenati, E., Chaudhry, Q.M., Craig, P., Frampton, G., Greiner, M., Hart, A., Hogstrand, C., Lambre, C., Luttik, R., Makowski, D., Siani, A., Wahlstroem, H., Aguilera, J., Dorne, J., Fernandez Dumont, A., Hempen, M., Valtueña Martínez, S., Martino, L., Smeraldi, C., Terron, A., Georgiadis, N., Younes, M., 2017. Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J. https://doi.org/10.2903/j.efsa.2017.4971

Heard, M.S., Baas, J., Dorne, J. Lou, Lahive, E., Robinson, A.G., Rortais, A., Spurgeon, D.J., Svendsen, C., Hesketh, H., 2017. Comparative toxicity of pesticides and environmental contaminants in bees: Are honey bees a useful proxy for wild bee species? Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2016.10.180

Hesketh, H., Lahive, E., Horton, A.A., Robinson, A.G., Svendsen, C., Rortais, A., Dorne, J.L., Baas, J., Spurgeon, D.J., Heard, M.S., 2016. Extending standard testing period in honeybees to predict lifespan impacts of pesticides and heavy metals using dynamic energy budget modelling. Sci. Rep. https://doi.org/10.1038/srep37655

Khan, P.M., Kar, S., Roy, K., 2020. Ecotoxicological QSARs of mixtures, in: Methods in Pharmacobgy and Toxicology. https://doi.org/10.1007/978-1-0716-0150-1_19

Kienzler, A., Berggren, E., Bessems, J., Bopp, S., Van Der Linden, S., Worth, A., 2014. Assessment of mixtures - review of regulatory requirements and guidance. JRC Science and Policy Report. European Commission, Joint Research Center, Ispra, Italy. doi:10.2788/84264.

Kienzler, A., Connors, K.A., Bonnell, M., Barron, M.G., Beasley, A., Inglis, C.G., Norberg-King, T.J., Martin, T., Sanderson, H., Vallotton, N., Wilson, P., Embry, M.R., 2019. Mode of Action Classifications in the EnviroTox Database: Development and Implementation of a Consensus MOA Classification. Environ. Toxicol. Chem. https://doi.org/10.1002/etc.4531

Kienzler, A., Barron, M.G., Belanger, S.E., Beasley, A., Embry, M.R., 2017. Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.7b02337

Kirkland, D., Reeve, L., Gatehouse, D., Vanparys, P., 2011. A core in vitro genotoxicity battery comprising the Ames test plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. Mutat. Res. - Genet. Toxicol. Environ. Mutagen. https://doi.org/10.1016/j.mrgentox.2010.12.015

Kirkland, D., Uno, Y., Luijten, M., Beevers, C., van Benthem, J., Burlinson, B., Dertinger, S., Douglas, G.R., Hamada, S., Horibata, K., Lovell, D.P., Manjanatha, M., Martus, H.J., Mei, N., Morita, T., Ohyama, W., Williams, A., 2019. In vivo genotoxicity testing strategies: Report from the 7th International workshop on genotoxicity testing (IWGT), in: Mutation Research - Genetic Toxicology and Environmental Mutagenesis. https://doi.org/10.1016/j.mrgentox.2019.03.008

Klüver, N., Bittermann, K., Escher, B.I., 2019. QSAR for baseline toxicity and classification of specific modes of action of ionizable organic chemicals in the zebrafish embryo toxicity test. Aquat. Toxicol. https://doi.org/10.1016/j.aquatox.2018.12.003

Kim, J., Fischer, M., Helms, V., 2018. Prediction of Synergistic Toxicity of Binary Mixtures to Vibrio fischeri Based on Biomolecular Interaction Networks. Chem. Res. Toxicol. https://doi.org/10.1021/acs.chemrestox.8b00164

Jacobs, M.N., Colacci, A., Corvi, R., Vaccari, M., Aguila, M.C., Corvaro, M., Delrue, N., Desaulniers, D., Ertych, N., Jacobs, A., Luijten, M., Madia, F., Nishikawa, A., Ogawa, K., Ohmori, K., Paparella, M., Sharma, A.K., Vasseur, P., 2020. Chemical carcinogen safety testing: OECD expert group international consensus on the development of an integrated approach for the testing and assessment of chemical non-genotoxic carcinogens. Arch. Toxicol. https://doi.org/10.1007/s00204-020-02784-5

Legradi, J.B., Di Paolo, C., Kraak, M.H.S., van der Geest, H.G., Schymanski, E.L., Williams, A.J., Dingemans, M.M.L., Massei, R., Brack, W., Cousin, X., Begout, M.L., van der Oost, R., Carion, A., Suarez-Ulloa, V., Silvestre, F., Escher, B.I., Engwall, M., Nilén, G., Keiter, S.H., Pollet, D., Waldmann, P., Kienle, C., Werner, I., Haigis, A.C., Knapen, D., Vergauwen, L., Spehr, M., Schulz, W., Busch, W., Leuthold, D., Scholz, S., vom Berg, C.M., Basu, N., Murphy, C.A., Lampert, A., Kuckelkorn, J., Grummt, T., Hollert, H., 2018. An ecotoxicological view on neurotoxicity assessment. Environ. Sci. Eur. https://doi.org/10.1186/s12302-018-0173-x

Luechtefeld, T., Hartung, T., 2017. Computational approaches to chemical hazard assessment. ALTEX. https://doi.org/10.14573/altex.1710141

Luijten, M., Corvi, R., Mehta, J., Corvaro, M., Delrue, N., Felter, S., Haas, B., Hewitt, N.J., Hilton, G., Holmes, T., Jacobs, M.N., Jacobs, A., Lamplmair, F., Lewis, D., Madia, F., Manou, I., Melching-Kollmuss, S., Schorsch, F., Schütte, K., Sewell, F., Strupp, C., van der Laan, J.W., Wolf, D.C., Wolterink, G., Woutersen, R., Zvonar, Z., Heusinkveld, H., Braakhuis, H., 2020. A comprehensive view on mechanistic approaches for cancer risk assessment of non-genotoxic

agrochemicals. Regul. Toxicol. Pharmacol. https://doi.org/10.1016/j.yrtph.2020.104789

Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J., 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. J. Cheminform. https://doi.org/10.1186/s13321-018-0263-1

Marques, G.M., Augustine, S., Lika, K., Pecquerie, L., Domingos, T., Kooijman, S.A.L.M., 2018. The AmP project: Comparing species on the basis of dynamic energy budget parameters. PLoS Comput. Biol. https://doi.org/10.1371/journal.pcbi.1006100

Medrzycki, P., Giffard, H., Aupinel, P., Belzunces, L.P., Chauzat, M.P., Claßen, C., Colin, M.E., Dupont, T., Girolami, V., Johnson, R., Conte, Y. Le, Lückmann, J., Marzaro, M., Pistorius, J., Porrini, C., Schur, A., Sgolastra, F., Delso, N.S., Van Der Steen, J.J.M., Wallner, K., Alaux, C., Biron, D.G., Blot, N., Bogo, G., Brunet, J.L., Delbac, F., Diogon, M., El Alaoui, H., Provost, B., Tosi, S., Vidau, C., 2013. Standard methods for toxicology research in Apis mellifera. J. Apic. Res. https://doi.org/10.3896/IBRA.1.52.4.14

Meek, M.E., Boobis, A., Cote, I., Dellarco, V., Fotakis, G., Munn, S., Seed, J., Vickers, C., 2014. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J. Appl. Toxicol. https://doi.org/10.1002/jat.2949

Mitchell, E.A.D., Mulhauser, B., Mulot, M., Aebi, A., 2017. A worldwide survey of neonicotinoids in honey. Science (80-.). 111, 109–111. https://doi.org/10.1126/science.aan3684

Moon, J., Lee, B., Ra, J.S., Kim, K.T., 2020. Predicting PBT and CMR properties of substances of very high concern (SVHCs) using QSAR models, and application for K-REACH. Toxicol. Reports. https://doi.org/10.1016/j.toxrep.2020.08.014

More, S.J., Bampidis, V., Benford, D., Bennekou, S.H., Bragard, C., Halldorsson, T.I., Hernández-Jerez, A.F., Koutsoumanis, K., Naegeli, H., Schlatter, J.R., Silano, V., Nielsen, S.S., Schrenk, D., Turck, D., Younes, M., Benfenati, E., Castle, L., Cedergreen, N., Hardy, A., Laskowski, R., Leblanc, J.C., Kortenkamp, A., Ragas, A., Posthuma, L., Svendsen, C., Solecki, R., Testai, E., Dujardin, B., Kass, G.E., Manini, P., Jeddi, M.Z., Dorne, J.C., Hogstrand, C., 2019. Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals. EFSA J. 17. https://doi.org/10.2903/j.efsa.2019.5634.

Morita, T., Shigeta, Y., Kawamura, T., Fujita, Y., Honda, H., Honma, M., 2019. In silico prediction of chromosome damage: Comparison of three (Q)SAR models. Mutagenesis. https://doi.org/10.1093/mutage/gey017

Muratov, E.N., Varlamova, E. V., Artemenko, A.G., Polishchuk, P.G., Kuz'Min, V.E., 2012. Existing and developing approaches for QSAR analysis of mixtures. Mol. Inform. https://doi.org/10.1002/minf.201100129

OECD (Organisation for Economic Co-operation and Development), 2007. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, ENV/JM/MONO(2007)2. OECD Publishing, Paris. Available at https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm

OECD (Organisation for Economic Co-operation and Development), 2019a. Synthesis Report: OECD Workshop on Approaches to Support Substitution and Alternatives Assessment,

ENV/JM/MONO(2019)3. OECD Publishing, Paris. Available at http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2019)3&doclanguage=en

OECD (Organisation for Economic Co-operation and Development). 2020a. Overview of Concepts and Available Guidance related to Integrated Approaches to Testing and Assessment (IATA), OECD Series on Testing and Assessment, No. 329, Environment, Health and Safety, Environment Directorate, OECD. Available at: http://www.oecd.org/chemicalsafety/risk-assessment/concepts-and-available-guidance-related-to-integrated-approaches-to-testing-and-assessment.pdf

OECD (Organisation for Economic Co-operation and Development), 2019b. Guiding Principles and Key Elements for Establishing a Weight of Evidence for Chemical Assessment, Series on Testing and Assessment No. 311, Environment, Health and Safety Division, Environment Directorate. Available at https://www.oecd.org/chemicalsafety/risk-assessment/guiding-principles-and-key-elements-for-establishing-a-weight-of-evidence-for-chemical-assessment.pdf

OECD (Organisation for Economic Co-operation and Development), 2018a. Test No. 451. Carcinogenicity Studies, OECD Guidelines for the Testing of Chemicals, Section, vol. 4, OECD Publishing, Paris. Available at: https://www.oecd-ilibrary.org/environment/test-no-451-carcinogenicity-studies_9789264071186-en

OECD (Organisation for Economic Co-operation and Development), 2018b. Considerations for Assessing the Risks of Combined Exposure to Multiple Chemicals, Series on Testing and Assessment No. 296, Environment, Health and Safety Division, Environment Directorate. Available at: https://www.oecd.org/chemicalsafety/risk-assessment/considerations-for-assessing-the-risks-of-combined-exposure-to-multiple-chemicals.pdf

OECD (Organisation for Economic Co-operation and Development), 2016. Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA). ENV/JM/MONO(2016)67. OECD Publishing, Paris. Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2016)67/&doclanguage=en

OECD (Organisation for Economic Co-operation and Development), 2020b. The OECD QSAR Toolbox, available at: https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm. Accessed on 16th January 2020

Perkins, E.J., Ashauer, R., Burgoon, L., Conolly, R., Landesmann, B., Mackay, C., Murphy, C.A., Pollesch, N., Wheeler, J.R., Zupanic, A., Scholz, S., 2019. Building and Applying Quantitative Adverse Outcome Pathway Models for Chemical Hazard and Risk Assessment. Environ. Toxicol. Chem. https://doi.org/10.1002/etc.4505

Pizzo, F., Lombardo, A., Manganaro, A., Cappelli, C.I., Petoumenou, M.I., Albanese, F., Roncaglioni, A., Brandt, M., Benfenati, E., 2016. Integrated in silico strategy for PBT assessment and prioritization under REACH. Environ. Res. https://doi.org/10.1016/j.envres.2016.08.014

Pradeep, P., Povinelli, R.J., White, S., Merrill, S.J., 2016. An ensemble model of QSAR tools for regulatory risk assessment. J. Cheminform. https://doi.org/10.1186/s13321-016-0164-0

Prado, A., Pioz, M., Vidau, C., Requier, F., Jury, M., Crauser, D., Brunet, J.-L., Le Conte, Y., Alaux, C., 2019. Exposure to pollen-bound pesticide mixtures induces longer-lived but less efficient

honey bees. Sci. Total Environ. 650, 1250–1260. https://doi.org/10.1016/j.scitotenv.2018.09.102.

Qin, L.T., Chen, Y.H., Zhang, X., Mo, L.Y., Zeng, H.H., Liang, Y.P., 2018. QSAR prediction of additive and non-additive mixture toxicities of antibiotics and pesticide. Chemosphere. https://doi.org/10.1016/j.chemosphere.2018.01.142

Robinson, C., Portier, C.J., Ki, A.Č.A.V.O.Š., Mesnage, R., Roger, A., Clausing, P., Whaley, P., Muilerman, H., Lyssimachou, A., 2020. Achieving a High Level of Protection from Pesticides in Europe : Problems with the Current Risk Assessment Procedure and Solutions. Eur. J. Risk Regul. 1–31. https://doi.org/10.1017/err.2020.18

Sanchez-Bayo, F., 2012. Insecticides Mode of Action in Relation to Their Toxicity to Non-Target Organisms. J. Environ. Anal. Toxicol. https://doi.org/10.4172/2161-0525.s4-002

Schultz, T.W., Diderich, R., Kuseva, C.D., Mekenyan, O.G., 2018. The OECD QSAR toolbox starts its second decade, in: Methods in Molecular Biology. https://doi.org/10.1007/978-1-4939-7899-1_2

Simon-Delso, N., Amaral-Rogers, V., Belzunces, L.P., Bonmatin, J.M., Chagnon, M., Downs, C., Furlan, L., Gibbons, D.W., Giorio, C., Girolami, V., Goulson, D., Kreutzweiser, D.P., Krupke, C.H., Liess, M., Long, E., Mcfield, M., Mineau, P., Mitchell, E.A., Morrissey, C.A., Noome, D.A., Pisa, L., Settele, J., Stark, J.D., Tapparo, A., Van Dyck, H., Van Praagh, J., Van Der Sluijs, J.P., Whitehorn, P.R., Wiemers, M., 2015. Systemic insecticides (Neonicotinoids and fipronil): Trends, uses, mode of action and metabolites. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-014-3470-y

Spinu, N., Cronin, M. T., Enoch, S. J., Madden, J. C., & Worth, A. P., 2020) Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. Arch. of toxicol., 94, 1497-1510. https://link.springer.com/article/10.1007/s00204-020-02774-7

Spurgeon, D., Hesketh, H., Lahive, E., Svendsen, C., Baas, J., Robinson, A., Horton, A., Heard, M., 2016. Chronic oral lethal and sub-lethal toxicities of different binary mixtures of pesticides and contaminants in bees (Apis mellifera, Osmia bicornis and Bombus terrestris). EFSA Support. Publ. 13. https://doi.org/10.2903/sp.efsa.2016.en-1076.

Tian, D., Lin, Z., Yin, D., 2013. Quantitative structure activity relationships (QSAR) for binary mixtures at non-equitoxic ratios based on toxic ratios-effects curves. Dose-Response. https://doi.org/10.2203/dose-response.11-042.Lin

Tickner, J., Jacobs, M.M., Mack, N.B., 2019. Alternatives assessment and informed substitution: A global landscape assessment of drivers, methods, policies and needs. Sustain. Chem. Pharm. https://doi.org/10.1016/j.scp.2019.100161

Thomas, R.S., Paules, R.S., Simeonov, A., Fitzpatrick, S.C., Crofton, K.M., Casey, W.M., Mendrick, D.L., 2018. The US Federal Tox21 Program: A strategic and operational plan for continued leadership. ALTEX. https://doi.org/10.14573/altex.1803011

Topping, C.J., Aldrich, A., Berny, P., 2020. Overhaul environmental risk assessment for pesticides. Science (80-.). https://doi.org/10.1126/science.aay1144

Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2012. CORAL: Models of toxicity of binary mixtures. Chemom. Intell. Lab. Syst. https://doi.org/10.1016/j.chemolab.2012.10.001

Toropova, A.P., Toropov, A.A., Carnesecchi, E., Benfenati, E., Dorne, J. Lou, 2019. The index of ideality of correlation: models for flammability of binary liquid mixtures. Chem. Pap. https://doi.org/10.1007/s11696-019-00903-w

Tosi, S., Costa, C., Vesco, U., Quaglia, G., Guido, G., 2018. A 3-year survey of Italian honey bee-collected pollen reveals widespread contamination by agricultural pesticides. Sci. Total Environ. 615, 208–218. https://doi.org/10.1016/j.scitotenv.2017.09.226.

Triebe, J., Worth, A., Janusch Roi, A. and Coe, A., 2017. JRC QSAR Model Database: EURL ECVAM DataBase service on ALternative Methods to animal experimentation: To promote the development and uptake of alternative and advanced methods in toxicology and biomedical sciences: User Support & Tutorial, EUR 28713 EN, Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-71406-1, doi:10.2760/905519, JRC107491

United Nations Environment Programme (UNEP), 2019. Global Chemicals Outlook II: From Legacies to Innovative Solutions - Implementing the 2030 Agenda for Sustainable Development, The Global Chemicals Outlook. ISBN No: 978-92-807-3745-5. Available at: https://www.unenvironment.org/resources/report/global-chemicals-outlook-ii-legacies-innovative-solutions

USEPA (Environmental Protection Agency), 2018. Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program. June 22, 2018. U.S. Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention, Washington, DC. Available at: https://www.epa.gov/sites/production/files/2018-06/documents/epa_alt_strat_plan_6-20-18_clean_final.pdf

Villaverde, J.J., Sevilla-Morán, B., López-Goti, C., Alonso-Prados, J.L., Sandín-España, P., 2020. QSAR/QSPR models based on quantum chemistry for risk assessment of pesticides according to current European legislation. SAR QSAR Environ. Res. https://doi.org/10.1080/1062936X.2019.1692368

Van Bossuyt, M., Raitano, G., Honma, M., Van Hoeck, E., Vanhaecke, T., Rogiers, V., Mertens, B., Benfenati, E., 2020. New QSAR models to predict chromosome damaging potential based on the in vivo micronucleus test. Toxicol. Lett. https://doi.org/10.1016/j.toxlet.2020.04.016

Wang, T., Tang, L., Luan, F., Cordeiro, M.N.D.S., 2018. Prediction of the toxicity of binary mixtures by qsar approach using the hypothetical descriptors. Int. J. Mol. Sci. https://doi.org/10.3390/ijms19113423

Wang, M., Braasch, T., Dietrich, C., 2020. Reduction of variability for the assessment of side effects of toxicants on honeybees and understanding drivers for colony development. PLoS One. https://doi.org/10.1371/journal.pone.0229295

WHO (World Health Organisation), 2009. Principles and Methods for the Risk Assessment of Chemicals in Food Chapter 5: Dose-response assessment and derivation of health-based guidance values. Princ. Methods Risk Assess. Chem. Food. Available at: https://apps.who.int/iris/bitstream/handle/10665/44065/WHO_EHC_240_8_eng_Chapter5.pdf;jsessionid=C8BEFE468F684EACAEFD5C821378311F?sequence=8

Zhang, S., Wang, N., Su, L., Xu, X., Li, C., Qin, W., Zhao, Y., 2020. MOA-based linear and nonlinear QSAR models for predicting the toxicity of organic chemicals to Vibrio fischeri. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-019-06681-y

Nederlandse samenvatting

Mensen, dieren en het milieu worden dagelijks blootgesteld aan duizenden chemicaliën. Blootstelling aan deze stoffen kan leiden tot nadelige gezondheidseffecten. Om de risico's op nadelige gezondheidseffecten na blootstelling in kaart te brengen, voeren overheidsinstanties zoals de Europese Autoriteit voor Voedselveiligheid (EFSA) risicobeoordelingen uit. Van oudsher gebruiken deze instanties hiervoor testresultaten uit dierproeven. Dieren zoals ratten worden in deze proeven blootgesteld aan verschillende doses van de stof om na te gaan bij welke dosis een nadelig gezondheidseffect optreedt. Vervolgens gaan we ervan uit dat een gevoelig mens of dier veilig blootgesteld kan worden aan een honderdste van de hoogste dosis waar geen gezondheidseffect optreedt in proefdieren.

Deze toxiciteitstesten worden beschouwd als 'black-box' benaderingen omdat de moleculaire mechanismen waarmee chemicaliën schadelijke gezondheidseffecten veroorzaken en de mate waarin effecten en dosisniveaus relevant zijn voor andere diersoorten en individuen slecht gekarakteriseerd worden. De onzekerheden rondom de vertaalslag van proefdier naar mens, dier en milieu roepen veel ethische vragen op. De hoeveelheid proefdieren die gebruikt worden in vereiste toxiciteitstesten zijn aanzienlijk. In 2017 werden er meer dan twee miljoen dieren in toxiciteitstesten gebruikt in de EU.

De zogenaamde Tox21-strategie van de Amerikaanse overheid biedt een basis om toxicologische risicobeoordelingen uit te voeren middels *in vitro*- en *in silico*-onderzoeken. Een *in vitro* toxiciteitstest houdt in dat weefsel, afkomstig van mens of dier, in een kweekschaaltje wordt blootgesteld aan chemicaliën om de toxische dosis te achterhalen. *In silico* testen zijn testen waarbij computeralgoritmen worden gebruikt om, op basis van de chemische structuur van stoffen, de toxiciteit van deze stoffen te organiseren, analyseren en voorspellen. Deze *in vitro* en *in silico* testmethodes, samen ook wel 'new approach methodologies' (NAMs) genoemd, bieden de risicobeoordelaar een meer moleculair mechanistische kijk op de toxiciteit van stoffen.

Dit proefschrift beschrijft hoe deze NAMs adviesorganen zoals EFSA kunnen helpen bij het opstellen van robuuste dierproefvrije risicobeoordelingen. Het proefschrift richt zich specifiek op één type *in silico* testmethode, namelijk quantitative structure-activity relationships (QSAR) ofwel kwantitatieve structuur-activiteitsrelaties. QSAR's zijn wiskundige formules die de chemische structuur en (mate van) toxische activiteit op een kwantitatieve manier relateren voor een reeks stoffen. Hiervoor worden verschillende regressie- en patroonherkenningstechnieken gebruikt. Er zijn tot nu toe nog weinig QSAR's ontwikkeld die mengseltoxicologie en toxicologische effecten op sleutelsoorten (keystone species) zoals bijen voorspellen. Dit proefschrift presenteert daarom praktische voorbeelden van QSAR-modeltoepassingen om mengseltoxiciteit in mens en bij te schatten. Hoofstuk 1 introduceert de principes die gebruikt worden in toxicologische risicobeoordelingen.

Hoofdstuk 2 evalueert de toepasbaarheid van vrij beschikbare QSAR-modellen in het VEGA-HUB-platform om de genotoxiciteit en carcinogeniteit van stoffen in mensen te schatten met behulp van drie verschillende EFSA-databestanden. Hoofdstuk 3 beschrijft de eerste uitgebreide meta-analyse van toxiciteit van binaire mengsels in bijen. Hoofdstuk 4 rapporteert de ontwikkeling van een op classificatie gebaseerd QSAR-modelom acute toxiciteit van bestrijdingsmiddelen in honingbijen (*Apis mellifera*) te voorspellen. Vervolgens presenteert hoofdstuk 5 drie nieuwe QSAR-modellen die CORAL-software gebruiken om de acute toxiciteit van mengsels van pesticides in honingbijen te voorspellen. Hoofdstuk 6 integreert classificaties van werkingsmechanismen ('Mode of Action', MoA) van gifstoffen in QSAR-modellen voor de honingbij. Het laatste hoofdstuk, hoofdstuk 7, beschrijft hoe de studies in de voorgaande hoofdstukken kunnen bijdragen aan het realiseren van de Europese Chemische Strategie voor Duurzaamheid (als onderdeel van de Europese 'Green Deal').

Supplementary materials - Appendices

The supplementary material of each chapter is publicly available on the corresponding Journal web site. In addition, supplementary materials of chapter 3 and 6 are also available on EFSA Knowledge Junction as excel spreadsheets under the corresponding DOI.

Chapter 2

Table S2.1. Number and percentage of compounds already present in the dataset of the models.

	% of experimental data on 628 negative compounds (number of not concordant values)	% of experimental data on 97 'cleaned' positive compounds (number of not concordant values)	% of experimental data on 74 additional positive compounds (number of not concordant values)
CAESAR	9.1% (6)	38.1% (5)	8.1% (0)
ISS	6.8% (12)	47.4% (2)	5.4% (0)
Antares	11.9% (6)	52.6% (6)	14.9% (2)
ISSCAN-CGX	8.9% (13)	53.6% (2)	9.5% (0)
OSF	11.5% (5)	46.4% (9)	10.8% (0)

Table S2.2. Number and percentage of compounds already present in the dataset of the models.

	Number of carcinogenic predictions for the 26 non genotoxic carcinogens in the 'cleaned' dataset (n = 94 positive)	Number of carcinogenic predictions for the 13 non genotoxic carcinogens among the 74 additional positive compounds
CAESAR	7	11
ISS	16	6
Antares	13	0
ISSCAN-CGX	14	0
OSF	12	11

Chapter 3

Table S3.1. Study on combined toxicity effects (expressed as LD50/24h) following acute contact exposure to binary mixtures (PPPs – synergists) in different honey bee sub-species (A. mellifera carnica, A. mellifera ligustica, A. mellifera primorski) (Rinkevich et al., 2015). EM_A and EM_M values are expressed as ng insectide/mg bee (Rinkevich et al., 2015). Different letters for EM_M values in the column referred to the same substance indicate significant differences.

ID_Study	Chemical A		Binary Mixture (A+B)							
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	CI_{EMR} (95%)	Slope (±SE)
Apis mellifera carnica										
Study_001	Phenothrin	0.84 (0.81-0.87)	Amitraz	0.01	0.36[b]	0.31	0.39	2.4	2.07-2.62	4.1±0.5
Study_002			Coumaphos	0.25	0.14[a]	0.12	0.16	6	5.17-6.95	4.4±0.5
Study_003			Pyperol butoxide	0.26	0.24[b]	0.20	0.27	3.6	3.01-4.07	3.4±0.5
A. mellifera ligustica										
Study_004	Phenothrin	0.80 (0.77-0.83)	Amitraz	0.01	0.56[a]	0.52	0.60	1.4	1.32-1.54	3.7±0.4
Study_005			Coumaphos	0.25	0.16 [a]	0.13	0.21	5	3.85-6.46	3.0±0.5
Study_006			Pyperol butoxide	0.26	0.32 [a]	0.29	0.35	2.5	2.25-2.76	3.2±0.3
A. mellifera primoski										
Study_007	Phenothrin	1.06 (1.00-1.12)	Amitraz	0.01	0.36[b]	0.26	0.43	2.9	2.29-3.76	2.3±0.5
Study_008			Coumaphos	0.25	0.15[a]	0.14	0.17	7	6.28-7.91	7.8±0.9
Study_009			Pyperol butoxide	0.26	0.24[b]	0.21	0.26	4.5	3.91-4.96	3.2±0.3

Table S3.2. Study on combined toxicity effects (LD50/48h) following acute contact exposure to binary mixtures (PPPs) in different bee species (Apis mellifera and Osmia cornifrons) (Biddinger et al., 2013). *= Statistically significant experiments according to the author. NA = Not Applicable

ID_Study	**Chemical A**		**Binary Mixture (A+B)**							
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	CI_{EMR}	Slope (±SE)
Osmia cornifrons										
Study_010	Acetamiprid	4.0 (1.1-7.1)	Indar	0.18	2.1	1.1	3.2	1.9*	0.95-3.9	1.94±0.42
Study_011	Imidacloprid	3.8 (1.7-12.6)	Indar	0.18	6.6	1.4	9.6	0.57	NA	3.11±1.09
Apis mellifera										
Study_012	Acetamiprid	64.6 (38.1-252)	Indar	0.18	14.3	8.5	30.8	4.5*	2.5-8.2	3.13±0.53
Study_013	Imidacloprid	0.2 (0.1-0.3)	Indar	0.18	0.3	0.1	0.4	0.66	NA	1.84±0.33

Table S3.3. Study on combined toxity effects (expressed as LD50/24h) following acute contact exposure to binary mixtures (PPPs, sinergists, veterinary products) in honey bees (A. mellifera). All the values are expressed in µg/bee. The slope is also reported as provided by the author in the experimental study. [1]= Iwasa et al. (2004). [2]= Johnson et al. (2009), treatments with non-overlapping confidence intervals are considered significantly different; treatments different from control are indicated with an asterisk (*). [3]= Johnson et al. (2013), significant differences compared to the control treatment are indicated with a superscript letter "a" = significant pre-treatment effect, "b" = significant pre-treatment*acaricide dose effect. [4]= Johnson et al. (2006), treatments with non-overlapping confidence intervals are considered significantly different; treatments different from control are indicated with an asterisk (*); dagger (ǂ) indicates nonoverlap of LD_{50} 95% CIs between significant inhibitor treatments. [5]= Ellis et al. 1997. ** = two different effect concentrations of tau-fluvalinate (EM_A = 19.8 µg/bee; EM_A = 8.98 µg/bee) are reported in the original publication (Johnson et al., 2013). NA = Not Applicable.

ID_Study	Chemical A		Binary Mixture (A+B)								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	EMR (-)	CI_{EMR}	Slope(±SE)
Study_014	Acetamiprid[1]	7.07 (4.57-11.2)	Piperonyl butoxide	0.03	1.17	0.342	3.79	6.04		4.29-8.51	1.55±0.181
Study_015			S,S,S-tributyl phosphorotrithioate	NA	2.39	0.278	12.40	2.96*		1.83-4.76	2.96±0.736
Study_016			Diethyl maleate	NA	6.94	4.100	13.20	1.02		0.783-1.33	1.46±0.140
Study_017			triflumizole	0.50	0.03	0.008	0.10	235		171-347	1.91±0.240
Study_018			propiconazole	0.10	0.07	0.023	0.20	101		76.7-143	2.30±0.242
Study_019			triadimefon	NA	0.08	0.043	0.18	84		64.2-110	2.05±0.198
Study_020			Epoxiconazole	NA	0.50	0.156	1.66	14		10.0-20.0	2.74±0.404
Study_021			uniconazole-P	0.50	1.12	0.270	4.96	6.3		4.22-9.45	2.05±0.349
Study_022	Imidaclo prid[1]	0.0179 (0.0092 - 0.0315)	Piperonyl butoxide	0.03	0.011	0.006	0.02	1.7		1.29-2.26	1.66±0.112
Study_023			triflumizole	0.50	0.010	0.005	0.02	1.85		1.67-3.09	2.76±0.284
Study_024			propiconazole	0.10	0.012	0.004	0.03	1.52		1.04-2.24	2.12±0.272
Study_025	Thiaclopr id[1]	14.6 (9.53-25.4)	Piperonyl butoxide	0.03	0.09	0.041	0.21	162		115-207	1.64±0.134
Study_026			triflumizole	0.50	0.01	0.003	0.04	1460		752-1740	2.32±0.363
Study_027			propiconazole	0.10	0.03	0.008	0.07	486		388-811	2.27±0.298
Study_028	Coumaphos[2]	20.29 (14.88-29.44)	Diethyl maleate	NA	19.92	10.48	53.45	1.02		NA	1.95±0.19
Study_029			S,S,S-tributyl phosphorotrithioate	NA	7.29	4.88	9.22	2.8*		1.48-4.08	4.77±0.52

ID_Study	Chemical A		Binary Mixture (A+B)								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	EMR (-)	CI_{EMR}	Slope(±SE)
Study_030	Coumaphos[2]	20.29 (14.88-29.44)	Piperonyl butoxide	0.03	5.04	3.34	7.01	4.03*		1.96-6.08	3.85±0.32
Study_031			Tau-fluvalinate	0.01	20.97	12.90	47.86	0.97		0.09-1.97	2.39±0.28
Study_032			Tau-fluvalinate	0.03	12.77	10.63	15.59	1.59		0.94-2.23	3.34±0.29
Study_033			Tau-fluvalinate	0.08	6.05	4.29	8.68	3.35		1.64-5.06	1.5±0.14
Study_034			Tau-fluvalinate	0.25	6.06	5.15	6.94	3.35		2.04-4.64	2.38±0.27
Study_035	Tau-fluvalinate[2]	6.75 (6.24-7.33)	Coumaphos	0.005	6.14	5.03	7.71	1.10		0.84-1.35	1.86±0.14
Study_036			Coumaphos	0.01	3.29	2.15	4.87	2.05*		1.18-2.91	1.42±0.12
Study_037			Coumaphos	0.05	2.68	2.25	3.28	2.52*		1.99-3.04	1.89±0.13
Study_038			Coumaphos	0.15	1.53	0.72	2.60	4.41*		1.67-7.14	1.56±0.11
Study_039			Coumaphos	0.49	0.21	0.15	0.29	32.14*		21.11-43.16	1.61±0.12
Study_040	Tau-fluvalinate[3]	19.8** (16.3-22.4)	Coumaphos	0.15	0.78	0.13	3.05	25.38[a]		NA	0.93±0.28

ID_Study	Chemical A		Binary Mixture (A+B)								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	EMR (-)	CI_{EMR}	Slope(±SE)
Study_041			Fenpyroximate	0.06	2.40	1.45	3.65	8.25[a]		NA	1.76 ± 0.25
Study_042			Amitraz	0.02	3.74	2.14	7.08	5.29[a]		1.70-8.88	1.61 ± 0.32
Study_043			thymol	NA	10.20	7.85	14.00	1.94[a]		1.28-2.59	1.35 ± 0.13
Study_044			oxalic acid	NA	7.05	5.67	8.98	2.81[a]		2.01-3.59	1.41 ± 0.13
Study_045			pyraclostrobin+boscalid	0.30	5.95	4.48	8.09	3.33[ab]		2.19-4.45	2.88±0.47
Study_046			pyraclostrobin	0.10	4.43	0.67	61.40	4.44 [a]		NA	1.96±0.66
Study_047			boscalid	0.10	11.60	7.43	19.90	1.71		0.75-2.66	2.58±0.51
Study_048			Chlorothalonil	0.16	7.24	3.96	12.90	2.73[a]		0.99-4.47	1.77±0.38
Study_049			Prochloraz	0.07	0.01	0.01	0.02	1980[a]		944-3015	1.94±0.32
Study_050			Diethyl maleate	NA	8.26	7.57	9.03	2.40		1.97-2.23	NA
Study_051			S,S,S-tributyl phosphorotrithioate	NA	1.96	0.83	4.17	10.1		1.35-18.84	NA
Study_052			Piperonyl butoxide	0.03	0.01	0.01	0.02	1980		944-3015	NA
Study_053	Coumaphos[3]	31.2 (22.2-49.6)	tau-fluvalinate	0.08	6.50	4.98	8.57	4.80[ab]		2.31-7.28	NA
Study_054			Fenpyroximate	0.06	4.12	3.35	5.06	7.57[a]		3.89-11.25	1.80±0.17

ID_Study	Chemical A		Binary Mixture (A+B)								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	EMR (-)	CI_{EMR}	Slope(±SE)
Study_055			Amitraz	0.02	9.20	1.12	25.10	3.39[a]		1.27-8.05	1.70±0.51
Study_056			thymol	NA	23.10	14.7	34.50	1.35		0.52-2.17	4.51±1.00
Study_057			oxalic acid	NA	14.70	10.8	22.10	2.12		0.88-3.36	1.75±0.25
Study_058			pyraclostrobin+boscalid	0.10	25.90	19.9	34.60	1.20		0.57-1.83	4.52±1.20
Study_059			boscalid	0.10	22.60	15.3	32.40	1.38		0.58-2.18	3.23±0.56
Study_060			Chlorothalonil	0.16	16.60	6.77	85.60	1.88		NA	1.81±0.52
Study_061			Prochloraz	0.07	0.44	0.38	0.50	70.91[a]		38.3-103.5	3.11±0.27
Study_062			Diethyl maleate	NA	19.90	10.5	53.50	1.57		0.26-3.39	NA
Study_063			S,S,S-tributyl phosphorotrithioate	NA	7.29	4.9	9.22	4.28		2.01-6.54	NA
Study_064			Piperonyl butoxide	0.03	5.04	3.3	7.01	6.19		2.64-9.73	NA
Study_065	Fenpyroximate[3]	6.65 (4-12)	tau-fluvalinate	0.08	5.54	3.1	12.80	1.20		0.15-3.78	1.67±0.39
Study_066			Coumaphos	0.15	2.03	1.3	4.46	3.28[a]		0.63-8.84	2.18±0.19
Study_067			Amitraz	0.02	1.80	1.6	2.04	3.69[a]		1.42-5.96	4.07±0.49
Study_068			thymol	NA	3.69	2.8	4.98	1.80		0.59-3.00	2.26±0.26

ID_Study	Chemical A		Binary Mixture (A+B)								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	EMR (-)	CI_{EMR}	Slope(±SE)
Study_069			oxalic acid	NA	1.50	0.8	3.06	4.43[a]		0.15-8.70	1.87±0.46
Study_070			pyraclostrobin+boscalid	0.3	3.16	2.6	3.92	2.10		0.76-3.44	3.09±0.32
Study_071			pyraclostrobin	0.10	2.09	0.5	4.24	3.18[a]		0.35-10.40	2.27±0.72
Study_072			boscalid	0.10	5.64	2.9	17.20	1.18		NA	3.30±0.96
Study_073			Chlorothalonil	0.16	6.41	5.6	7.36	1.04		0.39-1.67	3.13±0.32
Study_074			Prochloraz	0.07	0.25	0.17	0.34	26.60[a]		8.22-44.97	2.17±0.28
Study_075			Diethyl maleate	NA	4.38	1.90	8.80	1.52		0.01-3.02	1.60±0.17
Study_076			S,S,S-tributyl phosphorotrithioate	NA	1.26	0.10	15.20	5.28[a]		NA	2.44±0.74
Study_077			Piperonyl butoxide	0.03	0.27	0.12	0.75	24.63[a]		NA	1.90±0.46
Study_078	Amitraz[3]	3.66 (2.26-5.56)	tau-fluvalinate	0.08	4.87	2.38	8.31	0.75		0.32-2.33	3.10 ± 0.76
Study_079			Coumaphos	0.15	2.73	1.82	3.73	1.34		0.57-2.10	2.45 ± 0.41

ID_Study	Chemical A		Binary Mixture (A+B)								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	EMR (-)	CI_{EMR}	Slope(±SE)
Study_080			Fenpyroximate	0.06	4.57	2.78	6.48	0.80		0.49-2.00	2.25 ± 0.42
Study_081			thymol	NA	3.91	2.98	5.52	0.94		0.47-1.66	2.78±0.35
Study_082			oxalic acid	NA	14.60	8.66	38.70	0.25[a]	3.99	NA	2.81±0.78
Study_083			pyraclostrobin+boscalid	0.30	4.04	2.25	10.40	0.91		NA	3.04±0.28
Study_084			pyraclostrobin	0.10	1.64	0.90	2.51	2.23		0.74-3.71	1.96±0.36
Study_085			boscalid	0.10	4.82	2.83	6.74	0.76		0.51-2.11	3.89±0.81
Study_086			Chlorothalonil	0.16	3.34	1.48	8.89	1.10		NA	1.98±0.60
Study_087			Prochloraz	0.07	2.48	1.45	3.74	1.48		0.52-2.42	2.66±0.57
Study_088			Diethyl maleate	NA	2.30	0.31	4.24	1.59		0.05-3.12	3.10±0.86
Study_089			S,S,S-tributyl phosphorotrithioate	NA	2.17	1.63	2.87	1.69		0.78-2.58	1.99±0.23
Study_090			Piperonyl butoxide	0.03	2.41	0.92	6.35	1.52		NA	3.05±0.93
Study_091	Thymol[β]	55.1 (42.1-70)	tau-fluvalinate	0.08	16.10	11.2	21.40	3.42[a]		2.03-4.81	2.11±0.29
Study_092			Coumaphos	0.15	20.40	10.4	38.00	2.70[a]		0.75-4.65	3.33±0.87
Study_093			Fenpyroximate	0.06	34.90	23.9	47.90	1.58		0.90-2.25	2.09±0.27
Study_094			Amitraz	0.02	43.20	25.3	61.00	1.28		0.65-1.89	3.83±0.89

ID_Study	Chemical A		Binary Mixture (A+B)								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	EMR (-)	CI_{EMR}	Slope(±SE)
Study_095			oxalic acid	NA	30.70	23.4	39.90	1.79[a]		1.13-2.45	2.92±0.42
Study_096			pyraclostrobin+boscalid	0.30	31.90	16.9	44.70	1.73		0.85-2.59	2.77±0.71
Study_097			pyraclostrobin	0.10	28.20	4.96	57.90	1.95		0.05-3.85	1.44±0.42
Study_098			boscalid	0.10	47.10	35.4	62.10	1.17		0.72-1.61	2.50±0.38
Study_099			Chlorothalonil	0.16	29.80	21.1	39.90	1.85[a]		1.10-2.59	2.25±0.33
Study_100			Prochloraz	0.07	39.00	33.2	45.10	1.41[b]		0.99-1.83	2.66±0.20
Study_101			Diethyl maleate	NA	64.00	42.6	91.60	0.86		0.62-1.69	2.47±0.44
Study_102			S,S,S-tributyl phosphorotrithioate	NA	35.10	21.9	52.00	1.57		0.78-2,3.5	2.48±0.48
Study_103			Piperonyl butoxide	0.03	32.40	19.8	49.60	1.70		0.80-2.59	2.23±0.41
Study_104	Tau-fluvalinate[3]	8.98** (7.85-10.4	Piperonyl butoxide	0.0001	11.0	6.97	23.90	0.82[b]	1.22	0.26-2.18	1.62±0.37
Study_105			Prochloraz	0.0003	11.1	9.94	12.40	0.81		1.01-1.45	2.98±0.26
Study_106			propiconazole	0.0003	13.9	11.9	16.60	0.65[a]	1.55	1.20-1.88	3.24±0.32
Study_107			fenbuconazole	0.006	10.5	8.59	13.40	0.86		0.85-1.48	2.76±0.34
Study_108			metconazole	0.0003	12.7	8.55	22.50	0.71[a]	1.41	0.61-2.21	3.54±0.74
Study_109			myclobutanil	0.001	12.9	10.70	16.00	0.70[a]	1.44	1.07-1.79	3.08±0.35
Study_110			Piperonyl butoxide	0.001	8.48	6.42	11.40	1.06		0.71-1.40	4.23±0.55
Study_111			Prochloraz	0.003	4.64	3.19	7.15	1.94[a]		1.06-2.80	2.17±0.44
Study_112			propiconazole	0.003	5.97	4.67	7.71	1.50[a]		1.06-1.94	2.46±0.36
Study_113			fenbuconazole	0.06	6.21	4.94	7.80	1.45[a]		1.05-1.83	2.50±0.33
Study_114			metconazole	0.003	5.21	4.21	6.38	1.72[a]		1.28-2.15	2.41±0.31

ID_Study	Chemical A		Binary Mixture (A+B)									
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	EMR (-)	CI_{EMR}	Slope(±SE)
Study_115			myclobutanil	0.01	3.75	3.06	4.71	2.39[ab]		1.76-3.02	1.92±0.20
Study_116			Piperonyl butoxide	0.01	1.54	1.03	4.05	5.83[ab]		0.05-11.60	1.69±0.35
Study_117			Prochloraz	0.03	0.22	0.10	0.52	41[a]		1.42-80.20	2.00±0.51
Study_118			propiconazole	0.03	0.74	0.62	0.88	12[a]		9.39-14.87	3.37±0.37
Study_119			fenbuconazole	0.61	0.55	0.46	0.66	16[a]		12.56-20.09	2.52±0.24
Study_120			metconazole	0.03	0.45	0.36	0.56	20[a]		14.69-25.21	2.19±0.22
Study_121			myclobutanil	0.07	0.12	0.09	0.16	74[a]		50.55-99.10	2.71±0.28
Study_122	Cyfluthrin[4]	0.067 (0.062-0.075)	Diethyl maleate	NA	0.055	0.04	0.09	1.2		0.57-1.48	3.28±0.34
Study_123			S,S,S-tributyl phosphorotrithioate	NA	0.029	0.02	0.05	2.3[*]		0.99-3.14	3.47±0.39
Study_124			Piperonyl butoxide	0.34	0.0022	0.0015	0.0042	30[*‡]		NA	2.22±0.23
Study_125	λ-cyhalothrin[4]	0.102 (0.073-0.133)	Diethyl maleate	NA	0.04	0.03	0.05	2.7[*]		1.56-3.53	2.42±0.28
Study_126			S,S,S-tributyl phosphorotrithioate	NA	0.03	0.01	0.07	4.00		NA	2.57±0.29
Study_127			Piperonyl butoxide	0.34	0.0013	0.0011	0.0015	80[*‡]		52.4-104.5	2.64±0.27
Study_128	Tau-fluvalinate[4]	9.45 (7.480-12.00)	Diethyl maleate	NA	8.26	7.6	9.0	1.1		0.85-1.43	3.08±0.22

ID_Study	Chemical A		Binary Mixture (A+B)								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	EMR (-)	CI_{EMR}	Slope(±SE)
Study_129			S,S,S-tributyl phosphorotrithioate	NA	1.96	0.8	4.2	4.8*		0.48-9.15	1.57±0.14
Study_130			Piperonyl butoxide	0.34	0.010	0.007	0.015	980*‡		504-1385	1.38±0.09
Study_131	Bifenthrin[5]	0.034 (0.023-0.058)	Apistan®	NA	0.018	0.02	0.02	1.89		0.91-2.86	NA
Study_132	Carbaryl[5]	0.232 (0.19-0.279)	Apistan®	NA	0.175	0.14	0.21	1.33		0.95-1.69	NA
Study_133	Methyl Parathion[5]	0.041 (0.037-0.044)	Apistan®	NA	0.039	0.03	0.04	1.05		0.88-1.21	NA

Table S3.4. Summary table for acute oral toxicity (LD50, 24h) studies in bee species. No. = number of case studies. EMR (+) = Estimated Mean Ratios describing increased toxicity. EMR (-) = Estimated Mean Ratio describing decreased toxicity. ♀= female individuals. Na= Not available. NA = Not Applicable.

Substance	No.	Spp.	EMR (+)	EMR (-)
Endpoint= mortality (LD_{50})				
PPP + Vet. product	26	*A. mellifera*	1.1 – 201	1.14 – 1.70
Vet. products	4	*A. mellifera*	1.8 – 4.1	NA
Endpoint= mortality (%)				
PPP + PPP	5	*O. bicornis*♀	6.7 – 8.7	NA
PPP + PPP	2	*A. mellifera*	4 – 7.12	NA
PPP + Vet. product	3	*A. mellifera*	3.3 – 4	NA
PPP + PPP	1	*B. terrestris*	NA	NA

Table S3.5. Combined toxicity effects (expressed as LC50/24h, µg/ml) following acute oral exposure to binary mixtures (veterinary products) in A. mellifera (Guseman et al., 2016). Verapamil, Pristine, Fumagillin and Quercetin were present in 30% sucrose syrup. * = mixture of boscalid and pyraclostrobin.

Study_ID	**Chemical A**		**Binary mixture (A+B)**						
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	CI_{EMR}
Study_134	Ivermectim	1.57 (1.38–1.80)	Verapamil	NA	0.38	0.01	0.67	4.13	0.5-7.7
Study_135			Pristine*	NA	0.57	0.34	0.74	2.75	1.7-3.7
Study_136			Fumagillin	NA	0.86	0.47	1.25	1.83	0.9-2.6
Study_137			Quercetin	NA	0.61	0.41	0.77	2.57	1.7-3.4

Table S3.6. Combined toxicity effects (expressed as % of mortality) following acute oral exposure to binary mixtures (PPP - veterinary product) in A. mellifera (Guseman et al., 2016). [1]= EMR is represented as Standardized Mortality Ratio (SMR= ratio of observed deaths in the binary mixture study group to expected deaths in the compound-only study) (Everitt and Skrondal, 2010). Saple size N= 15. Honey bees fed 0.1 µg/µl acetamiprid with 1 mM solutions of inhibitors in sucrose syrup. EM_A (Estimated Mean) for compound A and EM_M (Estimated Mean mixture) values are expressed as % of mortality (±SE).

Study_ID	**Chemical A**		**Binary mixture (A+B)**					
	Name	EM_A (±SE)	Chemical B	TU_B	EM_M	±SE	SMR^1 (+)	CI_{EMR}
Study_138	Acetamiprid	0.24 (±0.09)	Verapamil	NA	0.95	0.05	3.95	NA
Study_139			Fumagillin	NA	0.80	0.03	3.33	NA
Study_140			Quercetin	NA	0.98	0.02	4.08	NA

Table S3.7. Combined toxicity effects (expressed as % of mortality) following acute oral exposure in different bee species (Apis mellifera, Bombus terrestris, Osmia bicornis) at various assessment times (Sgolastra et al., 2017). *= EMR was reported as Standardized Mortality Ratio (SMR= ratio of observed deaths in the binary mixture study group to expected deaths in the compound-only study). **= Not Applicable: mortality provided by the author is 0%. EM_A values for compound A and EM_M values are expressed as % of mortality. Survival probability is also reported as provided by the author. ♀= female individuals.

Study_ID	**Chemical A**		**Binary mixture (A+B)**				
	Name	EM_A	Chemical B	DoseB (TU_B)	Survival Probability	EM_M	SMR* (+)
A. mellifera, n=210 (4h)							
Study_141	Clothianidin	0.07	Propiconazole	0.07	0.50	0.50	7.14
A. mellifera, n=210 (24h)							
Study_142	Clothianidin	0.14	Propiconazole	0.07	0.43	0.57	4.07
B. terrestris, n=212(4h)							
Study_143	Clothianidin	1.00	Propiconazole	0.07	0.64	0.36	NA**
O. bicornis♀, n= 179 (4h)							
Study_144	Clothianidin	0.03	Propiconazole	0.07	0.74	0.26	8.67
O. bicornis♀, n= 179 (24h)							
Study_145	Clothianidin	0.06	Propiconazole	0.07	0.60	0.40	6.67
O. bicornis♀, n= 179 (48h)							
Study_146	Clothianidin	0.06	propiconazole	0.07	0.50	0.50	8.33
O. bicornis♀, n= 179 (72h)							
Study_147	Clothianidin	0.06	Propiconazole	0.07	0.50	0.50	8.33
O. bicornis♀, n= 179 (96h)							
Study_148	Clothianidin	0.12	Propiconazole	0.07	0.50	0.50	4.00

Table S3.8. Studies on combined toxicity effects (expressed as LD50/24h) of binary mixtures (PPPs, veterinary produts) following acute oral exposure to chemical B in honey bees (Johnson et al., 2013; 2012; Wilkins et al. 2013). [1]= Johnson et al. 2013 (tau-fluvalinate + sucrose): significant differences

compared to the respective treatment are indicated with a superscript letter "a" = significant pre-treatment effect, "b" = significant pre-treatment*acaricide dose effect. [2]= Johnson et al. 2012: treatments with non-overlapping 95% confidence interval are considered significantly different. [3]= Wilkins et al. 2013 (study on larvae). NA = Not Applicable.

Study_ID	**Chemical A**		**Binary Mixture (A+B)**								
	Name	EM_A	Chemical B	TU_B	EM_M	CI_1 (95%)	CI_2 (95%)	EMR (+)	EMR (-)	CI_{EMR}	Slope(±SE)
Study_149	Tau-fluvalinate[1]	9.2 (7.95-10.8)	Oxytetracycline	NA	8.43	7.34	9.80	1.09[b]		0.85-1.32	2.68±0.21
Study_150			Tylosin	NA	10.5	8.10	14.90	0.88	1.13	0.73-1.55	2.32±0.34
Study_151			Fumagillin	NA	4.80	3.74	6.32	1.92[a]		1.32-2.51	1.99±0.22
Study_152	Coumaphos[1]	26 (19.5-39.5)	Oxytetracycline	NA	20	15.10	27.60	1.30		0.65-1.94	2.47±0.36
Study_153			Tylosin	NA	25.7	17.90	43.00	1.01		0.38-1.64	3.50±0.72
Study_154			Fumagillin	NA	33.3	25.50	49.20	0.78		0.60-1.95	2.13±0.28
Study_155	Fenpyroximate[1]	3.24 (2.67-3.88)	Oxytetracycline	NA	4.71	3.90	5.65	0.69[a]	1.45	1.07-1.83	3.50±0.37
Study_156			Tylosin	NA	4.05	3.59	4.55	0.80	1.25	0.97-1.52	2.60±0.18
Study_157			Fumagillin	NA	5.53	4.41	6.87	0.59[a]	1.71	1.21-2.20	2.78±0.32
Study_158	Amitraz[1]	5.47 (4.12-7.08)	Oxytetracycline	NA	3.66	2.99	4.65	1.5		0.96-2.02	4.23±0.54
Study_159			Tylosin	NA	4.53	3.83	5.32	1.2		0.82-1.58	3.29±0.29
Study_160			Fumagillin	NA	3.88	2.94	5.15	1.4		0.85-1.96	3.75±0.52
Study_161	Thymol[1]	38.1 (27.3-49.6)	Oxytetracycline	NA	27.5	15.40	45.10	1.39		0.53-2.23	3.58±0.85
Study_162			Tylosin	NA	32.3	14.50	47.70	1.18		0.48-1.87	3.12±0.75
Study_163			Fumagillin	NA	25.3	21.30	29.70	1.51		0.99-2.01	4.02±0.41
Study_164	Tau-Fluvalinate[2]	8.050 (7.210-8.990)	Phenobarbital	NA	0.19	0.12	0.31	42.37*		20.67-64.06	1.46±0.12
Study_165			Xanthotoxin	NA	0.04	0.001	0.13	201.25*		NA	0.34±0.09
Study_166			Quercetin	NA	11.4	9.74	13.86	0.71	1.40	1.11-1.71	2.98±0.40
Study_167			Salicylic acid	NA	4.45	2.18	8.56	1.81		0.49-3.12	1.56±0.33
Study_168			Indole-3-carbinol	NA	8.34	5.92	10.93	0.97		0.70-1.36	2.53±0.67
Study_169	λ-cyhalothrin[2]	0.0475 (0.0343-0.0675)	Phenobarbital	NA	0.02	0.005	0.025	2.81*		0.67-4.07	2.95±0.39
Study_170	Aldrin[2]	0.0605	Phenobarbital	NA	0.04	0.03	0.047	1.57*		1.07-1.95	3.91±0.36

		(0.0527-0.071)									
Study_171	Dieldrin[2]	0.0372 (0.0319-0.0465)	Phenobarbital	NA	0.0207	0.014	0.025	1.80*		0.86-2.85	3.46±0.30
Study_172	Dimethoate[3]	0.5-12.8	Acetone (1.5%)	NA	0.87	0.00	0.00	NA		NA	NA
Study_173	Dimethoate[3]	0.5-12.8	Acetone (10%)	NA	0.25	0.05	0.62	NA		NA	NA
Study_174	Dimethoate[3]	0.5-12.8	Acetone (5%)	NA	0.15	0.04	0.22	NA		NA	NA

Table S3.9. Study on combined chronic toxicity (expressed as LC50/96h/240h, µg/bee) following oral exposure to binary mixtures (PPPs) in different bee species (A. mellifera, B. terrestris, and O. bicornis) (Spurgeon et al., 2016; Robinson et al., 2017). * = Statistical significant mixture effects; [1] = Spurgeon et al., 20016; [2] = Robinson et al., 2017; ♀ = female individuals; ♂ = male individuals; NA = Not Applicable.

Study_ID	Chemical A		Binary mixture (A+B)							
	Name	EM_A (95% CI)	Chemical B	TU_B	EM_M	CI_1 (95 %)	CI_2 (95 %)	EMR (+)	EMR (-)	CI_{EMR}
A. mellifera, LC_{50} (96 h), n=30										
Study_175	Clothianidin[1]	0.128 (0.109-0.148)	Propiconazole	(low)	0.143	0.120	0.166	0.90		0.70- 1.11
Study_176			Propiconazole	(high)	0.122	0.101	0.143	1.05		0.82-1.31
Study_177	Clothianidin[1]	0.160 (0.131-0.189)	Tau-fluvalinate	(low)	0.125	0.102	0.148	1.28		0.96-1.63
Study_178			Tau-fluvalinate	(high)	0.129	0.104	0.153	1.24		0.93-1.59
Study_179	Dimethoate[1]	1.158 (0.970-1.345)	Propiconazole	(low)	1.550	1.305	1.795	0.75	1.33	0.58-0.92
Study_180			Propiconazole	(high)	1.353	1.125	1.581	0.86	1.16	0.66-1.07
A. mellifera, LC_{50} (240 h) n=30										
Study_181	Clothianidin[1]	0.070 (0.057-0.082)	Propiconazole	(low)	0.054	0.044	0.064	1.30		0.98-1.65
Study_182			Propiconazole	(high)	0.053	0.042	0.064	1.32		0.98-1.71
Study_183	Clothianidin[1]	0.074 (0.057-0.091)	Tau-fluvalinate	(low)	0.066	0.053	0.080	1.06		0.79-1.49
Study_184			Tau-fluvalinate	(high)	0.075	0.052	0.098	0.93		0.64-1.42
Study_185	Dimethoate[1]	0.624 (0.525-0.723)	Propiconazole	(low)	0.508	0.407	0.609	1.23		0.93-1.57
Study_186			Propiconazole	(high)	0.504	0.433	0.647	1.24		0.93-1.60

Study_ID	Chemical A		Binary mixture (A+B)							
	Name	EM_A (95% CI)	Chemical B	TU_B	EM_M	CI_1 (95 %)	CI_2 (95 %)	EMR (+)	EMR (-)	CI_{EMR}
B. terrestris, LC_{50} (96 h), n= 12										
Study_187	Clothianidin[1]	0.018 (0.013-0.023)	Propiconazole	(high)	0.013	0.009	0.016	1.38		0.81-1.97
Study_188			Tau-fluvalinate	(low)	0.018	0.014	0.021	1.00		0.67-1.36
Study_189			Tau-fluvalinate	(high)	0.012	0.007	0.017	1.50		0.84-2.44
B. terrestris, LC_{50} (240h), n= 12										
Study_190	Clothianidin[1]	0.016 (0.011-0.021)	Propiconazole	(high)	0.012	0.009	0.016	1.33		0.81-1.97
Study_191			Tau-fluvalinate	(low)	0.015	0.012	0.019	1.07		0.67-1.51
Study_192			Tau-fluvalinate	(high)	0.012	0.009	0.016	1.33		0.81- 1.97
O. bicornis ♂, LC_{50} (96 h), n= 5										
Study_193	Dimethoate[2]	0.600 (0.306-0.894)	Propiconazole	(low)	NA	NA	NA	NA		NA
Study_194			Propiconazole	(high)	NA	NA	NA	NA		NA
O. bicornis ♀, LC_{50} (96 h), n= 5										
Study_195	Dimethoate[2]	1.011 (0.563-1.460)	Propiconazole	(low)	0.678	0.419	0.938	1.49		0.70-2.51
Study_196			Propiconazole	(high)	NA	NA	NA	NA		NA
O. bicornis ♂, LC_{50} (240 h) n= 5										
Study_197	Dimethoate[2]	0.600	Propiconazole	(low)	0.435	0.255	0.615	1.38		0.59- 2.43

Study_198		(0.306-0.894)	Propiconazole	(high)	0.368	0.236	0.501	1.63		0.72- 2.76
O. bicornis ♂, LC_{50} (96 h), n= 5										
Study_199	Clothianidin[2]	0.172 (0.101-0.242)	Propiconazole	(low)	0.084	0.042	0.126	2.05		0.90- 3.78
Study_200			Propiconazole	(high)	0.050	0.025	0.076	3.44*		1.50- 6.4
O. bicornis ♀ 5 , LC_{50} (96 h), n= 5										
Study_201	Clothianidin[2]	0.046 (0.024-0.068	Propiconazole	(low)	0.042	0.025	0.059	1.10		0.48- 1.90
Study_202			Propiconazole	(high)	0.048	0.030	0.067	0.96		0.42- 1.64
O. bicornis ♂ LC_{50} (240 h), n= 5										
Study_203	Clothianidin[2]	0.058 (0.017-0.098)	Propiconazole	(low)	0.056	0.021	0.091	1.04		0.22- 2.36
Study_204			Propiconazole	(high)	0.025	0.007	0.044	2.32		0.39- 6.04
O. bicornis ♀, *LC_{50} (240 h), n= 5*										
Study_205	Clothianidin[2]	0.036 (0.021-0.050)	Propiconazole	(low)	0.031	0.016	0.046	1.16		0.52- 2.10
Study_206			Propiconazole	(high)	0.036	0.021	0.050	1.00		0.49- 1.68

Table S3.10. Studies on combined chronic toxicity effects (expressed as % of mortality) of imidacloprid with 6 different PPPs (acephate, λ-cyhalothrin, oxamyl, tetraconazole, glyphosate and sulfoxaflor) in honey bee (A. mellifera) following chronic oral exposure (one or two weeks of feeding on PPPs at residue concentrations) (Zhu et al. 2017). [1]= EMR is reported as Standardized Mortality Ratio (SMR= ratio of observed deaths in the binary mixture study group to expected deaths in the compound-only study). EM_A and EM_M values are expressed as % of mortality (+ related Standard Error). DoseB (= ratio between the field use concentration and feeding concentration expressed as mg/L). NA = Not Applicable.

Study_ID	Chemical A		Binary Mixture (A+B)						
	Name	EM_A (±SE)	Chemical B	TU_B (- fold)	EM_M	±SEM	SMR^1 (+)	CI_{EMR}	SMR^1 (-)
Mortality after 1 week of feeding									
Study_207	Imidacloprid	22.55 (±5.11)	Acephate	49036	18.35	6.02	0.81	0.18- 1.45	1.23
Study_208			λ-cyhalothrin	213	20.75	4.21	0.92	0.37- 1.47	1.09
Study_209			Oxamyl	57531	20.15	5.71	0.89	0.26- 1.53	1.12
Study_210			Tetraconazole	40119	18.65	3.91	0.83	0.33- 1.33	1.21
Study_211			Glyphosate	2675	23.16	6.62	1.03	0.28 - 1.67	NA
Study_212			Sulfoxaflor	235	36.69	9.62	1.63	0.20 – na	NA
Mortality after 2 weeks of feeding									
Study_213	Imidacloprid	36.51 (±5.58)	Acephate	49036	32.09	6.74	0.88	0.43 - 1.33	1.14
Study_214			λ-cyhalothrin	213	46.98	3.02	1.29	0.52 - na	NA
Study_215			Oxamyl	57531	40.23	7.21	1.10	0.49- 1.33	NA
Study_216			Tetraconazole	40119	41.63	3.72	1.14	0.57 - 1.18	NA
Study_217			Glyphosate	2675	45.81	8.14	1.25	0.43 - na	NA
Study_218			Sulfoxaflor	235	54.42	7.67	1.49	0.40 - na	NA

Table S3.11. Study on combined effects following acute oral exposure to binary mixtures (PPP + nutritional stress) in honey bees (Tosi et al., 2017). All the values are expressed in muA/min/mg of protein referring to the Glucose oxidase activity. EM (Estimated Mean/Median Lethal Times (LT) for the Chemical A; in parenthesis the 25% and 75% LT, respectively, and the sucrose intake mean ± SE), DoseB (nuritional stress dose is reported as

sucrose intake, w/w), EMMEMm (Estimated Mean/Median, 25%, and 75% Lethal Times for the binary mixture), Max EMR (maximum Estimated Mean Ratio between expected and observed mortality of binary mixture, and in parenthesis the time after exposure of maximum EMR). [1]= TU_B is expressed as sucrose consumed in the experiment (mg/bee/day); mean ± SE are reported when a range of consumptions were tested. EM_A values for compound A and EM_M values for binary mixture are expressed as Lethal Times (LT, 25%, 50%, and 75%). [2]= NA (Not Applicable) is stated when the respective LT was not reached because the treatment did not cause the mortality of the respective percentage of bees (either 25, 50, or 75%). [1]= EMR is reported as Standardized Mortality Ratio (SMR= ratio of observed deaths in the binary mixture study group to expected deaths in the compound-only study).

Study_ID	Treatment A		Binary Mixture (treatment A+B)						
	Name	EM_A	Treatment B (Nutritional stress)	TU_B[1]	EM_M	25%	75%	Max EMR (increase)	Type of interaction
Study_219	Clothianidin	96 (72, >96h)[2] (40±3)	*Ad libitum*, *poor* (15%) quality	7 ± 1	8	3	48	17 (2h)	Synergism
Study_220	Clothianidin	96 (72, >96h) (40 ± 3)	*Ad libitum*, *intermediate* (33%) quality	23 ± 1	48	24	NA[2]	12 (9-10h)	Synergism
Study_221	Clothianidin	96 (72, >96h)[2] (40 ± 3)	*Limited*, *poor* (15%) quality	2	3	2	5	6 (2h)	Synergism
Study_222	Clothianidin	96 (72, >96h)[2] (40 ± 3)	*Limited*, *rich* (50%) quality	6	4	4	5	36 (4h)	Synergism
Study_223	Thiamethoxam	NA (96, NA)[2] (43 ± 2)	*Ad libitum*, *poor* quality (15%)	9 ± 1	48	4	72	14 (7h)	Synergism
Study_224	Thiamethoxam	NA (96, NA)[2] (43 ± 2)	*Ad libitum*, *intermediate* (33%) quality	22 ± 1	72	48	96	3 (6-7h)	Synergism
Study_225	Thiamethoxam	NA (96, NA)[2] (43 ± 2)	*Limited*, *poor* (15%) quality	2	3	2	5	6 (3h)	Synergism

Study_226	Thiamethoxam	NA (96, NA)[2] (43 ± 2)	*Limited*, *rich* (50%) quality	6	4	4	5	10 (3h)	Synergism

Table S3.12. Study on combined toxicity effects following chronic oral exposure to binary mixtures (PPP + pathogen) in honey bees (Alaux et al., 2010). All the values are expressed in muA/min/mg of protein referring to the Glucose oxidase activity. *= TU_B is expressed as number of spores applied in the experiment.

Study_ID	**Treatment A**		**Binary mixture (treatment A+B)**					
	Name	EM_A	Treatment B (Pathogen)	TU_B*	EM_M	25th percentile	75th percentile	EMR (+)
Study_227	Imidacloprid	1218.75 (927.08-1677.08)	N. ceranae	200000	979.17	760.42	1260.42	1.24
Study_228	Imidacloprid	117.04 (60.84-161.29	N. ceranae	200000	97.84	49.18	195.11	1.19

Table S3.13. Summary tables for Estimated Mean Ratio (EMR) and Toxic Unit values for chemical B (TUB) resulting from acute contact toxicity studies (LD50, 24h). No. = number of case studies. EMR (+) = Estimated Mean Ratios describing increased toxicity. EMR (-) = Estimated Mean Ratio describing decreased toxicity. NA= Not available.

Substance	No.	Spp.	TU_B (values)	No.	EMR (+)	EMR (-)
PPP + PPP/synergist	122	*A.m. mellifera*	TU ≤ 0.10	63	1.50 - 1980	1.44 – 1.55
			0.11 ≤ TU ≤ 0.30	17	2.10 - 25.40	NA
			0.31 < TU ≤ 0.61	9	16.0 - 980	NA
			NA	33	NA	NA
Vet. Product + PPP	9	*A. m. primoski* *A. m. ligustica* *A. m. carnica*	≤ 0.10	3	1.40 – 2.90	NA
			0.11 ≤ TU ≤ 0.30	6	2.50 – 7.00	NA
			0.31 < TU ≤ 0.61	NA	NA	NA

Table S3.14. Summary table for Estimated Mean Ratio (EMR) and Toxic Unit values for chemical B (TUB) resulting from chronic oral toxicity studies in bee species (LD50, 96h 240h). No. = number of case studies. EMR (+) = Estimated Mean Ratios describing increased toxicity. EMR (-) = Estimated Mean Ratio describing decreased toxicity. Na= Not Applicable.

Substance	No.	TU_B (values)	Spp.	EMR (+)	EMR (-)
	Endpoint= mortality (LD_{50})				
PPPs	14	*NA*	*O. bicornis*	1.10 – 3.44	NA
	24	NA	*A. mellifera*	1.07 – 1.32	1.16
	6	NA	*B. terrestris*	1.07 – 1.50	NA
	Endpoint= mortality (%)				
	2	NA	*O. bicornis*	4 – 8.3	NA
PPP + nutritional stressor /pathogen	10	NA	*A. mellifera*	1.2 – 36	NA

Table S3.15. Results of the analyses for Model Deviation Ratio (MDR) and Estimated Mean Ratio (EMR) for acute contact toxicity studies in honey bees according to Toxic Unit (TU) approach and MIXTOX model (Jonker et al., 2005). Mode of action (MoA) for chemical A and B are reported according results (see chapter 3.1). A ranking of combined toxicity for binary mixtures of pesticides in honey bees (expressed as LD50 µg/bee) following acute contact exposure is provided according to refined MDR thresholds. Binary mixtures effects are classified as synergistic "SYN" (MDR> 1.25), antagonistic "ANG" (MDR< 0.83) or additive "ADD" (0.83≤MDR≤1.25) according to three different coulours. Statistical significance of the combined toxicity has been estimated using non-overlapping 95% confidence intervals (95% CI of the EMA vs 95% CI of the EMM for chemical A+B) as described in Johnson et al. (2012, 2013). [1]= Iwasa et al. (2004). [2]= Johnson et al. (2009). [3]= Johnson et al. (2013). [4]= Johnson et al. (2006). [5]= Biddinger et al. 2013. NA = Not Applicable.

Study_ID	Chemical A							Chemical B					Binary mixture (A + B)							
	Name	Class	MoA	EM_A	EM_A (CI_1 95 %)	EM_A (CI_2 95%)	TU_A	Name	Class	MoA	EM_B	TU_B	EM_M	EM_M (CI_1 95%)	EM_M (CI_2 95%)	TU_m	MDR	Effect	EMR	Stat. significant
Study_052	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	19.8	16.3	22.4	0.001	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.03	0.01	0.006	0.015	0.031	32.26	SYN	1980	YES
Study_025	Thiacloprid[1]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	14.6	9.53	25.4	0.01	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.03	0.09	0.041	0.21	0.04	25.00	SYN	162.22	YES
Study_117	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.02	Prochloraz	amide/conazole fungicide	Sterol biosynthesis in membranes	141.28	0.03	0.22	0.1	0.52	0.05	20.00	SYN	40.82	YES
Study_049	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	19.8	16.3	22.4	0.001	Prochloraz	amide/conazole fungicide	Sterol biosynthesis in membranes	141.28	0.07	0.01	0.006	0.017	0.07	14.29	SYN	1980	YES
Study_077	Fenpyroximate[3]	pyrazole	Mitochondrial complex I electron transport inhibitors	6.65	4	12	0.04	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.03	0.27	0.12	0.75	0.07	14.29	SYN	24.63	YES
Study_121	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.013	myclobutanil	conazole	Sterol biosynthesis in membranes	39.6	0.07	0.12	0.09	0.16	0.08	12.50	SYN	74.83	YES

Study_061	Coumaphos[3]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	31.2	22.2	49.6	0.014	Prochloraz	amide/conazole fungicide	Sterol biosynthesis in membranes	141.28	0.07	0.44	0.38	0.5	0.08	12.50	SYN	70.91	YES
Study_120	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.05	metconazole	conazole	Sterol biosynthesis in membranes	>100	0.03	0.45	0.36	0.56	0.08	12.50	SYN	19.96	YES
Study_027	Thiacloprid[11]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	14.6	9.53	25.4	0.002	propiconazole	triazoles	Sterol biosynthesis in membranes	>100	0.1	0.03	0.008	0.07	0.1	10.00	SYN	486.67	YES
Study_018	Acetamiprid[1]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	7.07	4.57	11.2	0.0099009	propiconazole	triazoles	Sterol biosynthesis in membranes	>100	0.1	0.07	0.023	0.2	0.11	9.09	SYN	101	YES
Study_074	Fenpyroximate[3]	pyrazole	Mitochondrial complex I electron transport inhibitors	6.65	4	12	0.04	Prochloraz	amide/conazole fungicide	Sterol biosynthesis in membranes	141.28	0.07	0.25	0.17	0.34	0.11	9.09	SYN	26.6	YES
Study_118	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.08	propiconazole	triazoles	Sterol biosynthesis in membranes	>100	0.03	0.74	0.62	0.88	0.11	9.09	SYN	12.14	YES
Study_020	Acetamiprid[1]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	7.07	4.57	11.2	0.071	Epoxiconazole	triazoles	Sterol biosynthesis in membranes	>100	0.1	0.5	0.156	1.66	0.17	5.88	SYN	14.14	YES

Study_ID	Chemical A							Chemical B					Binary mixture (A + B)							
	Name	Class	MoA	EM_A	EM_A (CI_1 95 %)	EM_A (CI_2 95%)	TU_A	Name	Class	MoA	EM_B	TU_B	EM_M	EM_M (CI_1 95%)	EM_M (CI_2 95%)	TU_m	MDR	Effect	EMR	Stat. significant
Study_041	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	19.8	16.3	22.4	0.121	Fenpyroximate	pyrazole	Mitochondrial complex I electron transport inhibitors	15.8	0.06	2.4	1.45	3.65	0.18	5.56	SYN	8.25	YES
Study_116	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.17	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.01	1.54	1.03	4.05	0.18	5.56	SYN	5.83	YES
Study_040	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	19.8	16.3	22.4	0.039	Coumaphos	organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.39	0.15	0.78	0.13	3.05	0.19	5.26	SYN	25.38	YES
Study_054	Coumaphos[3]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	31.2	22.2	49.6	0.132	Fenpyroximate	pyrazole	Mitochondrial complex I electron transport inhibitors	15.8	0.06	4.12	3.35	5.06	0.19	5.26	SYN	7.57	YES

Study_ID	Chemical A							Chemical B					Binary mixture (A + B)							
	Name	Class	MoA	EM_A	EM_A (CI_1 95 %)	EM_A (CI_2 95%)	TU_A	Name	Class	MoA	EM_B	TU_B	EM_M	EM_M (CI_1 95%)	EM_M (CI_2 95%)	TU_m	MDR	Effect	EMR	Stat. significant
Study_064	Coumaphos[3]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	31.2	22.2	49.6	0.16	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.03	5.04	3.3	7.01	0.19	5.26	SYN	6.19	YES
Study_014	Acetamiprid[1]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	7.07	4.57	11.2	0.17	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.03	1.17	0.342	3.79	0.2	5.00	SYN	6.04	YES
Study_042	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	19.8	16.3	22.4	0.19	Amitraz	Amidine	Octopamine receptor agonists	50	0.02	3.74	2.14	7.08	0.21	4.76	SYN	5.29	YES
Study_030	Coumaphos[2]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.29	14.88	29.44	0.25	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.03	5.04	3.34	7.01	0.28	3.57	SYN	4.03	YES
Study_053	Coumaphos[3]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	31.2	22.2	49.6	0.208	tau-fluvalinate	pyrethroid	Sodium channel modulator	12	0.08	6.5	4.98	8.57	0.29	3.45	SYN	4.8	YES
Study_067	Fenpyroximate[3]	pyrazole	Mitochondrial complex I electron transpor	6.65	4	12	0.27	Amitraz	Amidine	Octopamine receptor agonists	50	0.02	1.8	1.6	2.04	0.29	3.45	SYN	3.69	YES

			t inhibitors																	
Study_055	Coumaphos[3]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	31.2	22.2	49.6	0.29	Amitraz	Amidine	Octopamine receptor agonists	50	0.02	9.2	1.12	25.1	0.31	3.23	SYN	3.39	NO
Study_046	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	19.8	16.3	22.4	0.224	pyraclostrobin	strobilurin, pyrazole, phenylpyrazole	Inhibit mitochondrial respiration, blocking the cytochrome bc1 complex	>100	0.1	4.43	0.67	61.4	0.32	3.13	SYN	4.47	NO
Study_130	Tau-fluvalinate[4]	Pyrethroids/Pyrethrins	Sodium channel modulators	9.45	7.48	12	0.001	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.34	0.01	0.007	0.015	0.34	2.94	SYN	945	YES
Study_127	λ-cyhalothrin[4]	Pyrethroids/Pyrethrins	Sodium channel modulators	0.102	0.073	0.133	0.013	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.34	0.0013	0.0011	0.0015	0.35	2.86	SYN	78.46	YES
Study_124	Cyfluthrin[4]	Pyrethroids/Pyrethrins	Sodium channel modulators	0.062	0.057	0.075	0.032	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.34	0.002	0.0015	0.0042	0.37	2.70	SYN	31	YES

Study_ID	Chemical A							Chemical B					Binary mixture (A + B)							
	Name	Class	MoA	EM_A	EM_A (CI_1 95 %)	EM_A (CI_2 95%)	TU_A	Name	Class	MoA	EM_B	TU_B	EM_M	EM_M (CI_1 95%)	EM_M (CI_2 95%)	TU_m	MDR	Effect	EMR	Stat. significant
Study_091	Thymol[3]	monoterpene phenol	na	55.1	42.1	70	0.292	tau-fluvalinate	pyrethroid	Sodium channel modulator	12	0.08	16.1	11.2	21.4	0.37	2.70	SYN	3.42	YES
Study_038	Tau-fluvalinate[2]	Pyrethroids/Pyrethrins	Sodium channel modulators	6.75	6.24	7.33	0.227	Coumaphos	organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.39	0.15	1.53	0.72	2.6	0.38	2.63	SYN	4.41	YES
Study_033	Coumaphos[2]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.29	14.88	29.44	0.298	Tau-fluvalinate	pyrethroid	Sodium channel modulator	12	0.08	6.05	4.29	8.68	0.38	2.63	SYN	3.35	YES
Study_012	Acetamiprid[5]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	64.6	38.1	252	0.221	Indar (fenbuconazole)	triazoles	Sterol biosynthesis in membranes	5.5	0.18	14.3	8.5	30.8	0.4	2.50	SYN	4.52	YES
Study_071	Fenpyroximate[3]	pyrazole	Mitochondrial complex I electron transport inhibitors	6.65	4	12	0.314	pyraclostrobin	strobilurin, pyrazole, phenylpyrazole	Inhibit mitochondrial respiration, blocking the cytochrome bc1 complex	>100	0.1	2.09	0.5	4.24	0.41	2.44	SYN	3.18	NO
Study_115	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.42	myclobutanil	conazole	Sterol biosynthesis in membranes	39.6	0.01	3.75	3.06	4.71	0.43	2.33	SYN	2.39	YES

Study_037	Tau-fluvalinate[2]	Pyrethroids/Pyrethrins	Sodium channel modulators	6.75	6.24	7.33	0.4	Coumaphos	organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.39	0.05	2.68	2.25	3.28	0.45	2.22	SYN	2.52	YES
Study_066	Fenpyroximate[3]	pyrazole	Mitochondrial complex I electron transport inhibitors	6.65	4	12	0.305	Coumaphos	organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.39	0.15	2.03	1.3	4.46	0.46	2.17	SYN	3.28	NO
Study_026	Thiaclorid1	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	14.6	9.53	25.4	0.001	triflumizole	imidazole	Sterol biosynthesis in membranes	20	0.5	0.01	0.003	0.04	0.5	2.00	SYN	1460	YES
Study_017	Acetamiprid[1]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	7.07	4.57	11.2	0.0042432 8	triflumizole	imidazole	Sterol biosynthesis in membranes	20	0.5	0.03	0.008	0.1	0.5	2.00	SYN	235.67	YES
Study_036	Tau-fluvalinate[2]	Pyrethroids/Pyrethrins	Sodium channel modulators	6.75	6.24	7.33	0.49	Coumaphos	organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.39	0.01	3.29	2.15	4.87	0.5	2.00	SYN	2.05	YES
Study_039	Tau-fluvalinate[2]	Pyrethroids/Pyrethrins	Sodium channel modulators	6.75	6.24	7.33	0.031	Coumaphos	organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.39	0.49	0.21	0.15	0.29	0.52	1.92	SYN	32.14	YES
Study_092	Thymol[3]	monoterpene phenol	na	55.1	42.1	70	0.37	Coumaphos	organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.39	0.15	20.4	10.4	38	0.52	1.92	SYN	2.7	YES

Study_ID	Chemical A							Chemical B					Binary mixture (A + B)							
	Name	Class	MoA	EM_A	EM_A (CI_1 95 %)	EM_A (CI_2 95%)	TU_A	Name	Class	MoA	EM_B	TU_B	EM_M	EM_M (CI_1 95%)	EM_M (CI_2 95%)	TU_m	MDR	Effect	EMR	Stat. significant
Study_111	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.52	Prochloraz	amide/conazole fungicide	Sterol biosynthesis in membranes	141.28	0.003	4.64	3.19	7.15	0.52	1.92	SYN	1.94	YES
Study_048	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	19.8	16.3	22.4	0.366	Chlorothalonil	Chloronitrile, aromatic fungicide	Multi-site activity	>63	0.16	7.24	3.96	12.9	0.53	1.89	SYN	2.73	YES
Study_034	Coumaphos[2]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.29	14.88	29.44	0.299	Tau-fluvalinate	pyrethroid	Sodium channel modulator	12	0.25	6.06	5.15	6.94	0.55	1.82	SYN	3.35	YES
Study_084	Amitraz[3]	Amidine	Octopamine receptor agonists	3.66	2.26	5.56	0.448	pyraclostrobin	strobilurin, pyrazole, phenylpyrazole	Inhibit mitochondrial respiration, blocking the cytochrome bc1 complex	>100	0.1	1.64	0.9	2.51	0.55	1.82	SYN	2.23	NO
Study_114	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.58	metconazole	conazole	Sterol biosynthesis in membranes	>100	0.003	5.21	4.21	6.38	0.58	1.72	SYN	1.72	YES
Study_045	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	19.8	16.3	22.4	0.301	pyraclostrobin+boscalid	strobilurin, pyrazole, phenylpyrazole + anilide	Inhibit mitochondrial respiration, blocking the cytochr	>100/>200	0.3	5.95	4.48	8.09	0.6	1.67	SYN	3.33	YES

										ome bc1 complex										
Study_097	Thymol[3]	monoterpene phenol	NA	55.1	42.1	70	0.512	pyraclostrobin	strobilurin, pyrazole, phenylpyrazole	Inhibit mitochondrial respiration, blocking the cytochrome bc1 complex	>100	0.1	28.2	4.96	57.9	0.61	1.64	SYN	1.95	NO
Study_022	Imidacloprid[1]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	0.0179	0.0092	0.032	0.59	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.03	0.0105	0.006	0.02	0.62	1.61	SYN	1.7	NO
Study_103	Thymol[3]	monoterpene phenol	na	55.1	42.1	70	0.59	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.03	32.4	19.8	49.6	0.62	1.61	SYN	1.7	NO
Study_021	Acetamiprid[1]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	7.07	4.57	11.2	0.158	uniconazole-P	triazoles	Sterol biosynthesis in membranes	>20	0.5	1.12	0.27	4.96	0.66	1.52	SYN	6.31	NO
Study_032	Coumaphos[2]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.29	14.88	29.44	0.63	Tau-fluvalinate	pyrethroid	Sodium channel modulator	12	0.03	12.77	10.63	15.59	0.66	1.52	SYN	1.59	NO

Study_ID	Chemical A							Chemical B					Binary mixture (A + B)							
	Name	Class	MoA	EM_A	EM_A (CI_1 95 %)	EM_A (CI_2 95%)	TU_A	Name	Class	MoA	EM_B	TU_B	EM_M	EM_M (CI_1 95%)	EM_M (CI_2 95%)	TU_m	MDR	Effect	EMR	Stat. significant
Study_119	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.061	fenbuconazole	conazole	Sterol biosynthesis in membranes	5.5	0.61	0.55	0.46	0.66	0.67	1.49	SYN	16.33	YES
Study_112	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.66	propiconazole	triazoles	Sterol biosynthesis in membranes	>100	0.003	5.97	4.67	7.71	0.67	1.49	SYN	1.5	YES
Study_060	Coumaphos[3]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	31.2	22.2	49.6	0.532	Chlorothalonil	Chloronitrile, aromatic fungicide	Multi-site activity	>63	0.16	16.6	6.77	85.6	0.69	1.45	SYN	1.88	NO
Study_047	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	19.8	16.3	22.4	0.586	boscalid	anilide	Succinate DeHydrogenase Inhibitor	>200	0.1	11.6	7.43	19.9	0.69	1.45	SYN	1.71	NO
Study_093	Thymol[3]	monoterpene phenol	na	55.1	42.1	70	0.633	Fenpyroximate	pyrazole	Mitochondrial complex I electron transport inhibitors	15.8	0.06	34.9	23.9	47.9	0.69	1.45	SYN	1.58	NO
Study_090	Amitraz[3]	Amidine	Octopamine receptor agonists	3.66	2.26	5.56	0.66	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.03	2.41	0.92	6.35	0.69	1.45	SYN	1.52	NO

Study_043	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	19.8	16.3	22.4	0.515	thymol	monoterpene phenol	na	55	0.18	10.2	7.85	14	0.7	1.43	SYN	1.94	YES
Study_099	Thymol[3]	monoterpene phenol	NA	55.1	42.1	70	0.541	Chlorothalonil	Chloronitrile, aromatic fungicide	Multi-site activity	>63	0.16	29.8	21.1	39.9	0.7	1.43	SYN	1.85	YES
Study_068	Fenpyroximate[3]	pyrazole	Mitochondrial complex I electron transport inhibitors	6.65	4	12	0.555	thymol	monoterpene phenol	na	55	0.18	3.69	2.8	4.98	0.74	1.35	SYN	1.8	NO
Study_087	Amitraz[3]	Amidine	Octopamine receptor agonists	3.66	2.26	5.56	0.678	Prochloraz	amide/conazole fungicide	Sterol biosynthesis in membranes	141.28	0.07	2.48	1.45	3.74	0.75	1.33	SYN	1.48	NO
Study_113	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.692	fenbuconazole	conazole	Sterol biosynthesis in membranes	5.5	0.06	6.21	4.94	7.8	0.75	1.33	SYN	1.45	YES
Study_024	Imidacloprid[1]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	0.0179	0.0092	0.032	0.67	propiconazole	triazoles	Sterol biosynthesis in membranes	>100	0.1	0.012	0.004	0.03	0.77	1.30	SYN	1.49	NO

Study_ID	Chemical A							Chemical B					Binary mixture (A + B)							
	Name	Class	MoA	EM_A	EM_A (CI_1 95 %)	EM_A (CI_2 95%)	TU_A	Name	Class	MoA	EM_B	TU_B	EM_M	EM_M (CI_1 95%)	EM_M (CI_2 95%)	TU_m	MDR	Effect	EMR	Stat. significant
Study_070	Fenpyroximate[3]	pyrazole	Mitochondrial complex I electron transport inhibitors	6.65	4	12	0.475	pyraclostrobin+boscalid	strobilurin, pyrazole, phenylpyrazole + anilide	Inhibit mitochondrial respiration, blocking the cytochrome bc1 complex	>100/>200	0.3	3.16	2.6	3.92	0.78	1.28	SYN	2.1	YES
Study_100	Thymol[3]	monoterpene phenol	na	55.1	42.1	70	0.708	Prochloraz	amide/conazole fungicide	Sterol biosynthesis in membranes	141.28	0.07	39	33.2	45.1	0.78	1.28	SYN	1.41	NO
Study_094	Thymol[3]	monoterpene phenol	na	55.1	42.1	70	0.78	Amitraz	Amidine	Octopamine receptor agonists	50	0.02	43.2	25.3	61	0.8	1.25	ADD	1.28	NO
Study_059	Coumaphos[3]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	31.2	22.2	49.6	0.724	boscalid	anilide	Succinate DeHydrogenase Inhibitor	>200	0.1	22.6	15.3	32.4	0.82	1.22	ADD	1.38	NO
Study_096	Thymol[3]	monoterpene phenol	NA	55.1	42.1	70	0.579	pyraclostrobin+boscalid	strobilurin, pyrazole, phenylpyrazole + anilide	Inhibit mitochondrial respiration, blocking the cytochrome bc1 complex	>100/>200	0.3	31.9	16.9	44.7	0.88	1.14	ADD	1.73	NO

Study_079	Amitraz[3]	Amidine	Octopamine receptor agonists	3.66	2.26	5.56	0.746	Coumaphos	organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.39	0.15	2.73	1.82	3.73	0.9	1.11	ADD	1.34	NO
Study_065	Fenpyroximate[3]	pyrazole	Mitochondrial complex I electron transport inhibitors	6.65	4	12	0.833	tau-fluvalinate	pyrethroid	Sodium channel modulator	12	0.08	5.54	3.1	12.8	0.91	1.10	ADD	1.2	NO
Study_035	Tau-fluvalinate[2]	Pyrethroids/Pyrethrins	Sodium channel modulators	6.75	6.24	7.33	0.91	Coumaphos	organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.39	0.005	6.14	5.03	7.71	0.91	1.10	ADD	1.1	NO
Study_056	Coumaphos[3]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	31.2	22.2	49.6	0.74	thymol	monoterpene phenol	na	55	0.18	23.1	14.7	34.5	0.92	1.09	ADD	1.35	NO
Study_058	Coumaphos[3]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	31.2	22.2	49.6	0.83	pyraclostrobin+boscalid	strobilurin, pyrazole, phenylpyrazole + anilide	Inhibit mitochondrial respiration, blocking the cytochrome bc1 complex	>100/>200	0.1	25.9	19.9	34.6	0.93	1.08	ADD	1.2	NO
Study_072	Fenpyroximate[3]	pyrazole	Mitochondrial complex I electron transport inhibitors	6.65	4	12	0.848	boscalid	anilide	Succinate DeHydrogenase Inhibitor	>200	0.1	5.64	2.9	17.2	0.95	1.05	ADD	1.18	NO

Study_ID	Chemical A							Chemical B					Binary mixture (A + B)							
	Name	Class	MoA	EM_A	EM_A (CI_1 95 %)	EM_A (CI_2 95%)	TU_A	Name	Class	MoA	EM_B	TU_B	EM_M	EM_M (CI_1 95%)	EM_M (CI_2 95%)	TU_m	MDR	Effect	EMR	Stat. significant
Study_098	Thymol[3]	monoterpene phenol	na	55.1	42.1	70	0.855	boscalid	anilide	Succinate DeHydrogenase Inhibitor	>200	0.1	47.1	35.4	62.1	0.95	1.05	ADD	1.17	NO
Study_110	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	0.94	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.001	8.48	6.42	11.4	0.95	1.05	ADD	1.06	NO
Study_023	Imidacloprid[1]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	0.0179	0.0092	0.032	0.542	triflumizole	imidazole	Sterol biosynthesis in membranes	20	0.5	0.0097	0.005	0.02	1.04	0.96	ADD	1.85	NO
Study_031	Coumaphos[2]	Organothiophosphate, phosphorothioates	Acetylcholinesterase (AChE) inhibitor.	20.29	14.88	29.44	1.03	Tau-fluvalinate	pyrethroid	Sodium channel modulator	12	0.01	20.97	12.9	47.86	1.04	0.96	ADD	0.97	NO
Study_086	Amitraz[3]	Amidine	Octopamine receptor agonists	3.66	2.26	5.56	0.913	Chlorothalonil	Chloronitrile, aromatic fungicide	Multi-site activity	>63	0.16	3.34	1.48	8.89	1.07	0.93	ADD	1.1	NO
Study_073	Fenpyroximate[3]	pyrazole	Mitochondrial complex I electron transport	6.65	4	12	0.964	Chlorothalonil	Chloronitrile, aromatic fungicide	Multi-site activity	>63	0.16	6.41	5.6	7.36	1.12	0.89	ADD	1.04	NO

			inhibitors																	
Study_107	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	1.17	fenbuconazole	conazole	Sterol biosynthesis in membranes	5.5	0.006	10.5	8.59	13.4	1.18	0.85	ADD	0.86	NO
Study_104	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	1.22	Piperonyl butoxide	synergist-cyclic aromatic	P450-dependent monooxygenase inhibitor	294	0.0001	11	6.97	23.9	1.23	0.81	ANG	0.82	NO
Study_105	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	1.24	Prochloraz	amide/conazole fungicide	Sterol biosynthesis in membranes	141.28	0.0003	11.1	9.94	12.4	1.24	0.81	ANG	0.81	NO
Study_081	Amitraz[3]	Amidine	Octopamine receptor agonists	3.66	2.26	5.56	1.068	thymol	monoterpene phenol	na	55	0.18	3.91	2.98	5.52	1.25	0.80	ANG	0.94	NO
Study_080	Amitraz[3]	Amidine	Octopamine receptor agonists	3.66	2.26	5.56	1.249	Fenpyroximate	pyrazole	Mitochondrial complex I electron transport inhibitors	15.8	0.06	4.57	2.78	6.48	1.31	0.76	ANG	0.8	NO
Study_083	Amitraz[3]	Amidine	Octopamine receptor agonists	3.66	2.26	5.56	1.104	pyraclostrobin+boscalid	strobilurin, pyrazole, phenylpyrazole + anilide	Inhibit mitochondrial respiration, blocking the cytochrome bc1 complex	>100/>200	0.3	4.04	2.25	10.4	1.4	0.71	ANG	0.91	NO

Study_ID	Chemical A							Chemical B					Binary mixture (A + B)							
	Name	Class	MoA	EM_A	EM_A (CI_1 95 %)	EM_A (CI_2 95%)	TU_A	Name	Class	MoA	EM_B	TU_B	EM_M	EM_M (CI_1 95%)	EM_M (CI_2 95%)	TU_m	MDR	Effect	EMR	Stat. significant
Study_078	Amitraz[3]	Amidine	Octopamine receptor agonists	3.66	2.26	5.56	1.331	tau-fluvalinate	pyrethroid	Sodium channel modulator	12	0.08	4.87	2.38	8.31	1.41	0.71	ANG	0.75	NO
Study_108	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	1.41	metconazole	conazole	Sterol biosynthesis in membranes	>100	0.0003	12.7	8.55	22.5	1.41	0.71	ANG	0.71	NO
Study_085	Amitraz[3]	Amidine	Octopamine receptor agonists	3.66	2.26	5.56	1.317	boscalid	anilide	Succinate DeHydrogenase Inhibitor	>200	0.1	4.82	2.83	6.74	1.42	0.70	ANG	0.76	NO
Study_109	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	1.44	myclobutanil	conazole	Sterol biosynthesis in membranes	39.6	0.001	12.9	10.7	16	1.44	0.69	ANG	0.7	YES
Study_106	Tau-fluvalinate[3]	Pyrethroids/Pyrethrins	Sodium channel modulators	8.98	7.85	10.4	1.55	propiconazole	conazole	Sterol biosynthesis in membranes	>100	0.0003	13.9	11.9	16.6	1.55	0.65	ANG	0.65	YES
Study_013	Imidacloprid[5]	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) agonists	0.2	0.1	0.3	1.5	Indar (fenbuconazole)	triazoles	Sterol biosynthesis in membranes	5.5	0.18	0.3	0.1	0.4	1.68	0.60	ANG	0.67	NO

Chapter 4

Appendix 4A. Supplementary data. Source od the data (i.e. database) are reported as follows: D= Demetra, E= EFSA, TB= OECD QSAR ToolBox. Chemical identifier (ID), training set (TR) and TS (test set) are reported. Available at http://dx.doi.org/10.1016/j.chemosphere.2016.09.092

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
14	TR	CC1(C)C(C=C(Br)Br)C1C(=O)OC(C#N)c1cccc(Oc2ccccc2)c1	D	52918-63-5	0,0015
43	TR	CCC1CCCC(OC2CCC(C(C)O2)N(C)C)C(C)C(=O)C2=CC3C4CC(CC4C=CC3C2CC(=O)O1)OC1OC(C)C(OC)C(OC)C1OC	D	131929-60-7	0,0029
122	TR	CCOc1ccc(cc1)C(C)(C)COCc1cccc(Oc2ccccc2)c1	TB	80844-07-1	0,0145
48	TR	CC1(C)C(C=C(Cl)Cl)C1C(=O)OC(C#N)c1cccc(Oc2ccccc2)c1	D	66841-24-5	0,023
54	TR	CC(C)=CC1C(C(=O)OC2CC(=O)C(CC#C)=C2C)C1(C)C	D	23031-36-9	0,028
4	TR	CC1(C)C(C=C(Cl)Cl)C1C(=O)OC(C#N)c1ccc(F)c(Oc2ccccc2)c1	D	68359-37-5	0,037
50	TR	CC1(C)C(C=C(Cl)C(F)(F)F)C1C(=O)OC(C#N)c1cccc(Oc2ccccc2)c1	D	91465-08-6	0,038
87	TR	CCOP(=S)(NC(C)C)Oc1ccccc1C(=O)OC(C)C	TB	25311-71-1	0,049
69	TR	COC(=O)C=C(C)OP(=O)(OC)OC	D	7786-34-7	0,07
33	TR	[O-][N+](=O)N=C1NCCN1Cc1ccc(Cl)nc1	D	105827-78-9	0,078
24	TR	CCOC(=O)C(C)Oc1ccc(Oc2nc3ccc(Cl)cc3o2)cc1	D	66441-23-4	0,1
51	TR	[O-][N+](=O)c1c(Cl)c(Cl)c(Cl)c(Cl)c1Cl	D	82-68-8	0,1
49	TR	CCOP(=S)(OCC)Oc1nc(Cl)c(Cl)cc1Cl	TB	2921-88-2	0,114
191	TR	CN(C)S(=O)(=O)n1c(nc(Cl)c1-c1ccc(C)cc1)C#N	TB	120116-88-3	0,118
165	TR	CS(=O)(=O)c1cc(ccc1C(=O)c1cnoc1C1CC1)C(F)(F)F	TB	141112-29-0	0,12
45	TR	CC1(C)C(C(Br)C(Br)(Br)Br)C1C(=O)OC(C#N)c1cccc(Oc2ccccc2)c1	D	66841-25-6	0,129
8	TR	CNC(=O)Oc1cccc2CC(C)(C)Oc12	D	1563-66-2	0,16
17	TR	CNC(=O)CSP(=S)(OC)OC	D	60-51-5	0,16
142	TR	CNC(=O)ON=C(C)SC	E + D	16752-77-5	0,16
35	TR	CCOP(=S)(OCC)Oc1ccc(cc1)[N+]([O-])=O	D	56-38-2	0,175

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
139	TR	CCOC(=O)CC(SP(=S)(OC)OC)C(=O)OCC	E + D	121-75-5	0,18
190	TR	OC(=O)C1(CC1)C(=O)Nc1ccc(Cl)cc1Cl	TB	113136-77-9	0,18
59	TR	CCOP(=S)(OCC)Oc1cc(C)nc(n1)C(C)C	D	333-41-5	0,2
41	TR	COC1=NN(CSP(=S)(OC)OC)C(=O)S1	TB	950-37-8	0,236
1	TR	CNC(=O)ON=CC(C)(C)SC	D	116-06-3	0,285
141	TR	CNC(=O)Oc1cc(C)c(SC)c(C)c1	E + D	2032-65-7	0,3025
118	TR	CCOC(=O)C(SP(=S)(OC)OC)c1ccccc1	TB	254642	0,306
25	TR	COP(=S)(OC)Oc1ccc(SC)c(C)c1	D	55-38-9	0,308
55	TR	COP(=S)(OC)Oc1ccc(c(C)c1)[N+]([O-])=O	TB	122-14-5	0,383
53	TR	CCN(CC)c1nc(C)cc(OP(=S)(OC)OC)n1	D	29232-93-7	0,39
56	TR	COP(=S)(OC)SCN1N=Nc2ccccc2C1=O	D	86-50-0	0,42
83	TR	CNC(=O)Oc1cccc2OC(C)(C)Oc12	TB	22781-23-3	0,428
15	TR	COP(=O)(OC)OC=C(Cl)Cl	D	62-73-7	0,5
29	TR	ClC1C(Cl)C(Cl)C(Cl)C(Cl)C1Cl	D	58-89-9	0,56
140	TR	COP(=S)(OC)SCN1C(=O)c2ccccc2C1=O	E + D	732-11-6	0,64
57	TR	CCOP(=S)(OCC)SCCSCC	TB	298-04-4	0,96
125	TR	CC(C)C(Nc1ccc(cc1Cl)C(F)(F)F)C(=O)OC(C#N)c1cccc(Oc2ccccc2)c1	TB	69409-94-5	1,1
7	TR	CNC(=O)Oc1cccc2ccccc12	D	63-25-2	1,3
54	TR	CNC(=O)Oc1ccccc1OC(C)C	TB	114-26-1	1,35
65	TR	COP(=O)(OC)OC(=CCl)c1cc(Cl)c(Cl)cc1Cl	D	22248-79-9	1,37

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
66	TR	COP(N)(=O)SC	D	10265-92-6	1,37
82	TR	CCOP(=O)(NC(C)C)Oc1ccc(SC)c(C)c1	TB	22224-92-6	1,87
21	TR	ClC1=C(Cl)C2(Cl)C3C4CC(C5OC45)C3C1(Cl)C2(Cl)Cl	D	72-20-8	2,02
215	TR	CCOP(=O)(OCC)SCCSCC	TB	8065-48-3	2,6
158	TR	CCOC(=O)C(Cl)Cc1cc(N2N=C(C)N(C(F)F)C2=O)c(F)cc1Cl	TB	128639-02-1	2,66
70	TR	CC(C)=CC1C(C(=O)OC2CC(=O)C(CC=C)=C2C)C1(C)C	D	28434-00-6	3,4
178	TR	COc1cc(OC)nc(Oc2cccc(Oc3nc(OC)cc(OC)n3)c2C(O)=O)n1	TB	125401-92-5	3,94
79	TR	CCCSP(=O)(OCC)SCCC	TB	13194-48-4	4,09
61	TR	ClC1=C(Cl)C2(Cl)C3COS(=O)OCC3C1(Cl)C2(Cl)Cl	D	115-29-7	4,5
132	TR	COC(=O)c1ccccc1N	TB	134-20-3	7,8
66	TR	CCOP(=S)(CC)Sc1ccccc1	TB	944-22-9	8,68
42	TR	CCCC(=NOCC)C1=C(O)CC(CC(C)SCC)CC1=O	D	74051-80-2	10
39	TR	CCOP(=S)(OCC)SCSCC	D	298-02-2	10,07
75	TR	SC(=S)NCCNC(S)=S	D	142-59-6	12,09
20	TR	CCCCCCCCCCCCNC(N)=N	D	2439-10-3	12,1
27	TR	ClC(Cl)(Cl)SN1C(=O)c2ccccc2C1=O	D	133-07-3	12,1
11	TR	CCOC(=O)c1ccccc1S(=O)(=O)NC(=O)Nc1nc(Cl)cc(OC)n1	D	90982-32-4	12,5
14	TR	COC(=O)c1sccc1S(=O)(=O)NC(=O)Nc1nc(C)nc(OC)n1	TB	79277-27-3	12,5
141	TR	Nc1cc(Cl)nc(C(O)=O)c1Cl	TB	150114-71-9	12,5
6	TR	Oc1c(Br)cc(cc1Br)C#N	D	1689-84-5	14,5

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
47	TR	CCCCN(CC)c1c(cc(cc1[N+]([O-])=O)C(F)(F)F)[N+]([O-])=O	D	1861-40-1	14,5
55	TR	OC(=O)CCCOc1ccc(Cl)cc1Cl	D	94-82-6	14,5
64	TR	Nc1c(Cl)c(Cl)nc(C(O)=O)c1Cl	D	1918-02-1	14,5
52	TR	CN(C)C(=O)Oc1nc(nc(C)c1C)N(C)C	D	23103-98-2	18,72
23	TR	CCOP(=S)(OCC)SCSP(=S)(OCC)OCC	D	563-12-2	20,55
80	TR	COc1ccc(cc1)C(c1ccc(OC)cc1)C(Cl)(Cl)Cl	D	72-43-5	23,57
3	TR	CC(C)OP(=S)(OC(C)C)SCCNS(=O)(=O)c1ccccc1	D	741-58-2	24
48	TR	CCCN(CCC)c1c(cc(cc1[N+]([O-])=O)C(F)(F)F)[N+]([O-])=O	TB	1582-09-8	24,2
148	TR	Fc1cc2OCC(=O)N(CC#C)c2cc1N1C(=O)C2=C(CCCC2)C1=O	TB	103361-09-7	28
68	TR	CCSC(=O)N(CC(C)C)CC(C)C	D	2008-41-5	29
143	TR	CC(C)(C)c1ccc(OC2CCCCC2OS(=O)OCC#C)cc1	E + D	2312-35-8	31,46
12	TR	O[Sn](C1CCCCC1)(C1CCCCC1)C1CCCCC1	D	13121-70-5	35,9
168	TR	CCC(C)(NC(=O)c1cc(Cl)c(C)c(Cl)c1)C(=O)CCl	TB	156052-68-5	35,9
40	TR	COc1nc(NC(C)C)nc(NC(C)C)n1	D	1610-18-0	36
180	TR	COc1cccc(C(=O)NN(C(=O)c2cc(C)cc(C)c2)C(C)(C)C)c1C	TB	161050-58-4	37,8
37	TR	CCC(CC)Nc1c(cc(C)c(C)c1[N+]([O-])=O)[N+]([O-])=O	D	40487-42-1	49,8
139	TR	CCCCOC(=O)C(C)Oc1ccc(Oc2ccc(cc2F)C#N)cc1	TB	122008-85-9	51
46	TR	CCCCCCCCSCCO	D	3547-33-9	56,9
67	TR	COP(=O)(OC)C(O)C(Cl)(Cl)Cl	D	52-68-6	59,8
7	TR	CSC1=NN=C(C(=O)N1N)C(C)(C)C	TB	21087-64-9	60,4

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
10	TR	Cc1ccc2nc3SC(=O)Sc3nc2c1	D	2439-01-2	66,47
128	TR	CNC(S)=S	TB	137-42-8	66,5
30	TR	CC1(C)CNC(NC1)=NN=C(C=Cc1ccc(cc1)C(F)(F)F)C=Cc1ccc(cc1)C(F)(F)F	D	67485-29-4	67
44	TR	CN(C)C(=S)SSC(=S)N(C)C	D	137-26-8	74
18	TR	CSC(=O)c1c(nc(c(C(=O)SC)c1CC(C)C)C(F)(F)F)C(F)F	D	97886-45-8	81
138	TR	COCC(C)N(C(=O)CCl)c1c(C)csc1C	E + D	87674-68-8	94
147	TR	COC(=O)c1c(Cl)nn(C)c1S(=O)(=O)NC(=O)Nc1nc(OC)cc(OC)n1	TB	100784-20-1	94
2	TR	CCNc1nc(NC(C)C)nc(SC)n1	D	834-12-8	100
71	TR	C1C[n+]2ccccc2-c2cccc[n+]12	D	2764-72-9	100
79	TR	COC(=O)c1ccccc1S(=O)(=O)NC(=O)N(C)c1nc(C)nc(OC)n1	D	101200-48-0	100
52	TR	C1CN2ccccc2c2ccccN12	TB	85-00-7	100
74	TR	CN(C)C(=O)Nc1ccc(Cl)cc1	D	150-68-5	110
32	TR	OC(=O)c1ccccc1C(=O)Nc1cccc2ccccc12	D	132-66-1	113,2
26	TR	O[Sn](c1ccccc1)(c1ccccc1)c1ccccc1	D	76-87-9	114,8
19	TR	CN(C)C(=O)Nc1ccc(Cl)c(Cl)c1	D	330-54-1	145,03
57	TR	Clc1c(Cl)c(C#N)c(Cl)c(C#N)c1Cl	D	1897-45-6	181,29
5	TR	CCC(C)N1C(=O)NC(C)=C(Br)C1=O	D	314-40-9	193,38
58	TR	CCNc1nc(Cl)nc(NC(C)(C)C#N)n1	D	21725-46-2	193,38
154	TR	COCC(=O)N(C(C)C(=O)OC)c1c(C)cccc1C	TB	70630-17-0	198
38	TR	COC(=O)Nc1cccc(OC(=O)Nc2cccc(C)c2)c1	D	13684-63-4	241,72

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
62	TR	CP(O)(=O)CCC(N)C(O)=O	D	51276-47-2	345,5
140	TR	CN(Cc1ccc(Cl)nc1)C(C)=NC#N	TB	135410-20-7	346
10	TR	CCOCN(C(=O)CCl)c1c(C)cccc1CC	TB	34256-82-1	1720
206	TR	COC(=O)c1ccc(CNS(C)(=O)=O)cc1S(=O)(=O)NC(=O)Nc1nc(OC)cc(OC)n1	TB	208465-21-8	3980
217	TR	CC(C)(C[Sn](CC(C)(C)c1ccccc1)(CC(C)(C)c1ccccc1)O[Sn](CC(C)(C)c1ccccc1)(CC(C)(C)c1ccccc1)CC(C)(C)c1ccccc1)c1ccccc1	TB	13356-08-6	3980
83	TR	COP(=S)(OC)Oc1c(Cl)cc(C)cc1Cl	E	57018-04-9	> 100
85	TR	COc1cc(OC)nc(NC(=O)NS(=O)(=O)N(C)S(C)(=O)=O)n1	E	120923-37-7	> 100
89	TR	Nc1c(Cl)c(Oc2ccccc2)ccc1[N+]([O-])=O	E	74070-46-5	> 100
94	TR	COc1nc(C)nc(NC(=O)NS(=O)(=O)c2ccccc2Cl)n1	E	64902-72-3	> 100
95	TR	CC(C)Oc1cc(N2N=C(OC2=O)C(C)(C)C)c(Cl)cc1Cl	E	19666-30-9	> 100
96	TR	Cc1cnc2c(C(O)=O)c(Cl)ccc2c1	E	90717-03-6	> 100
98	TR	CCCC(=NOCC)C1=C(O)CC(CC1=O)C1CCCSC1	E	101205-02-1	> 100
99	TR	COC(=O)C(C)Oc1ccc(Oc2ccc(Cl)cc2Cl)cc1	E	51338-27-3	> 100
101	TR	CON=C(C(=O)OC)c1ccccc1COc1ccccc1C	E	143390-89-0	> 100
102	TR	CC1COC(Cn2cncn2)(O1)c1ccc(Oc2ccc(Cl)cc2)cc1Cl	E	119446-68-3	> 100
104	TR	FC(F)(F)c1cnc(CCNC(=O)c2ccccc2C(F)(F)F)c(Cl)c1	E	658066-35-4	> 100
105	TR	COC(=O)C(C)N(C(=O)Cc1ccccc1)c1c(C)cccc1C	E	98243-83-5	> 100
106	TR	COc1cc(OC)n2nc(NS(=O)(=O)c3c(OC)nccc3C(F)(F)F)nc2n1	E	422556-08-9	> 100
107	TR	CCOC(=O)OC1=C(C(=O)NC11CCC(CC1)OC)c1cc(C)ccc1C	E	203313-25-1	> 100
109	TR	CCc1cc(C)cc(CC)c1C1=C(OC(=O)C(C)(C)C)N2CCOCCN2C1=O	E	243973-20-8	> 100

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
110	TR	Cc1cc(C)cc(c1)C(=O)N(NC(=O)c1ccc2OCCCc2c1C)C(C)(C)C	E	143807-66-3	> 100
111	TR	CSC(=O)c1cccc2nnsc12	E	135158-54-2	> 100
113	TR	COc1nc(C)nc(NC(=O)NS(=O)(=O)c2ccccc2CCC(F)(F)F)n1	E	94125-34-5	> 100
116	TR	Cc1cc(C)nc(Nc2ccccc2)n1	E	53112-28-0	> 100
117	TR	CC(Oc1ccc(Oc2ncc(cc2Cl)C(F)(F)F)cc1)C(O)=O	E	95977-29-0	> 100
119	TR	FC1(F)Oc2cccc(c2O1)-c1c[nH]cc1C#N	E	131341-86-1	> 100
120	TR	CCCCCCCCCCCCC1=C(OC(C)=O)C(=O)c2ccccc2C1=O	E	57960-19-7	> 100
122	TR	CC(C)C(O)(c1ccc(OC(F)(F)F)cc1)c1cncnc1	E	56425-91-3	> 100
123	TR	COC(=O)c1cccc(C)c1S(=O)(=O)NC(=O)Nc1nc(OCC(F)(F)F)nc(n1)N(C)C	E	126535-15-7	> 100
124	TR	COc1cnc(OC)n2nc(NS(=O)(=O)c3c(OCC(F)F)cccc3C(F)(F)F)nc12	E	219714-96-2	> 100
125	TR	CC1=CC(=O)NO1	E	10004-44-1	> 100
127	TR	CCOc1ccc(NC(=O)OC(C)C)cc1OCC	E	87130-20-9	> 100
129	TR	CC(C1CC1)C(O)(Cn1cncn1)c1ccc(Cl)cc1	E	94361-06-5	> 100
130	TR	CCOc1cc(Oc2ccc(cc2Cl)C(F)(F)F)ccc1[N+]([O-])=O	E	42874-03-3	> 100
131	TR	COCCOc1cccc2N(N=C(C(O)=O)C(=O)c12)c1ccc(Cl)cc1	E	130561-48-7	> 100
133	TR	Fc1ccc2N=C(N(C(=O)c2c1)c1ccc(Cl)cc1Cl)n1cncn1	E	136426-54-5	> 100
134	TR	CC(C)C1CCC(Cc2ccc(Cl)cc2)C1(O)Cn1cncn1	E	125225-28-7	> 100
135	TR	FC(F)(F)c1cccc(OCCCOc2c(Cl)cc(OCC=C(Cl)Cl)cc2Cl)n1	E	179101-81-6	> 100
137	TR	CN(C)S(=O)(=O)n1cnc(n1)S(=O)(=O)n1c(C)c(Br)c2ccc(F)cc12	E	348635-87-0	> 100
81	TR	CON=C(C1=NOCCO1)c1ccccc1Oc1ncnc(Oc2ccccc2Cl)c1F	E	361377-29-9	> 200

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
84	TR	Cc1cc(C)c(C2=C(OC(=O)CC(C)(C)C)C3(CCCC3)OC2=O)c(C)c1	E	283594-90-1	> 200
86	TR	COC(=O)c1cc(Oc2ccc(Cl)cc2Cl)ccc1[N+]([O-])=O	E	42576-02-3	> 200
88	TR	CC(C)Oc1cccc(NC(=O)c2ccccc2C(F)(F)F)c1	E	66332-96-5	> 200
91	TR	CC(C)(C)C(O)C(Oc1ccc(Cl)cc1)n1cncn1	E	55219-65-3	> 200
92	TR	[O-][n+]1nc(nc2ccc(Cl)cc12)-n1ccnc1	E	72459-58-6	> 200
93	TR	FC(F)(F)Oc1ccc(NC(=O)NC(=O)c2ccccc2Cl)cc1	E	64628-44-0	> 200
112	TR	CC1(CCCCC1)C(=O)Nc1ccc(O)c(Cl)c1Cl	E	126833-17-8	> 200
118	TR	NC1=C(Cl)C(=O)N(N=C1)c1ccccc1	E	1698-60-8	> 200
136	TR	Cc1cc(ccc1NC(=O)c1cccc(I)c1C(=O)NC(C)(C)CS(C)(=O)=O)C(F)(C(F)(F)F)C(F)(F)F	E	272451-65-7	> 200
128	TR	CCOC(=O)NCCOc1ccc(Oc2ccccc2)cc1	E	72490-01-8	> 204
100	TR	CCc1ccc(cc1)C(=O)NN(C(=O)c1cc(C)cc(C)c1)C(C)(C)C	E	112410-23-8	> 234
8	TR	CNC(=O)N(C)c1nnc(s1)C(C)(C)C	TB	34014-18-1	>100
12	TR	CN(C)C1=NC(=O)N(C2CCCCC2)C(=O)N1C	TB	51235-04-2	>100
19	TR	CCS(=O)(=O)c1cccnc1S(=O)(=O)NC(=O)Nc1nc(OC)cc(OC)n1	TB	122931-48-0	>100
40	TR	O=Cc1ccco1	TB	98-01-1	>100
86	TR	COC(=O)NC(=S)Nc1ccccc1NC(=S)NC(=O)OC	TB	23564-05-8	>100
88	TR	ClC(Cl)(Cl)C(NC=O)N1CCN(CC1)C(NC=O)C(Cl)(Cl)Cl	TB	26644-46-2	>100
89	TR	CN(C=Nc1ccc(C)cc1C)C=Nc1ccc(C)cc1C	TB	33089-61-1	>100
91	TR	O=C(Nc1cnns1)Nc1ccccc1	TB	51707-55-2	>100
100	TR	CC(CN1CC(C)OC(C)C1)Cc1ccc(cc1)C(C)(C)C	TB	67564-91-4	>100

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
103	TR	CC1(C)CON(Cc2ccccc2Cl)C1=O	TB	81777-89-1	>100
111	TR	CCCCC(CC)Cc1c(Cl)ccc(OOC(C)=O)c1Cl	TB	1928-43-4	>100
114	TR	CSSC	TB	624-92-0	>100
126	TR	CC(C)C1(C)N=C(NC1=O)c1ncccc1C(O)=O	TB	81334-34-1	>100
131	TR	C[As](C)(O)=O	TB	75-60-5	>100
137	TR	COc1ccc(cc1)C(O)(C1CC1)c1cncnc1	TB	12771-68-5	>100
157	TR	CCOC(=O)CSc1nc(nn1C(=O)N(C)C)C(C)(C)C	TB	112143-82-5	>100
163	TR	COc1cc(OC)nc(Sc2cccc(Cl)c2C(O)=O)n1	TB	123343-16-8	>100
164	TR	CC1(OC(=O)N(Nc2ccccc2)C1=O)c1ccc(Oc2ccccc2)cc1	TB	131807-57-3	>100
167	TR	CCCCC(CC)COC(=O)c1nc(Cl)c(Cl)c(N)c1Cl	TB	26952-20-5	>100
171	TR	COC(=O)c1ccccc1S(=O)(=O)NC(=O)Nc1nc(C)cc(C)n1	TB	74222-97-2	>100
173	TR	CC(C)(C)N(NC(=O)c1ccc(Cl)cc1)C(=O)c1ccccc1	TB	112226-61-6	>100
177	TR	CC1=NNC(=O)N(C1)N=Cc1cccnc1	TB	123312-89-0	>100
182	TR	CCc1cnc(C2=NC(C)(C(C)C)C(=O)N2)c(c1)C(O)=O	TB	81335-77-5	>100
184	TR	CC(COc1ccc(Oc2ccccc2)cc1)Oc1ccccn1	TB	95737-68-1	>100
188	TR	CCOP(O)(=O)C(N)=O	TB	25954-13-6	>100
194	TR	CC(C)C1=NN(C(=O)NC(C)(C)C)C(=O)N1N	TB	129909-90-6	>100
201	TR	OC(CN1NC=NC1=S)(Cc1ccccc1Cl)C1(Cl)CC1	TB	178928-70-6	>100
203	TR	CCCOC1=NN(C(O)=NS(=O)(=O)c2ccccc2C(=O)OC)C(=O)N1C	TB	181274-15-7	>100
208	TR	COc1ccc(Br)c(C)c1C(=O)c1c(C)cc(OC)c(OC)c1OC	TB	220899-03-6	>100

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
209	TR	FC(F)(F)c1cnc(CNC(=O)c2c(Cl)cccc2Cl)c(Cl)c1	TB	239110-15-7	>100
212	TR	CS(=O)(=O)c1ccc(C(=O)C2C(=O)CCCC2=O)c(Cl)c1COCC(F)(F)F	TB	335104-84-2	>100
214	TR	COc1cc(OC)[n+]2[nH]c(NS(=O)(=O)c3c(OC)nccc3C(F)(F)F)nc2n1	TB	422556-08-9	>100
11	TR	CCc1cccc(C)c1N(C(C)COC)C(=O)CCl	TB	51218-45-2	>110
90	TR	Fc1cccc(F)c1C(=O)NC(=O)Nc1ccc(Cl)cc1	TB	35367-38-5	>115
1	TR	Clc1ccccc1Nc1nc(Cl)nc(Cl)n1	TB	101-05-3	>117
68	TR	Clc1cccc(Cl)c1C#N	TB	1194-65-6	>120
32	TR	Nc1c(Cl)cc(cc1Cl)[N+]([O-])=O	TB	99-30-9	>181
5	TR	CC1=C(Cl)C(=O)N(C(=O)N1)C(C)(C)C	TB	5902-51-2	>193
169	TR	CCCCOCCOC(=O)C(C)Oc1ccc(Cl)cc1Cl	TB	53404-31-2	>1E3
112	TR	CC(C)N1C(SCN(C1=O)c1ccccc1)=NC(C)(C)C	TB	69327-76-0	>200
159	TR	CN1C(=O)ON(C1=O)c1ccc(Cl)c(Cl)c1	TB	20354-26-1	>200
166	TR	CCS(=O)(=O)c1[nH]c2cccc[n+]2c1S(=O)(=O)NC(=O)Nc1nc(OC)cc(OC)n1	TB	141776-32-1	>200
185	TR	COc1cc(OC)nc(NC(=O)NS(=O)(=O)c2ncccc2C(F)(F)F)n1	TB	104040-78-0	>200
195	TR	CN1C(=O)N(C(=O)C=C1C(F)(F)F)c1ccc(Cl)c(c1)C(=O)OC(C)(C)C(=O)OCC=C	TB	134605-64-4	>200
197	TR	COC(=O)c1ccc(I)cc1S(=O)(=O)NC(=O)Nc1nc(C)nc(OC)n1	TB	144550-36-7	>200
199	TR	CCOc1nc(F)cc2[nH]c(n[n+]12)S(=O)(=O)Nc1c(Cl)cccc1C(=O)OC	TB	147150-35-4	>200
200	TR	CON(C(=O)OC)c1ccccc1COc1ccn(n1)-c1ccc(Cl)cc1	TB	175013-18-0	>200
202	TR	FC(F)(F)c1ccc(OCCCOc2c(Cl)cc(OCC=C(Cl)Cl)cc2Cl)nc1	TB	179101-81-6	>200
213	TR	COc1cc(CCNC(=O)C(OCC#C)c2ccc(Cl)cc2)ccc1OCC#C	TB	374726-62-2	>200

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
216	TR	CCC(C)(C)C(=O)OC1=C(C(=O)OC11CCCCC1)c1cc(Cl)cc(Cl)c1	TB	148477-71-8	>200
24	TR	ClC(Cl)(Cl)SN1C(=O)C2CC=CCC2C1=O	TB	133-06-2	>215
183	TR	OC(=O)c1c(Cl)ccc2cc(Cl)cnc12	TB	84087-01-4	>215
189	TR	Clc1ccc(CN2CCSC2=NC#N)cn1	TB	111988-49-9	>400
198	TR	COc1ncc(F)c2[nH]c(n[n+]12)S(=O)(=O)Nc1c(F)cccc1F	TB	145701-23-1	>500
76	TS	CC1(C)C(C=C(Cl)Cl)C1C(=O)OCc1cccc(Oc2ccccc2)c1	D	52645-53-1	0,024
107	TS	CC(C)=CC1C(C(=O)OCc2coc(Cc3ccccc3)c2)C1(C)C	TB	10453-86-8	0,063
36	TS	COP(=S)(OC)Oc1ccc(cc1)[N+]([O-])=O	D	298-00-0	0,111
16	TS	ClC1=C(Cl)C2(Cl)C3C4CC(C5OC45)C3C1(Cl)C2(Cl)Cl	D	60-57-1	0,139
73	TS	COC(=O)N(C(=O)N1COC2(Cc3cc(Cl)ccc3C2=N1)C(=O)OC)c1ccc(OC(F)(F)F)cc1	D	173584-44-6	0,18
22	TS	CCOP(=S)(Oc1ccc(cc1)[N+]([O-])=O)c1ccccc1	D	2104-64-5	0,245
74	TS	CNC(=O)Oc1cc(C)c(SC)c(C)c1	TB	2032-65-7	0,375
58	TS	COP(=O)(OC)OC(Br)C(Cl)(Cl)Br	TB	300-76-5	0,48
144	TS	FC(F)(F)Oc1ccc(NC(=O)NN=C(Cc2ccc(cc2)C#N)c2cccc(c2)C(F)(F)F)cc1	E + D	139968-49-3 / 58-89-9	1,105
63	TS	CCN(CC)C(=O)C(Cl)=C(C)OP(=O)(OC)OC	D	13171-21-6	1,46
59	TS	CCS(=O)CCSP(=O)(OC)OC	TB	301-12-2	3
28	TS	Clc1ccc(cc1)C(c1ccc(Cl)cc1)C(Cl)(Cl)Cl	TB	50-29-3	5,15
34	TS	CNC(=O)ON=C(SC)C(=O)N(C)C	D	23135-22-0	10,32
34	TS	OC(c1ccc(Cl)cc1)(c1ccc(Cl)cc1)C(Cl)(Cl)Cl	TB	115-32-2	12,2
28	TS	CNC(=O)Oc1cccc(c1)N=CN(C)C	D	22259-30-9	14,27

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
175	TS	FC(OC(F)(F)F)C(F)(F)Oc1ccc(NC(=O)NC(=O)c2c(F)cccc2F)cc1Cl	TB	116714-46-6	14,5
77	TS	CC(Cl)(Cl)C(O)=O	D	127-20-8	24,2
23	TS	CCC(C)c1cc(cc(c1O)[N+]([O-])=O)[N+]([O-])=O	TB	88-85-7	32,3
49	TS	CCOC(=O)C1CC(=O)C(=C(O)C2CC2)C(=O)C1	D	95266-40-3	47
31	TS	CSC1=NN=C(C(=O)N1N)C(C)(C)C	D	21087-64-9	60,4
65	TS	CCCN(CCC)C(=O)SCC	TB	759-94-4	72,5
41	TS	CSc1nc(NC(C)C)nc(NC(C)C)n1	D	7287-19-6	96,69
72	TS	CN(C)c1ccc(cc1)N=NS(O)(=O)=O	D	140-56-7	102
9	TS	CC1=C(SCCO1)C(=O)Nc1ccccc1	D	5234-68-4	181
13	TS	CN(C)NC(=O)CCC(O)=O	D	1596-84-5	205,46
60	TS	CN(C)C(=O)C(c1ccccc1)c1ccccc1	D	957-51-7	2431,7
87	TS	Fc1ccc(NC(=O)c2cccnc2Oc2cccc(c2)C(F)(F)F)c(F)c1	E	83164-33-4	> 100
97	TS	COc1cc(OC)n2nc(nc2n1)S(=O)(=O)Nc1c(Cl)ccc(C)c1Cl	E	139528-85-1	> 100
103	TS	Fc1cc(Oc2ccc(cc2Cl)C(F)(F)F)ccc1NC(=O)NC(=O)c1c(F)cccc1F	E	101463-69-8	> 100
108	TS	COC(=O)CC(NC(=O)C(NC(=O)OC(C)C)C(C)C)c1ccc(Cl)cc1	E	283159-90-0	> 100
114	TS	COc1ncc(F)c2nc(nn12)S(=O)(=O)Nc1c(F)cccc1F	E	145701-23-1	> 100
121	TS	Clc1ccc(c(Cl)c1)C1(Cn2cncn2)CC(Br)CO1	E	116255-48-2	> 100
126	TS	CCC(C)(CC)c1cc(NC(=O)c2c(OC)cccc2OC)on1	E	82558-50-7	> 100
132	TS	COc1c(Cl)ccc(Cl)c1C(O)=O	E	1918-00-9	> 100
82	TS	CN(C)S(=O)(=O)N(SC(F)(Cl)Cl)c1ccc(C)cc1	E	731-27-1	> 196

ID	Set	SMILES	Source	CAS_NO	Exp. Value (µg/bee)
90	TS	COCCN(C(=O)CCl)c1c(C)cccc1C	E	50563-36-5	> 200
115	TS	CC(Oc1ccc(Cl)cc1Cl)C(O)=O	E	15165-67-0	> 200
174	TS	COCc1cnc(C2=NC(C)(C(C)C)C(=O)N2)c(c1)C(O)=O	TB	114311-32-9	>1.83E3
73	TS	CC1CCCCC1NC(=O)Nc1ccccc1	TB	1982-49-6	>100
94	TS	CCCCCCCCSC(=O)Oc1cc(Cl)nnc1-c1ccccc1	TB	55512-33-9	>100
119	TS	CC(Oc1ccc(Oc2ncc(Cl)cc2F)cc1)C(=O)OCC#C	TB	105512-06-9	>100
161	TS	CCCN(CC)CC1COC2(CCC(CC2)C(C)(C)C)O1	TB	118134-30-8	>100
172	TS	CCC1OC2(CCC1C)CC1CC(CC=C(C)CC(C)C=CC=C3COC4C(O)C(C)=CC(C(=O)O1)C34O)O2	TB	51596-11-3	>100
186	TS	CCCCOCCOC(=O)COc1nc(Cl)c(Cl)cc1Cl	TB	64700-56-7	>100
207	TS	COc1cc(OC)nc(NC(=O)NS(=O)(=O)Nc2ccccc2C(=O)N(C)C)n1	TB	213464-77-8	>100
134	TS	ClCC1(CCl)C(=C)C2(Cl)C(Cl)C(Cl)C1(Cl)C2(Cl)Cl	TB	8001-35-2	>105
60	TS	CON(C)C(=O)Nc1ccc(Cl)c(Cl)c1	TB	330-55-2	>121
124	TS	[O-][N+](=O)c1cc(c(Cl)c(c1Nc1ncc(cc1Cl)C(F)(F)F)[N+]([O-])=O)C(F)(F)F	TB	79622-59-6	>200
196	TS	CON=C(C(=O)OC)c1ccccc1CON=C(C)c1cccc(c1)C(F)(F)F	TB	141517-21-7	>200
205	TS	Clc1ccc(cc1)-c1ccccc1NC(=O)c1cccnc1Cl	TB	188425-85-6	>200
109	TS	Clc1ccc(CCC(Cn2cncn2)(C#N)c2ccccc2)cc1	TB	114369-43-6	>292

Chapter 5

Table S5.1. Overall dataset (n= 123) on organic binary mixtures toxicity used to develop regression model according to approach A.

Set	Quasi-SMILES	Observed	Predicted	Applicability
+	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	-0.0680	0.7610	YES
+	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl^CCCCSP(=O)(SCCCC)SCCCC	-0.3780	-0.0773	YES
*	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl^CCOC(=O)C=CC(=O)OCC	-0.8410	-0.2057	YES
-	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl^CCCOCC(=NC1=C(C=C(C=C1)Cl)C(F)(F)F)N2C=CN=C2	1.5230	1.1636	YES
+	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl	1.1550	0.0315	No
-	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl^CC(C)(C)C(=O)C(N1C=NC=N1)OC2=CC=C(C=C2)Cl	1.0970	0.1204	YES
-	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl^C1=CC=C(C(=C1)C2C(O2)(CN3C=NC=N3)C4=CC=C(C=C4)F)Cl	0.3010	-0.0513	YES
*	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl^CC(C)(C)C(C(=CC1=CC=C(C=C1)Cl)N2C=NC=N2)O	-0.0490	-0.2814	YES

+	C1CN(C(=N1)N[N+](=O)[O-])CC2=CN=C(C=C2)Cl^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	1.9790	1.7027	No
+	C1CN(C(=N1)N[N+](=O)[O-])CC2=CN=C(C=C2)Cl^CCCOCC(=NC1=C(C=C(C=C1)Cl)C(F)(F)F)N2C=CN=C2	2.0130	2.1053	No
*	C1CN(C(=N1)N[N+](=O)[O-])CC2=CN=C(C=C2)Cl^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl	1.9210	0.9732	No
#	C1CSC(=NC#N)N1CC2=CN=C(C=C2)Cl^CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl	1.0460	1.8104	YES
*	C1CSC(=NC#N)N1CC2=CN=C(C=C2)Cl^CCCOCC(=NC1=C(C=C(C=C1)Cl)C(F)(F)F)N2C=CN=C2	2.0000	2.0162	YES
-	C1CSC(=NC#N)N1CC2=CN=C(C=C2)Cl^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl	1.5230	0.8841	No

Set	Quasi-SMILES	Observed	Predicted	Applicability
-	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl^C1=CC=C(C=C1)C(CCC2=CC=C(C=C2)Cl)(CN3C=NC=N3)C#N	-1.1550	-0.5744	YES
-	C1CN(C(=N1)N[N+](=O)[O-])CC2=CN=C(C=C2)Cl^C1=CC=C(C=C1)C(CCC2=CC=C(C=C2)Cl)(CN3C=NC=N3)C#N	0.5230	0.3673	No
*	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CCOC(=O)C=CC(=O)OCC	-1.2990	-1.0684	YES
#	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CCCCSP(=O)(SCCCC)SCCCC	-0.8630	-0.9401	YES
+	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	-0.7020	-0.1017	YES
+	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl	-1.3220	-0.6177	YES

-	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl	-1.1060	-0.6177	YES
*	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl	-0.7820	-0.6177	YES
*	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl	-0.7820	-0.6177	YES
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	-0.7880	-0.7883	YES

Set	Quasi-SMILES	Observed	Predicted	Applicability
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	-0.5170	-0.7883	YES
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	-0.4280	-0.7883	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	-0.1850	-0.7883	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	0.6780	-0.7883	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	0.1080	-0.7883	YES

#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C	-0.3800	-0.0072	YES
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C	-0.5730	-0.6364	YES
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CC1=CC(=C(C=C1)C(C)C)O	-1.0090	-1.6733	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^C(=O)(C(=O)O)O	-0.8480	-0.8146	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^COC(=O)N(C1=CC=CC=C1COC2=NN(C=C2)C3=CC=C(C=C3)Cl)OC	-0.6460	0.0562	YES

Set	Quasi-SMILES	Observed	Predicted	Applicability
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^C1=CC=C(C(=C1)C2=CC=C(C=C2)Cl)NC(=O)C3=C(N=CC=C3)Cl	-1.0640	-0.5966	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^C(#N)C1=C(C(=C(C(=C1Cl)Cl)Cl)C#N)Cl	-0.8600	-0.8483	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C2	2.0000	0.0662	YES
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCOC(=O)C=CC(=O)OCC	-0.9170	-0.9186	YES
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCCSP(=O)(SCCCC)SCCCC	-0.2920	-0.7902	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	2.0000	0.0482	YES
#	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl	-0.8130	-0.6177	YES

*	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C	-0.6150	-0.1570	YES
*	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C	-0.9640	-0.7863	YES
#	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CC1=CC(=C(C=C1)C(C)C)O	-1.3640	-1.8232	YES
#	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^C(=O)(C(=O)O)O	-1.1670	-0.9645	YES
#	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^C1=CC=C(C(=C1)C2=CC=C(C=C2)Cl)NC(=O)C3=C(N=CC=C3)Cl	-1.3540	-0.7465	YES
+	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^C(#N)C1=C(C(=C(C(=C1Cl)Cl)Cl)C#N)Cl	-1.2200	-0.9982	YES
*	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C2	0.3570	-0.0837	YES

Set	Quasi-SMILES	Observed	Predicted	Applicability
-	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CCOC(=O)C=CC(=O)OCC	-1.2990	-1.0684	YES
-	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CCCCSP(=O)(SCCCC)SCCCC	-0.8630	-0.9401	YES
+	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	-0.7020	-0.1017	YES
*	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl	-0.7440	0.1634	YES
*	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	-0.3070	-0.1570	YES
#	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C	-0.2550	-0.0051	YES
-	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^CC1=CC(=C(C=C1)C(C)C)O	-0.5670	-1.0421	YES

-	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^C(=O)(C(=O)O)O	-0.1760	-0.1833	YES
*	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^COC(=O)N(C1=CC=CC=C1COC2=NN(C=C2)C3=CC=C(C=C3)Cl)OC	-0.3200	0.6875	YES
#	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^C1=CC=C(C(=C1)C2=CC=C(C=C2)Cl)NC(=O)C3=C(N=CC=C3)Cl	-0.7510	0.0346	YES
#	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^C(#N)C1=C(C(=C(C(=C1Cl)Cl)Cl)C#N)Cl	-0.8070	-0.2170	YES
#	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C2	0.6020	0.6974	YES
#	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^CCOC(=O)C=CC(=O)OCC	-0.6410	-0.2873	YES
*	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^CCCCSP(=O)(SCCCC)SCCCC	-0.1000	-0.1589	YES

Set	Quasi-SMILES	Observed	Predicted	Applicability
*	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	0.5690	0.6794	YES
*	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl	-0.6880	-0.4658	YES
#	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	-0.4360	-0.7863	YES
#	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C	-0.6600	-0.0051	YES
-	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^CC1=CC(=C(C=C1)C(C)C)O	-0.5920	-1.6713	YES
#	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^C(=O)(C(=O)O)O	-1.1640	-0.8126	YES
#	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^COC(=O)N(C1=CC=CC=C1COC2=NN(C=C2)C3=CC=C(C=C3)Cl)OC	-0.2150	0.0583	YES

+	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^C1=CC=C(C(=C1)C2=CC=C(C=C2)Cl)NC(=O)C3=C(N=CC=C3)Cl	-0.6830	-0.5946	YES
*	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^C(#N)C1=C(C(=C(C(=C1Cl)Cl)Cl)C#N)Cl	-0.5240	-0.8463	YES
#	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C2	-0.3940	0.0682	YES
-	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^CCOC(=O)C=CC(=O)OCC	-0.3620	-0.9165	YES
#	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^CCCCSP(=O)(SCCCC)SCCCC	-0.3360	-0.7882	YES
#	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	-0.3820	0.0502	YES
*	CC1=CC(=C(C=C1)C(C)C)O^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl	-1.2070	-1.8518	YES

Set	Quasi-SMILES	Observed	Predicted	Applicability
#	CC1=CC(=C(C=C1)C(C)C)O^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	-1.3100	-2.1723	YES
-	CC1=CC(=C(C=C1)C(C)C)O^CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C	-1.5430	-1.3911	YES
*	CC1=CC(=C(C=C1)C(C)C)O^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C	-1.6350	-2.0203	YES
+	CC1=CC(=C(C=C1)C(C)C)O^C(=O)(C(=O)O)O	-1.4870	-2.1985	YES
+	CC1=CC(=C(C=C1)C(C)C)O^COC(=O)N(C1=CC=CC=C1COC2=NN(C=C2)C3=CC=C(C=C3)Cl)OC	-1.4500	-1.3277	YES
+	CC1=CC(=C(C=C1)C(C)C)O^C1=CC=C(C(=C1)C2=CC=C(C=C2)Cl)NC(=O)C3=C(N=CC=C3)Cl	-1.6730	-1.9806	YES
#	CC1=CC(=C(C=C1)C(C)C)O^C(#N)C1=C(C(=C(C(=C1Cl)Cl)Cl)C#N)Cl	-1.4740	-2.2322	YES
+	CC1=CC(=C(C=C1)C(C)C)O^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C2	-1.5910	-1.3178	YES
*	CC1=CC(=C(C=C1)C(C)C)O^CCOC(=O)C=CC(=O)OCC	-1.8060	-2.3025	YES
*	CC1=CC(=C(C=C1)C(C)C)O^CCCCSP(=O)(SCCCC)SCCCC	-1.5450	-2.1741	YES

+	CC1=CC(=C(C=C1)C(C)C)O^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	-1.5110	-1.3358	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	-1.0410	0.0482	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C2	-1.0450	0.0662	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl	-1.1430	-0.6813	No
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^C1=CC=C(C=C1)C(CCC2=CC=C(C=C2)Cl)(CN3C=NC=N3)C#N	-1.0210	-1.2872	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CC1(CCC(C1(CN2C=NC=N2)O)CC3=CC=C(C=C3)Cl)C	-1.1040	-0.4853	No
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCCC(CN1C=NC=N1)(C#N)C2=CC=C(C=C2)Cl	-1.1110	-0.3309	YES

Set	Quasi-SMILES	Observed	Predicted	Applicability
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	-0.9280	0.0482	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C2	-0.6670	0.0662	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl	-0.7760	-0.6813	No
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^C1=CC=C(C=C1)C(CCC2=CC=C(C=C2)Cl)(CN3C=NC=N3)C#N	-0.7930	-1.2872	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CC1(CCC(C1(CN2C=NC=N2)O)CC3=CC=C(C=C3)Cl)C	-0.7170	-0.4853	No
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCCC(CN1C=NC=N1)(C#N)C2=CC=C(C=C2)Cl	-0.5740	-0.3309	YES
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	-0.1880	0.0482	YES

#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C2	0.6580	0.0662	YES
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl	0.1310	-0.6813	No
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^C1=CC=C(C=C1)C(CCC2=CC=C(C=C2)Cl)(CN3C=NC=N3)C#N	0.2600	-1.2872	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CC1(CCC(C1(CN2C=NC=N2)O)CC3=CC=C(C=C3)Cl)C	0.3470	-0.4853	No
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCCC(CN1C=NC=N1)(C#N)C2=CC=C(C=C2)Cl	0.9210	-0.3309	YES
+	CC1(C(C1C(=O)OC(C#N)C2=CC(=C(C=C2)F)OC3=CC=CC=C3)C=C(Cl)Cl)C^CCOC(=O)C=CC(=O)OCC	1.2220	1.2749	YES
+	CC1(C(C1C(=O)OC(C#N)C2=CC(=C(C=C2)F)OC3=CC=CC=C3)C=C(Cl)Cl)C^CCCCSP(=O)(SCCCC)SCCCC	1.5230	1.4033	YES

Set	Quasi-SMILES	Observed	Predicted	Applicability
-	CC1(C(C1C(=O)OC(C#N)C2=CC(=C(C=C2)F)OC3=CC=CC=C3)C=C(Cl)Cl)C^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	2.6990	2.2417	YES
+	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)C=C(C(F)(F)F)Cl)C^CCOC(=O)C=CC(=O)OCC	1.3980	1.0080	YES
+	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)C=C(C(F)(F)F)Cl)C^CCCCSP(=O)(SCCCC)SCCCC	1.5230	1.1364	YES
+	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)C=C(C(F)(F)F)Cl)C^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	2.8860	1.9748	YES
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCOC(=O)C=CC(=O)OCC	-0.9170	-0.9186	YES
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCCSP(=O)(SCCCC)SCCCC	-0.2920	-0.7902	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	2.0000	0.0482	YES
-	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C	0.4440	0.2257	YES

+	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)CCCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	0.8540	1.0758	YES
*	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	0.6200	0.9103	YES
-	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C	0.2520	0.2257	YES
*	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)CCCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	0.7960	1.0758	YES
*	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO2	0.4950	0.9103	YES
+	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C	0.4440	0.2257	YES
#	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)CCCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C	0.8240	1.0758	YES

#	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C^CCCC OCCOCCOCC1=CC2=C(C=C1CCC)OCO2	0.6200	0.9103	YES

Table S5.2. Collection of SMILES attributes (SAk) which are promoters of increase or decrease of binary mixtures toxicity ($pLD_{50\text{-}mix}$) according to Approach A.

SA_k	**CWs Run 1**	**CWs Run 2**	**CWs Run 3**	**N1**	**N2**	**N3**	**Dj**
1...(.....	0.47979	0.07282	0.50700	31	30	28	0.0016
C...(...=...	0.10946	2.36483	0.40467	31	31	31	0.0000
C...(...C...	0.20715	0.30194	0.11965	31	31	31	0.0000
C...1.....	0.49844	0.25787	1.43446	31	31	31	0.0000
C...C.....	0.42647	0.65328	0.02769	31	31	31	0.0000
2.....	0.73809	0.90239	0.40871	28	31	31	0.0016
C...2.....	0.23117	0.74069	0.37256	28	31	31	0.0016
=...2.....	0.09720	0.48016	0.25940	26	29	31	0.0028
C...2...=...	0.21369	0.16348	0.37200	26	29	31	0.0028
N...(.....	0.32901	0.53943	1.14972	26	29	25	0.0006

SA_k	CWs Run 1	CWs Run 2	CWs Run 3	N1	N2	N3	Dj
N.....	0.05329	0.24784	0.24532	26	29	25	0.0006
O...=...(...	1.04615	0.70385	0.49177	26	27	29	0.0018
O...=.....	0.54781	0.69784	0.28764	26	27	29	0.0018
(.....	-0.44876	-0.28851	-0.09192	31	31	31	0.0000
(...C...(...	-0.11656	-0.09905	-0.11495	31	31	31	0.0000
C.....	-0.07742	-0.18506	-0.43105	31	31	31	0.0000
C...=.....	-0.06136	-0.14651	-0.30507	31	31	31	0.0000
C...#.....	-0.24834	-0.01388	-0.08974	22	21	12	0.0095
=...C...3...	-0.23690	-0.23720	-0.33687	20	24	21	0.0008
(...C...#...	-0.46572	-0.03278	-0.21581	18	18	11	0.0078
C...N...(...	-0.07335	-0.49241	-0.06609	15	22	21	0.0054
F...(...(...	-0.02799	-0.29201	-0.21691	15	16	11	0.0050
N...(...2...	-0.28489	-0.08615	-0.74869	11	15	10	0.0015
O...(...1...	-0.21704	-0.49633	-0.40564	11	16	10	0.0015
P...O...C...	-0.08276	-0.01255	-0.42379	6	6	11	0.0095

(...N...(...	-0.06555	-0.02475	-0.07730	4	4	2	0.0108
N...(...Cl..	-0.37220	-3.78634	-0.23783	4	0	3	0.0046
N...#...(...	-0.46982	-0.03538	-0.35218	3	0	1	0.0161

Table S5.3. Collection of SMILES attributes (SAk) which are promoters of increase or decrease for bee toxicity of binary mixtures toxicity ($pLD_{50\text{-}mix}$) according to approach B.

No.	SAk	CWs Run 1	CWs Run 2	CWs Run 3	N1	N2	N3	Defect[SAk]
1	=...(.....	0.20138	0.21425	0.18951	25	24	24	0
2	=.....	0.18692	0.31074	0.58725	25	24	24	0
3	C...(.....	0.20504	0.30947	0.72085	25	24	24	0
4	C...(...=...	0.4104	2.25731	0.17718	25	24	24	0
5	C...1.....	0.32041	1.45988	1.71818	25	24	24	0
6	C...2.....	0.05908	0.06229	0.83235	25	24	24	0
7	C...=.....	0.54288	0.75392	0.24836	25	24	24	0
8	C...=...1...	0.06045	0.39778	0.4424	25	24	24	0
9	^.....	3.4187	3.40145	8.0686	25	24	24	0
10	^...C.....	5.95474	2.47254	1.8062	25	24	24	0
11	C...1...(...	0.25914	0.16566	2.44348	24	21	23	0
12	C...C...=...	0.48705	0.29644	0.18291	24	23	24	0.0008
13	C...2...=...	0.28302	0.31196	1.70791	23	23	23	0.0008
14	^...C...C...	0.17446	2.02675	2.15176	23	20	17	0.0053
15	O...C.....	0.43319	0.15619	0.20863	22	24	20	0.0011
1	(.....	-0.73322	-0.49071	-0.55293	25	24	24	0
2	=...C...(...	-0.73622	-0.60929	-0.74267	25	24	24	0
3	C.....	-0.39311	-0.80465	-0.33338	25	24	24	0
4	3.....	-0.28168	-0.3011	-0.3781	16	18	20	0.0054
5	?.....	-0.06173	-0.27796	-0.19033	11	8	12	0.0026
6	Cl..(...2...	-1.24011	-1.56015	-1.02412	11	8	12	0.0026
7	P...(.....	-0.02228	-0.01292	-0.50825	9	6	5	0.0108
8	%52.....	-0.73116	-0.80873	-0.68492	8	2	1	0.0309
9	X...?.....	-0.08846	-0.60153	-0.29484	8	5	8	0.0008
10	%51.....	-0.54416	-0.11808	-0.35323	6	8	7	0.004
11	Cl..(...3...	-0.67348	-0.02014	-0.53636	3	4	5	0.011
12	Y.....	-0.77726	-1.31583	-0.95171	3	3	4	0.0067
13	%15.....	-0.30397	-0.55685	-1.11566	2	1	2	0.0008

14	%16.....	-0.79527	-0.555	-0.24782	2	2	2	0.0008
15	1...(...(...	-0.09767	-0.19633	-0.561	2	3	1	0.0128

Table S5.4. Full list of correlation weights (CW) for the calculation of the regression-based model on binary mixtures toxicity (pLD_{50mix}) according to Approach B (Toxic Unit values) (Toropova et al., 2019).

SAk	CW(SAk)	N1	N2	N3	dj
#.....	0.41229	14	14	13	0.0007
$.....	2.00779	14	16	12	0.0023
%11.....	0.85318	6	3	1	0.0283
%12.....	2.19106	2	0	0	1.0000
%13.....	0.0	0	1	2	0.0000
%14.....	-0.44712	3	2	1	0.0196
%15.....	-0.86965	2	1	2	0.0008
%16.....	1.72772	2	2	2	0.0008
%17.....	2.48608	1	1	0	1.0000
%18.....	0.0	0	2	1	0.0000
%19.....	0.0	0	1	1	0.0000
%20.....	0.0	0	1	1	0.0000
%21.....	0.0	0	1	2	0.0000
%22.....	0.32101	1	0	3	0.0213
%23.....	-0.69886	2	1	0	1.0000
%24.....	0.0	0	2	1	0.0000
%25.....	-0.92466	1	1	1	0.0008
%26.....	1.92161	1	0	0	1.0000
%27.....	0.0	0	0	1	0.0000
%28.....	0.0	0	0	1	0.0000
%29.....	0.0	0	2	0	0.0000
%31.....	-1.45641	1	0	1	0.0008
%33.....	0.0	0	0	1	0.0000
%34.....	-0.69679	1	0	0	1.0000
%35.....	0.0	0	1	0	0.0000
%36.....	0.0	0	0	2	0.0000
%38.....	-0.11404	1	1	0	1.0000
%40.....	-0.05689	1	1	0	1.0000
%51.....	-1.28250	6	8	7	0.0040
%51.N...#...	0.0	0	0	1	0.0000
%51.X...$...	-0.18294	1	7	5	0.0281
%52.....	1.02031	8	2	1	0.0309
%52.X...$...	0.48101	5	2	0	1.0000

%53.....	0.0	0	2	1	0.0000
%53.X...$...	0.0	0	1	1	0.0000
%54.....	-0.71908	3	3	1	0.0196
%54.X...$...	0.09030	3	1	0	1.0000
%55.....	2.30565	2	2	4	0.0144
%55.X...$...	2.56765	2	0	0	1.0000
%58.....	2.23094	1	3	2	0.0144
%58.X...$...	0.0	0	1	1	0.0000
%59.....	2.16997	2	0	4	0.0144
%59.N...#...	-1.43435	1	0	1	0.0008
%59.X...$...	0.0	0	0	2	0.0000
%63.....	-0.36612	2	2	2	0.0008
%63.X...$...	0.69999	2	2	2	0.0008
%67.....	2.70537	1	1	0	1.0000
%67.X...$...	2.45431	1	1	0	1.0000
%75.....	0.0	0	1	2	0.0000
%75.X...$...	0.0	0	1	1	0.0000
(...#.....	0.0	0	2	1	0.0000
(...(.....	0.17079	17	20	19	0.0031
(.....	0.15482	25	24	24	0.0000
(...C...#...	-0.63049	10	13	12	0.0045
(...C...%11.	1.98200	2	1	0	1.0000
(...C...%13.	0.0	0	0	2	0.0000
(...C...%14.	-0.11214	2	0	0	1.0000
(...C...%15.	0.33269	2	1	0	1.0000
(...C...%16.	-1.17003	2	1	2	0.0008
(...C...%17.	-1.24955	1	1	0	1.0000
(...C...%18.	0.0	0	1	0	0.0000
(...C...%19.	0.0	0	1	0	0.0000
(...C...%21.	0.0	0	1	1	0.0000
(...C...%23.	-0.63232	2	0	0	1.0000
(...C...%24.	0.0	0	1	1	0.0000
(...C...%25.	0.00262	1	1	1	0.0008

SAk	CW(SAk)	N1	N2	N3	dj
(...C...%27.	0.0	0	0	1	0.0000
(...C...%29.	0.0	0	1	0	0.0000
(...C...%31.	-0.50415	1	0	1	0.0008
(...C...%35.	0.0	0	1	0	0.0000
(...C...%36.	0.0	0	0	2	0.0000
(...C...%51.	2.11754	4	3	4	0.0008
(...C...%53.	0.0	0	2	1	0.0000
(...C...%58.	0.08434	1	1	1	0.0008
(...C...%63.	3.61513	1	0	1	0.0008
(...C...%75.	0.0	0	0	1	0.0000
(...C...(...	0.18973	24	24	24	0.0008
(...F...(...	0.32159	10	12	11	0.0028
(...Cl..%11.	4.53209	4	2	1	0.0237
(...Cl..%12.	2.05065	2	0	0	1.0000
(...Cl..%13.	0.0	0	1	0	0.0000
(...Cl..%14.	0.09118	1	2	1	0.0008
(...Cl..%15.	0.0	0	0	2	0.0000
(...Cl..%18.	0.0	0	0	1	0.0000
(...Cl..%19.	0.0	0	0	1	0.0000
(...Cl..%20.	0.0	0	1	1	0.0000
(...Cl..%21.	0.0	0	0	1	0.0000
(...Cl..%22.	0.0	0	0	3	0.0000
(...Cl..%23.	0.0	0	1	0	0.0000
(...Cl..%28.	0.0	0	0	1	0.0000
(...Cl..%29.	0.0	0	1	0	0.0000
(...Cl..%33.	0.0	0	0	1	0.0000
(...Cl..%34.	-0.71050	1	0	0	1.0000
(...Cl..%38.	-0.94922	1	1	0	1.0000
(...Cl..%40.	-0.82218	1	1	0	1.0000
(...Cl..%51.	-0.09423	1	3	1	0.0008
(...Cl..%52.	-0.66768	2	0	0	1.0000
(...Cl..%54.	2.26407	1	1	1	0.0008
(...Cl..%55.	1.36780	2	1	3	0.0090

(...Cl..%58.	0.0	0	2	1	0.0000
(...Cl..%63.	0.0	0	1	0	0.0000
(...Cl..(...	-0.10981	16	13	15	0.0005
(...N...#...	-0.39227	13	14	13	0.0008
(...N...(...	1.52191	3	1	3	0.0008
(...O...%16.	0.0	0	1	0	0.0000
(...O...%18.	0.0	0	1	0	0.0000
(...O...%22.	-0.66185	1	0	0	1.0000
(...O...%24.	0.0	0	1	0	0.0000
(...O...%26.	0.33845	1	0	0	1.0000
(...O...%59.	1.42795	1	0	3	0.0213
(...O...%75.	0.0	0	0	1	0.0000
(...O...(...	-1.63538	1	0	1	0.0008
+.....	-0.84734	1	0	1	0.0008
+...[...(...	2.15304	1	0	1	0.0008
-.....	2.12227	1	0	1	0.0008
-...[...(...	1.63054	1	0	1	0.0008
1...(...(...	1.61845	2	3	1	0.0128
1...(.....	0.68833	24	21	24	0.0008
1.....	-0.46880	25	24	24	0.0000
1...C...(...	0.05230	12	16	14	0.0040
1...Cl..(...	3.37342	3	4	2	0.0073
1...N...(...	0.64065	1	1	0	1.0000
1...O...(...	1.33002	2	2	0	1.0000
2...%51.....	-1.83821	1	2	1	0.0008
2...%52.....	0.04629	6	2	1	0.0283
2...%54.....	1.34144	2	2	0	1.0000
2...%63.....	1.31679	1	1	1	0.0008
2...%67.....	4.32087	1	1	0	1.0000
2...%75.....	0.0	0	1	0	0.0000
2...(...(...	0.22327	1	0	0	1.0000
2...(.....	-0.11304	23	23	24	0.0017
2.....	-1.25250	25	24	24	0.0000

SAk	CW(SAk)	N1	N2	N3	dj
2...C...(...	-0.74708	2	2	4	0.0144
2...N...(...	1.22066	3	3	2	0.0073
2...O...(...	0.48666	10	6	5	0.0128
3...(.....	0.37534	16	18	20	0.0054
3.....	0.07177	16	18	20	0.0054
3...C...(...	-0.79758	2	4	4	0.0144
4...(.....	2.30591	1	0	0	1.0000
4.....	-0.11336	1	0	0	1.0000
4...C...(...	-0.70517	1	0	0	1.0000
=...(.....	0.85635	25	24	24	0.0000
=.....	0.32245	25	24	24	0.0000
=...1.....	-0.47351	25	24	24	0.0000
=...2.....	2.02923	23	23	23	0.0008
=...3.....	0.13259	14	18	19	0.0070
=...4.....	-0.99013	1	0	0	1.0000
=...C...(...	-0.06556	25	24	24	0.0000
=...C...1...	-0.58230	25	24	24	0.0000
=...C...2...	2.31672	20	21	24	0.0045
=...C...3...	0.21155	16	18	20	0.0054
=...C...4...	-0.74890	1	0	0	1.0000
=...N...(...	-0.86286	4	5	8	0.0144
=...N...1...	-0.55719	1	1	2	0.0144
=...N...2...	-0.82965	3	2	2	0.0073
=...N...3...	0.51440	2	0	2	0.0008
=...O...(...	-0.72876	18	23	21	0.0040
=...S...(...	-0.09012	9	6	5	0.0108
?.....	-0.30680	11	8	12	0.0026
?...X...%51.	-0.47079	3	0	0	1.0000
?...X...%52.	0.30222	3	0	1	0.0196
?...X...%53.	0.0	0	1	0	0.0000
?...X...%54.	0.0	0	2	1	0.0000
?...X...%55.	0.0	0	1	3	0.0000
?...X...%58.	1.99389	1	1	0	1.0000

?...X...%59.	2.06820	1	0	2	0.0144
?...X...%75.	0.0	0	0	1	0.0000
?...Y...%51.	-0.78098	2	1	2	0.0008
?...Y...%55.	0.0	0	1	1	0.0000
?...Y...%58.	0.0	0	1	1	0.0000
?...Y...%59.	2.21072	1	0	0	1.0000
C...#.....	0.34144	14	14	13	0.0007
C...%11.....	1.22421	2	1	0	1.0000
C...%13.....	0.0	0	0	2	0.0000
C...%14.....	-0.52265	2	0	0	1.0000
C...%15.....	0.41114	2	1	0	1.0000
C...%16.....	-0.96811	2	1	2	0.0008
C...%17.....	-1.20137	1	1	0	1.0000
C...%18.....	0.0	0	1	0	0.0000
C...%19.....	0.0	0	1	0	0.0000
C...%21.....	0.0	0	1	1	0.0000
C...%23.....	-0.72830	2	0	0	1.0000
C...%24.....	0.0	0	1	1	0.0000
C...%25.....	-0.60978	1	1	1	0.0008
C...%27.....	0.0	0	0	1	0.0000
C...%29.....	0.0	0	1	0	0.0000
C...%31.....	0.23461	1	0	1	0.0008
C...%35.....	0.0	0	1	0	0.0000
C...%36.....	0.0	0	0	2	0.0000
C...%51.....	0.73402	4	3	4	0.0008
C...%53.....	0.0	0	2	1	0.0000
C...%55.....	0.0	0	1	1	0.0000
C...%58.....	2.86269	1	1	1	0.0008
C...%63.....	3.40910	1	0	1	0.0008
C...%75.....	0.0	0	0	1	0.0000
C...(...#...	0.0	0	2	1	0.0000
C...(...(...	0.08712	5	8	8	0.0103
C...(.....	-0.89423	25	24	24	0.0000

SAk	CW(SAk)	N1	N2	N3	dj
C...(...1...	0.01581	17	10	16	0.0004
C...(...2...	0.36116	7	10	10	0.0080
C...(...3...	1.28545	15	17	19	0.0056
C...(...=...	0.29512	25	24	24	0.0000
C...(...C...	0.62932	25	24	24	0.0000
C.....	0.23912	25	24	24	0.0000
C...1...(...	1.48263	24	21	23	0.0000
C...1.....	-0.16903	25	24	24	0.0000
C...1...=...	-1.51621	25	24	24	0.0000
C...1...C...	-0.23422	14	14	9	0.0080
C...2...%51.	-0.90302	1	0	1	0.0008
C...2...%54.	2.13472	2	2	0	1.0000
C...2...%75.	0.0	0	1	0	0.0000
C...2...(...	-0.18671	18	20	23	0.0058
C...2.....	-0.03387	25	24	24	0.0000
C...2...=...	0.18887	23	23	23	0.0008
C...2...C...	-1.69733	1	0	0	1.0000
C...3...(...	-0.70490	14	18	19	0.0070
C...3.....	0.09538	16	18	20	0.0054
C...3...=...	0.92662	14	18	19	0.0070
C...4...(...	-0.80502	1	0	0	1.0000
C...4.....	0.45273	1	0	0	1.0000
C...4...=...	-0.79991	1	0	0	1.0000
C...=...(...	1.47145	21	23	22	0.0018
C...=.....	0.37200	25	24	24	0.0000
C...=...1...	0.19041	25	24	24	0.0000
C...=...2...	1.07788	23	23	23	0.0008
C...=...3...	0.47290	14	18	19	0.0070
C...=...4...	1.25783	1	0	0	1.0000
C...=...C...	-0.54023	25	24	24	0.0000
C...C...(...	0.08720	25	22	23	0.0009
C...C.....	-1.06917	25	24	24	0.0000
C...C...1...	-0.26661	18	19	16	0.0016

C...C...2...	-0.97917	18	17	13	0.0058
C...C...3...	-2.71931	1	0	1	0.0008
C...C...=...	0.29079	24	23	24	0.0008
C...C...C...	-0.07584	15	12	7	0.0140
C...N...(...	0.24351	14	16	18	0.0059
C...N...1...	0.0	0	1	1	0.0000
C...N...2...	-0.13151	3	2	1	0.0196
C...N...3...	-1.44485	2	0	2	0.0008
C...N...=...	0.20381	14	9	13	0.0007
C...O...(...	1.37705	21	21	20	0.0002
C...O...2...	-0.84218	8	6	2	0.0237
C...O...C...	2.47209	13	12	4	0.0208
C...S...C...	-0.48592	1	1	0	1.0000
C...^...%11.	2.46946	6	3	1	0.0283
C...^...%12.	0.21859	2	0	0	1.0000
C...^...%13.	0.0	0	1	2	0.0000
C...^...%14.	-0.56611	3	2	1	0.0196
C...^...%15.	-1.00890	2	1	2	0.0008
C...^...%16.	2.46546	2	2	2	0.0008
C...^...%17.	-0.08949	1	1	0	1.0000
C...^...%18.	0.0	0	2	1	0.0000
C...^...%19.	0.0	0	1	1	0.0000
C...^...%20.	0.0	0	1	1	0.0000
C...^...%21.	0.0	0	1	2	0.0000
C...^...%22.	-0.65712	1	0	3	0.0213
C...^...%23.	-0.53637	2	1	0	1.0000
C...^...%24.	0.0	0	2	1	0.0000
C...^...%25.	-0.67503	1	1	1	0.0008
C...^...%26.	-0.51689	1	0	0	1.0000
C...^...%27.	0.0	0	0	1	0.0000
C...^...%28.	0.0	0	0	1	0.0000
C...^...%29.	0.0	0	2	0	0.0000
C...^...%31.	1.29572	1	0	1	0.0008

SAk	CW(SAk)	N1	N2	N3	dj
C...^...%33.	0.0	0	0	1	0.0000
C...^...%34.	-0.65911	1	0	0	1.0000
C...^...%35.	0.0	0	1	0	0.0000
C...^...%36.	0.0	0	0	2	0.0000
C...^...%38.	0.32401	1	1	0	1.0000
C...^...%40.	-1.72964	1	1	0	1.0000
F...(...(...	0.55967	8	12	11	0.0073
F...(.....	-0.09473	10	12	11	0.0028
F...(...2...	2.54267	1	0	0	1.0000
F...(...4...	0.81142	1	0	0	1.0000
F...(...C...	0.22811	8	12	11	0.0073
F...(...F...	0.00418	8	12	11	0.0073
F.....	0.35226	10	12	11	0.0028
Cl..%11.....	2.56657	4	2	1	0.0237
Cl..%12.....	1.53837	2	0	0	1.0000
Cl..%13.....	0.0	0	1	0	0.0000
Cl..%14.....	0.29031	1	2	1	0.0008
Cl..%15.....	0.0	0	0	2	0.0000
Cl..%18.....	0.0	0	0	1	0.0000
Cl..%19.....	0.0	0	0	1	0.0000
Cl..%20.....	0.0	0	1	1	0.0000
Cl..%21.....	0.0	0	0	1	0.0000
Cl..%22.....	0.0	0	0	3	0.0000
Cl..%23.....	0.0	0	1	0	0.0000
Cl..%28.....	0.0	0	0	1	0.0000
Cl..%29.....	0.0	0	1	0	0.0000
Cl..%33.....	0.0	0	0	1	0.0000
Cl..%34.....	-0.63429	1	0	0	1.0000
Cl..%38.....	-0.75272	1	1	0	1.0000
Cl..%40.....	-0.75242	1	1	0	1.0000
Cl..%51.....	-0.11390	1	3	1	0.0008
Cl..%52.....	-0.74584	2	0	0	1.0000
Cl..%54.....	-0.64237	1	1	1	0.0008

Cl..%55.....	1.05417	2	1	3	0.0090
Cl..%58.....	0.0	0	2	1	0.0000
Cl..%63.....	0.0	0	1	0	0.0000
Cl..(...(...	0.76304	1	0	2	0.0144
Cl..(.....	-0.47755	20	20	19	0.0002
Cl..(...1...	1.22115	3	1	2	0.0073
Cl..(...2...	-1.73883	11	8	12	0.0026
Cl..(...3...	-0.71327	3	4	5	0.0110
Cl..(...C...	2.14175	13	10	8	0.0089
Cl..(...F...	-0.91484	9	11	11	0.0049
Cl..(...Cl..	0.34661	6	6	2	0.0196
Cl.....	0.01504	20	20	19	0.0002
Cl..1.....	0.76465	3	4	2	0.0073
Cl..1...C...	0.66230	3	4	2	0.0073
N...#...(...	0.0	0	2	1	0.0000
N...#.....	0.11496	14	14	13	0.0007
N...#...C...	0.17968	14	14	13	0.0007
N...%51.....	0.0	0	0	1	0.0000
N...%59.....	-1.98573	1	0	1	0.0008
N...(.....	0.25436	20	21	22	0.0028
N...(...1...	1.58873	5	4	7	0.0076
N...(...2...	-0.98342	8	11	11	0.0073
N...(...C...	0.16155	19	20	22	0.0038
N...(...F...	0.0	0	1	0	0.0000
N...(...Cl..	0.0	0	3	5	0.0000
N...(...N...	2.04089	4	1	2	0.0128
N...+.....	0.35402	1	0	1	0.0008
N.....	0.11714	20	21	22	0.0028
N...1...(...	-0.54254	1	1	2	0.0144
N...1.....	0.03043	2	2	2	0.0008
N...1...C...	-0.85868	1	2	1	0.0008
N...2...(...	0.14967	3	2	2	0.0073
N...2.....	-0.86752	6	5	3	0.0128

SAk	CW(SAk)	N1	N2	N3	dj
N...2...C...	2.41908	6	5	3	0.0128
N...3...(...	-0.90351	2	0	2	0.0008
N...3.....	-0.04774	2	0	2	0.0008
N...3...C...	-0.95391	2	0	2	0.0008
N...=...(...	0.27412	5	1	3	0.0094
N...=.....	1.47890	15	14	18	0.0045
N...=...1...	0.48042	1	5	4	0.0253
N...=...2...	0.0	0	1	1	0.0000
N...=...C...	2.18270	15	14	17	0.0034
N...C...#...	1.70983	4	1	2	0.0128
N...C...(...	0.46770	5	4	7	0.0076
N...C.....	0.39939	18	17	20	0.0030
N...C...1...	8.58049	1	1	1	0.0008
N...C...2...	0.33982	4	4	6	0.0090
N...C...3...	0.02710	8	11	11	0.0073
N...C...=...	-0.49696	14	9	13	0.0007
N...C...C...	-0.66536	3	2	1	0.0196
N...N...(...	-0.13909	1	5	5	0.0281
N...N.....	1.14652	1	5	5	0.0281
N...N...=...	-0.44029	1	5	5	0.0281
N...O...C...	0.36998	1	5	4	0.0253
N...[...N...	0.76692	1	0	1	0.0008
O...%16.....	0.0	0	1	0	0.0000
O...%18.....	0.0	0	1	0	0.0000
O...%22.....	0.57125	1	0	0	1.0000
O...%24.....	0.0	0	1	0	0.0000
O...%26.....	1.08865	1	0	0	1.0000
O...%59.....	3.20793	1	0	3	0.0213
O...%75.....	0.0	0	0	1	0.0000
O...(...(...	-0.86867	9	6	5	0.0108
O...(.....	0.26585	25	23	23	0.0009
O...(...1...	-0.70404	8	11	11	0.0073
O...(...2...	4.05757	4	2	5	0.0054

O...(...C...	-0.71161	21	19	15	0.0060
O...(...F...	0.73149	1	0	0	1.0000
O...(...Cl..	0.0	0	1	1	0.0000
O...(...N...	-1.07718	3	3	2	0.0073
O...(...O...	0.72679	17	21	19	0.0031
O...-.....	2.72935	1	0	1	0.0008
O.....	0.03655	25	24	23	0.0009
O...1...(...	0.51694	2	2	0	1.0000
O...1.....	-0.84128	2	2	0	1.0000
O...2...%51.	0.0	0	2	0	0.0000
O...2...%52.	0.02999	6	2	1	0.0283
O...2...%63.	0.31138	1	1	1	0.0008
O...2...%67.	3.37455	1	1	0	1.0000
O...2...(...	0.21809	10	6	5	0.0128
O...2.....	0.45048	17	11	7	0.0162
O...=...(...	-0.05264	18	23	21	0.0040
O...=.....	-0.07135	18	23	21	0.0040
O...C...%55.	0.0	0	1	1	0.0000
O...C...(...	0.15757	11	15	15	0.0071
O...C.....	0.42172	22	24	20	0.0011
O...C...1...	-0.15152	13	11	7	0.0114
O...C...2...	0.04987	8	12	11	0.0073
O...C...3...	0.96259	5	7	7	0.0076
O...C...C...	2.22792	19	19	12	0.0084
O...C...O...	2.10077	8	6	2	0.0237
O...N.....	-0.10110	1	5	4	0.0253
O...N...=...	-0.94172	1	5	4	0.0253
O...P...(...	-0.75236	9	6	5	0.0108
O...[...(...	0.34041	1	0	1	0.0008
P...(.....	-0.57283	9	6	5	0.0108
P...(...=...	0.30762	9	6	5	0.0108
P.....	-0.63293	9	6	5	0.0108
P...O.....	-0.69725	9	6	5	0.0108

SAk	CW(SAk)	N1	N2	N3	dj
P...O...C...	0.14575	9	6	5	0.0108
S...(...(...	0.36804	9	6	5	0.0108
S...(.....	0.10584	9	6	5	0.0108
S.....	1.31315	10	7	5	0.0128
S...=...(...	-0.74849	9	6	5	0.0108
S...=.....	-0.70595	9	6	5	0.0108
S...C...(...	-2.04340	1	1	0	1.0000
S...C.....	0.44800	1	1	0	1.0000
S...C...1...	1.45404	1	1	0	1.0000
X...$.....	1.47967	14	16	12	0.0023
X...%51.....	-0.48599	4	7	5	0.0054
X...%51.2...	1.62944	1	1	1	0.0008
X...%51.C...	0.60586	3	3	3	0.0008
X...%51.Cl..	0.0	0	3	1	0.0000
X...%52.....	0.65780	8	2	1	0.0309
X...%52.2...	-0.14754	6	2	1	0.0283
X...%52.Cl..	-0.54874	2	0	0	1.0000
X...%53.....	0.0	0	2	1	0.0000
X...%53.C...	0.0	0	2	1	0.0000
X...%54.....	0.34323	3	3	1	0.0196
X...%54.2...	2.19572	2	2	0	1.0000
X...%54.Cl..	0.76379	1	1	1	0.0008
X...%55.....	0.64205	2	1	3	0.0090
X...%55.C...	0.0	0	1	1	0.0000
X...%55.Cl..	0.14102	2	0	2	0.0008
X...%58.....	1.67971	1	2	1	0.0008
X...%58.C...	1.37258	1	1	0	1.0000
X...%58.Cl..	0.0	0	1	1	0.0000
X...%59.....	1.75550	1	0	4	0.0253
X...%59.N...	-1.42622	1	0	1	0.0008
X...%59.O...	0.0	0	0	3	0.0000
X...%63.....	1.85320	2	2	2	0.0008
X...%63.2...	-1.63244	1	1	1	0.0008

X...%63.C...	3.97302	1	0	1	0.0008
X...%63.Cl..	0.0	0	1	0	0.0000
X...%67.....	2.49421	1	1	0	1.0000
X...%67.2...	2.26284	1	1	0	1.0000
X...%75.....	0.0	0	1	2	0.0000
X...%75.2...	0.0	0	1	0	0.0000
X...%75.C...	0.0	0	0	1	0.0000
X...%75.O...	0.0	0	0	1	0.0000
X..........	0.68550	22	21	20	0.0011
X...?.....	-1.76192	8	5	8	0.0008
Y...%51.....	-1.07626	2	1	2	0.0008
Y...%51.2...	0.0	0	1	0	0.0000
Y...%51.C...	-0.17014	1	0	1	0.0008
Y...%51.Cl..	-0.02652	1	0	0	1.0000
Y...%51.N...	0.0	0	0	1	0.0000
Y...%55.....	0.0	0	1	1	0.0000
Y...%55.Cl..	0.0	0	1	1	0.0000
Y...%58.....	0.0	0	1	1	0.0000
Y...%58.C...	0.0	0	0	1	0.0000
Y...%58.Cl..	0.0	0	1	0	0.0000
Y...%59.....	2.27548	1	0	0	1.0000
Y...%59.O...	-0.12456	1	0	0	1.0000
Y..........	0.15293	3	3	4	0.0067
Y...?.....	-0.16362	3	3	4	0.0067
[...(.....	-0.27091	1	0	1	0.0008
[...(...=...	-0.34181	1	0	1	0.0008
[...(...C...	1.37861	1	0	1	0.0008
[...(...O...	2.56124	1	0	1	0.0008
[...+.....	1.90619	1	0	1	0.0008
[...+...N...	1.56503	1	0	1	0.0008
[...-.....	1.32550	1	0	1	0.0008
[...-...O...	1.35714	1	0	1	0.0008
[..........	0.03871	1	0	1	0.0008

SAk	CW(SAk)	N1	N2	N3	dj
[...N...(...	0.24887	1	0	1	0.0008
[...N...+...	1.61213	1	0	1	0.0008
[...N.....	1.12090	1	0	1	0.0008
[...O...-...	0.15269	1	0	1	0.0008
[...O.....	0.25454	1	0	1	0.0008
^...%11.....	2.38098	6	3	1	0.0283
^...%11.C...	0.57789	2	1	0	1.0000
^...%11.Cl..	3.48758	4	2	1	0.0237
^...%12.....	0.20112	2	0	0	1.0000
^...%12.Cl..	-0.80679	2	0	0	1.0000
^...%13.....	0.0	0	1	2	0.0000
^...%13.C...	0.0	0	0	2	0.0000
^...%13.Cl..	0.0	0	1	0	0.0000
^...%14.....	-0.72540	3	2	1	0.0196
^...%14.C...	0.19651	2	0	0	1.0000
^...%14.Cl..	0.80833	1	2	1	0.0008
^...%15.....	-1.39717	2	1	2	0.0008
^...%15.C...	-0.87933	2	1	0	1.0000
^...%15.Cl..	0.0	0	0	2	0.0000
^...%16.....	-2.22432	2	2	2	0.0008
^...%16.C...	-1.69967	2	1	2	0.0008
^...%16.O...	0.0	0	1	0	0.0000
^...%17.....	0.64152	1	1	0	1.0000
^...%17.C...	0.39243	1	1	0	1.0000
^...%18.....	0.0	0	2	1	0.0000
^...%18.C...	0.0	0	1	0	0.0000
^...%18.Cl..	0.0	0	0	1	0.0000
^...%18.O...	0.0	0	1	0	0.0000
^...%19.....	0.0	0	1	1	0.0000
^...%19.C...	0.0	0	1	0	0.0000
^...%19.Cl..	0.0	0	0	1	0.0000
^...%20.....	0.0	0	1	1	0.0000
^...%20.Cl..	0.0	0	1	1	0.0000

^...%21.....	0.0	0	1	2	0.0000
^...%21.C...	0.0	0	1	1	0.0000
^...%21.Cl..	0.0	0	0	1	0.0000
^...%22.....	0.39405	1	0	3	0.0213
^...%22.Cl..	0.0	0	0	3	0.0000
^...%22.O...	-0.95634	1	0	0	1.0000
^...%23.....	0.28490	2	1	0	1.0000
^...%23.C...	0.40236	2	0	0	1.0000
^...%23.Cl..	0.0	0	1	0	0.0000
^...%24.....	0.0	0	2	1	0.0000
^...%24.C...	0.0	0	1	1	0.0000
^...%24.O...	0.0	0	1	0	0.0000
^...%25.....	2.05851	1	1	1	0.0008
^...%25.C...	-0.57287	1	1	1	0.0008
^...%26.....	-0.63100	1	0	0	1.0000
^...%26.O...	2.20411	1	0	0	1.0000
^...%27.....	0.0	0	0	1	0.0000
^...%27.C...	0.0	0	0	1	0.0000
^...%28.....	0.0	0	0	1	0.0000
^...%28.Cl..	0.0	0	0	1	0.0000
^...%29.....	0.0	0	2	0	0.0000
^...%29.C...	0.0	0	1	0	0.0000
^...%29.Cl..	0.0	0	1	0	0.0000
^...%31.....	1.23959	1	0	1	0.0008
^...%31.C...	0.01403	1	0	1	0.0008
^...%33.....	0.0	0	0	1	0.0000
^...%33.Cl..	0.0	0	0	1	0.0000
^...%34.....	0.07394	1	0	0	1.0000
^...%34.Cl..	-0.16578	1	0	0	1.0000
^...%35.....	0.0	0	1	0	0.0000
^...%35.C...	0.0	0	1	0	0.0000
^...%36.....	0.0	0	0	2	0.0000
^...%36.C...	0.0	0	0	2	0.0000

SAk	CW(SAk)	N1	N2	N3	dj
^...%38.....	-0.72052	1	1	0	1.0000
^...%38.Cl..	-0.29139	1	1	0	1.0000
^...%40.....	-1.26045	1	1	0	1.0000
^...%40.Cl..	-0.50756	1	1	0	1.0000
^..........	0.51914	25	24	24	0.0000
^...C...(...	0.0	0	2	1	0.0000
^...C.....	3.64507	25	24	24	0.0000
^...C...1...	-0.26741	2	1	5	0.0183
^...C...C...	0.33146	23	20	17	0.0053
^...C...O...	0.0	0	1	1	0.0000

Table S5.5. Categorical model for the synergism (1) / non-synergism (0). Class = true positive, true negative, false positive, false negative. Set = + training; Set = - invisible training; Set = # calibration set; Set = * validation set.

Set	Quasi-SMILES	Observed	Predicted	Class	Applicability
*	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl14%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO252%^	0.	-0.5573	TN	YES
-	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl11%^CCCOCC(=NC1=C(C=C(C=C1)Cl)C(F)(F)F)N2C=CN=C275%^	0.	0.1575	TN	YES
+	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl11%^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl55%^	0.	-0.0883	TN	YES
#	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl12%^C1=CC=C(C(=C1)C2C(O2)(CN3C=NC=N3)C4=CC=C(C=C4)F)Cl55%^	0.	0.2957	TN	YES
*	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl14%^CC(C)(C)C(C(=CC1=CC=C(C=C1)Cl)N2C=NC=N2)O75%^	0.	-0.1567	TN	YES
#	C1CN(C(=N1)N[N+](=O)[O-])CC2=CN=C(C=C2)Cl22%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO252%^	0.	-0.0124	TN	YES
+	C1CN(C(=N1)N[N+](=O)[O-])CC2=CN=C(C=C2)Cl21%^CCCOCC(=NC1=C(C=C(C=C1)Cl)C(F)(F)F)N2C=CN=C275%^	1.	0.6916	TP	YES
-	C1CN(C(=N1)N[N+](=O)[O-])CC2=CN=C(C=C2)Cl23%^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl55%^	0.	0.1808	TN	No
#	C1CSC(=NC#N)N1CC2=CN=C(C=C2)Cl11%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO252%^	0.	-0.7572	TN	YES
#	C1CSC(=NC#N)N1CC2=CN=C(C=C2)Cl11%^CCCOCC(=NC1=C(C=C(C=C1)Cl)C(F)(F)F)N2C=CN=C275%^	0.	-0.2195	TN	YES
*	C1CSC(=NC#N)N1CC2=CN=C(C=C2)Cl11%^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl55%^	0.	-0.4653	TN	YES
+	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl15%^C1=CC=C(C=C1)C(CCC2=CC=C(C=C2)Cl)(CN3C=NC=N3)C#N59%^	0.	0.1234	TN	YES
*	C1CN(C(=N1)N[N+](=O)[O-])CC2=CN=C(C=C2)Cl40%^C1=CC=C(C=C1)C(CCC2=CC=C(C=C2)Cl)(CN3C=NC=N3)C#N59%^	1.	0.7248	TP	YES
+	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C15%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO252%^	0.	0.1613	TN	YES
+	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C31%^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl51%^	1.	0.7991	TP	YES

+	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C23%^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl52%^	0.	0.0763	TN	No
+	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C16%^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl54%^	0.	0.0657	TN	YES
#	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C16%^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl63%^	0.	-0.1289	TN	YES
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl28%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C51%^	1.	0.1729	FN	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl20%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C51%^	0.	0.1065	TN	No
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl18%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C53%^	0.	0.1584	TN	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl15%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C58%^	0.	0.1893	TN	YES
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl11%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C75%^	0.	0.3915	TN	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl11%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C58%^	0.	0.0497	TN	YES
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl13%^CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C53%^	0.	-0.0273	TN	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl14%^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C51%^	0.	0.1060	TN	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl20%^CC1=CC(=C(C=C1)C(C)C)O59%^	0.	0.2823	TN	No
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl15%^COC(=O)N(C1=CC=CC=C1COC2=NN(C=C2)C3=CC=C(C=C3)Cl)OC55%^	0.	-0.1468	TN	YES
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl22%^C1=CC=C(C(=C1)C2=CC=C(C=C2)Cl)NC(=O)C3=C(N=CC=C3)Cl55%^	0.	0.2778	TN	YES
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl18%^C(#N)C1=C(C(=C(C(=C1Cl)Cl)Cl)C#N)Cl58%^	0.	-0.0508	TN	YES

Set	Quasi-SMILES	Observed	Predicted	Class	Applicability
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl11%^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C254%^	0.	-0.1243	TN	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl11%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO252%^	0.	0.0086	TN	YES
*	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C15%^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl54%^	0.	0.2428	TN	YES
+	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C13%^CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C53%^	0.	-0.0142	TN	YES
#	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C16%^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C51%^	0.	0.2587	TN	YES
-	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C25%^CC1=CC(=C(C=C1)C(C)C)O59%^	1.	0.6854	TP	YES
*	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C25%^C1=CC=C(C(=C1)C2=CC=C(C=C2)Cl)NC(=O)C3=C(N=CC=C3)Cl55%^	1.	0.8201	TP	YES
-	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C21%^C(#N)C1=C(C(=C(C(=C1Cl)Cl)Cl)C#N)Cl58%^	0.	0.0486	TN	YES
+	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C11%^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C254%^	0.	-0.1113	TN	YES
-	CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C14%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO252%^	0.	-0.1554	TN	YES
*	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C27%^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl54%^	1.	0.6278	TP	No
#	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C16%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C58%^	0.	-0.0019	TN	YES
+	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C16%^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C51%^	0.	0.2314	TN	YES
#	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C21%^CC1=CC(=C(C=C1)C(C)C)O59%^	0.	0.4438	TN	YES
+	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C17%^COC(=O)N(C1=CC=CC=C1COC2=NN(C=C2)C3=CC=C(C=C3)Cl)OC55%^	0.	-0.2308	TN	No
-	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C27%^C1=CC=C(C(=C1)C2=CC=C(C=C2)Cl)NC(=O)C3=C(N=CC=C3)Cl55%^	1.	0.9820	TP	No

#	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C29%^C(#N)C1=C(C(=C(C(=C1Cl)Cl)Cl)C#N)Cl58%^	1.	0.2682	FN	YES
-	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C11%^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C254%^	0.	-0.1385	TN	YES
-	CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C11%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO252%^	0.	-0.0056	TN	YES
-	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C36%^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl54%^	1.	0.6569	TP	YES
#	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C25%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C58%^	1.	0.5747	TP	YES
-	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C35%^CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C53%^	1.	0.7984	TP	YES
*	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C31%^CC1=CC(=C(C=C1)C(C)C)O59%^	1.	0.8404	TP	YES
-	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C19%^COC(=O)N(C1=CC=CC=C1COC2=NN(C=C2)C3=CC=C(C=C3)Cl)OC55%^	0.	0.0235	TN	YES
*	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C36%^C1=CC=C(C(=C1)C2=CC=C(C=C2)Cl)NC(=O)C3=C(N=CC=C3)Cl55%^	1.	1.0111	TP	YES
*	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C28%^C(#N)C1=C(C(=C(C(=C1Cl)Cl)Cl)C#N)Cl58%^	1.	0.2371	FN	YES
#	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C24%^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C254%^	0.	0.2464	TN	YES
*	CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C23%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO252%^	0.	0.0542	TN	No
+	CC1=CC(=C(C=C1)C(C)C)O16%^CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl54%^	0.	-0.0609	TN	YES
-	CC1=CC(=C(C=C1)C(C)C)O18%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C58%^	0.	-0.0018	TN	YES
-	CC1=CC(=C(C=C1)C(C)C)O23%^CC1=NN(C(=C1C=NOCC2=CC=C(C=C2)C(=O)OC(C)(C)C)OC3=CC=CC=C3)C53%^	0.	-0.0387	TN	No
-	CC1=CC(=C(C=C1)C(C)C)O26%^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C51%^	1.	0.6429	TP	YES

Set	Quasi-SMILES	Observed	Predicted	Class	Applicability
-	CC1=CC(=C(C=C1)C(C)C)O20%^COC(=O)N(C1=CC=CC=C1COC2=NN(C=C2)C3=CC=C(C=C3)Cl)OC55%^	0.	-0.1561	TN	No
+	CC1=CC(=C(C=C1)C(C)C)O27%^C1=CC=C(C(=C1)C2=CC=C(C=C2)Cl)NC(=O)C3=C(N=CC=C3)Cl55%^	1.	1.0051	TP	No
*	CC1=CC(=C(C=C1)C(C)C)O21%^C(#N)C1=C(C(=C(C(=C1Cl)Cl)Cl)C#N)Cl58%^	0.	0.0445	TN	YES
+	CC1=CC(=C(C=C1)C(C)C)O24%^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C254%^	0.	0.0931	TN	YES
#	CC1=CC(=C(C=C1)C(C)C)O22%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO252%^	0.	-0.0003	TN	YES
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl34%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO251%^	1.	1.0713	TP	No
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl34%^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C251%^	1.	1.0282	TP	No
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl40%^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl51%^	1.	0.4607	FN	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl33%^C1=CC=C(C=C1)C(CCC2=CC=C(C=C2)Cl)(CN3C=NC=N3)C#N51%^	1.	0.9218	TP	No
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl38%^CC1(CCC(C1(CN2C=NC=N2)O)CC3=CC=C(C=C3)Cl)C51%^	1.	0.7306	TP	YES
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl38%^CCCCC(CN1C=NC=N1)(C#N)C2=CC=C(C=C2)Cl51%^	1.	0.6500	TP	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl29%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO251%^	1.	0.4928	FN	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl21%^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C251%^	0.	0.2029	TN	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl23%^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl51%^	0.	-0.0113	TN	No
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl24%^C1=CC=C(C=C1)C(CCC2=CC=C(C=C2)Cl)(CN3C=NC=N3)C#N53%^	0.	0.2500	TN	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl22%^CC1(CCC(C1(CN2C=NC=N2)O)CC3=CC=C(C=C3)Cl)C51%^	0.	-0.0710	TN	YES

-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl19%^CCCCC(CN1C=NC=N1)(C#N)C2=CC=C(C=C2)Cl51%^	0.	0.1626	TN	YES
-	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl14%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO251%^	0.	0.0691	TN	YES
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl11%^CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C252%^	0.	-0.0345	TN	YES
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl12%^CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl52%^	0.	-0.2958	TN	YES
+	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl12%^C1=CC=C(C=C1)C(CCC2=CC=C(C=C2)Cl)(CN3C=NC=N3)C#N81%^	0.	-0.1868	TN	No
#	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl11%^CC1(CCC(C1(CN2C=NC=N2)O)CC3=CC=C(C=C3)Cl)C52%^	0.	-0.1421	TN	YES
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl11%^CCCCC(CN1C=NC=N1)(C#N)C2=CC=C(C=C2)Cl54%^	0.	-0.3125	TN	YES
-	CC1(C(C1C(=O)OC(C#N)C2=CC(=C(C=C2)F)OC3=CC=CC=C3)C=C(Cl)Cl)C11%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO267%^	0.	-0.0764	TN	No
*	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)C=C(C(F)(F)F)Cl)C11%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO267%^	0.	-0.2233	TN	No
*	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl11%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO267%^	0.	0.2567	TN	No
#	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C19%^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C51%^	0.	0.3385	TN	YES
*	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C14%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C63%^	0.	-0.3572	TN	YES
-	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C16%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO263%^	0.	-0.2752	TN	YES
*	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C24%^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C51%^	0.	0.3993	TN	YES
*	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C14%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C63%^	0.	-0.3572	TN	YES
*	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C18%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO263%^	0.	-0.1756	TN	YES

Set	Quasi-SMILES	Observed	Predicted	Class	Applicability
-	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C17%^CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C51%^	0.	0.2607	TN	No
#	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C13%^CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C63%^	0.	-0.2765	TN	YES
#	CC(=CC1C(C1(C)C)C(=O)OCC2=CC(=CC=C2)OC3=CC=CC=C3)C15%^CCCCOCCOCCOCC1=CC2=C(C=C1CCC)OCO263%^	0.	-0.0981	TN	YES

Table S5.6. Full list of correlation weights (CW) for the calculation of the classification model for synergism/non-synergism (Approach B) via semi-correlations (Toropova et al., 2019).

SAk	CW(SAk)	N1	N2	N3	d
#.....	-0.37199	16	10	14	0.0028
%.....	0.99911	24	24	24	0.0000
(...#.....	0.0	0	1	2	0.0000
(...(.....	-0.67628	22	18	18	0.0042
(.....	-0.10765	24	24	24	0.0000
+.....	0.49466	1	1	1	0.0000
-.....	0.21235	1	1	1	0.0000
0...%.....	0.61890	1	2	1	0.0000
0.....	1.17261	1	2	1	0.0000
1...%.....	1.65318	12	13	10	0.0038
1...(.....	0.07970	23	22	22	0.0009
1.....	-0.52381	24	24	24	0.0000
1...1.....	0.66820	3	5	3	0.0000
2...%.....	-0.42299	4	3	7	0.0114
2...(.....	-0.50714	23	23	23	0.0000
2.....	0.15384	24	24	24	0.0000
2...0.....	-0.82330	1	2	0	1.0000
2...1.....	0.01675	3	1	3	0.0000
2...2.....	0.32367	1	0	3	0.0208
3...%.....	-0.76470	4	4	4	0.0000
3...(.....	-0.48366	20	18	17	0.0034
3.....	0.82031	20	18	17	0.0034
3...1.....	1.10961	2	0	1	0.0139
3...2.....	-0.64967	2	2	0	1.0000
3...3.....	1.73579	1	0	0	1.0000
4...%.....	0.19659	5	5	2	0.0179
4...(.....	0.0	0	0	1	0.0000
4.....	0.43877	5	5	4	0.0046
4...0.....	0.0	0	0	1	0.0000
4...1.....	0.0	0	3	0	0.0000
4...2.....	0.90473	1	0	2	0.0139

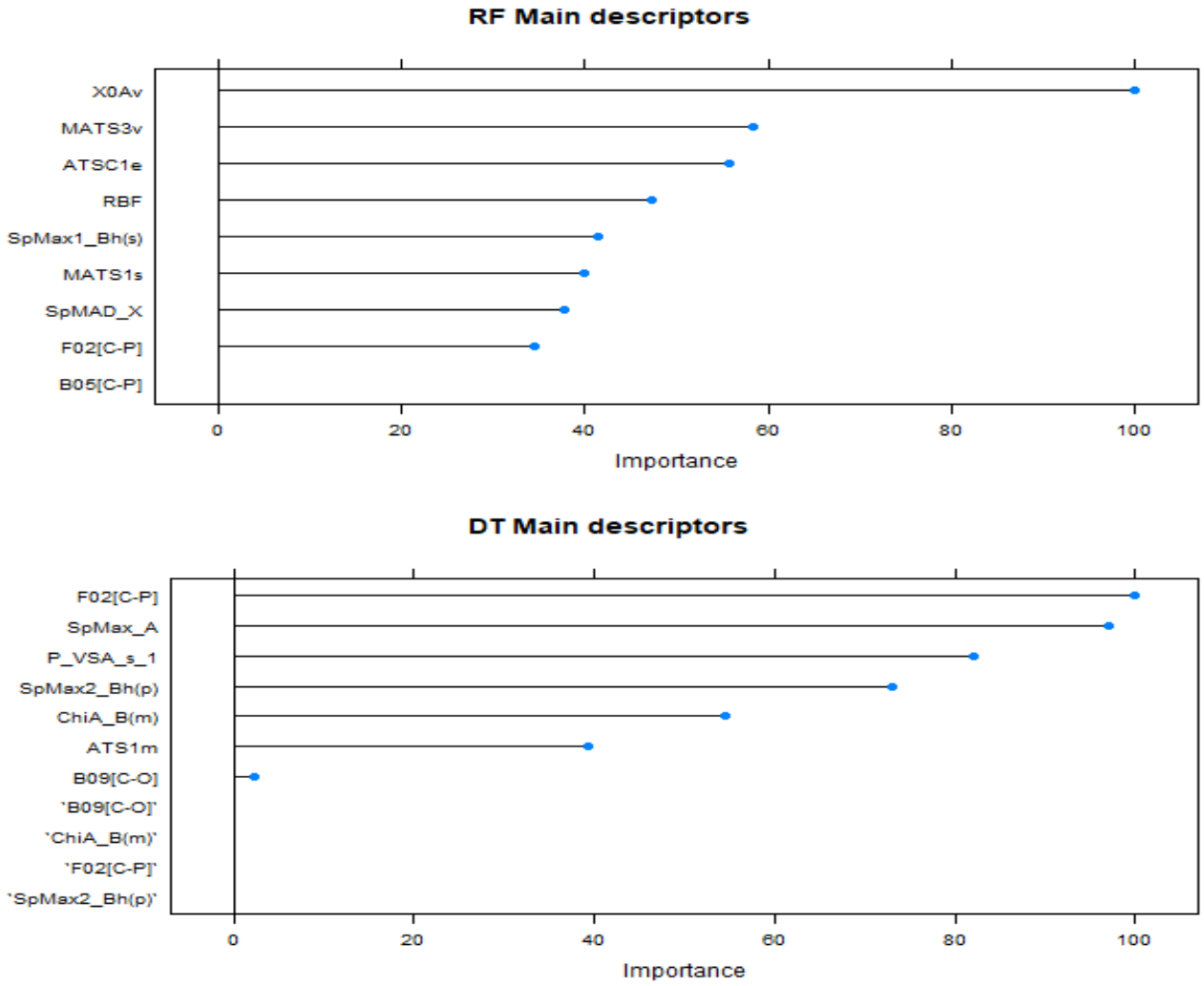
4...3.....	2.01614	1	0	0	1.0000
5...%.....	0.82772	8	7	5	0.0096
5.....	0.91470	23	22	22	0.0009
5...1.....	0.97180	11	7	7	0.0093
5...2.....	0.81079	6	7	7	0.0032
5...3.....	1.79941	1	2	1	0.0000
5...4.....	-0.68363	4	2	1	0.0250
5...5.....	0.66382	4	4	2	0.0139
6...%.....	1.27549	3	3	3	0.0000
6.....	1.19957	3	4	5	0.0104
6...1.....	-0.92332	3	1	3	0.0000
6...2.....	0.0	0	3	1	0.0000
6...3.....	0.0	0	2	3	0.0000
7...%.....	1.29900	2	3	0	1.0000
7.....	1.39751	3	4	1	0.0208
7...1.....	-0.44092	1	1	0	1.0000
7...2.....	1.09809	2	2	1	0.0139
7...5.....	1.26300	1	1	1	0.0000
7...6.....	0.0	0	1	0	0.0000
8...%.....	0.76050	2	2	7	0.0231
8.....	0.64354	3	2	7	0.0167
8...1.....	0.80120	1	1	1	0.0000
8...2.....	0.0	0	0	1	0.0000
8...3.....	0.0	0	0	2	0.0000
8...5.....	0.14491	2	2	4	0.0139
9...%.....	1.24211	2	4	3	0.0083
9.....	1.52442	2	4	3	0.0083
9...1.....	0.0	0	2	1	0.0000
9...2.....	0.0	0	1	1	0.0000
9...5.....	0.28565	2	1	1	0.0139
=...(.....	0.61935	24	24	24	0.0000
=.....	0.08002	24	24	24	0.0000
=...1.....	1.43017	24	23	24	0.0000

SAk	CW(SAk)	N1	N2	N3	d
=...2.....	-0.22890	22	23	23	0.0009
=...3.....	-0.02044	19	18	16	0.0036
=...4.....	0.0	0	0	1	0.0000
C...#.....	-0.42419	16	10	14	0.0028
C...(.....	0.61505	24	24	24	0.0000
C.....	-0.44876	24	24	24	0.0000
C...1.....	-0.42305	24	24	24	0.0000
C...2.....	0.52485	24	24	24	0.0000
C...3.....	0.21057	20	18	17	0.0034
C...4.....	0.0	0	0	1	0.0000
C...5.....	-0.30458	7	9	7	0.0000
C...6.....	0.0	0	0	1	0.0000
C...=.....	0.06767	24	24	24	0.0000
C...C.....	0.61080	24	24	24	0.0000
F...(.....	-0.08248	15	9	12	0.0046
F.....	0.23542	15	9	12	0.0046
Cl..(.....	0.26766	23	18	20	0.0029
Cl.....	-0.15999	23	18	20	0.0029
Cl..1.....	-0.40034	10	7	8	0.0046
Cl..2.....	-0.42781	5	3	4	0.0046
Cl..3.....	2.81789	2	0	2	0.0000
Cl..4.....	0.0	0	0	1	0.0000
Cl..5.....	1.66655	7	5	7	0.0000
Cl..6.....	0.0	0	0	1	0.0000
N...#.....	-0.36539	16	10	14	0.0028
N...(.....	-0.49911	23	20	21	0.0019
N...+.....	0.13744	1	1	1	0.0000
N.....	0.39346	23	20	21	0.0019
N...1.....	0.36062	1	2	4	0.0250
N...2.....	0.00400	8	3	6	0.0060
N...3.....	0.25943	3	0	2	0.0083
N...5.....	0.32164	2	0	1	0.0139
N...8.....	0.09245	1	0	0	1.0000

N...=.....	0.25980	16	14	18	0.0025
N...C.....	0.38652	21	15	18	0.0032
N...N.....	-0.52629	4	7	3	0.0060
O...(.....	-0.06360	23	23	23	0.0000
O...-.....	-0.36148	1	1	1	0.0000
O.....	0.09349	23	24	24	0.0009
O...1.....	-0.07301	3	2	2	0.0083
O...2.....	1.67920	10	14	11	0.0020
O...5.....	1.07952	1	1	1	0.0000
O...=.....	-0.53253	22	22	20	0.0020
O...C.....	0.28645	22	23	23	0.0009
O...N.....	-0.85222	3	5	3	0.0000
P...(.....	-0.32524	8	5	6	0.0060
P.....	-0.32840	8	5	6	0.0060
P...O.....	0.29233	8	5	6	0.0060
S...(.....	0.13730	8	5	6	0.0060
S.....	0.21811	8	5	8	0.0000
S...=.....	-0.51398	8	5	6	0.0060
S...C.....	0.0	0	0	2	0.0000
[...(.....	0.10747	1	1	1	0.0000
[...+.....	0.12550	1	1	1	0.0000
[...-.....	0.25444	1	1	1	0.0000
[.....	0.32371	1	1	1	0.0000
[...N.....	0.01292	1	1	1	0.0000
[...O.....	0.09802	1	1	1	0.0000
^...%.....	0.66801	24	24	24	0.0000
^.....	0.84936	24	24	24	0.0000
^...C.....	1.35534	24	24	24	0.0000

Chapter 6

Figure S6.1. Variable importace of classification models. RF= random forest; DT= decision tree; KNN= k-nearest neighbors.



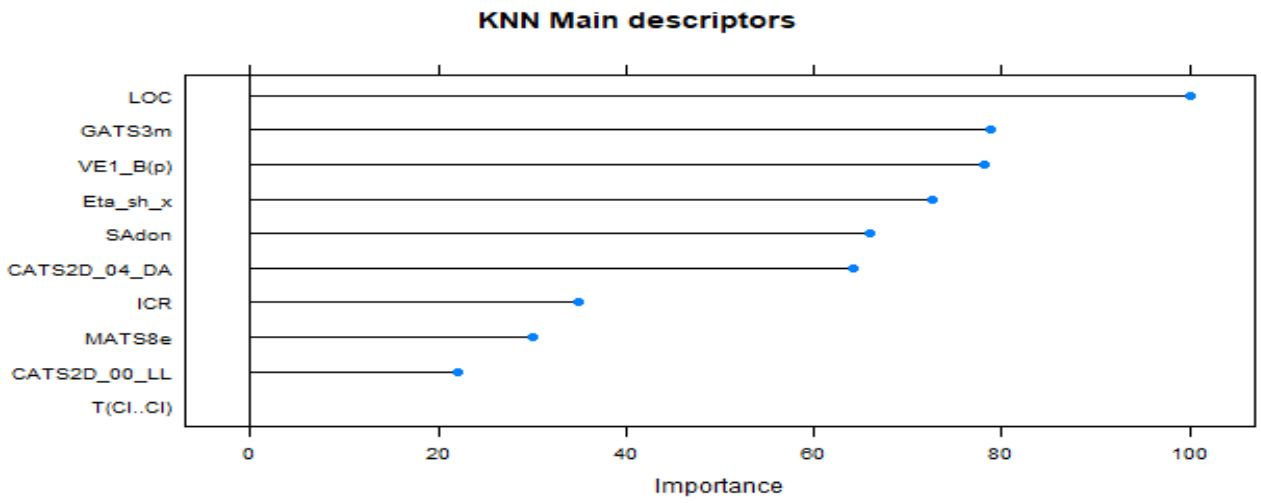
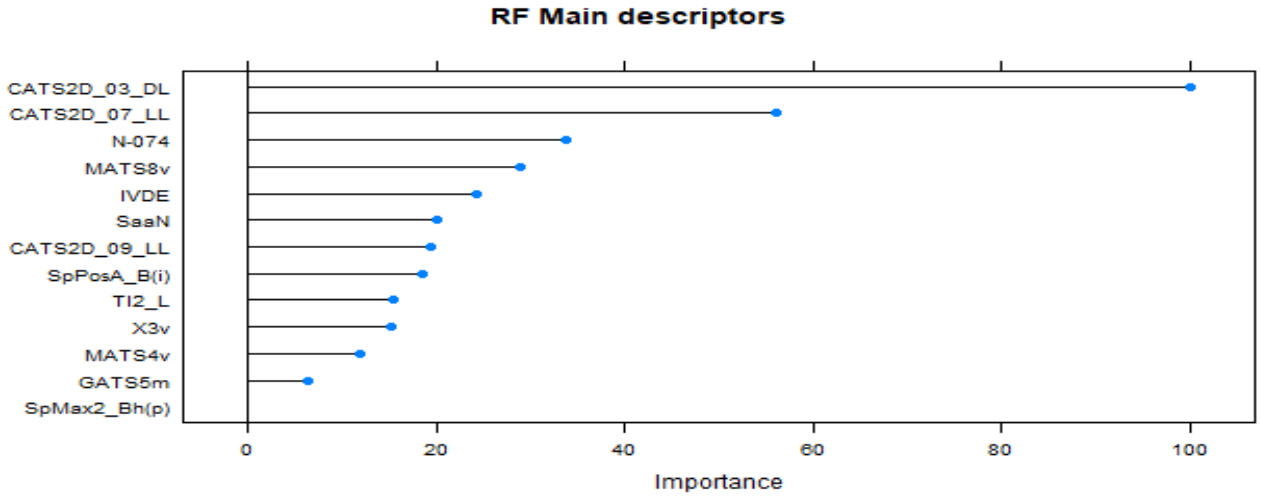
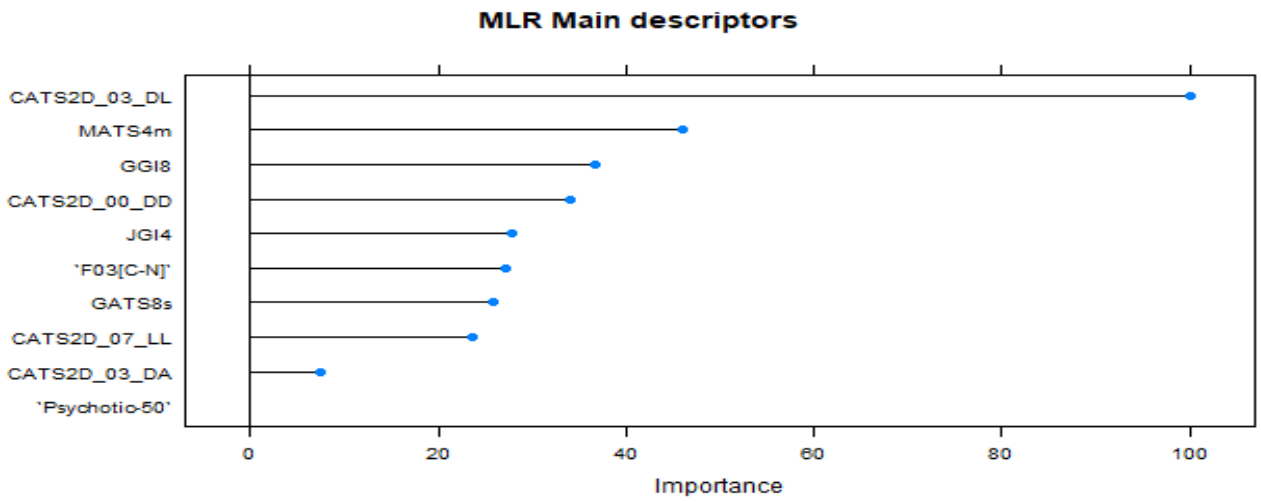
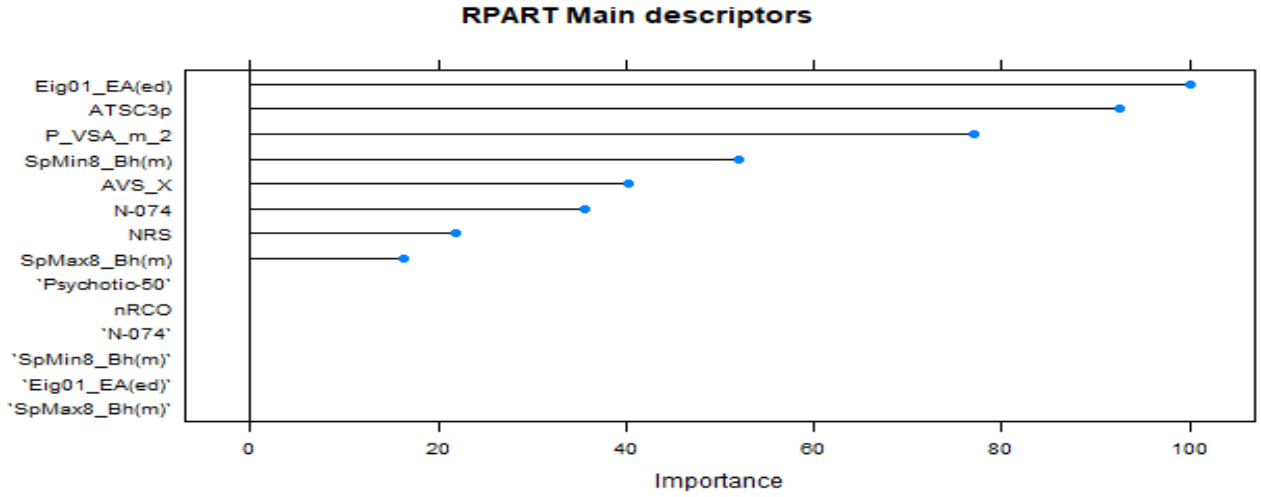


Figure S6.2. Variable importance of regression models. RF= random forest; MLR= multiple linear regression; RPART= recursive partitioning and regression trees; KNN= KNN= k-nearest neighbors.




MLR Main descriptors
CATS2D_03_DL
MATS4m
GGI8
CATS2D_00_DD
JGI4
`F03[C-N]`
GATS8s
CATS2D_07_LL
CATS2D_03_DA
`Psychotic-50`
0
20
40
60
80
100
Importance


RPART Main descriptors
Eig01_EA(ed)
ATSC3p
P_VSA_m_2
SpMin8_Bh(m)
AVS_X
N-074
NRS
SpMax8_Bh(m)
`Psychotic-50`
nRCO
`N-074`
`SpMin8_Bh(m)`
`Eig01_EA(ed)`
`SpMax8_Bh(m)`
0
20
40
60
80
100
Importance

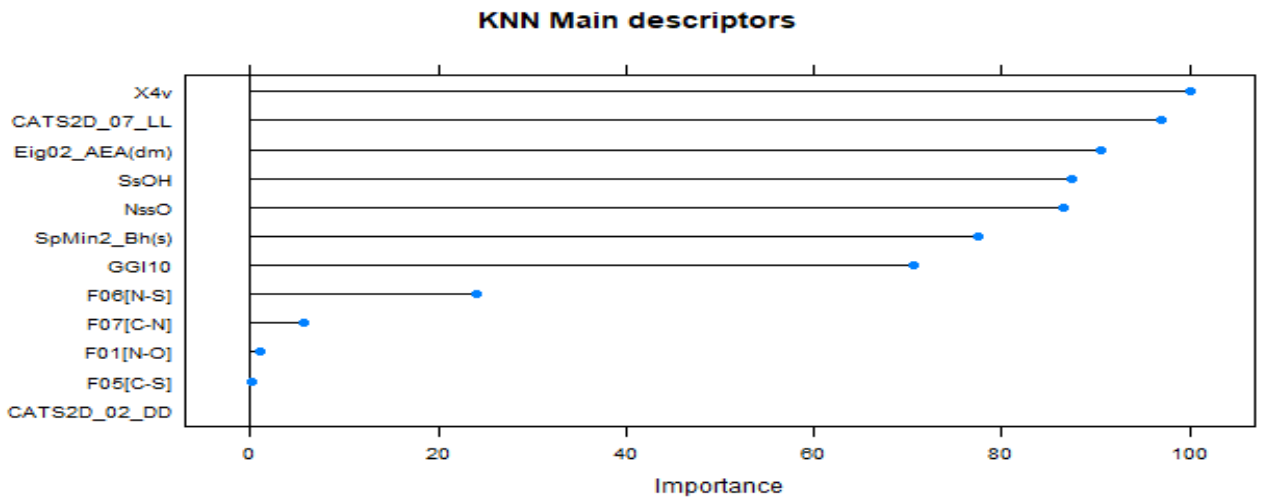
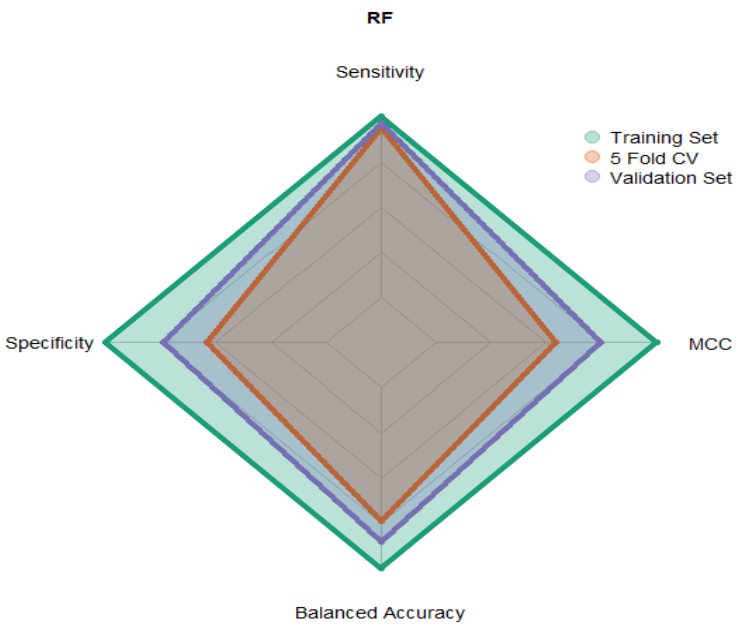
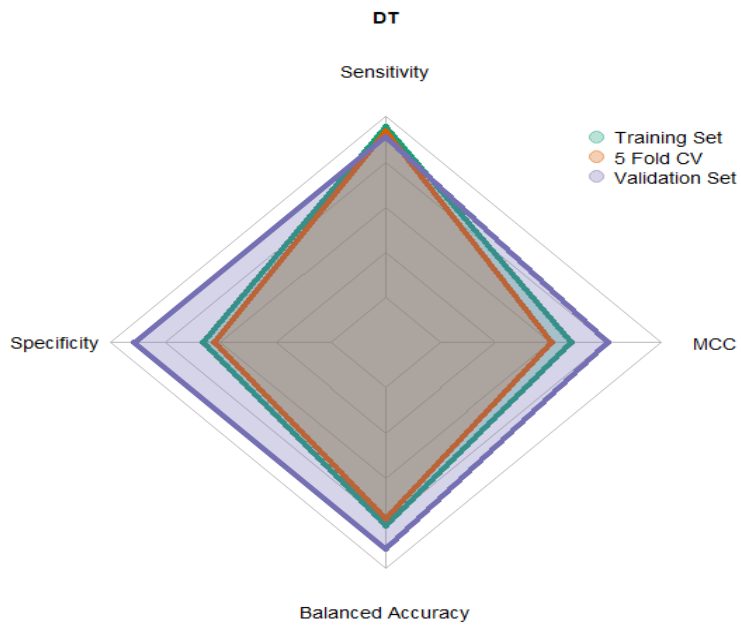


Figure S6.3. Radar Plots for Classification Models. RF= random forest; DT= decision tree; KNN= k-nearest neighbors.


RF
Sensitivity
Specificity
MCC
Balanced Accuracy
Training Set
5 Fold CV
Validation Set


DT
Sensitivity
Specificity
MCC
Balanced Accuracy
Training Set
5 Fold CV
Validation Set

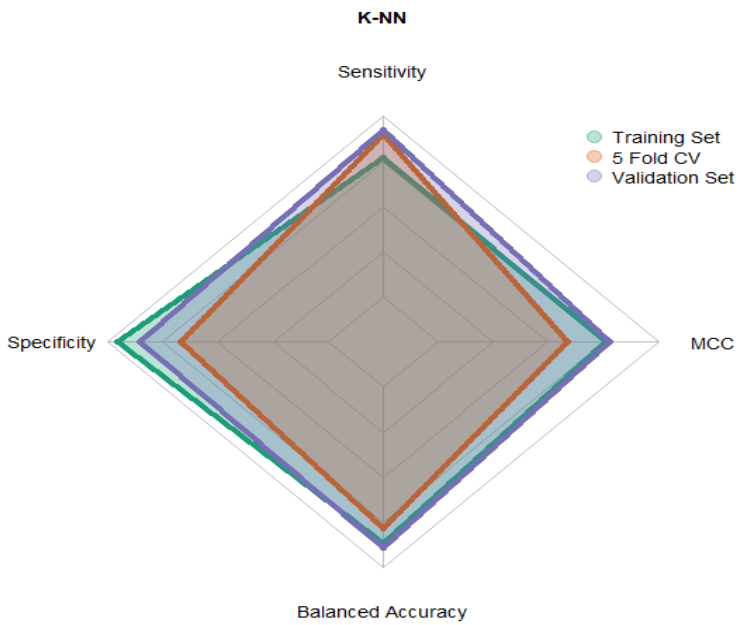
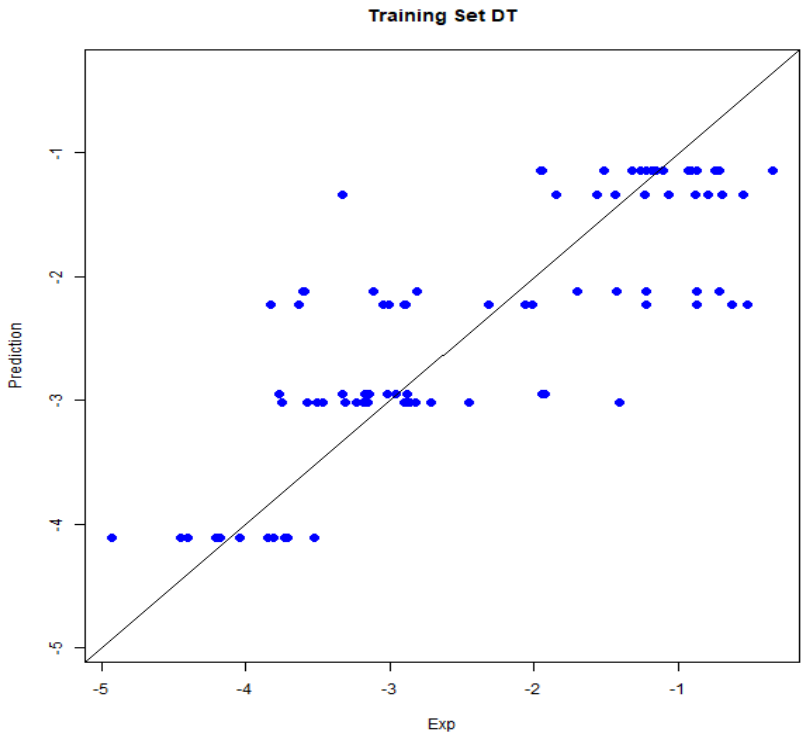
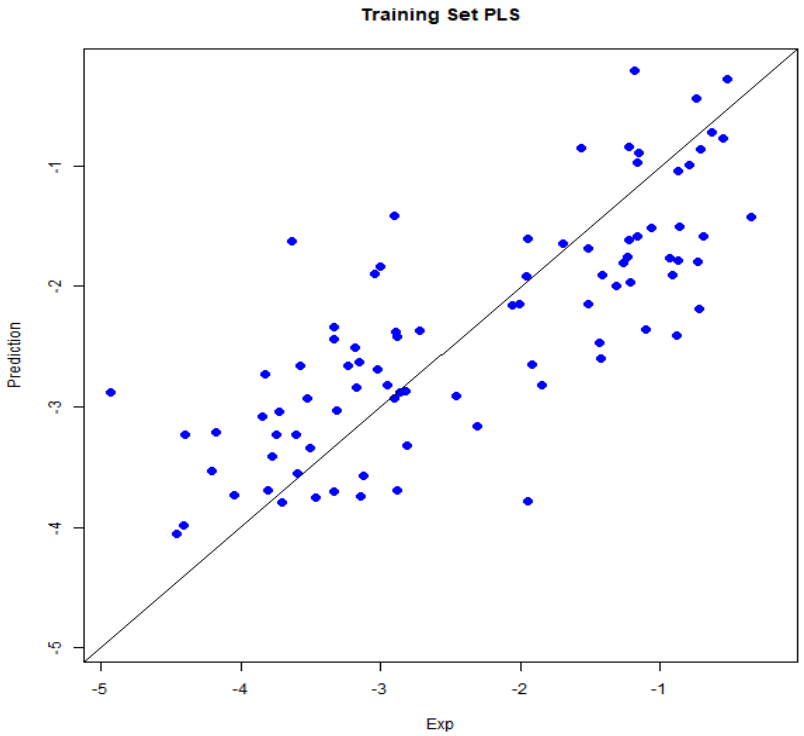
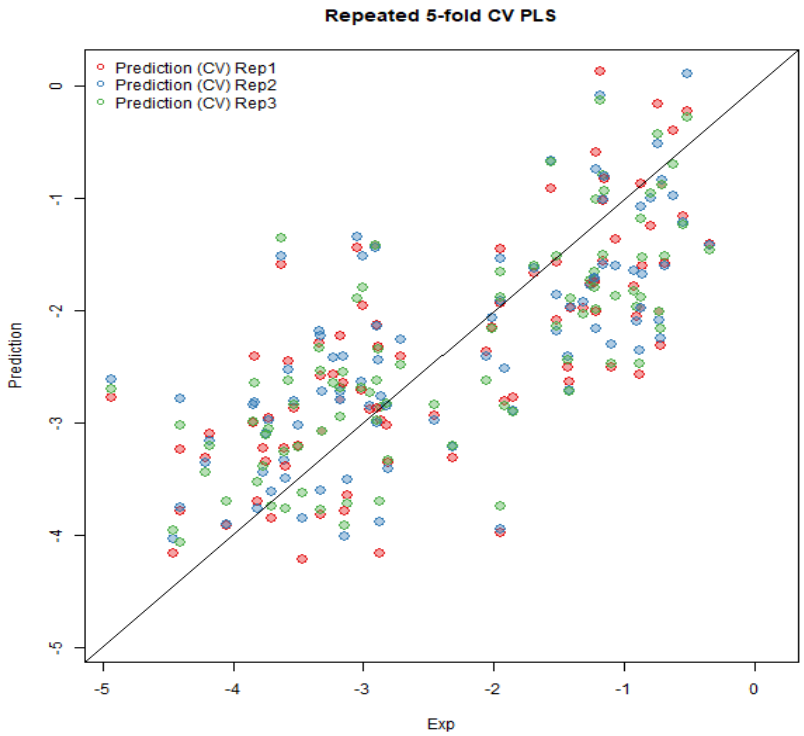
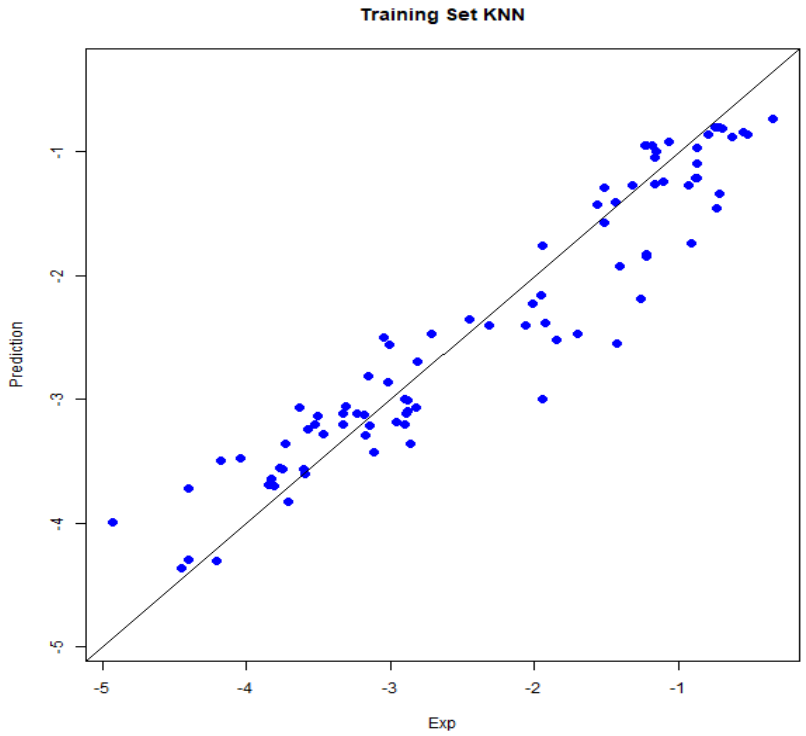


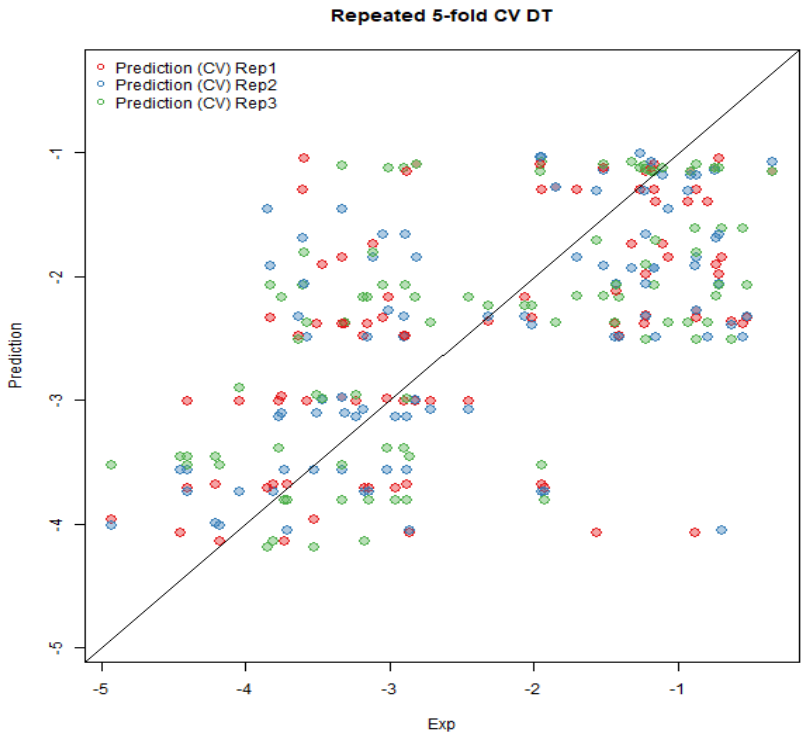
Figure S6.4. Scatter plots for Regression Model4. Training and validation sets for decision tree (DT), partial least squares (PLS), k-nearest neighbors (KNN) and cross validation partial least squares regression (CV-PLS) are reported.


Training Set DT
Prediction
Exp
-1
-2
-3
-4
-5


Training Set PLS
Prediction
Exp
-1
-2
-3
-4
-5


Repeated 5-fold CV PLS
Prediction (CV) Rep1
Prediction (CV) Rep2
Prediction (CV) Rep3
Prediction
Exp
0
-1
-2
-3
-4
-5


Training Set KNN
Prediction
Exp
-1
-2
-3
-4
-5



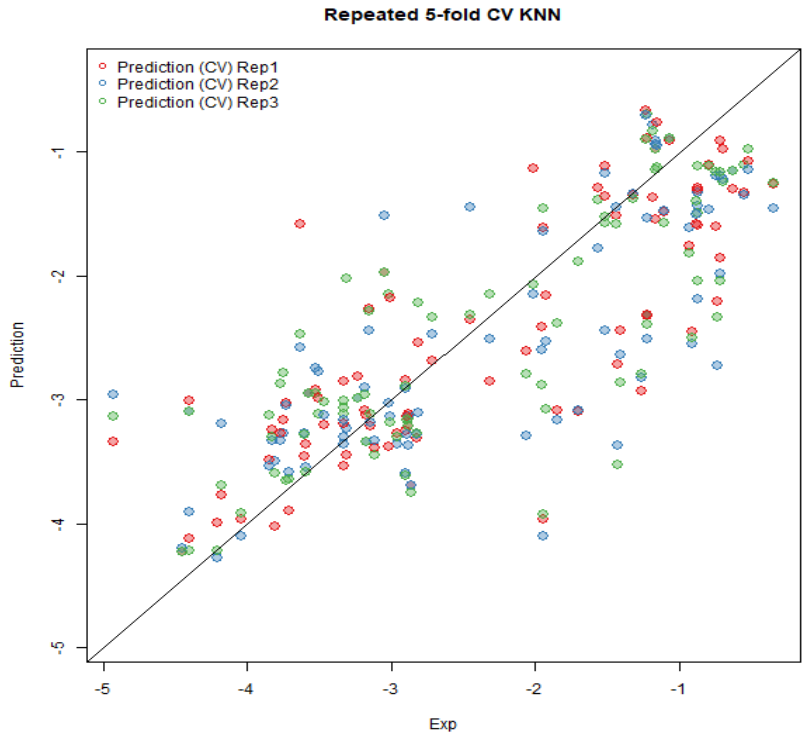


Table S6.1. Split train and test for classification model.

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
chloropicrin	0	353	Train	O=[N+]([O-])C(Cl)(Cl)Cl	76-06-2
(4R)-1-Methyl-4-(1-methylethenyl)cyclohexene	0	3	Train	C=C(C)C1CC=C(C)CC1	5989-27-5
amitrole	0	405	Train	n1c[nH]nc1N	61-82-5
Ethanedioic acid	0	144	Train	O=C(O)C(=O)O	144-62-7
Dimethyl disulfide	0	5	Train	CSSC	624-92-0
fluroxypyr	0	157	Train	O=C(O)COc1nc(F)c(c(N)c1Cl)Cl	69377-81-7
propamocarb hydrochloride| propamocarb	0	239	Train	O=C(OCCC)NCCCN(C)C	25606-41-1| 24579-73-5
fluensulfone	0	348	Train	O=S(=O)(c1ncc(Cl)s1)CCC(F)=C(F)F	318290-98-1
2,4,5,6-Tetrachloro-1,3-benzenedicarbonitrile	0	44	Train	N#Cc1c(c(C#N)c(c(c1Cl)Cl)Cl)Cl	1897-45-6
mesotrione	0	272	Train	O=C(c1ccc(cc1[N+](=O)[O-])S(=O)(=O)C)C2C(=O)CCCC2(=O)	104206-82-8
maleic hydrazide	0	292	Train	O=C1C=CC(=O)NN1	123-33-1
ametoctradin	0	412	Train	n1cnn2c1nc(c(c2(N))CCCCCCCC)CC	865318-97-4
metaldehyde	0	66	Train	O1C(OC(OC(OC1C)C)C)C	108-62-3
4,5-Dihydro-3-methoxy-4-methyl-5-oxo-N-[[2-(trifluoromethoxy)phenyl]sulfonyl]-1H-1,2,4-triazole-1-carboxamide, Sodium salt (1:1)	0	114	Train	O=C(NS(=O)(=O)c1ccccc1(OC(F)(F)F))N2N=C(OC)N(C2(=O))C	181274-17-9
N-[5-(1,1-Dimethylethyl)-1,3,4-thiadiazol-2-yl]-N,N'-dimethylurea	0	104	Train	O=C(NC)N(c1nnc(C(C)(C)C)s1)C	34014-18-1
cycloxydim| Cycloxydim	0	285	Train	O=C1C(=C(O)CC(C1)C2CCCSC2)C(=NOCC)CCC	101205-02-1
P-(Aminocarbonyl)phosphonic acid monoethyl ester ammonium salt (1:1)	0	85	Train	O=C(N)P(=O)(O)OCC	25954-13-6

metrafenone| (3-Bromo-6-methoxy-2-methylphenyl)(2,3,4-trimethoxy-6-methylphenyl)methanone	0	267	Train	O=C(c1c(OC)c(OC)c(OC)cc1C)c2c(OC)ccc(c2C)Br	220899-03-6
hexazinone| 3-Cyclohexyl-6-(dimethylamino)-1-methyl-1,3,5-triazine-2,4-(1H,3H)dione	0	297	Train	O=C1N=C(N(C(=O)N1C2CCCCC2)C)N(C)C	51235-04-2
mepiquat chloride| mepiquat	0	6	Train	C[N+]1(C)(CCCCC1)	24307-26-4| 15302-91-7
diuron| N'-(3,4-Dichlorophenyl)-N,N-dimethylurea	0	121	Train	O=C(Nc1ccc(c(c1)Cl)Cl)N(C)C	330-54-1
fenpropimorph| (2R,6S)-rel-4-[3-[4-(1,1-Dimethylethyl)phenyl]-2-methylpropyl]-2,6-dimethylmorpholine	0	65	Train	O1C(C)CN(CC1C)CC(C)Cc2ccc(cc2)C(C)(C)C	67564-91-4
tri-allate	0	71	Train	O=C(N(C(C)C)C(C)C)SCC(=C(Cl)Cl)Cl	2303-17-5
amidosulfuron| Amidosulfuron	0	338	Train	O=S(=O)(NC(O)=Nc1nc(OC)cc(n1)OC)N(C)S(=O)(=O)C	120923-37-7
captan| 3a,4,7,7a-Tetrahydro-2-[(trichloromethyl)thio]-1H-isoindole-1,3-(2H)-dione	0	293	Train	O=C1N(C(=O)C2CC=CCC12)SC(Cl)(Cl)Cl	133-06-2
propineb	0	54	Train	N(C(=S)S)CC(NC(=S)S)C	12071-83-9
Mono(2,2-dimethylhydrazide)butanedioic acid	0	150	Train	O=C(O)CCC(=O)NN(C)C	1596-84-5
pyroxasulfone	0	334	Train	O=S(=O)(C1=NOC(C)(C)C1)Cc2c(OC(F)F)n(nc2C(F)(F)F)C	447399-55-5

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
benthiavalicarb isopropyl	0	183	Train	O=C(OC(C)C)NC(C(=O)NC(c2nc1ccc(F)cc1s2)C)C(C)C	177406-68-7
tolclofos-methyl| Tolclofos-methyl	0	61	Train	O(c1c(cc(cc1Cl)C)Cl)P(OC)(OC)=S	57018-04-9
sodium o-nitrophenolate	0	369	Train	O=[N+]([O-])c1ccccc1(O)	824-39-5
methiozolin	0	21	Train	Fc1cccc(F)c1COCC3(ON=C(c2c(ccs2)C)C3)(C)	403640-27-7
N,N'-[1,4-Piperazinediylbis(2,2,2-trichloroethylidene)]bis-formamide	0	323	Train	O=CNC(N1CCN(CC1)C(NC=O)C(Cl)(Cl)Cl)C(Cl)(Cl)Cl	26644-46-2
prothioconazole	0	378	Train	OC(Cc1ccccc1Cl)(CN2NC=NC2=S)C3(CC3)Cl	178928-70-6
pinoxaden| Pinoxaden	0	218	Train	O=C(OC1=C(C(=O)N2N1CCOCC2)c3c(cc(cc3CC)C)CC)C(C)(C)C	243973-20-8
aminocyclopyrachlor	0	170	Train	O=C(O)c1nc(nc(N)c1Cl)C2CC2	858956-08-8
2-Furancarboxaldehyde	0	324	Train	O=Cc1occc1	98-01-1
fuberidazole	0	404	Train	n1c3ccccc3([nH]c1c2occc2)	3878-19-1
bixafen	0	24	Train	Fc2ccc(N=C(O)c1cn(nc1C(F)F)C)c(c2)c3ccc(c(c3)Cl)Cl	581809-46-3
1-naphthylacetic acid	0	159	Train	O=C(O)Cc2cccc1ccccc12	86-87-3
glufosinate-ammonium| glufosinate| Glufosinate-ammonium| 2-Amino-4-(hydroxymethylphosphinyl)butanoic acid monoammonium salt	0	145	Train	O=C(O)C(N)CCP(=O)(O)C	77182-82-2| 51276-47-2
fludioxonil	0	46	Train	N#Cc1c[nH]cc1c3cccc2OC(F)(F)Oc23	131341-86-1
2,6-Dinitro-N1,N1-dipropyl-4-(trifluoromethyl)-1,3-benzenediamine	0	358	Train	O=[N+]([O-])c1cc(c(N)c(c1N(CCC)CCC)[N+](=O)[O-])C(F)(F)F	29091-21-2
lactofen	0	179	Train	O=C(OC(C(=O)OCC)C)c2cc(Oc1ccc(cc1Cl)C(F)(F)F)ccc2[N+](=O)[O-]	77501-63-4

indaziflam| N2-[(1R,2S)-2,3-Dihydro-2,6-dimethyl-1H-inden-1-yl]-6-(1-fluoroethyl)-1,3,5-triazine-2,4-diamine	0	16	Train	FC(c1nc(nc(n1)NC3c2cc(ccc2CC3(C))C)N)C	950782-86-2
zoxamide| 3,5-Dichloro-N-(3-chloro-1-ethyl-1-methyl-2-oxopropyl)-4-methylbenzamide	0	97	Train	O=C(NC(C(=O)CCl)(C)CC)c1cc(c(c(c1)Cl)C)Cl	156052-68-5
cymoxanil	0	29	Train	N#CC(=NOC)C(=O)NC(=O)NCC	57966-95-7
4-Amino-N-(1,1-dimethylethyl)-4,5-dihydro-3-(1-methylethyl)-5-oxo-1H-1,2,4-triazole-1-carboxamide	0	99	Train	O=C(NC(C)(C)C)N1N=C(N(N)C1(=O))C(C)C	129909-90-6
2-(Difluoromethyl)-5-(4,5-dihydro-2-thiazolyl)-4-(2-methylpropyl)-6-(trifluoromethyl)-3-pyridinecarboxylic acid methyl ester	0	216	Train	O=C(OC)c2c(nc(c(C1=NCCS1)c2CC(C)C)C(F)(F)F)C(F)F	117718-60-2
fenoxycarb| [2-(4-Phenoxyphenoxy)ethyl]carbamic acid ethyl ester	0	236	Train	O=C(OCC)NCCOc2ccc(Oc1ccccc1)cc2	79127-80-3| 72490-01-8
6,7-Dihydrodipyrido[1,2-a:2',1'-c]pyrazinediium bromide (1:2)	0	390	Train	c1cc[n+]3c(c1)c2cccc[n+]2CC3	85-00-7
carfentrazone-ethyl	0	229	Train	O=C(OCC)C(Cc1cc(c(F)cc1Cl)N2N=C(N(C2(=O))C(F)F)C)Cl	128639-02-1
hymexazol	0	291	Train	O=C1C=C(ON1)C	10004-44-1
thifluzamide	0	8	Train	FC(F)(F)Oc2cc(c(N=C(O)c1c(nc(C)s1)C(F)(F)F)c(c2)Br)Br	130000-40-7

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
prohexadione| prohexadione-calcium| 3,5-Dioxo-4-(1-oxopropyl)cyclohexanecarboxylic acid calcium ion(1-), Calcium salt (2:1:1)	0	148	Train	O=C(O)C1CC(=O)C(C(=O)C1)=C(O)CC	88805-35-0| 127277-53-6
pyridate| Carbonothioic acid, O-(6-Chloro-3-phenyl-4-pyridazinyl) S-octyl ester	0	256	Train	O=C(Oc1cc(nnc1c2ccccc2)Cl)SCCCCCCCC	55512-33-9
chlorthal-dimethyl	0	206	Train	O=C(OC)c1c(c(c(C(=O)OC)c(c1Cl)Cl)Cl)Cl	1861-32-1
flumioxazin| 2-[7-Fluoro-3,4-dihydro-3-oxo-4-(2-propynyl)-2H-1,4-benzoxazin-6-yl]-4,5,6,7-tetrahydro-1H-isoindole-1,3(2H)-dione	0	289	Train	O=C1C4=C(C(=O)N1c3cc2c(OCC(=O)N2CC#C)cc3(F))CCCC4	103361-09-7
chlorsulfuron| Chlorsulfuron	0	140	Train	O=C(Nc1nc(nc(n1)C)OC)NS(=O)(=O)c2ccccc2Cl	64902-72-3
bispyribac-sodium	0	176	Train	O=C(O)c3c(Oc1nc(OC)cc(n1)OC)cccc3(Oc2nc(OC)cc(n2)OC)	125401-92-5
prochloraz	0	279	Train	O=C(n1cncc1)N(CCOc2c(cc(cc2Cl)Cl)Cl)CCC	67747-09-5
dithianon	0	40	Train	N#CC2=C(C#N)SC=1C(=O)c3ccccc3(C(=O)C=1S2)	3347-22-6
mefenpyr diethyl	0	232	Train	O=C(OCC)C2=NN(c1ccc(cc1Cl)Cl)C(C(=O)OCC)(C)C2	135590-91-9
2-[[4-Chloro-6-(ethylamino)-1,3,5-triazin-2-yl]amino]-2-methylpropanenitrile	0	31	Train	N#CC(Nc1nc(nc(n1)Cl)NCC)(C)C	21725-46-2
5-(trifluoromethyl)-1H-pyrazole-3-carboxylic acid	0	167	Train	O=C(O)c1n[nH]c(c1)C(F)(F)F	129768-28-1
acibenzolar-S-methyl| Acibenzolar-S-methyl| 1,2,3-Benzothiadiazole-7-carbothioic acid S-methyl ester	0	278	Train	O=C(c2cccc1nnsc12)SC	135158-54-2

cyflufenamid	0	87	Train	O=C(NC(=NOCC1CC1)c2c(F)c(F)ccc2C(F)(F)F)Cc3ccccc3	180409-60-3
tembotrione	0	270	Train	O=C(c1ccc(c(c1Cl)COCC(F)(F)F)S(=O)(=O)C)C2C(=O)CCCC2(=O)	335104-84-2
thiobencarb	0	76	Train	O=C(N(CC)CC)SCc1ccc(cc1)Cl	28249-77-6
asulam| Asulam Sodium	0	205	Train	O=C(OC)NS(=O)(=O)c1ccc(N)cc1	3337-71-1| 2302-17-2
(2S)-2-Hydroxypropanoic acid	0	146	Train	O=C(O)C(O)C	79-33-4
5-Bromo-6-methyl-3-(1-methylpropyl)-2,4(1H,3H)-pyrimidinedione	0	299	Train	O=C1NC(=C(C(=O)N1C(C)CC)Br)C	314-40-9
triazoxide| Triazoxide	0	374	Train	O=[N+]2[N-]C(=Nc1ccc(cc12)Cl)n3cncc3	72459-58-6
fluthiacet methyl| [[2-Chloro-4-fluoro-5-[(tetrahydro-3-oxo-1H,3H-[1,3,4]thiadiazolo[3,4-a]pyridazin-1-ylidine)amino]phenyl]thioacetic acid methyl ester	0	201	Train	O=C(OC)CSc3cc(N=C1N2N(C(=O)S1)CCCC2)c(F)cc3Cl	117337-19-6
fenhexamid| Fenhexamid| N-(2,3-Dichloro-4-hydroxyphenyl)-1-methylcyclohexanecarboxamide	0	117	Train	O=C(Nc1ccc(O)c(c1Cl)Cl)C2(C)(CCCCC2)	126833-17-8
amisulbrom	0	349	Train	O=S(=O)(c1ncn(n1)S(=O)(=O)N(C)C)n3c2cc(F)ccc2c(c3C)Br	348635-87-0
imazosulfuron	0	339	Train	O=S(=O)(NC(O)=Nc1nc(OC)cc(n1)OC)c3c(nc2ccccn23)Cl	122548-33-8
chloridazon| 5-Amino-4-chloro-2-phenyl-3(2H)-pyridazinone	0	282	Train	O=C1C(=C(N)C=NN1c2ccccc2)Cl	1698-60-8
isofetamid	0	98	Train	O=C(NC(C(=O)c1ccc(OC(C)C)cc1C)(C)C)c2c(ccs2)C	875915-78-9

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
3,6-Dichloro-2-pyridinecarboxylic acid compd. with 2-aminoethanol (1:1)	0	169	Train	O=C(O)c1nc(ccc1Cl)Cl	57754-85-5
metalaxyl| metalaxyl-M	0	196	Train	O=C(OC)C(N(C(=O)COC)c1c(cccc1C)C)C	57837-19-1| 70630-17-0
bromoxynil heptanoate	0	47	Train	N#Cc1cc(c(OC(=O)CCCCCC)c(c1)Br)Br	56634-95-8
triafamone	0	277	Train	O=C(c1nc(nc(n1)OC)OC)c2cccc(F)c2N(C)S(=O)(=O)C(F)F	874195-61-6
flurtamone	0	310	Train	O=C2C(c1cccc(c1)C(F)(F)F)=C(OC2c3ccccc3)NC	96525-23-4
oxadiazon| Oxadiazon	0	300	Train	O=C1OC(=NN1c2cc(OC(C)C)c(cc2Cl)Cl)C(C)(C)C	19666-30-9
quizalofop-P-tefuryl	0	238	Train	O=C(OCC1OCCC1)C(Oc4ccc(Oc2nc3ccc(cc3(nc2))Cl)cc4)C	119738-06-6
N-Phenyl-N'-1,2,3-thiadiazol-5-yl-urea	0	133	Train	O=C(Nc1ccccc1)Nc2cnns2	51707-55-2
profoxydim	0	284	Train	O=C1C(=C(O)CC(C1)C2CCCSC2)C(=NOCC(Oc3ccc(cc3)Cl)C)CCC	139001-49-3
bistrifluron	0	91	Train	O=C(NC(=O)c1c(F)cccc1(F))Nc2cc(cc(c2Cl)C(F)(F)F)C(F)(F)F	201593-84-2
spirotetramat| Spirotetramat	0	219	Train	O=C(OC1=C(C(=O)NC12(CCC(OC)CC2))c3cc(ccc3C)C)OCC	203313-25-1
naptalam| 2-[(1-Naphthalenylamino)carbonyl] benzoic acid	0	177	Train	O=C(O)c3ccccc3(C(=O)Nc2cccc1ccccc12)	132-66-1
bitertanol	0	377	Train	OC(C(Oc1ccc(cc1)c2ccccc2)n3ncnc3)C(C)(C)C	55179-31-2
etoxazole| 2-(2,6-Difluorophenyl)-4-[4-(1,1-dimethylethyl)-2-ethoxyphenyl]-4,5-dihydrooxazole	0	20	Train	Fc1cccc(F)c1C3=NC(c2ccc(cc2(OCC))C(C)(C)C)CO3	153233-91-1

flutianil	0	27	Train	N#CC(=C2N(c1ccccc1(OC))CCS2)Sc3cc(ccc3(F))C(F)(F)F	958647-10-4
N'-(2,4-Dimethylphenyl)-N-[[(2,4-dimethylphenyl)imino]methyl]-N-methylmethanimidanide	0	53	Train	N(=CN(C=Nc1ccc(cc1C)C)C)c2ccc(cc2C)C	33089-61-1
penoxsulam| 2-(2,2-Difluoroethoxy)-N-(5,8-dimethoxy[1,2,4]triazolo[1,5-c]pyrimidin-2-yl)-6-(trifluoromethyl)benzenesulfonamide	0	344	Train	O=S(=O)(Nc1nc2c(OC)cnc(OC)n2(n1))c3c(OCC(F)F)cccc3C(F)(F)F	219714-96-2
penthiopyrad| N-[2-(1,3-Dimethylbutyl)-3-thienyl]-1-methyl-3-(trifluoromethyl)-1H-pyrazole-4-carboxamide	0	134	Train	O=C(Nc1ccsc1C(C)CC(C)C)c2cn(nc2C(F)(F)F)C	183675-82-3
2,4-D	0	155	Train	O=C(O)COc1ccc(cc1Cl)Cl	94-75-7
triflusulfuron-methyl	0	209	Train	O=C(OC)c1cccc(c1S(=O)(=O)NC(=O)Nc2nc(nc(n2)N(C)C)OCC(F)(F)F)C	126535-15-7
dicloran| Dicloran| 2,6-Dichloro-4-nitrobenzenamine	0	357	Train	O=[N+]([O-])c1cc(c(N)c(c1)Cl)Cl	99-30-9
isoxaflutole| (5-Cyclopropyl-4-isoxazolyl)[2-(methylsulfonyl)-4-(trifluoromethyl)phenyl]methanone	0	276	Train	O=C(c1cnoc1C2CC2)c3ccc(cc3S(=O)(=O)C)C(F)(F)F	141112-29-0
flupyradifurone	0	304	Train	O=C1OCC(=C1)N(Cc2cnc(cc2)Cl)CC(F)F	951659-40-8
cyazofamid| 4-Chloro-2-cyano-N,N-dimethyl-5-(4-methylphenyl)-1H-imidazole-1-sulfonamide	0	51	Train	N#Cc2nc(c(c1ccc(cc1)C)n2S(=O)(=O)N(C)C)Cl	120116-88-3
aclonifen| Aclonifen	0	373	Train	O=[N+]([O-])c2ccc(Oc1ccccc1)c(c2(N))Cl	74070-46-5

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
dimethomorph	0	70	Train	O=C(C=C(c1ccc(cc1)Cl)c2ccc(OC)c(OC)c2)N3CCOCC3	110488-70-5
pyriofenone	0	269	Train	O=C(c1c(ncc(c1C)Cl)OC)c2c(OC)c(OC)c(OC)cc2C	688046-61-9
florpyrauxifen-benzyl	0	246	Train	O=C(OCc1ccccc1)c2nc(c(F)c(N)c2Cl)c3ccc(c(OC)c3(F))Cl	1390661-72-9
phenmedipham| (3-Methylphenyl)carbamic acid 3-[(methoxycarbonyl)amino]phenyl ester	0	259	Train	O=C(Oc1cccc(c1)NC(=O)OC)Nc2cccc(c2)C	13684-63-4
N-[[4-(Dimethylamino)phenyl]imino]sulfamic acid sodium salt (1:1)	0	346	Train	O=S(=O)(O)N=Nc1ccc(cc1)N(C)C	140-56-7
fluquinconazole	0	317	Train	O=C2c4cc(F)ccc4(N=C(n1ncnc1)N2c3ccc(cc3Cl)Cl)	136426-54-5
flufenacet	0	83	Train	O=C(N(c1ccc(F)cc1)C(C)C)COc2nnc(C(F)(F)F)s2	142459-58-3
dimefuron	0	301	Train	O=C1OC(=NN1c2ccc(N=C(O)N(C)C)cc2Cl)C(C)(C)C	34205-21-5
carboxin| 5,6-Dihydro-2-methyl-N-phenyl-1,4-oxathiin-3-carboxamide	0	130	Train	O=C(Nc1ccccc1)C2=C(OCCS2)C	5234-68-4
pyrimethanil| 4,6-Dimethyl-N-phenyl-2-pyrimidinamine	0	395	Train	n1c(nc(cc1C)C)Nc2ccccc2	53112-28-0
fenpyrazamine	0	283	Train	O=C1C(=C(N)N(C(=O)SCC=C)N1C(C)C)c2ccccc2C	473798-59-3
fluazinam| 3-Chloro-N-[3-chloro-2,6-dinitro-4-(trifluoromethyl)phenyl]-5-(trifluoromethyl)-2-pyridinamine	0	370	Train	O=[N+]([O-])c2cc(c(c(c2(Nc1ncc(cc1Cl)C(F)(F)F))[N+](=O)[O-])Cl)C(F)(F)F	79622-59-6

imazapyr	0	175	Train	O=C(O)c2cccnc2(C1=NC(C(=O)N1)(C)C(C)C)	81334-34-1
bentazone| 3-(1-Methylethyl)-1H-2,1,3-benzothiadiazin-4(3H)-one 2,2-dioxide sodium salt (1:1)	0	308	Train	O=C1c2ccccc2(NS(=O)(=O)N1C(C)C)	25057-89-0| 50723-80-3
flurochloridone	0	312	Train	O=C2N(c1cccc(c1)C(F)(F)F)CC(CCl)C2Cl	61213-25-0
valifenalate| Valiphenal	0	182	Train	O=C(OC(C)C)NC(C(=O)NC(c1ccc(cc1)Cl)CC(=O)OC)C(C)C	283159-90-0
cinidon-ethyl	0	228	Train	O=C(OCC)C(=Cc1cc(ccc1Cl)N2C(=O)C3=C(C2(=O))CCCC3)Cl	142891-20-1
fenpicoxamid	0	244	Train	O=C(OCOc3c(OC)ccnc3(C(=O)NC1C(=O)OC(C)C(OC(=O)C(C)C)C(C(=O)OC1)Cc2ccccc2))C(C)C	517875-34-2
fenarimol| alpha-(2-Chlorophenyl)-alpha-(4-chlorophenyl)-5-pyrimidinemethanol	0	382	Train	OC(c1cncnc1)(c2ccc(cc2)Cl)c3ccccc3Cl	60168-88-9
thiencarbazone-methyl| Thiencarbazone-methyl	0	215	Train	O=C(OC)c1csc(c1S(=O)(=O)NC(=O)N2N=C(OC)N(C2(=O))C)C	317815-83-1
clomazone| 2-[(2-Chlorophenyl)methyl]-4,4-dimethyl-3-isoxazolidinone	0	296	Train	O=C1N(OCC1(C)C)Cc2ccccc2Cl	81777-89-1
oxathiapiprolin	0	86	Train	O=C(N4CCC(c1nc(cs1)C3=NOC(c2c(F)cccc2(F))C3)CC4)Cn5nc(cc5C)C(F)(F)F	1003318-67-9
fluoxastrobin| Fluoxastrobin| (1E)-[2-[[6-(2-Chlorophenoxy)-5-fluoro-4-pyrimidinyl]oxy]phenyl](5,6-dihydro-1,4,2-dioxazin-3-yl)-methanone, O-Methyloxime	0	23	Train	Fc2c(ncnc2(Oc1ccccc1Cl))Oc4ccccc4(C(=NOC)C3=NOCCO3)	361377-29-9

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
cyflumetofen	0	30	Train	N#CC(C(=O)OCCOC)(C(=O)c1ccccc1C(F)(F)F)c2ccc(cc2)C(C)(C)C	400882-07-7
glyphosate trimesium| glyphosate| N-(Phosphonomethyl)glycine compd. with 2-propanamine (1:1)	0	153	Train	O=C(O)CNCP(=O)(O)O	81591-81-3| 1071-83-6| 38641-94-0
buprofezin	0	313	Train	O=C2N(c1ccccc1)CSC(=NC(C)(C)C)N2C(C)C	69327-76-0
isoxaben	0	141	Train	O=C(Nc1onc(c1)C(C)(CC)CC)c2c(OC)cccc2(OC)	82558-50-7
metamitron	0	288	Train	O=C1C(=NN=C(N1(N))C)c2ccccc2	41394-05-2
dicamba	0	161	Train	O=C(O)c1c(OC)c(ccc1Cl)Cl	1918-00-9
metosulam| Metosulam	0	343	Train	O=S(=O)(Nc1c(ccc(c1Cl)C)Cl)c2nc3nc(OC)cc(OC)n3(n2)	139528-85-1
halosulfuron-methyl| 3-Chloro-5-[[[[(4,6-dimethoxy-2-pyrimidinyl)amino]carbonyl]amino]sulfonyl]-1-methyl-1H-pyrazole-4-carboxylic acid, Methyl ester	0	207	Train	O=C(OC)c1c(nn(c1S(=O)(=O)NC(=O)Nc2nc(OC)cc(n2)OC)C)Cl	100784-20-1
florasulam| Florasulam| N-(2,6-Difluorophenyl)-8-fluoro-5-methoxy-[1,2,4]triazolo[1,5-c]pyrimidine-2-sulfonamide	0	341	Train	O=S(=O)(Nc1c(F)cccc1(F))c2nc3c(F)cnc(OC)n3(n2)	145701-23-1
flubendiamide| N2-[1,1-Dimethyl-2-(methylsulfonyl)ethyl]-3-iodo-N1-[2-methyl-4-[1,2,2,2-tetrafluoro-1-(trifluoromethyl)ethyl]phenyl]-1,2-benzenedicarboxamide	0	126	Train	O=C(Nc1ccc(cc1C)C(F)(C(F)(F)F)C(F)(F)F)c2cccc(c2(C(=O)NC(C)(C)CS(=O)(=O)C))I	272451-65-7

3-(3,5-Dichlorophenyl)-5-ethenyl-5-methyl-2,4-oxazolidinedione	0	302	Train	O=C1OC(C=C)(C(=O)N1c2cc(cc(c2)Cl)Cl)C	50471-44-8
metazachlor	0	78	Train	O=C(N(c1c(cccc1C)C)Cn2nccc2)CCl	67129-08-2
tralkoxydim| 2-[1-(Ethoxyimino)propyl]-3-hydroxy-5-(2,4,6-trimethylphenyl)-2-cyclohexen-1-one	0	309	Train	O=C2C(=C(O)CC(c1c(cc(cc1C)C)C)C2)C(=NOCC)CC	87820-88-0
pyridalyl	0	11	Train	FC(F)(F)c2cnc(OCCCOc1c(cc(OCC=C(Cl)Cl)cc1Cl)Cl)cc2	179101-81-6
kresoxim-methyl| Kresoxim-methyl	0	195	Train	O=C(OC)C(=NOC)c1ccccc1COc2ccccc2C	143390-89-0
pyraclostrobin| [2-[[[1-(4-Chlorophenyl)-1H-pyrazol-3-yl]oxy]methyl]phenyl]methoxycarbamic acid methyl ester	0	203	Train	O=C(OC)N(OC)c1ccccc1COc2nn(cc2)c3ccc(cc3)Cl	175013-18-0
quinmerac| Quinmerac	0	162	Train	O=C(O)c1c2ncc(cc2(ccc1Cl))C	90717-03-6
picoxystrobin	0	193	Train	O=C(OC)C(=COC)c1ccccc1COc2nc(ccc2)C(F)(F)F	117428-22-5
tebufenozide| Tebufenozide	0	110	Train	O=C(NN(C(=O)c1cc(cc(c1)C)C)C(C)(C)C)c2ccc(cc2)CC	112410-23-8
2-Chloro-5-[3,6-dihydro-3-methyl-2,6-dioxo-4-(trifluoromethyl)-1(2H)-pyrimidinyl]benzoic acid 1,1-dimethyl-2-oxo-2-(2-propenyloxy)ethyl ester	0	180	Train	O=C(OC(C(=O)OCC=C)(C)C)c1cc(ccc1Cl)N2C(=O)C=C(N(C2(=O))C)C(F)(F)F	134605-64-4

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
topramezone| Topramezone| [3-(4,5-Dihydro-3-isoxazolyl)-2-methyl-4-(methylsulfonyl)phenyl](5-hydroxy-1-methyl-1H-pyrazol-4-yl)methanone	0	275	Train	O=C(c1cnn(c1(O))C)c3ccc(c(C2=NOCC2)c3C)S(=O)(=O)C	210631-68-8
N-[[[3,5-Dichloro-4-(1,1,2,2-tetrafluoroethoxy)phenyl]amino]carbonyl]-2,6-difluorobenzamide	0	90	Train	O=C(NC(=O)c1c(F)cccc1(F))Nc2cc(c(OC(F)(F)C(F)F)c(c2)Cl)Cl	86479-06-3
cyhalofop-butyl| (2R)-2-[4-(4-Cyano-2-fluorophenoxy)phenoxy]propanoic acid, Butyl ester	0	50	Train	N#Cc2ccc(Oc1ccc(OC(C(=O)OCCCC)C)cc1)c(F)c2	122008-85-9
diflufenican| Diflufenican	0	115	Train	O=C(Nc1ccc(F)cc1(F))c3cccnc3(Oc2cccc(c2)C(F)(F)F)	83164-33-4
orthosulfamuron| Orthosulfamuron| 2-[[[[[(4,6-Dimethoxy-2-pyrimidinyl)amino]carbonyl]amino]sulfonyl]amino]-N,N-dimethylbenzamide	0	135	Train	O=C(Nc1nc(OC)cc(n1)OC)NS(=O)(=O)Nc2ccccc2(C(=O)N(C)C)	213464-77-8
ethoxysulfuron	0	347	Train	O=S(=O)(Oc1ccccc1(OCC))NC(O)=Nc2nc(OC)cc(n2)OC	126801-58-9
thiophanate-methyl| N,N'-[1,2-Phenylenebis(iminocarbonothioyl)]bis-C,C'- Dimethyl ester carbamic acid	0	204	Train	O=C(OC)NC(Nc1ccccc1(NC(NC(=O)OC)=S))=S	23564-05-8
tricyclazole	0	413	Train	n1nc3n(c1)c2c(cccc2s3)C	41814-78-2
flonicamid| N-(Cyanomethyl)-4-(trifluoromethyl)-3-pyridinecarboxamide	0	41	Train	N#CCNC(=O)c1cnccc1C(F)(F)F	158062-67-0
mandestrobin	0	102	Train	O=C(NC)C(OC)c1ccccc1COc2cc(ccc2C)C	173662-97-0

propiconazole	0	411	Train	n1cnn(c1)CC2(OCC(O2)CCC)c3ccc(cc3Cl)Cl	60207-90-1
spiromesifen| Spiromesifen| 3,3-Dimethylbutanoic acid 2-oxo-3-(2,4,6-trimethylphenyl)-1-oaxspiro[4.4]non-3-en-4-yl ester	0	303	Train	O=C1OC3(C(OC(=O)CC(C)(C)C)=C1c2c(cc(cc2C)C)C)(CCCC3)	283594-90-1
pydiflumetofen	0	274	Train	O=C(c1cn(nc1C(F)F)C)N(OC)C(C)Cc2c(cc(cc2Cl)Cl)Cl	1228284-64-7
pyraflufen-ethyl| Pyraflufen-ethyl| 2-[2-Chloro-5-[4-chloro-5-(difluoromethoxy)-1-methyl-1H-pyrazol-3-yl]-4-fluorophenoxy]acetic acid ethyl ester	0	234	Train	O=C(OCC)COc1cc(c(F)cc1Cl)c2nn(c(OC(F)F)c2Cl)C	129630-19-9
diphenamid| N,N-Dimethyl-alpha-phenylbenzeneacetamide	0	73	Train	O=C(N(C)C)C(c1ccccc1)c2ccccc2	957-51-7
2,6-Dichlorobenzonitrile	0	45	Train	N#Cc1c(cccc1Cl)Cl	1194-65-6
N-[2,4-Dichloro-5-[4-(difluoromethyl)-4,5-dihydro-3-methyl-5-oxo-1H-1,2,4-triazol-1-yl]phenyl]methanesulfonamide	0	295	Train	O=C1N(N=C(N1C(F)F)C)c2cc(NS(=O)(=O)C)c(cc2Cl)Cl	122836-35-5
pymetrozine| 4,5-Dihydro-6-methyl-4-[(E)-(3-pyridinylmethylene)amino]-1,2,4-triazin-3(2H)-one	0	314	Train	O=C2NN=C(C)CN2(N=Cc1cnccc1)	123312-89-0
tolylfluanid| Tolylfluanid	0	335	Train	O=S(=O)(N(c1ccc(cc1)C)SC(F)(Cl)Cl)N(C)C	731-27-1
oxasulfuron	0	225	Train	O=C(OC1COC1)c2ccccc2S(=O)(=O)NC(=O)Nc3nc(cc(n3)C)C	144651-06-9

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
bicyclopyrone	0	273	Train	O=C(c1ccc(nc1COCCOC)C(F)(F)F)C=2C(=O)C3CCC(C=2(O))C3	352010-68-5
metobromuron	0	123	Train	O=C(Nc1ccc(cc1)Br)N(OC)C	3060-89-7
methoxyfenozide| Methoxyfenozide| 3-Methoxymethylbenzoic acid 2-(3,5-dimethylbenzoyl)-2-(1,1-dimethylethyl)hydrazide	0	111	Train	O=C(NN(C(=O)c1cc(cc(c1)C)C)C(C)(C)C)c2cccc(OC)c2C	161050-58-4
pethoxamid| Pethoxamid	0	72	Train	O=C(N(C(c1ccccc1)=C(C)C)CCOCC)CCl	106700-29-2
sodium 5-nitroguaiacolate	0	365	Train	O=[N+]([O-])c1ccc(OC)c(O)c1	67233-85-6
fluopicolide	0	109	Train	O=C(NCc1ncc(cc1Cl)C(F)(F)F)c2c(cccc2Cl)Cl	239110-15-7
tritosulfuron	0	337	Train	O=S(=O)(N=C(O)N=C1N=C(N=C(OC)N1)C(F)(F)F)c2ccccc2C(F)(F)F	142469-14-5
MCPB	0	151	Train	O=C(O)CCCOc1ccc(cc1C)Cl	94-81-5
fluopyram| Fluopyram| N-[2-[3-Chloro-5-(trifluoromethyl)-2-pyridinyl]ethyl]-2-(trifluoromethyl)benzamide	0	9	Train	FC(F)(F)c1cnc(c(c1)Cl)CCN=C(O)c2ccccc2C(F)(F)F	658066-35-4
oxyfluorfen	0	372	Train	O=[N+]([O-])c2ccc(Oc1ccc(cc1Cl)C(F)(F)F)cc2(OCC)	42874-03-3
N2-Ethyl-N4-(1-methylethyl)-6-(methylthio)-1,3,5-triazine-2,4-diamine	0	403	Train	n1c(nc(nc1NC(C)C)SC)NCC	834-12-8
isoproturon	0	124	Train	O=C(Nc1ccc(cc1)C(C)C)N(C)C	34123-59-6
4,6-Dichloro-N-(2-chlorophenyl)-1,3,5-triazin-2-amine	0	397	Train	n1c(nc(nc1Cl)Cl)Nc2ccccc2Cl	101-05-3
flurprimidol| alpha-(1-Methylethyl)-alpha-[4-(trifluoromethoxy)phenyl]-5-pyrimidinemethanol	0	7	Train	FC(F)(F)Oc1ccc(cc1)C(O)(c2cncnc2)C(C)C	56425-91-3
2-naphthyloxyacetic acid	0	156	Train	O=C(O)COc1ccc2ccccc2(c1)	120-23-0

quinoxyfen| 5,7-Dichloro-4-(4-fluorophenoxy)quinoline	0	26	Train	Fc3ccc(Oc2ccnc1cc(cc(c12)Cl)Cl)cc3	124495-18-7
carbetamide	0	178	Train	O=C(OC(C(=O)NCC)C)Nc1ccccc1	16118-49-3
acequinocyl| 2-(Acetyloxy)-3-dodecyl-1,4-naphthalenedione	0	226	Train	O=C(OC=1C(=O)c2ccccc2(C(=O)C=1CCCCCCCCCCC))C	57960-19-7
pyroxsulam| Pyroxsulam| N-(5,7-Dimethoxy[1,2,4]triazolo[1,5-a]pyrimidin-2-yl)-2-methoxy-4-(trifluoromethyl)-3-pyridinesulfonamide	0	345	Train	O=S(=O)(Nc1nc2nc(OC)cc(OC)n2(n1))c3c(nccc3C(F)(F)F)OC	422556-08-9
triticonazole| Triticonazole	0	383	Train	OC2(C(=Cc1ccc(cc1)Cl)CCC2(C)(C))(Cn3ncnc3)	131983-72-7
hexythiazox	0	106	Train	O=C(NC1CCCCC1)N3C(=O)SC(c2ccc(cc2)Cl)C3C	78587-05-0
6-chloronicotinic acid	0	166	Train	O=C(O)c1cnc(cc1)Cl	5326-23-8
procymidone	0	311	Train	O=C2N(c1cc(cc(c1)Cl)Cl)C(=O)C3(C)(CC23(C))	32809-16-8
flutolanil| Flutolanil	0	127	Train	O=C(Nc1cccc(OC(C)C)c1)c2ccccc2C(F)(F)F	66332-96-5
cyromazine	0	398	Train	n1c(nc(nc1N)NC2CC2)N	66215-27-8
proquinazid	0	305	Train	O=C1c2cc(ccc2(N=C(OCCC)N1CCC))I	189278-12-4
pencycuron	0	131	Train	O=C(Nc1ccccc1)N(Cc2ccc(cc2)Cl)C3CCCC3	66063-05-6
boscalid| 2-Chloro-N-(4'-chloro[1,1'-biphenyl]-2-yl)-3-pyridinecarboxamide	0	143	Train	O=C(Nc2ccccc2(c1ccc(cc1)Cl))c3cccnc3Cl	188425-85-6
sulcotrione	0	271	Train	O=C(c1ccc(cc1Cl)S(=O)(=O)C)C2C(=O)CCCC2(=O)	99105-77-8

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
trifloxystrobin| Trifloxystrobin| (alphaE)-alpha-(Methoxyimino)-2-[[[(E)-[1-[3-(trifluoromethyl)phenyl]ethylidene]amino]oxy]methyl]benzeneacteic acid methylester	0	194	Train	O=C(OC)C(=NOC)c1ccccc1CON=C(c2cccc(c2)C(F)(F)F)C	141517-21-7
beflubutamid	0	25	Train	Fc2ccc(OC(C(O)=NCc1ccccc1)CC)cc2C(F)(F)F	113614-08-7
mandipropamid	0	107	Train	O=C(NCCc1ccc(OCC#C)c(OC)c1)C(OCC#C)c2ccc(cc2)Cl	374726-62-2
iprodione| Iprodione	0	100	Train	O=C(NC(C)C)N2C(=O)N(c1cc(cc(c1)Cl)Cl)C(=O)C2	36734-19-7
mefentrifluconazole	0	10	Train	FC(F)(F)c2cc(Oc1ccc(cc1)Cl)ccc2C(O)(C)Cn3ncnc3	1417782-03-6
primisulfuron	0	165	Train	O=C(O)c1ccccc1S(=O)(=O)NC(=O)Nc2nc(OC(F)F)cc(n2)OC(F)F	113036-87-6
penflufen| Penflufen	0	22	Train	Fc2c(C(O)=Nc1ccccc1C(C)CC(C)C)c(nn2C)C	494793-67-8
propisochlor	0	80	Train	O=C(N(c1c(cccc1CC)C)COC(C)C)CCl	86763-47-5
rimsulfuron| N-[[(4,6-Dimethoxy-2-pyrimidinyl)amino]carbonyl]-3-(ethylsulfonyl)-2-pyridine sulfonamide	0	138	Train	O=C(Nc1nc(OC)cc(n1)OC)NS(=O)(=O)c2ncccc2S(=O)(=O)CC	122931-48-0
azoxystrobin	0	52	Train	N#Cc3ccccc3(Oc2ncnc(Oc1ccccc1C(=COC)C(=O)OC)c2)	131860-33-8
diethofencarb	0	185	Train	O=C(OC(C)C)Nc1ccc(OCC)c(OCC)c1	87130-20-9
1-naphthylacetamide	0	84	Train	O=C(N)Cc2cccc1ccccc12	86-86-2
[(3,5,6-Trichloro-2-pyridinyl)oxy]acetic acid, 2-Butoxyethylester	0	240	Train	O=C(OCCOCCCC)COc1nc(c(cc1Cl)Cl)Cl	64700-56-7

2-(4-Chlorophenyl)-1-ethyl-1,4-dihydro-6-methyl-4-oxo-3-pyridinecarboxylic acid, Potassium salt	0	149	Train	O=C(O)C=2C(=O)C=C(N(C=2(c1ccc(cc1)Cl))CC)C	81052-29-1
saflufenacil| 2-Chloro-5-[3,6-dihydro-3-methyl-2,6-dioxo-4-(trifluoromethyl-1(2H)-pyrimidinyl]-4-fluoro-N-[[methyl(1-methylethyl)amino]sulfonyl]benzamide	0	290	Train	O=C1C=C(N(C(=O)N1c2cc(C(O)=NS(=O)(=O)N(C)C(C)C)c(cc2(F))Cl)C)C(F)(F)F	372137-35-4
terbacil| 5-Chloro-3-(1,1-dimethylethyl)-6-methyl-2,4(1H,3H)-pyrimidinedione	0	298	Train	O=C1NC(=C(C(=O)N1C(C)(C)C)Cl)C	5902-51-2
epoxiconazole	0	19	Train	Fc1ccc(cc1)C3(OC3(c2ccccc2Cl))(Cn4ncnc4)	135319-73-2
iodosulfuron-methyl-sodium| Iodosulfuron-methyl-sodium	0	208	Train	O=C(OC)c1ccc(cc1S(=O)(=O)NC(=O)Nc2nc(nc(n2)C)OC)I	144550-36-7
difenoconazole| Difenoconazole| 1-[[2-[2-Chloro-4-(4-chlorophenoxyl)phenyl]-4-methyl-1,3-dioxolan-2-yl]methyl]-1H-1,2,4-triazole	0	410	Train	n1cnn(c1)CC2(OCC(O2)C)c4ccc(Oc3ccc(cc3)Cl)cc4Cl	119446-68-3
propoxycarbazone-sodium| 2-[[[(4,5-Dihydro-4-methyl-5-oxo-3-propoxy-1H-1,2,4-triazol-1-yl)carbonyl]amino]sulfonyl]benzoic acid, Methyl ester, Sodium salt (1:1)	0	211	Train	O=C(OC)c1ccccc1S(=O)(=O)NC(=O)N2N=C(OCCC)N(C2(=O))C	181274-15-7
napropamide	0	75	Train	O=C(N(CC)CC)C(Oc1cccc2ccccc12)C	15299-99-7

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
triadimenol| Triadimenol	0	376	Train	OC(C(Oc1ccc(cc1)Cl)n2ncnc2)C(C)(C)C	55219-65-3
butralin	0	361	Train	O=[N+]([O-])c1cc(cc(c1(NC(C)CC))[N+](=O)[O-])C(C)(C)C	33629-47-9
imazaquin	0	174	Train	O=C(O)c2cc3ccccc3(nc2(C1=NC(C(=O)N1)(C)C(C)C))	81335-37-7
haloxyfop-P-methyl	0	199	Train	O=C(OC)C(Oc2ccc(Oc1ncc(cc1Cl)C(F)(F)F)cc2)C	72619-32-0
2-[(1E)-1-[[[(2E)-3-Chloro-2-propenyl]oxy]imino]propyl]-5-[2-(ethylthio)propyl]-3-hydroxy-2-cyclohexen-1-one	0	287	Train	O=C1C(=C(O)CC(C1)CC(C)SCC)C(=NOCC=CCl)CC	99129-21-2
picolinafen	0	116	Train	O=C(Nc1ccc(F)cc1)c3nc(Oc2cccc(c2)C(F)(F)F)ccc3	137641-05-5
spinosad| Spinosad	1	319	Train	O=C5OC(CC)CCCC(OC1OC(C)C(N(C)C)CC1)C(C(=O)C6CC4C(CCC3CC(OC2CC(C)C(OC)C(OC)C2(OC))CC34)C6(C5))C	168316-95-8
1,3-dichloropropene	1	1	Train	C(=CCl)CCl	542-75-6
Phosphoramidothioic acid, O,S-Dimethyl ester	1	325	Train	O=P(OC)(N)SC	10265-92-6
pentachlorophenol	1	387	Train	Oc1c(c(c(c(c1Cl)Cl)Cl)Cl)Cl	87-86-5
clothianidin| Clothianidin| [C(E)]-N-[(2-Chloro-5-thiazolyl)methyl]-N'-methyl-N''-nitroguanidine	1	355	Train	O=[N+]([O-])N=C(NC)NCc1cnc(Cl)s1	210880-92-5
N,N-Bis(2-methylpropyl)carbamothioic acid S-ethyl ester	1	74	Train	O=C(N(CC(C)C)CC(C)C)SCC	2008-41-5
Toxaphene	1	4	Train	C=C1C(CCl)(CCl)C2(C(C(C(C1(C2(Cl)Cl)Cl)Cl)Cl)Cl)Cl	8001-35-2
fluchloralin	1	363	Train	O=[N+]([O-])c1cc(cc(c1N(CCC)CCCl)[N+](=O)[O-])C(F)(F)F	33245-39-5

imiprothrin	1	243	Train	O=C(OCN1C(=O)N(CC#C)CC1(=O))C2C(C=C(C)C)C2(C)(C)	72963-72-5
prometon| 6-Methoxy-N2,N4-bis(1-methylethyl)-1,3,5-triazine-2,4-diamine	1	401	Train	n1c(nc(nc1NC(C)C)NC(C)C)OC	1610-18-0
3-[(Dimethoxyphosphinyl)oxy]-2-butenoic acid, Methyl ester	1	200	Train	O=C(OC)C=C(OP(=O)(OC)OC)C	7786-34-7
resmethrin	1	247	Train	O=C(OCc1coc(c1)Cc2ccccc2)C3C(C=C(C)C)C3(C)(C)	10453-86-8
2-[1-(Ethoxyimino)butyl]-5-[2-(ethylthio)propyl]-3-hydroxy-2-cyclohexen-1-one	1	286	Train	O=C1C(=C(O)CC(C1)CC(C)SCC)C(=NOCC)CCC	74051-80-2
diazinon| O,O-Diethyl O-[6-methyl-2-(1-methylethyl)-4-pyrimidinyl] ester phosphorothioic acid	1	394	Train	n1c(OP(OCC)(OCC)=S)cc(nc1C(C)C)C	333-41-5
bensulide| Phosphorodithioic acid, O,O-Bis(1-methylethyl) S-[2-[(phenylsulfonyl)amino]ethyl] ester	1	340	Train	O=S(=O)(NCCSP(OC(C)C)(OC(C)C)=S)c1ccccc1	741-58-2
Phosphoric acid 1,2-dibromo-2,2-dichloroethyl dimethyl ester	1	327	Train	O=P(OC)(OC)OC(C(Cl)(Cl)Br)Br	300-76-5
diniconazole	1	375	Train	OC(C(=Cc1ccc(cc1Cl)Cl)n2ncnc2)C(C)(C)C	83657-24-3
lindane| (1alpha,2alpha,3beta,4alpha,5alpha,6beta)-1,2,3,4,5,6-Hexachlorocyclohexane	1	2	Train	C1(C(C(C(C(C1Cl)Cl)Cl)Cl)Cl)Cl	58-89-9
2-(Octylthio)ethanol	1	386	Train	OCCSCCCCCCCC	3547-33-9

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
pyridaben| 4-Chloro-2-(1,1-dimethylethyl)-5-[[[4-(1,1-dimethylethyl)phenyl]methyl]thio]-3(2H)pyridazinene	1	280	Train	O=C1C(=C(C=NN1C(C)(C)C)SCc2ccc(cc2)C(C)(C)C)Cl	96489-71-3
triazamate	1	235	Train	O=C(OCC)CSc1nc(nn1(C(=O)N(C)C))C(C)(C)C	112143-82-5
malathion| Malathion	1	233	Train	O=C(OCC)CC(C(=O)OCC)SP(OC)(OC)=S	121-75-5
phosalone| Phosalone	1	315	Train	O=C2Oc1cc(ccc1N2CSP(OCC)(OCC)=S)Cl	2310-17-0
3-Amino-2,5-dichlorobenzoic acid ammonium salt (1:1)	1	164	Train	O=C(O)c1cc(cc(N)c1Cl)Cl	1076-46-6
DDT	1	389	Train	c1cc(ccc1C(c2ccc(cc2)Cl)C(Cl)(Cl)Cl)Cl	50-29-3
dimethoate| Phosphorodithioic acid, O,O-Dimethyl S-[2-(methylamino)-2-oxoethyl] ester	1	103	Train	O=C(NC)CSP(OC)(OC)=S	60-51-5
Tetrahydro-5,5-dimethyl-2(1H)-pyrimidio[3-[4-trifluoromethyl)phenyl]-1-[2-[4-(trifluoromethyl)phenyl]ethenyl]-2-propenylidene]hydrazone	1	12	Train	FC(F)(F)c3ccc(C=CC(=NN=C1NCC(C)(C)CN1)C=Cc2ccc(cc2)C(F)(F)F)cc3	67485-29-4
tribufos	1	333	Train	O=P(SCCCC)(SCCCC)SCCCC	78-48-8
methabenzthiazuron	1	105	Train	O=C(NC)N(c2nc1ccccc1s2)C	18691-97-9
Phosphoric acid, 2-Chloro-3-(diethylamino)-1-methyl-3-oxo-1-propen-1-yl dimethyl ester	1	69	Train	O=C(C(=C(OP(=O)(OC)OC)C)Cl)N(CC)CC	13171-21-6
N''-Methyl-N-nitro-N'-[(tetrahydro-3-furanyl)methyl]guanidine	1	354	Train	O=[N+]([O-])N=C(NC)NCC1COCC1	165252-70-0
phosmet| Phosmet| S-[(1,3-Dihydro-1,3-dioxo-2H-	1	306	Train	O=C1c2ccccc2(C(=O)N1CSP(OC)(OC)=S)	732-11-6

isoindol-2-yl)methyl]O,O-dimethyl ester, Phosphorodithioic acid					
3,5,6-trichloro-2-pyridinol	1	388	Train	Oc1nc(c(cc1Cl)Cl)Cl	6515-38-4
P-Phenylphosphonothioic acid O-ethyl O-(4-nitrophenyl) ester	1	371	Train	O=[N+]([O-])c2ccc(OP(OCC)(c1ccccc1)=S)cc2	2104-64-5
6,7,8,9,10,10-Hexachloro-1,5,5a,6,9,9a-hexahydro-6,9-methano-2,4,3-benzodioxathiepin 3-oxide	1	352	Train	O=S1OCC2C(CO1)C3(C(=C(C2(C3(Cl)Cl)Cl)Cl)Cl)Cl	115-29-7
(6R,25R)-5-O-Demethyl-28-deoxy-6,28-epoxy-25-ethylmilbemycin B	1	318	Train	O=C3OC2CC(OC1(OC(CC)C(C)CC1)C2)CC=C(C)CC(C=CC=C4COC5C(O)C(=CC3C45(O))C)C	51596-11-3
chlorbromuron	1	119	Train	O=C(Nc1ccc(c(c1)Cl)Br)N(OC)C	13360-45-7
thiofanox	1	253	Train	O=C(ON=C(CSC)C(C)(C)C)NC	39196-18-4
methiocarb| Methiocarb| 3,5-Dimethyl-4-(methylthio)phenol, Methylcarbamate	1	255	Train	O=C(Oc1cc(c(c(c1)C)SC)C)NC	2032-65-7
profenofos| Phosphorothioic acid O-(4-bromo-2-chlorophenyl) O-ethyl S-propyl ester	1	332	Train	O=P(Oc1ccc(cc1Cl)Br)(OCC)SCCC	41198-08-7
tau-fluvalinate| Tau-fluvalinate| N-[2-Chloro-4-(trifluoromethyl)phenyl]-DL-valine cyano(3-phenoxyphenyl)methyl ester	1	32	Train	N#CC(OC(=O)C(Nc1ccc(cc1Cl)C(F)(F)F)C(C)C)c3cccc(Oc2ccccc2)c3	102851-06-9| 69409-94-5
nicosulfuron	1	136	Train	O=C(Nc1nc(OC)cc(n1)OC)NS(=O)(=O)c2ncccc2(C(=O)N(C)C)	111991-09-4

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
oxydemeton-methyl| Phosphorothioic acid, S-[2-(Ethylsulfinyl)ethyl] O,O-dimethyl ester	1	350	Train	O=S(CC)CCSP(=O)(OC)OC	301-12-2
fipronil| 5-Amino-1-[2,6-dichloro-4-(trifluoromethyl)phenyl]-4-[(trifluoromethyl)sulfinyl]-1H-pyrazole-3-carbonitrile	1	48	Train	N#Cc1nn(c(N)c1S(=O)C(F)(F)F)c2c(cc(cc2Cl)C(F)(F)F)Cl	120068-37-3
spiroxamine| 8-(1,1-Dimethylethyl)-N-ethyl-N-propyl-1,4-dioxaspiro[4.5]decane-2-methanamine	1	68	Train	O1CC(OC12(CCC(CC2)C(C)(C)C))CN(CC)CCC	118134-30-8
prallethrin| 2,2-Dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid 2-methyl-4-oxo-3-(2-propynyl)-2-cyclopenten-1-yl ester	1	220	Train	O=C(OC1C(=C(C(=O)C1)CC#C)C)C2C(C=C(C)C)C2(C)(C)	23031-36-9
ethametsulfuron-methyl	1	214	Train	O=C(OC)c1ccccc1S(=O)(=O)NC(=O)Nc2nc(nc(n2)NC)OCC	97780-06-8
imidacloprid| (2E)-1-[(6-Chloro-3-pyridinyl)methyl]-N-nitro-2-imidazolidinimine| Imidacloprid	1	356	Train	O=[N+]([O-])N=C1NCCN1Cc2cnc(cc2)Cl	138261-41-3
2-(Difluoromethyl)-4-(2-methylpropyl)-6-(trifluoromethyl)-3,5-pyridinedicarbothioic acid, S3,S5-Dimethyl ester	1	268	Train	O=C(c1c(nc(c(C(=O)SC)c1CC(C)C)C(F)(F)F)C(F)F)SC	97886-45-8
Phosphorodithioic acid, S,S'-Methylene O,O,O',O'-tetraethyl ester	1	59	Train	O(CC)P(OCC)(=S)SCSP(OCC)(OCC)=S	563-12-2

dichlofluanid	1	336	Train	O=S(=O)(N(c1ccccc1)SC(F)(Cl)Cl)N(C)C	1085-98-9
fenpyroximate	1	181	Train	O=C(OC(C)(C)C)c1ccc(cc1)CON=Cc3c(nn(c3(Oc2ccccc2))C)C	134098-61-6
4-Bromo-2-(4-chlorophenyl)-1-(ethoxymethyl)-5-(trifluoromethyl)-1H-pyrrole-3-carbonitrile	1	49	Train	N#Cc2c(c1ccc(cc1)Cl)n(c(c2Br)C(F)(F)F)COCC	122453-73-0
formetanate hydrochloride| N,N-Dimethyl-N'-[3[[(methylamino)carbonyl]oxy]phenyl]methanimidamide hydrochloride (1:1)	1	258	Train	O=C(Oc1cccc(N=CN(C)C)c1)NC	23422-53-9
4-(2,4-Dichlorophenoxy)butanoic acid	1	152	Train	O=C(O)CCCOc1ccc(cc1Cl)Cl	94-82-6
alpha-[(Dimethoxyphosphinothioyl)thio] benzeneacetic acid, Ethyl ester	1	231	Train	O=C(OCC)C(c1ccccc1)SP(OC)(OC)=S	2597-03-7
nuarimol	1	18	Train	Fc1ccc(cc1)C(O)(c2cncnc2)c3ccccc3Cl	63284-71-9
tefluthrin	1	245	Train	O=C(OCc1c(F)c(F)c(c(F)c1(F))C)C2C(C=C(C(F)(F)F)Cl)C2(C)(C)	79538-32-2
fosthiazate	1	294	Train	O=C1N(CCS1)P(=O)(OCC)SC(C)CC	98886-44-3
triazophos	1	407	Train	n1cn(nc1OP(OCC)(OCC)=S)c2ccccc2	24017-47-8
chlorethoxyfos| Phosphorothioic acid O,O-diethyl-O-(1,2,2,2-tetrachloroethyl) ester	1	60	Train	O(CC)P(OCC)(OC(C(Cl)(Cl)Cl)Cl)=S	54593-83-8
azinphos-methyl| O,O-Dimethyl S-[(4-oxo-1,2,3-benzotriazin-3(4H)-yl)methyl] ester, Phosphorodithioic acid	1	307	Train	O=C1c2ccccc2(N=NN1CSP(OC)(OC)=S)	86-50-0

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
benfuracarb| Benfuracarb	1	264	Train	O=C(Oc2cccc1c2(OC(C)(C)C1))N(C)SN(CCC(=O)OCC)C(C)C	82560-54-1
indoxacarb| (4aS)-7-Chloro-2,5-dihydro-2-[[(methoxycarbonyl)[4-(trifluoromethoxy)phenyl]amino]carbonyl]indeno[1,2-e][1,3,4]oxadiazine-4a(3H)-carboxylic acid methyl ester	1	202	Train	O=C(OC)N(C(=O)N2N=C3c1ccc(cc1CC3(OC2)(C(=O)OC))Cl)c4ccc(OC(F)(F)F)cc4	173584-44-6
alachlor	1	82	Train	O=C(N(c1c(cccc1CC)CC)COC)CCl	15972-60-8
1,1'-(2,2,2-Trichloroethylidene)bis[4-methoxybenzene]	1	63	Train	O(c1ccc(cc1)C(c2ccc(OC)cc2)C(Cl)(Cl)Cl)C	72-43-5
Bifenazate| 2-(4-Methoxy[1,1'-biphenyl]-3-yl)hydrazinecarboxylic acid, 1-Methylethyl ester	1	184	Train	O=C(OC(C)C)NNc1cc(ccc1(OC))c2ccccc2	149877-41-8
chlorthiamid	1	55	Train	NC(c1c(cccc1Cl)Cl)=S	1918-13-4
thiodicarb| Thiodicarb	1	251	Train	O=C(ON=C(C)SC)N(C)SN(C(=O)ON=C(C)SC)C	59669-26-0
propham	1	187	Train	O=C(OC(C)C)Nc1ccccc1	122-42-9
sulfoxaflor	1	43	Train	N#CN=S(=O)(C)C(c1cnc(cc1)C(F)(F)F)C	946578-00-3
6-Methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one	1	316	Train	O=C2Sc1nc3ccc(cc3(nc1S2))C	2439-01-2
imazalil	1	406	Train	n1ccn(c1)CC(OCC=C)c2ccc(cc2Cl)Cl	35554-44-0
oxamyl| 2-(Dimethylamino)-N-[[(methylamino)carbonyl]oxy]-2-oxo-ethanimidothioic acid methyl ester	1	250	Train	O=C(ON=C(C(=O)N(C)C)SC)NC	23135-22-0
fenazaquin| Fenazaquin| 4-[2-[4-(1,1-	1	408	Train	n1cnc3ccccc3(c1OCCc2ccc(cc2)C(C)(C)C)	120928-09-8

Dimethylethyl)phenyl]ethoxy] quinazoline					
(1aR,2R,2aS,3S,6R,6aR,7S,7aS)-rel-3,4,5,6,9,9-Hexachloro-1a,2,2a,3,6,6a,7,7a-octahydro-2,7:3,6-dimethanonaphth[2,3-b]oxirene| 3,4,5,6,9,9-Hexachloro-1a,2,2a,3,6,6a,7,7a-octahydro-[2,7:3,6-dimethanonaphth[2,3-b]oxirene,[1a alpha,2 beta,2a beta,3 alpha,6 alpha,6a beta,7 beta,7a alpha]	1	67	Train	O1C2C1C3CC2C4C3C5(C(=C(C4(C5(Cl)Cl)Cl)Cl)Cl)Cl	60-57-1| 72-20-8
2-[[Ethoxy[(1-methylethyl)amino]phosphinothioyl]oxy]benzoic acid 1-methylethyl ester	1	188	Train	O=C(OC(C)C)c1ccccc1(OP(OCC)(NC(C)C)=S)	25311-71-1
2-(1-Methylpropyl)-4,6-dinitrophenol	1	359	Train	O=[N+]([O-])c1cc(c(O)c(c1)C(C)CC)[N+](=O)[O-]	88-85-7
(1,3,4,5,6,7-Hexahydro-1,3-dioxo-2H-isoindol-2-yl)methyl ester 2,2-dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid	1	242	Train	O=C(OCN1C(=O)C2=C(C1(=O))CCCC2)C3C(C=C(C)C)C3(C)(C)	7696-12-0
metoxuron	1	118	Train	O=C(Nc1ccc(OC)c(c1)Cl)N(C)C	19937-59-8
2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate)	1	261	Train	O=C(Oc1cccc2OC(Oc12)(C)C)NC	22781-23-3

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
Dimethylcarbamic acid, 2-(Dimethylamino)-5,6-dimethyl-4-pyrimidinyl ester	1	263	Train	O=C(Oc1nc(nc(c1C)C)N(C)C)N(C)C	23103-98-2
P-(2,2,2-Trichloro-1-hydroxyethyl)phosphonic acid dimethyl ester	1	326	Train	O=P(OC)(OC)C(O)C(Cl)(Cl)Cl	52-68-6
oryzalin	1	362	Train	O=[N+]([O-])c1cc(cc(c1N(CCC)CCC)[N+](=O)[O-])S(=O)(=O)N	19044-88-3
tebufenpyrad	1	108	Train	O=C(NCc1ccc(cc1)C(C)(C)C)c2c(c(nn2C)CC)Cl	119168-77-3
esfenvalerate| (alphaS)-4-Chloro-alpha-(1-methylethyl)benzeneacetic acid (S)-cyano-(3-phenoxyphenyl)methyl ester| 4-Chloro-alpha-(1-methylethyl)benzeneacetic acid cyano(3-phenoxyphenyl)methyl ester	1	33	Train	N#CC(OC(=O)C(c1ccc(cc1)Cl)C(C)C)c3cccc(Oc2ccccc2)c3	66230-04-4| 51630-58-1
fenamiphos| Fenamiphos (aka phenamiphos)| (1-Methylethyl)phosphoramidic acid ethyl-3-methyl-4-(methylthio)phenyl ester	1	331	Train	O=P(Oc1ccc(c(c1)C)SC)(OCC)NC(C)C	22224-92-6
chlorimuron-ethyl| 2-[[[[(4-Chloro-6-methoxy-2-pyrimidinyl)amino]carbonyl]amino]sulfonyl]benzoic acid, Ethyl ester	1	237	Train	O=C(OCC)c1ccccc1S(=O)(=O)NC(=O)Nc2nc(OC)cc(n2)Cl	90982-32-4
O,O-Diethyl S-[(ethylthio)methyl])ester, Phosphorodithioic acid	1	58	Train	O(CC)P(OCC)(=S)SCSCC	298-02-2
etrimfos	1	391	Train	n1c(OCC)cc(nc1CC)OP(OC)(OC)=S	38260-54-7

tetraconazole	1	13	Train	FC(F)C(F)(F)OCC(c1ccc(cc1Cl)Cl)Cn2ncnc2	112281-77-3
Phosphoric acid 2,2-dichloroethenyl dimethyl ester	1	328	Train	O=P(OC=C(Cl)Cl)(OC)OC	62-73-7
aminocarb	1	257	Train	O=C(Oc1ccc(c(c1)C)N(C)C)NC	2032-59-9
2-Methyl-2-(methylthio)propanol O-[(methylamino)carbonyl]oxime	1	254	Train	O=C(ON=CC(C)(C)SC)NC	116-06-3
formothion	1	322	Train	O=CN(C(=O)CSP(OC)(OC)=S)C	2540-82-1
propargite| 2-[4-(1,1-Dimethylethyl)phenoxy]cyclohexyl-2-propynyl ester sulfurous acid	1	351	Train	O=S(OCC#C)OC2CCCCC2(Oc1ccc(cc1)C(C)(C)C)	2312-35-8
carbaryl	1	262	Train	O=C(Oc1cccc2ccccc12)NC	63-25-2
cadusafos	1	329	Train	O=P(OCC)(SC(C)CC)SC(C)CC	95465-99-9
permethrin| 3-(2,2-Dichloroethenyl)-2,2-dimethylcyclopropanecarboxylic acid, (3-Phenoxyphenyl)methyl ester	1	249	Train	O=C(OCc2cccc(Oc1ccccc1)c2)C3C(C=C(Cl)Cl)C3(C)(C)	52645-53-1
prometryn| N2,N4-Bis(1-methylethyl)-6-(methylthio)-1,3,5-triazine-2,4-diamine	1	402	Train	n1c(nc(nc1NC(C)C)SC)NC(C)C	7287-19-6
tralomethrin| 2,2-Dimethyl-3-(1,1,2,2-tetrabromoethyl)cyclopropanecarboxylic acid cyano(3-phenoxyphenyl)methyl ester	1	34	Train	N#CC(OC(=O)C1C(C(C(Br)(Br)Br)Br)C1(C)(C))c3cccc(Oc2ccccc2)c3	66841-25-6
carbofuran| 2,3-Dihydro-2,2-dimethyl-7-benzofuranol 7-(N-methylcarbamate)	1	266	Train	O=C(Oc2cccc1c2(OC(C)(C)C1))NC	1563-66-2
acetamiprid	1	42	Train	N#CN=C(N(C)Cc1cnc(cc1)Cl)C	135410-20-7

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
Etofenprox	1	64	Train	O(c1ccccc1)c2cccc(c2)COCC(c3ccc(OCC)cc3)(C)C	80844-07-1
O,O-Dimethyl O-[3-methyl-4-(methylthio)phenyl]ester phosphorothioic acid	1	62	Train	O(c1ccc(c(c1)C)SC)P(OC)(OC)=S	55-38-9
isopyrazam	0	14	Test	FC(F)c4nn(cc4(C(O)=Nc1cccc2c1C3CCC2C3C(C)C))C	881685-58-1
sedaxane	0	15	Test	FC(F)c4nn(cc4(C(O)=Nc1ccccc1C2CC2C3CC3))C	874967-67-6
fluxapyroxad	0	17	Test	Fc1cc(cc(F)c1(F))c3ccccc3(N=C(O)c2cn(nc2C(F)F)C)	907204-31-3
momfluorothrin	1	28	Test	N#CC(=CC2C(C(=O)OCc1c(F)c(F)c(c(F)c1(F))COC)C2(C)(C))C	609346-29-4
deltamethrin| (1R,3R)-3-(2,2-Dibromoethenyl)-2,2-dimethylcyclopropanecarboxylic acid (S)cyano(3-phenoxyphenyl)methyl ester	1	35	Test	N#CC(OC(=O)C1C(C=C(Br)Br)C1(C)(C))c3ccccc(Oc2ccccc2)c3	52918-63-5
gamma-cyhalothrin| lambda-cyhalothrin| Gamma-cyhalothrin| (1R,3R)-3-(2-Chloro-3,3,3-trifluoro-1-propenyl)-2,2-dimethylcyclopropanecarboxylic acid (S)-cyano(3-phenoxyphenyl)methyl ester| 3-[(1Z)-2-Chloro-3,3,3-trifluoro-1-propen-1-yl]-2,2-dimethylcyclopropanecarboxylic acid (1S,3S)-rel-(R)-cyano(3-phenoxyphenyl)methyl ester	1	36	Test	N#CC(OC(=O)C1C(C=C(C(F)(F)F)Cl)C1(C)(C))c3cccc(Oc2ccccc2)c3	76703-62-3| 91465-08-6
beta-cyfluthrin| cyfluthrin| 3-(2,2-Dichloroethenyl)-2,2-	1	37	Test	N#CC(OC(=O)C1C(C=C(Cl)Cl)C1(C)(C))c3ccc(F)c(Oc2ccccc2)c3	1820573-27-0| 68359-37-5

dimethyl cyclopropanecarboxylic acid cyano(4-fluoro-3-phenoxyphenyl)methyl ester					
beta-cypermethrin| zeta-cypermethrin| alpha-cypermethrin| cypermethrin| 3-(2,2-Dichloroethenyl)-2,2-dimethylcyclopropanecarboxylic acid cyano(3-phenoxyphenyl)methyl ester	1	38	Test	N#CC(OC(=O)C1C(C=C(Cl)Cl)C1(C)(C))c3cccc(Oc2ccccc2)c3	65731-84-2| 97955-44-7| 67375-30-8| 52315-07-8
acrinathrin	1	39	Test	N#CC(OC(=O)C1C(C=CC(=O)OC(C(F)(F)F)C(F)(F)F)C1(C)(C))c3cccc(Oc2ccccc2)c3	101007-06-1
disulfoton| Phosphorodithioic acid, O,O-Diethyl-S-[2-(ethylthio)ethyl] ester	1	56	Test	O(CC)P(OCC)(=S)SCCSCC	298-04-4
terbufos	1	57	Test	O(CC)P(OCC)(=S)SCSC(C)(C)C	13071-79-9
dimethachlor| Dimethachlor	0	77	Test	O=C(N(c1c(cccc1C)C)CCOC)CCl	50563-36-5
S-metolachlor| metolachlor| 2-Chloro-N-(2-ethyl-6-methylphenyl)-N-(2-methoxy-1-methylethyl)acetamide	0	79	Test	O=C(N(c1c(cccc1CC)C)C(C)COC)CCl	87392-12-9| 51218-45-2
acetochlor| 2-Chloro-N-(ethoxymethyl)-N-(2-ethyl-6-methylphenyl)acetamide	0	81	Test	O=C(N(c1c(cccc1CC)C)COCC)CCl	34256-82-1
teflubenzuron	0	88	Test	O=C(NC(=O)c1c(F)cccc1(F))Nc2cc(c(F)c(c2(F))Cl)Cl	83121-18-0
lufenuron	0	89	Test	O=C(NC(=O)c1c(F)cccc1(F))Nc2cc(c(OC(F)(F)C(F)C(F)(F)F)cc2Cl)Cl	103055-07-8

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
novaluron| N-[[[3-Chloro-4-[1,1,2-trifluoro-2-(trifluoromethoxy)ethoxy]phenyl]amino]carbonyl]-2,6-difluorobenzamide	0	92	Test	O=C(NC(=O)c1c(F)cccc1(F))Nc2ccc(OC(F)(F)C(F)OC(F)(F)F)c(c2)Cl	116714-46-6
N-[[(4-Chlorophenyl)amino]carbonyl]-2,6-difluorobenzamide	0	93	Test	O=C(NC(=O)c1c(F)cccc1(F))Nc2ccc(cc2)Cl	35367-38-5
flufenoxuron| Flufenoxuron	0	94	Test	O=C(NC(=O)c1c(F)cccc1(F))Nc3ccc(Oc2ccc(cc2Cl)C(F)(F)F)cc3(F)	101463-69-8
triflumuron| Triflumuron	0	95	Test	O=C(NC(=O)c1ccccc1Cl)Nc2ccc(OC(F)(F)F)cc2	64628-44-0
propyzamide	0	96	Test	O=C(NC(C#C)(C)C)c1cc(cc(c1)Cl)Cl	23950-58-5
dimoxystrobin	0	101	Test	O=C(NC)C(=NOC)c1ccccc1COc2cc(ccc2C)C	149961-52-4
chromafenozide| Chromafenozide	0	112	Test	O=C(NN(C(=O)c1cc(cc(c1)C)C)C(C)(C)C)c3ccc2OCCCc2c3C	143807-66-3
halofenozide| 2-Benzoyl-2-(1,1-dimethylethyl)hydrazide-4-chlorobenzoic acid	0	113	Test	O=C(NN(C(=O)c1ccccc1)C(C)(C)C)c2ccc(cc2)Cl	112226-61-6
propanil	0	120	Test	O=C(Nc1ccc(c(c1)Cl)Cl)CC	709-98-8
linuron| N'-(3,4-Dichlorophenyl)-N-methoxy-N-methylurea	0	122	Test	O=C(Nc1ccc(c(c1)Cl)Cl)N(OC)C	330-55-2
monuron| N'-(4-chlorophenyl)-N,N-dimethylurea	0	125	Test	O=C(Nc1ccc(cc1)Cl)N(C)C	150-68-5
fluometuron	0	128	Test	O=C(Nc1cccc(c1)C(F)(F)F)N(C)C	2164-17-2
oxycarboxin	0	129	Test	O=C(Nc1ccccc1)C2=C(OCCS2(=O)(=O))C	5259-88-1
siduron| N-(2-Methylcyclohexyl)-N'-phenylurea	0	132	Test	O=C(Nc1ccccc1)NC2CCCCC2(C)	1982-49-6

flazasulfuron| N-[[4,6-Dimethoxy-2-pyrimidinyl)amino]carbonyl]-3-(trifluoromethyl)-2-pyridinesulfonamide	0	137	Test	O=C(Nc1nc(OC)cc(n1)OC)NS(=O)(=O)c2ncccc2C(F)(F)F	104040-78-0
prosulfuron| Prosulfuron| N-[[(4-Methoxy-6-methyl-1,3,5-triazin-2-yl)amino]carbonyl]-2-(3,3,3-trifluoropropyl)benzenesulfonamide	0	139	Test	O=C(Nc1nc(nc(n1)C)OC)NS(=O)(=O)c2ccccc2CCC(F)(F)F	94125-34-5
chloroxuron	1	142	Test	O=C(Nc2ccc(Oc1ccc(cc1)Cl)cc2)N(C)C	1982-47-4
mecoprop| mecoprop-P	0	147	Test	O=C(O)C(Oc1ccc(cc1C)Cl)C	7085-19-0| 16484-77-8
MCPA	0	154	Test	O=C(O)COc1ccc(cc1C)Cl	94-74-6
triclopyr| 2-[(3,5,6-Trichloro-2-pyridinyl)oxy]acetic acid	0	158	Test	O=C(O)COc1nc(c(cc1Cl)Cl)Cl	55335-06-3
benthiavalicarb	0	160	Test	O=C(O)NC(C(O)=NC(c2nc1ccc(F)cc1s2)C)C(C)C	413615-35-7
quinclorac	0	163	Test	O=C(O)c1c2ncc(cc2(ccc1Cl))Cl	84087-01-4
Aminopyralid| 4-Amino-3,6-dichloro-2-pyradinecarboxylic acid	0	168	Test	O=C(O)c1nc(cc(N)c1Cl)Cl	150114-71-9
acifluorfen-sodium	0	171	Test	O=C(O)c2cc(Oc1ccc(cc1Cl)C(F)(F)F)ccc2[N+](=O)[O-]	62476-59-9
imazapic	0	172	Test	O=C(O)c2cc(cnc2(C1=NC(=O)C(N1)(C)C(C)C))C	104098-48-8
2-[4,5-Dihydro-4-methyl-4-(1-methylethyl)-5-oxo-1H-imidazol-2-yl]-5-ethyl-3-pyridinecarboxylic acid	0	173	Test	O=C(O)c2cc(cnc2(C1=NC(C(=O)N1)(C)C(C)C))CC	81335-77-5
chlorpropham	1	186	Test	O=C(OC(C)C)Nc1cccc(c1)Cl	101-21-3

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
fluroxypyr-meptyl	0	189	Test	O=C(OC(C)CCCCCC)COc1nc(F)c(c(N)c1Cl)Cl	81406-37-3
pyrethrins (cinerin Ⅱ)	1	190	Test	O=C(OC)C(=CC2C(C(=O)OC1C(=C(C(=O)C1)CC=CC)C)C2(C)(C))C	121-20-0
pyrethrins (pyrethrin Ⅱ)	1	191	Test	O=C(OC)C(=CC2C(C(=O)OC1C(=C(C(=O)C1)CC=CC=C)C)C2(C)(C))C	121-29-9
pyrethrins (jasmolin Ⅱ)	1	192	Test	O=C(OC)C(=CC2C(C(=O)OC1C(=C(C(=O)C1)CC=CCC)C)C2(C)(C))C	1172-63-0
benalaxyl-M| Benalaxyl-M	0	197	Test	O=C(OC)C(N(C(=O)Cc1ccccc1)c2c(cccc2C)C)C	98243-83-5
diclofop-methyl| Diclofop-methyl	0	198	Test	O=C(OC)C(Oc2ccc(Oc1ccc(cc1Cl)Cl)cc2)C	51338-27-3
2-[[[[(4-Methoxy-6-methyl-1,3,5-triazin-2-yl)methylamino]carbonyl]amino]sulfonyl]benzoic acid methyl ester	0	210	Test	O=C(OC)c1ccccc1S(=O)(=O)NC(=O)N(c2nc(nc(n2)C)OC)C	101200-48-0
2-[[[[[4,6-Bis(difluoromethoxy)-2-pyrimidinyl]amino]carbonyl]amino]sulfonyl]benzoic acid, Methyl ester	0	212	Test	O=C(OC)c1ccccc1S(=O)(=O)NC(=O)Nc2nc(OC(F)F)cc(n2)OC(F)F	86209-51-0
sulfometuron-methyl| 2-[[[[(4,6-Dimethyl-2-pyrimidinyl)amino]carbonyl]amino]sulfonyl]benzoic acid, Methyl ester	0	213	Test	O=C(OC)c1ccccc1S(=O)(=O)NC(=O)Nc2nc(cc(n2)C)C	74222-97-2
bifenox| Bifenox	0	217	Test	O=C(OC)c2cc(Oc1ccc(cc1Cl)Cl)ccc2[N+](=O)[O-]	42576-02-3
bioallethrin| 2,2-Dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid 2-methyl-4-oxo-3-(2-	1	221	Test	O=C(OC1C(=C(C(=O)C1)CC=C)C)C2C(C=C(C)C)C2(C)(C)	584-79-2

propenyl)-2-cyclopenten-1-yl ester					
pyrethrins (cinerin I)	1	222	Test	O=C(OC1C(=C(C(=O)C1)CC=CC)C)C2C(C=C(C)C)C2(C)(C)	25402-06-6
pyrethrins (pyrethrin I)	1	223	Test	O=C(OC1C(=C(C(=O)C1)CC=CC=C)C)C2C(C=C(C)C)C2(C)(C)	121-21-1
pyrethrins (jasmolin I)	1	224	Test	O=C(OC1C(=C(C(=O)C1)CC=CCC)C)C2C(C=C(C)C)C2(C)(C)	4466-14-2
dichlorprop-P 2-ethylyhexyl ester	0	227	Test	O=C(OCC(CC)CCCC)C(Oc1ccc(cc1Cl)Cl)C	865363-39-9
quizalofop-P-ethyl	0	230	Test	O=C(OCC)C(Oc3ccc(Oc1nc2ccc(cc2(nc1))Cl)cc3)C	100646-51-3
propaquizafop	0	241	Test	O=C(OCCON=C(C)C)C(Oc3ccc(Oc1nc2ccc(cc2(nc1))Cl)cc3)C	111479-05-1
2,2-Dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid (3-phenoxyphenyl)methyl ester	1	248	Test	O=C(OCc2cccc(Oc1ccccc1)c2)C3C(C=C(C)C)C3(C)(C)	26002-80-2
methomyl| Methomyl| N-[[(Methylamino)carbonyl]oxy] ethanimidothioic acid methyl ester	1	252	Test	O=C(ON=C(C)SC)NC	16752-77-5
desmedipham	0	260	Test	O=C(Oc1cccc(c1)NC(=O)OCC)Nc2ccccc2	13684-56-5
carbosulfan| Carbosulfan	1	265	Test	O=C(Oc2cccc1c2(OC(C)(C)C1))N(C)SN(CCCC)CCCC	55285-14-8
norflurazon| 4-Chloro-5-(methylamino)-2-[3-(trifluoromethyl)phenyl]-3(2H)-pyridazinone	0	281	Test	O=C1C(=C(C=NN1c2cccc(c2)C(F)(F)F)NC)Cl	27314-13-2
ethoprophos	1	330	Test	O=P(OCC)(SCCC)SCCC	13194-48-4

Chemical Name	CLASS	CLASS_ID	Set	SMILES (VEGA)	CAS Number
flumetsulam| N-(2,6-Difluorophenyl)-5-methyl-[1,2,4]triazolo[1,5-a]pyrimidine-2-sulfonamide	0	342	Test	O=S(=O)(Nc1c(F)cccc1(F))c2nc3nc(ccn3(n2))C	98967-40-9
pendimethalin	0	360	Test	O=[N+]([O-])c1cc(c(c(c1(NC(CC)CC))[N+](=O)[O-])C)C	40487-42-1
sodium p-nitrophenolate	0	364	Test	O=[N+]([O-])c1ccc(O)cc1	824-78-2
parathion-methyl| O,O-Dimethyl O-(4-nitrophenyl) ester phosphorothioic acid	1	366	Test	O=[N+]([O-])c1ccc(OP(OC)(OC)=S)cc1	298-00-0
fenitrothion| Phosphorothioic acid O,O-dimethyl O-(3-methyl-4-nitrophenyl)ester	1	367	Test	O=[N+]([O-])c1ccc(OP(OC)(OC)=S)cc1C	122-14-5
Phosphorothioic acid, O,O-Diethyl-O-(4-nitrophenyl)ester	1	368	Test	O=[N+]([O-])c1ccc(OP(OCC)(OCC)=S)cc1	56-38-2
tebuconazole| Tebuconazole	0	379	Test	OC(Cn1ncnc1)(CCc2ccc(cc2)Cl)C(C)(C)C	107534-96-3
cyproconazole	0	380	Test	OC(c1ccc(cc1)Cl)(Cn2ncnc2)C(C)C3CC3	94361-06-5
ancymidol| alpha-Cyclopropyl-alpha-(4-methoxyphenyl)-5-pyrimidinemethanol	0	381	Test	OC(c1cncnc1)(c2ccc(OC)cc2)C3CC3	12771-68-5
ipconazole	0	384	Test	OC3(Cn1ncnc1)(C(Cc2ccc(cc2)Cl)CCC3(C(C)C))	125225-28-7
metconazole	0	385	Test	OC3(Cn1ncnc1)(C(Cc2ccc(cc2)Cl)CCC3(C)(C))	125116-23-6
chlorpyrifos-methyl	1	392	Test	n1c(OP(OC)(OC)=S)c(cc(c1Cl)Cl)Cl	5598-13-0
chlorpyrifos	1	393	Test	n1c(OP(OCC)(OCC)=S)c(cc(c1Cl)Cl)Cl	2921-88-2
cyprodinil| Cyprodinil| 4-Cyclopropyl-6-methyl-N-phenyl-2-pyrimidinamine	0	396	Test	n1c(nc(cc1C)C2CC2)Nc3ccccc3	121552-61-2
propazine	1	399	Test	n1c(nc(nc1NC(C)C)Cl)NC(C)C	139-40-2

atrazine	0	400	Test	n1c(nc(nc1NC(C)C)Cl)NCC	1912-24-9
bromuconazole	0	409	Test	n1cnn(c1)CC2(OCC(C2)Br)(c3ccc(cc3Cl)Cl)	116255-48-2

Table S6.2. Split train and test for regression model.

Chemical Name	Exp_log	ID_QSAR	Set	SMILES	CAS Number
1,3-dichloropropene	-1,22173	1	Train	C(=CCl)CCl	542-75-6
lindane	-2,90345	2	Train	C1(C(C(C(C(C1Cl)Cl)Cl)Cl)Cl)Cl	58-89-9
Toxaphene	-0,90803	3	Train	C=C1C(CCl)(CCl)C2(C(C(C1(C2(Cl)Cl)Cl)Cl)Cl)Cl	8001-35-2
Tetrahydro-5,5-dimethyl-2(1H)-pyrimidio[3-[4-trifluoromethyl)phenyl]-1-[2-[4-(trifluoromethyl)phenyl]ethenyl]-2-propenylidene]hydrazone	-0,86782	4	Train	FC(F)(F)c3ccc(C=CC(=NN=C1NCC(C)(C)CN1)C=Cc2ccc(cc2)C(F)(F)F)cc3	67485-29-4
tetraconazole	-0,77006	5	Test	FC(F)C(F)(F)OCC(c1ccc(cc1Cl)Cl)Cn2ncnc2	112281-77-3
nuarimol	-1,16256	6	Train	Fc1ccc(cc1)C(O)(c2cncnc2)c3ccccc3Cl	63284-71-9
momfluorothrin	-3,28458	7	Test	N#CC(=CC2C(C(=O)OCc1c(F)c(F)c(c(F)c1(F))COC)C2(C)(C))C	609346-29-4
2,2-Dimethyl-3-(1,1,2,2-tetrabromoethyl)cyclopropanecarboxylic acid cyano(3-phenoxyphenyl)methyl ester	-3,70781	8	Train	N#CC(OC(=O)C1C(C(C(Br)(Br)Br)Br)C1(C)(C))c3cccc(Oc2ccccc2)c3	66841-25-6
acrinathrin	-3,80902	9	Train	N#CC(OC(=O)C1C(C=CC(=O)OC(C(F)(F)F)C(F)(F)F)C1(C)(C))c3cccc(Oc2ccccc2)c3	101007-06-1
acetamiprid	-1,43854	10	Train	N#CN=C(N(C)Cc1cnc(cc1)Cl)C	135410-20-7
sulfoxaflor	-2,86392	11	Train	N#CN=S(=O)(C)C(c1cnc(cc1)C(F)(F)F)C	946578-00-3
5-Amino-1-[2,6-dichloro-4-(trifluoromethyl)phenyl]-4-[(trifluoromethyl)sulfinyl]-1H-pyrazole-3-carbonitrile	-4,93172	12	Train	N#Cc1nn(c(N)c1S(=O)C(F)(F)F)c2c(cc(cc2Cl)C(F)(F)F)Cl	120068-37-3
4-Bromo-2-(4-chlorophenyl)-1-(ethoxymethyl)-5-(trifluoromethyl)-1H-pyrrole-3-carbonitrile	-3,52931	13	Train	N#Cc2c(c1ccc(cc1)Cl)n(c(c2Br)C(F)(F)F)COCC	122453-73-0

chlorthiamid	-0,42347	14	Test	NC(c1c(cccc1Cl)Cl)=S	1918-13-4
disulfoton	-2,45552	15	Train	O(CC)P(OCC)(=S)SCCSCC	298-04-4
terbufos	-1,84668	16	Test	O(CC)P(OCC)(=S)SCSC(C)(C)C	13071-79-9
O,O-Diethyl S-[(ethylthio)methyl])ester, Phosphorodithioic acid	-1,41197	17	Train	O(CC)P(OCC)(=S)SCSCC	298-02-2
Phosphorodithioic acid, S,S'-Methylene O,O,O',O'-tetraethyl ester	-1,27151	18	Test	O(CC)P(OCC)(=S)SCSP(OCC)(OCC)=S	563-12-2
chlorethoxyfos	-3,74545	19	Train	O(CC)P(OCC)(OC(C(Cl)(Cl)Cl)Cl)=S	54593-83-8
O,O-Dimethyl O-[3-methyl-4-(methylthio)phenyl]ester phosphorothioic acid	-2,95553	20	Train	O(c1ccc(c(c1)C)SC)P(OC)(OC)=S	55-38-9
1,1'-(2,2,2-Trichloroethylidene)bis[4-methoxybenzene]	-1,16422	21	Train	O(c1ccc(cc1)C(c2ccc(OC)cc2)C(Cl)(Cl)Cl)C	72-43-5
Etofenprox	-4,39933	22	Train	O(c1ccccc1)c2cccc(c2)COCC(c3ccc(OCC)cc3)(C)C	80844-07-1
spiroxamine	-1,84887	23	Train	O1CC(OC12(CCC(CC2)C(C)(C)C))CN(CC)CCC	118134-30-8
Phosphoric acid, 2-Chloro-3-(diethylamino)-1-methyl-3-oxo-1-propen-1-yl dimethyl ester	-2,31142	24	Train	O=C(C(=C(OP(=O)(OC)OC)C)Cl)N(CC)CC	13171-21-6
N,N-Bis(2-methylpropyl)carbamothioic acid S-ethyl ester	-0,87436	25	Train	O=C(N(CC(C)C)CC(C)C)SCC	2008-41-5
alachlor	-1,22582	26	Train	O=C(N(c1c(cccc1CC)CC)COC)CCl	15972-60-8
dimethoate	-3,1815	27	Train	O=C(NC)CSP(OC)(OC)=S	60-51-5
methabenzthiazuron	-2,44142	28	Test	O=C(NC)N(c2nc1ccccc1s2)C	18691-97-9
tebufenpyrad	-1,69658	29	Train	O=C(NCc1ccc(cc1)C(C)(C)C)c2c(c(nn2C)CC)Cl	119168-77-3
metoxuron	-1,15394	30	Test	O=C(Nc1ccc(OC)c(c1)Cl)N(C)C	19937-59-8

Chemical Name	Exp_log	ID_QSAR	Set	SMILES	CAS Number
chlorbromuron	-1,26121	31	Train	O=C(Nc1ccc(c(c1)Cl)Br)N(OC)C	13360-45-7
nicosulfuron	-0,73208	32	Train	O=C(Nc1nc(OC)cc(n1)OC)NS(=O)(=O)c2ncccc2(C(=O)N(C)C)	111991-09-4
chloroxuron	-1,2584	33	Test	O=C(Nc2ccc(Oc1ccc(cc1)Cl)cc2)N(C)C	1982-47-4
4-(2,4-Dichlorophenoxy)butanoic acid	-1,23308	34	Train	O=C(O)CCCOc1ccc(cc1Cl)Cl	94-82-6
3-Amino-2,5-dichlorobenzoic acid ammonium salt (1:1)	-1,18498	35	Train	O=C(O)c1cc(cc(N)c1Cl)Cl	1076-46-6
fenpyroximate	-1,42583	36	Train	O=C(OC(C)(C)C)c1ccc(cc1)CON=Cc3c(nn(c3(Oc2ccccc2))C)C	134098-61-6
2-(4-Methoxy[1,1'-biphenyl]-3-yl)hydrazinecarboxylic acid, 1-Methylethyl ester	-1,56658	37	Train	O=C(OC(C)C)NNc1cc(ccc1(OC))c2ccccc2	149877-41-8
chlorpropham	-0,34577	38	Train	O=C(OC(C)C)Nc1cccc(c1)Cl	101-21-3
propham	-1,04896	39	Test	O=C(OC(C)C)Nc1ccccc1	122-42-9
2-[[Ethoxy[(1-methylethyl)amino]phosphinothioyl]oxy]benzoic acid 1-methylethyl ester	-3,84777	40	Train	O=C(OC(C)C)c1ccccc1(OP(OCC)(NC(C)C)=S)	25311-71-1
pyrethrins (cinerin Ⅱ)	-4,44259	41	Test	O=C(OC)C(=CC2C(C(=O)OC1C(=C(C(=O)C1)CC=CC)C)C2(C)(C))C	121-20-0
pyrethrins (pyrethrin Ⅱ)	-4,45683	42	Train	O=C(OC)C(=CC2C(C(=O)OC1C(=C(C(=O)C1)CC=CC=C)C)C2(C)(C))C	121-29-9
pyrethrins (jasmolin Ⅱ)	-4,45917	43	Test	O=C(OC)C(=CC2C(C(=O)OC1C(=C(C(=O)C1)CC=CCC)C)C2(C)(C))C	1172-63-0
3-[(Dimethoxyphosphinyl)oxy]-2-butenoic acid, Methyl ester	-3,50524	44	Train	O=C(OC)C=C(OP(=O)(OC)OC)C	7786-34-7
indoxacarb	-3,60767	45	Train	O=C(OC)N(C(=O)N2N=C3c1ccc(cc1CC3(OC2)(C(=O)OC))Cl)c4ccc(OC(F)(F)F)cc4	173584-44-6
ethametsulfuron-methyl	-1,94825	46	Train	O=C(OC)c1ccccc1S(=O)(=O)NC(=O)Nc2nc(nc(n2)NC)OCC	97780-06-8

prallethrin	-4,04631	47	Train	O=C(OC1C(=C(C(=O)C1)CC#C)C)C2C(C=C(C)C)C2(C)(C)	23031-36-9
bioallethrin	-1,9488	48	Train	O=C(OC1C(=C(C(=O)C1)CC=C)C)C2C(C=C(C)C)C2(C)(C)	584-79-2
pyrethrins (cinerin I)	-4,38602	49	Test	O=C(OC1C(=C(C(=O)C1)CC=CC)C)C2C(C=C(C)C)C2(C)(C)	25402-06-6
pyrethrins (pyrethrin I)	-4,4022	50	Test	O=C(OC1C(=C(C(=O)C1)CC=CC=C)C)C2C(C=C(C)C)C2(C)(C)	121-21-1
pyrethrins (jasmolin I)	-4,40486	51	Train	O=C(OC1C(=C(C(=O)C1)CC=CCC)C)C2C(C=C(C)C)C2(C)(C)	4466-14-2
alpha-[(Dimethoxyphosphinothioyl)thio] benzeneacetic acid, Ethyl ester	-3,01947	52	Train	O=C(OCC)C(c1ccccc1)SP(OC)(OC)=S	2597-03-7
malathion	-3,31444	53	Train	O=C(OCC)CC(C(=O)OCC)SP(OC)(OC)=S	121-75-5
triazamate	-1,06576	54	Train	O=C(OCC)CSc1nc(nn1(C(=O)N(C)C))C(C)(C)C	112143-82-5
chlorimuron-ethyl	-1,52013	55	Train	O=C(OCC)c1ccccc1S(=O)(=O)NC(=O)Nc2nc(OC)cc(n2)Cl	90982-32-4
(1,3,4,5,6,7-Hexahydro-1,3-dioxo-2H-isoindol-2-yl)methyl ester 2,2-dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid	-3,32973	56	Train	O=C(OCN1C(=O)C2=C(C1(=O))CCCC2)C3C(C=C(C)C)C3(C)(C)	7696-12-0
imiprothrin	-2,90058	57	Train	O=C(OCN1C(=O)N(CC#C)CC1(=O))C2C(C=C(C)C)C2(C)(C)	72963-72-5
tefluthrin	-3,17408	58	Train	O=C(OCc1c(F)c(F)c(c(F)c1(F))C)C2C(C=C(C(F)(F)F)Cl)C2(C)(C)	79538-32-2
resmethrin	-3,72982	59	Train	O=C(OCc1coc(c1)Cc2ccccc2)C3C(C=C(C)C)C3(C)(C)	10453-86-8

Chemical Name	Exp_log	ID_QSAR	Set	SMILES	CAS Number
2,2-Dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid (3-phenoxyphenyl)methyl ester	-3,71823	60	Test	O=C(OCc2cccc(Oc1ccccc1)c2)C3C(C=C(C)C)C3(C)(C)	26002-80-2
permethrin	-4,21094	61	Train	O=C(OCc2cccc(Oc1ccccc1)c2)C3C(C=C(Cl)Cl)C3(C)(C)	52645-53-1
thiodicarb	-2,0577	62	Train	O=C(ON=C(C)SC)N(C)SN(C(=O)ON=C(C)SC)C	59669-26-0
methomyl	-3,00552	63	Train	O=C(ON=C(C)SC)NC	16752-77-5
thiofanox	-3,57525	64	Train	O=C(ON=C(CSC)C(C)(C)C)NC	39196-18-4
2-Methyl-2-(methylthio)propanol O-[(methylamino)carbonyl]oxime	-2,82409	65	Train	O=C(ON=CC(C)(C)SC)NC	116-06-3
methiocarb	-2,88446	66	Train	O=C(Oc1cc(c(c(c1)C)SC)C)NC	2032-65-7
aminocarb	-3,23553	67	Test	O=C(Oc1ccc(c(c1)C)N(C)C)NC	2032-59-9
2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate)	-2,71703	68	Train	O=C(Oc1cccc2OC(Oc12)(C)C)NC	22781-23-3
carbaryl	-3,15724	69	Train	O=C(Oc1cccc2ccccc12)NC	63-25-2
Dimethylcarbamic acid, 2-(Dimethylamino)-5,6-dimethyl-4-pyrimidinyl ester	-1,10453	70	Train	O=C(Oc1nc(nc(c1C)C)N(C)C)N(C)C	23103-98-2
benfuracarb	-3,33423	71	Train	O=C(Oc2cccc1c2(OC(C)(C)C1))N(C)SN(CCC(=O)OCC)C(C)C	82560-54-1
carbosulfan	-3,32475	72	Test	O=C(Oc2cccc1c2(OC(C)(C)C1))N(C)SN(CCCC)CCCC	55285-14-8
2-(Difluoromethyl)-4-(2-methylpropyl)-6-(trifluoromethyl)-3,5-pyridinedicarbothioic acid, S3,S5-Dimethyl ester	-0,69472	73	Train	O=C(c1c(nc(c(C(=O)SC)c1CC(C)C)C(F)(F)F)C(F)F)SC	97886-45-8

pyridaben	-4,18105	74	Train	O=C1C(=C(C=NN1C(C)(C)C)SCc2ccc(cc2)C(C)(C)C)Cl	96489-71-3
2-[1-(Ethoxyimino)butyl]-5-[2-(ethylthio)propyl]-3-hydroxy-2-cyclohexen-1-one	-1,5148	75	Train	O=C1C(=C(O)CC(C1)CC(C)SCC)C(=NOCC)CCC	74051-80-2
fosthiazate	-3,04362	76	Train	O=C1N(CCS1)P(=O)(OCC)SC(C)CC	98886-44-3
azinphos-methyl	-2,87782	77	Train	O=C1c2ccccc2(N=NN1CSP(OC)(OC)=S)	86-50-0
phosalone	-1,9212	78	Train	O=C2Oc1cc(ccc1N2CSP(OCC)(OCC)=S)Cl	2310-17-0
6-Methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one	-0,54658	79	Train	O=C2Sc1nc3ccc(cc3(nc1S2))C	2439-01-2
formothion	-3,23383	80	Train	O=CN(C(=O)CSP(OC)(OC)=S)C	2540-82-1
Phosphoramidothioic acid, O,S-Dimethyl ester	-2,0125	81	Train	O=P(OC)(N)SC	10265-92-6
P-(2,2,2-Trichloro-1-hydroxyethyl)phosphonic acid dimethyl ester	-0,63141	82	Train	O=P(OC)(OC)C(O)C(Cl)(Cl)Cl	52-68-6
Phosphoric acid 1,2-dibromo-2,2-dichloroethyl dimethyl ester	-2,896	83	Train	O=P(OC)(OC)OC(C(Cl)(Cl)Br)Br	300-76-5
Phosphoric acid 2,2-dichloroethenyl dimethyl ester	-2,64335	84	Test	O=P(OC=C(Cl)Cl)(OC)OC	62-73-7
cadusafos	-2,39808	85	Test	O=P(OCC)(SC(C)CC)SC(C)CC	95465-99-9
ethoprophos	-1,63884	86	Test	O=P(OCC)(SCCC)SCCC	13194-48-4
profenofos	-3,59274	87	Train	O=P(Oc1ccc(cc1Cl)Br)(OCC)SCCC	41198-08-7
tribufos	-0,85164	88	Test	O=P(SCCCC)(SCCCC)SCCCC	78-48-8
dichlofluanid	-1,31697	89	Train	O=S(=O)(N(c1ccccc1)SC(F)(Cl)Cl)N(C)C	1085-98-9
bensulide	-1,21865	90	Train	O=S(=O)(NCCSP(OC(C)C)(OC(C)C)=S)c1ccccc1	741-58-2

Chemical Name	Exp_log	ID_QSAR	Set	SMILES	CAS Number
6,7,8,9,10,10-Hexachloro-1,5,5a,6,9,9a-hexahydro-6,9-methano-2,4,3-benzodioxathiepin 3-oxide	-1,95297	91	Train	O=S1OCC2C(CO1)C3(C(=C(C2(C3(Cl)Cl)Cl)Cl)Cl)Cl	115-29-7
N''-Methyl-N-nitro-N'-[(tetrahydro-3-furanyl)methyl]guanidine	-3,63348	92	Train	O=[N+]([O-])N=C(NC)NCC1COCC1	165252-70-0
Clothianidin	-3,82889	93	Train	O=[N+]([O-])N=C(NC)NCc1cnc(Cl)s1	210880-92-5
2-(1-Methylpropyl)-4,6-dinitrophenol	-0,87168	94	Train	O=[N+]([O-])c1cc(c(O)c(c1)C(C)CC)[N+](=O)[O-]	88-85-7
oryzalin	-0,92853	95	Train	O=[N+]([O-])c1cc(cc(c1N(CCC)CCC)[N+](=O)[O-])S(=O)(=O)N	19044-88-3
fluchloralin	-0,56802	96	Test	O=[N+]([O-])c1cc(cc(c1N(CCC)CCCl)[N+](=O)[O-])C(F)(F)F	33245-39-5
fenitrothion	-3,04885	97	Test	O=[N+]([O-])c1ccc(OP(OC)(OC)=S)cc1C	122-14-5
Phosphorothioic acid, O,O-Diethyl-O-(4-nitrophenyl)ester	-3,2209	98	Test	O=[N+]([O-])c1ccc(OP(OCC)(OCC)=S)cc1	56-38-2
P-Phenylphosphonothioic acid O-ethyl O-(4-nitrophenyl) ester	-3,12009	99	Train	O=[N+]([O-])c2ccc(OP(OCC)(c1ccccc1)=S)cc2	2104-64-5
diniconazole	-1,21095	100	Test	OC(C(=Cc1ccc(cc1Cl)Cl)n2ncnc2)C(C)(C)C	83657-24-3
2-(Octylthio)ethanol	-0,52396	101	Train	OCCSCCCCCCCC	3547-33-9
pentachlorophenol	-0,74011	102	Train	Oc1c(c(c(c(c1Cl)Cl)Cl)Cl)Cl	87-86-5
3,5,6-trichloro-2-pyridinol	-0,71565	103	Train	Oc1nc(c(cc1Cl)Cl)Cl	6515-38-4
DDT	-2,81404	104	Train	c1cc(ccc1C(c2ccc(cc2)Cl)C(Cl)(Cl)Cl)Cl	50-29-3
etrimfos	-3,46548	105	Train	n1c(OCC)cc(nc1CC)OP(OC)(OC)=S	38260-54-7
chlorpyrifos-methyl	-3,33027	106	Train	n1c(OP(OC)(OC)=S)c(cc(c1Cl)Cl)Cl	5598-13-0
chlorpyrifos	-3,77188	107	Train	n1c(OP(OCC)(OCC)=S)c(cc(c1Cl)Cl)Cl	2921-88-2
diazinon	-3,14195	108	Train	n1c(OP(OCC)(OCC)=S)cc(nc1C(C)C)C	333-41-5

propazine	-1,15592	109	Train	n1c(nc(nc1NC(C)C)Cl)NC(C)C	139-40-2
prometon	-0,79619	110	Train	n1c(nc(nc1NC(C)C)NC(C)C)OC	1610-18-0
N2,N4-Bis(1-methylethyl)-6-(methylthio)-1,3,5-triazine-2,4-diamine	-0,39175	111	Test	n1c(nc(nc1NC(C)C)SC)NC(C)C	7287-19-6
imazalil	-0,8803	112	Train	n1ccn(c1)CC(OCC=C)c2ccc(cc2Cl)Cl	35554-44-0
triazophos	-0,71893	113	Train	n1cn(nc1OP(OCC)(OCC)=S)c2ccccc2	24017-47-8

Table S6.3. Tuning parameters for classification models.

algorithm	***RF***
mtry	5
splitrule	gini
min.node.size	1
algorithm	***DT***
cp	0,019
algorithm	***KNN***
kmax	9
distance	2
kernel	optimal

Table S6.4. Tuning parameters for regression models.

algorithm	***RF***
mtry	13
splitrule	extratrees
min.node.size	5
algorithm	***MLR***
intercept	TRUE
algorithm	***PLS***
ncomp	2
algorithm	***DT***
cp	0
algorithm	***KNN***
kmax	5
distance	2
kernel	optimal

Table S6.5. Random forest classification prediction and descriptors.

CLASS_ID	Exp	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	X0Av	F02[C-P]	ATSC1e	MATS3v	SpMax1_Bh(s)	SpMAD_X	RBF	B05[C-P]	MATS1s
1	1	0	0	0	2,38031	-0,31635	-0,92153	0,68527	-2,49935	3,24344	-2,03182	-0,24759	1,00046
2	1	1	1	1	2,73748	-0,31635	0,00645	0,34106	-2,40551	1,30288	-2,03182	-0,24759	1,09616
3	0	0	0	0	0,8564	-0,31635	-0,94944	-1,13331	-2,71126	-0,1202	-1,38304	-0,24759	-1,243
4	1	1	1	1	2,36841	-0,31635	0,33438	0,5017	-2,25819	0,23557	-2,03182	-0,24759	0,35188
5	0	0	0	1	5,7853	-0,31635	-1,00526	-4,47216	-3,38733	5,4104	-0,13669	-0,24759	-2,28499
6	0	0	1	0	1,52311	-0,31635	-0,8378	-1,51194	-3,00187	-0,99345	-2,03182	-0,24759	-1,60451
7	0	0	0	0	-0,95324	-0,31635	0,29949	-0,03183	1,13228	-0,76705	-0,23913	-0,24759	-0,3605
8	0	0	0	0	0,02301	-0,31635	3,29273	-0,4162	1,1353	-2,06076	-0,47815	-0,24759	-0,44556
9	0	0	0	0	-1,21516	-0,31635	2,27405	0,16896	1,13329	-1,09048	-0,23913	-0,24759	-0,08406
10	0	0	0	0	-0,89371	-0,31635	0,74603	-0,7432	1,12925	-0,92877	-0,08547	-0,24759	-0,26481
11	0	0	0	0	-0,2151	-0,31635	1,73681	0,62217	1,12925	-0,63768	1,46819	-0,24759	0,16049
12	1	0	0	0	-0,98896	-0,31635	0,56463	-1,32836	1,13329	-1,0258	-0,63181	-0,24759	0,0967
13	1	0	0	0	-0,72704	-0,31635	1,52051	-0,26704	1,02733	-0,54066	0,8877	-0,24759	0,37314
16	0	0	0	0	-0,4413	-0,31635	-0,88664	-0,2441	0,80433	-0,47597	-0,87084	-0,24759	-0,23291
18	1	0	0	0	-0,84609	-0,31635	-0,58662	0,00259	0,81442	0,52665	-0,61474	-0,24759	-0,14785
19	0	0	0	0	-0,82228	-0,31635	-0,49592	-0,32441	0,81442	0,13854	-0,27328	-0,24759	0,71339
20	0	0	0	0	-0,36987	-0,31635	-0,9006	-0,21541	0,8235	-0,60534	-0,35864	-0,24759	0,54326
21	0	0	0	1	-0,33415	-0,31635	-0,30753	-1,1161	0,8235	0,04151	-0,0001	-0,24759	0,90477
22	0	0	0	0	-0,03652	-0,31635	-0,76105	0,0829	0,81442	-0,573	-0,2562	-0,24759	0,29872
23	0	0	0	0	-0,96515	-0,31635	1,12281	1,06391	0,81442	0,97945	0,30721	-0,24759	0,80908
24	0	0	0	0	-0,70323	-0,31635	0,7879	-0,06052	0,97688	-0,08786	-0,78547	-0,24759	0,01164
25	0	0	0	0	-1,06039	-0,31635	0,42508	0,22059	1,12925	0,04151	0,35843	-0,24759	0,0329
26	0	0	0	0	-0,53655	-0,31635	-0,3773	0,31812	0,81442	-0,15254	-0,88791	-0,24759	0,36251
27	0	0	0	0	-0,56036	-0,31635	0,34135	0,58775	1,12925	0,13854	-0,47815	-0,24759	0,59643
29	0	0	0	0	-0,79847	-0,31635	0,0204	1,77528	0,05459	1,52928	0,1877	-0,24759	0,89414
30	0	0	0	0	-0,66751	-0,31635	-0,08426	-0,10641	1,12925	-0,41128	0,66575	-0,24759	0,46884
31	0	0	0	0	0,08254	-0,31635	-0,67733	0,0829	-0,58213	-0,44363	0,32429	-0,24759	0,51137
32	1	0	0	0	-0,78656	-0,31635	0,15297	0,00833	1,12925	-0,24957	0,58039	-0,24759	0,33061
33	1	0	0	0	-0,48893	-0,31635	-1,07503	0,03701	0,03845	0,52665	0,49502	-0,24759	0,89414
34	1	1	1	1	1,4874	-0,31635	-0,71919	-0,28425	0,03845	-0,50831	0,30721	-0,24759	0,90477

40	0	0	0	0	-0,63179	-0,31635	-0,64244	1,40812	0,04249	1,27054	-2,03182	-0,24759	1,59589
41	0	0	0	0	-1,59614	-0,31635	0,69719	0,04848	1,12925	0,04151	-0,47815	-0,24759	0,34125
42	1	1	1	1	-0,02461	-0,31635	-0,97037	-1,13905	-0,56699	0,88242	-0,0684	-0,24759	1,77664
43	1	0	0	0	-0,56036	-0,31635	0,24367	-0,23262	1,12925	-1,0258	-0,81962	-0,24759	0,24555
44	0	0	0	0	0,57067	-0,31635	-0,0424	0,94343	-0,57405	1,10882	-2,03182	-0,24759	2,60598
45	0	0	0	0	0,22541	-0,31635	-0,81687	0,83443	-0,5791	0,65602	-2,03182	-0,24759	1,57462
46	0	0	0	0	-1,56043	-0,31635	0,4809	0,31238	0,97789	-0,31426	-1,38304	-0,24759	0,10733
47	0	0	0	0	1,24928	-0,31635	-0,98433	1,50565	0,03845	1,20585	1,48527	-0,24759	1,68095
48	1	0	0	0	-0,92943	-0,31635	2,95085	-0,60552	1,13429	-1,54328	-0,92206	-0,24759	-0,07342
49	1	0	0	0	-0,07223	-0,31635	0,50881	1,27617	1,12925	-0,15254	-0,08547	-0,24759	0,44757
50	0	0	0	0	-0,72704	-0,31635	-0,60755	1,0008	0,81442	0,97945	1,22917	-0,24759	0,93667
51	0	0	0	0	0,07063	-0,31635	-0,37033	-0,82926	0,32502	-0,28191	-0,56352	-0,24759	-0,08406
52	0	0	0	0	-1,01277	-0,31635	-0,25171	0,24927	0,03946	1,23819	0,75112	-0,24759	0,74528
53	0	0	1	0	0,10635	-0,31635	-0,80292	-1,43163	-2,50642	-0,18489	-0,54645	-0,24759	0,19239
54	0	1	1	1	2,73748	-0,31635	-0,91455	-0,1351	-2,50339	-0,89643	0,52917	-0,24759	-1,55135
55	1	1	1	1	0,98736	-0,31635	-0,74012	0,00259	-2,18049	-0,70237	-0,97328	-0,24759	-2,1574
58	1	1	1	1	3,22561	2,92624	-0,60058	2,03344	-2,46605	1,10882	2,68039	4,02661	-2,82725
59	1	1	1	1	3,04702	6,16883	-0,08426	2,7276	-2,44789	0,39728	3,09015	-0,24759	-3,11433
60	1	1	1	1	2,72557	2,92624	0,82278	2,44075	-2,1926	-1,25219	2,06576	-0,24759	-2,34879
61	0	1	1	1	1,63026	2,92624	0,19483	-0,84073	-2,26929	0,49431	0,49502	4,02661	-2,08297
62	1	1	1	1	1,65407	2,92624	-0,39824	-0,15804	-2,31167	1,20585	0,71697	4,02661	-2,79535
63	1	0	0	0	0,43971	-0,31635	-0,75408	-0,04904	-2,25919	0,17088	-0,18791	-0,24759	-0,49872
64	1	1	1	1	-0,28653	-0,31635	-0,8378	-0,50225	-2,46908	0,26791	0,61453	-0,24759	0,11796
65	0	0	0	0	0,78497	-0,31635	-0,62151	-0,87515	-2,66988	-1,54328	-0,5123	-0,24759	-1,40249
66	0	0	0	0	0,4278	-0,31635	-0,78198	1,40812	-2,47211	-1,54328	-2,03182	-0,24759	-2,05108
67	1	1	1	1	1,24928	-0,31635	0,19483	1,13849	-2,23397	-1,0258	-2,03182	-0,24759	1,19185
68	1	1	1	1	0,92783	-0,31635	-0,5936	-0,34736	-2,57301	-0,24957	-0,23913	-0,24759	-0,96656
69	1	1	1	1	1,0588	2,92624	0,36229	1,24749	0,03441	1,78802	1,28039	-0,24759	-0,59442
70	0	0	0	0	-0,32225	-0,31635	-0,83083	0,48448	0,03441	1,20585	-0,35864	-0,24759	0,77718
71	0	1	1	1	2,10648	-0,31635	-0,86571	1,07538	0,0334	-0,67003	0,71697	-0,24759	0,50073

CLASS_ID	Exp	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	X0Av	F02[C-P]	ATSC1e	MATS3v	SpMax1_Bh(s)	SpMAD_X	RBF	B05[C-P]	MATS1s
72	0	0	0	0	0,35637	-0,31635	-1,03317	0,65085	0,0334	1,30288	0,40965	-0,24759	0,56453
73	0	0	0	0	-0,36987	-0,31635	-0,94944	-0,38752	0,0334	0,72071	-0,61474	-0,24759	0,61769
74	1	1	1	1	1,77313	-0,31635	-0,82385	-0,5711	0,0334	-0,54066	0,81941	-0,24759	0,44757
75	0	0	0	0	-0,13176	-0,31635	-0,94246	1,28191	0,0334	1,14116	-0,0001	-0,24759	0,47947
76	0	1	0	0	0,8445	-0,31635	-1,08201	0,90901	0,0334	1,27054	0,63161	-0,24759	0,72402
78	0	0	0	0	-0,08414	-0,31635	-0,88664	-0,43915	0,0334	0,75305	-0,61474	-0,24759	0,72402
80	0	0	0	0	0,58257	-0,31635	-1,02619	0,50743	0,0334	1,04414	0,05112	-0,24759	0,35188
82	1	0	0	0	0,45161	-0,31635	-1,08898	0,84591	0,0334	2,37019	0,22185	-0,24759	0,75592
83	0	0	0	0	-0,76275	-0,31635	1,10885	0,77133	1,12925	-0,76705	0,22185	-0,24759	0,34125
84	0	0	0	0	-0,88181	-0,31635	-0,96339	-0,0261	0,04349	0,04151	-0,71718	-0,24759	-0,13722
85	0	0	0	0	0,2135	0,76451	0,56463	0,58775	0,10908	-0,99345	1,16088	-0,24759	-2,35942
86	0	0	0	0	-0,8699	-0,31635	1,2484	0,30091	1,12925	-0,60534	-0,68303	-0,24759	0,79845
87	0	0	0	0	-1,25088	-0,31635	1,08792	0,58775	1,12925	-0,37894	0,46087	-0,24759	0,70275
90	0	0	0	0	-1,20326	-0,31635	3,61369	0,27796	1,02733	-1,09048	-0,23913	-0,24759	0,29872
91	0	0	0	0	-1,56043	-0,31635	4,24862	-0,16952	1,13429	-1,73734	-1,10986	-0,24759	0,28808
97	0	1	0	0	0,74925	-0,31635	-0,65639	0,0198	0,03946	0,81774	-0,13669	-0,24759	0,522
98	0	0	1	1	0,36827	-0,31635	-0,96339	-0,28425	0,03946	-0,92877	-0,01718	-0,24759	0,16049
99	0	0	0	0	0,11826	-0,31635	-0,41219	-1,44884	0,05157	-1,86671	-1,07572	-0,24759	-0,04153
100	0	0	0	0	-0,14367	-0,31635	-0,09124	0,15175	0,06166	-1,12282	-1,05864	-0,24759	0,58579
102	0	0	0	0	-0,17938	-0,31635	-0,97037	-0,38178	0,03441	1,20585	0,15355	-0,24759	0,37314
103	1	1	1	1	2,18982	2,92624	-0,20287	0,28943	0,03542	2,66127	1,67307	4,02661	-0,56252
104	0	1	1	0	0,8564	-0,31635	-0,83083	-1,32836	0,03643	-0,31426	-0,92206	-0,24759	0,54326
105	1	0	0	0	-0,04842	-0,31635	-0,91455	0,46727	0,03643	1,23819	-1,40011	-0,24759	0,27745
106	0	0	0	0	0,15397	-0,31635	-1,01921	0,63938	0,05056	0,1062	-1,29767	-0,24759	1,05363
107	0	0	0	0	-0,52464	-0,31635	-0,78896	-0,3072	0,03441	1,33522	1,24624	-0,24759	0,4901
108	1	1	1	1	0,39208	-0,31635	-0,7471	-0,50799	0,03542	-0,28191	-0,2562	-0,24759	0,35188
109	0	0	0	0	-0,47702	-0,31635	1,17165	0,01406	1,12925	-0,96111	-0,42693	-0,24759	0,22429
110	0	0	0	0	0,10635	-0,31635	-0,64244	-0,86368	0,04047	-0,47597	-0,47815	-0,24759	0,31998
111	0	0	0	0	0,10635	-0,31635	-0,74012	-0,9612	0,03946	-0,63768	-0,5123	-0,24759	0,29872
114	0	0	0	0	-1,06039	-0,31635	2,26708	-0,25557	1,13228	-1,12282	-0,23913	-0,24759	-0,28607

115	0	0	0	0	-1,53661	-0,31635	1,53447	-0,19246	1,12925	-0,54066	-0,35864	-0,24759	0,23492
116	0	0	0	0	-1,44137	-0,31635	0,84372	-0,06625	1,12925	-0,37894	-0,35864	-0,24759	0,10733
117	0	0	0	0	0,26112	-0,31635	-0,72617	0,04275	0,03441	0,49431	-1,10986	-0,24759	0,53263
118	1	0	0	0	0,18969	-0,31635	-0,70523	-1,02431	0,03643	0,1062	-0,81962	-0,24759	0,22429
119	1	1	1	1	0,95165	-0,31635	-0,21683	0,2378	0,03643	1,01179	-0,66596	-0,24759	0,26682
121	0	0	1	0	0,49923	-0,31635	-0,64244	-1,08168	0,03643	-0,05552	-1,31474	-0,24759	0,16049
123	0	0	0	0	0,65401	-0,31635	-0,4401	-0,33015	0,03643	1,36756	-0,66596	-0,24759	0,36251
124	0	1	1	1	0,27303	-0,31635	-0,9006	-0,85794	0,03643	-0,28191	-0,99035	-0,24759	0,28808
126	0	0	0	0	-0,17938	-0,31635	2,21126	-0,4621	1,15246	-2,28716	-0,0684	-0,24759	0,11796
127	0	0	0	0	-0,90562	-0,31635	-0,09821	0,46154	1,12925	-1,0258	-0,3245	-0,24759	0,0329
130	0	0	0	0	-0,13176	-0,31635	-0,9913	0,22059	0,03542	0,59134	-0,88791	-0,24759	0,56453
131	0	0	0	0	-0,31034	-0,31635	-0,94246	-0,12936	0,03542	0,559	-0,54645	-0,24759	1,06426
133	0	0	0	0	-0,71513	-0,31635	-0,50289	-0,47357	0,03744	0,59134	-0,61474	-0,24759	0,42631
134	0	0	0	0	-0,13176	-0,31635	-0,18892	-0,18673	1,12925	-1,28454	-0,13669	-0,24759	0,26682
135	0	0	0	0	-0,51274	-0,31635	0,66231	-0,27852	0,32906	-0,15254	0,35843	-0,24759	-0,47746
136	1	0	0	0	-0,53655	-0,31635	0,69719	-0,10067	0,32704	0,07386	0,15355	-0,24759	-0,40303
138	0	0	0	0	-0,36987	-0,31635	0,80883	0,88033	0,3331	0,07386	0,56331	-0,24759	-0,60505
140	0	0	0	0	-0,39368	-0,31635	0,37624	0,42138	0,32704	-0,63768	-0,13669	-0,24759	-0,71138
141	0	0	0	0	-0,10795	-0,31635	-0,57964	0,42712	0,03542	1,59396	0,40965	-0,24759	1,03236
143	0	0	0	0	-0,47702	-0,31635	-0,82385	0,21485	0,03542	0,49431	-0,64889	-0,24759	0,23492
144	0	0	0	0	-2,07236	-0,31635	0,27856	-2,58473	0,16256	-0,86408	0,40965	-0,24759	-5,62361
145	0	0	0	0	0,24922	0,76451	0,22972	-1,07594	0,08083	-1,67265	1,07551	-0,24759	-1,54071
146	0	0	0	0	-0,75085	-0,31635	-0,20287	-1,06447	0,08285	-0,86408	-0,47815	-0,24759	-1,91285
148	0	0	0	0	-0,84609	-0,31635	-0,08426	0,37548	0,08184	-0,02317	-0,7684	-0,24759	0,22429
149	0	0	0	0	-0,34606	-0,31635	-0,62151	0,98359	0,08688	-0,21723	-0,56352	-0,24759	0,1286
150	0	0	0	0	-0,28653	-0,31635	-0,34242	-1,24805	0,08184	-0,89643	1,07551	-0,24759	0,13923
151	0	0	0	0	-0,10795	-0,31635	-0,64942	-0,02036	0,08083	-0,18489	1,02429	-0,24759	0,17113
152	1	0	0	0	-0,0008	-0,31635	-0,20985	0,66806	0,08083	-0,18489	1,38283	-0,24759	0,1286
153	0	0	0	0	-0,41749	-0,31635	1,15769	-1,05873	0,08285	-2,5459	1,98039	-0,24759	-2,6146
155	0	0	0	0	-0,15557	-0,31635	0,15995	-0,38752	0,08083	-0,50831	0,66575	-0,24759	-0,57315

CLASS_ID	Exp	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	X0Av	F02[C-P]	ATSC1e	MATS3v	SpMax1_Bh(s)	SpMAD_X	RBF	B05[C-P]	MATS1s
156	0	0	0	0	-1,09611	-0,31635	-0,73314	-0,1351	0,08083	0,42963	-0,0684	-0,24759	-0,21165
157	0	0	0	0	-0,59608	-0,31635	1,06001	-0,28999	0,81644	0,39728	0,52917	-0,24759	-1,3068
159	0	0	0	0	-0,98896	-0,31635	-0,96339	0,03701	0,08083	0,04151	-0,66596	-0,24759	0,01164
161	0	0	0	0	0,04682	-0,31635	0,15995	0,1919	0,08184	0,559	-0,23913	-0,24759	-0,65821
162	0	0	0	0	-0,51274	-0,31635	-0,57964	-0,65715	0,08184	-0,21723	-1,31474	-0,24759	-0,3286
164	1	0	0	0	-0,15557	-0,31635	-0,00053	-0,92105	0,08184	-0,44363	-1,0245	-0,24759	-1,81716
165	0	0	0	0	-1,47709	-0,31635	3,38344	0,31812	0,98294	-1,38157	0,46087	-0,24759	-0,54125
166	0	0	0	0	-0,70323	-0,31635	-0,23776	-0,8981	0,08184	-0,1202	-0,81962	-0,24759	-1,03035
167	0	0	0	0	-1,98903	-0,31635	1,27631	-2,034	1,12925	-1,93139	-0,88791	-0,24759	-1,04098
169	0	0	0	0	-0,17938	-0,31635	0,10413	0,3468	0,08184	0,04151	-0,81962	-0,24759	-1,26427
170	0	0	0	0	-0,69132	-0,31635	-0,11915	2,09654	0,08184	-0,08786	-0,54645	-0,24759	-0,56252
174	0	0	0	0	-0,6556	-0,31635	-0,64942	0,11159	0,08184	-0,47597	-0,81962	-0,24759	-0,02026
175	0	0	0	0	-0,58417	-0,31635	-0,53778	0,04275	0,08184	-0,47597	-0,56352	-0,24759	-0,02026
176	0	0	0	0	-0,92943	-0,31635	1,31817	0,40991	0,08184	0,81774	0,97307	-0,24759	-0,67948
177	0	0	0	0	-1,16754	-0,31635	-0,73314	0,22633	0,08184	0,39728	-0,64889	-0,24759	-0,03089
178	0	0	0	0	-0,52464	-0,31635	-0,75408	0,72543	0,0445	0,559	0,56331	-0,24759	0,31998
179	0	0	0	0	-0,98896	-0,31635	2,40662	1,1098	1,12925	-0,573	1,22917	-0,24759	0,39441
180	0	0	0	0	-0,78656	-0,31635	1,71587	0,03701	1,12925	-0,60534	0,64868	-0,24759	0,13923
181	1	0	0	0	-0,36987	-0,31635	-0,49592	-0,78336	0,03946	-0,47597	0,52917	-0,24759	0,58579
182	0	0	0	0	0,04682	-0,31635	-0,56569	0,0198	0,04551	-0,15254	1,12673	-0,24759	0,51137
183	0	0	0	0	0,02301	-0,31635	-0,61453	0,40417	0,81442	-0,73471	0,30721	-0,24759	0,44757
184	1	0	0	0	-0,46511	-0,31635	-0,75408	0,31238	0,04148	0,62368	0,35843	-0,24759	0,23492
185	0	0	0	0	-0,02461	-0,31635	-0,89362	1,38517	0,04148	0,23557	0,956	-0,24759	0,1286
187	1	0	1	1	-0,32225	-0,31635	-1,03317	0,09438	0,04148	-0,31426	-0,0684	-0,24759	0,1286
188	1	1	1	1	0,86831	2,92624	-0,50987	1,17291	0,03946	-0,86408	1,31453	4,02661	-0,96656
193	0	0	0	0	-1,02468	-0,31635	0,69022	-0,0892	1,12925	0,49431	0,75112	-0,24759	0,07543
194	0	0	0	0	-0,89371	-0,31635	0,85767	-0,04331	1,12925	0,46197	0,75112	-0,24759	0,33061
195	0	0	0	0	-0,51274	-0,31635	-0,58662	-0,07199	0,03946	1,39991	0,75112	-0,24759	0,9792
196	0	0	0	0	0,04682	-0,31635	-0,97037	0,15175	0,04148	0,97945	0,46087	-0,24759	0,61769
199	0	0	0	0	-0,83419	-0,31635	1,36003	0,05422	1,12925	-0,15254	0,59746	-0,24759	-0,03089

200	1	1	1	1	0,51114	2,92624	0,78092	-0,75468	0,04047	1,36756	1,9121	-0,24759	-1,00909
201	0	0	0	0	0,03492	-0,31635	0,12506	1,25322	0,81442	0,52665	-0,0001	-0,24759	1,03236
202	1	0	0	0	-1,02468	-0,31635	2,53221	-0,3072	1,13228	-0,1202	-0,5123	-0,24759	0,23492
203	0	0	0	0	-0,58417	-0,31635	0,00645	0,00833	0,03946	1,01179	0,51209	-0,24759	0,522
204	0	0	0	0	-0,20319	-0,31635	0,18785	-0,46783	0,04349	0,78539	-0,13669	-0,24759	-0,14785
205	0	0	0	0	-0,51274	-0,31635	-0,07728	-0,84073	0,32704	-0,21723	0,01697	-0,24759	-1,10478
206	0	0	0	0	0,59448	-0,31635	1,05303	1,28764	0,04148	1,14116	0,81941	-0,24759	0,36251
207	0	0	0	0	-0,40559	-0,31635	1,44376	0,54185	0,32704	0,04151	0,68283	-0,24759	-0,48809
208	0	0	0	0	0,16588	-0,31635	0,69022	-0,03757	0,32704	-0,28191	0,40965	-0,24759	-0,47746
209	0	0	0	0	-0,76275	-0,31635	1,96008	-0,27278	1,12925	-1,05814	0,54624	-0,24759	-0,24354
211	0	0	0	0	-0,54845	-0,31635	0,71115	0,81722	0,32704	0,49431	0,56331	-0,24759	0,02227
214	1	0	0	0	-0,63179	-0,31635	0,7251	1,2647	0,32704	0,36494	0,87063	-0,24759	-0,44556
215	0	0	0	0	-0,20319	-0,31635	0,91349	-0,39325	0,32704	0,1062	0,10234	-0,24759	-0,10532
216	0	0	0	0	-0,53655	-0,31635	1,10885	0,80001	1,12925	-0,83174	-0,08547	-0,24759	0,42631
218	0	0	0	0	0,10635	-0,31635	-0,82385	0,3468	0,03946	-0,08786	-0,40986	-0,24759	0,63896
219	0	0	0	0	-0,227	-0,31635	-0,92851	0,38122	0,04551	0,30025	-0,20498	-0,24759	0,74528
220	1	1	1	1	0,14207	-0,31635	-0,96339	-1,47178	0,03946	-0,73471	-0,22206	-0,24759	0,73465
225	0	0	0	0	-0,66751	-0,31635	0,37624	0,40991	0,32704	-1,18751	0,10234	-0,24759	-0,20101
226	0	0	0	0	-0,13176	-0,31635	-0,90758	1,29912	0,04753	0,91477	1,60478	-0,24759	1,18122
228	0	0	0	0	-0,26272	-0,31635	-0,4401	1,24175	0,05359	0,49431	-0,13669	-0,24759	0,9473
229	0	0	0	0	-0,50083	-0,31635	1,42283	0,93196	0,97789	-0,28191	0,46087	-0,24759	0,58579
231	1	1	1	1	1,0707	2,92624	-0,40521	1,05243	0,03845	1,65865	1,75844	4,02661	-0,3605
232	0	0	0	0	0,05873	-0,31635	-0,16799	1,00654	0,04148	0,04151	0,75112	-0,24759	0,57516
233	1	1	1	1	1,285	2,92624	0,1111	1,56875	0,04249	1,27054	3,03893	4,02661	0,10733
234	0	1	1	1	-0,54845	-0,31635	1,92519	1,27043	0,97991	-0,18489	0,956	-0,24759	0,30935
235	1	1	1	1	0,57067	-0,31635	-0,69128	-1,03578	0,03946	-0,76705	0,35843	-0,24759	0,4901
236	0	0	0	0	-0,71513	-0,31635	-0,82385	0,97785	0,04148	1,01179	1,2121	-0,24759	0,40504
237	1	0	0	0	-0,48893	-0,31635	0,73906	1,17291	0,32704	-0,18489	0,75112	-0,24759	-0,37113
238	0	0	0	0	-0,6437	-0,31635	-0,37033	0,60496	0,03845	0,39728	0,49502	-0,24759	0,54326
239	0	0	1	0	0,60638	-0,31635	-0,95642	-0,45636	0,04148	1,27054	1,70722	-0,24759	0,81971

CLASS_ID	Exp	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	X0Av	F02[C-P]	ATSC1e	MATS3v	SpMax1_Bh(s)	SpMAD_X	RBF	B05[C-P]	MATS1s
240	0	0	0	0	0,36827	-0,31635	0,07622	1,53433	0,03845	0,97945	2,57795	-0,24759	0,75592
242	1	1	1	0	-0,08414	-0,31635	-1,12387	-0,92678	0,05459	-0,67003	-0,35864	-0,24759	0,92604
243	1	1	1	1	-0,227	-0,31635	-0,91455	-1,93073	0,05359	-0,76705	0,1877	-0,24759	0,79845
244	0	0	0	0	-0,53655	-0,31635	0,28554	0,16322	0,04753	0,36494	0,61453	-0,24759	0,60706
245	1	0	0	0	-0,82228	-0,31635	2,69967	-1,58078	1,12925	-0,96111	-0,0001	-0,24759	0,75592
246	0	0	0	0	-0,66751	-0,31635	0,6344	-0,03757	0,81745	0,78539	0,23892	-0,24759	0,17113
247	1	1	1	1	-0,11985	-0,31635	-0,9006	-1,65536	0,03845	-0,50831	0,22185	-0,24759	0,71339
249	1	1	1	1	-0,07223	-0,31635	-0,94246	-1,04152	0,03845	-0,41128	0,46087	-0,24759	0,54326
250	1	0	0	0	0,48733	-0,31635	-0,20287	-0,6342	0,04249	1,36756	0,59746	-0,24759	0,58579
251	1	1	1	1	1,10642	-0,31635	0,27158	-1,91926	0,04249	1,56162	1,57063	-0,24759	0,85161
253	1	1	1	1	1,21357	-0,31635	-0,69826	-1,75863	0,04148	0,26791	0,71697	-0,24759	0,85161
254	1	1	1	1	1,04689	-0,31635	-0,5936	-2,26347	0,04148	0,91477	0,6999	-0,24759	0,77718
255	1	0	1	0	0,6421	-0,31635	-1,0541	0,27796	0,04148	0,46197	-0,3245	-0,24759	0,19239
256	0	0	0	0	0,15397	-0,31635	-0,94246	1,19586	0,03845	0,85008	1,79259	-0,24759	1,51083
257	1	1	1	1	0,08254	-0,31635	-1,04014	-0,49078	0,04148	0,07386	-0,37572	-0,24759	0,23492
258	1	1	1	1	-0,31034	-0,31635	-0,88664	-0,88662	0,04148	0,62368	0,17063	-0,24759	0,25619
259	0	0	0	0	-0,82228	-0,31635	-0,50289	-0,7432	0,04349	0,1062	0,15355	-0,24759	0,17113
261	1	0	0	0	-0,48893	-0,31635	-0,4401	0,47875	0,04148	-0,41128	-0,88791	-0,24759	-0,31797
262	1	0	0	0	-0,82228	-0,31635	-0,97037	0,70249	0,04148	1,17351	-0,7684	-0,24759	0,25619
263	1	0	0	0	0,32065	-0,31635	-0,94944	-1,46031	0,03946	-0,41128	-0,56352	-0,24759	0,20302
264	1	0	0	0	0,33256	-0,31635	-0,79594	0,00833	0,04148	-0,41128	0,85356	-0,24759	0,51137
266	1	0	0	0	-0,26272	-0,31635	-0,88664	-0,61126	0,04148	-0,41128	-0,97328	-0,24759	0,19239
267	0	0	0	0	0,51114	-0,31635	-0,71221	0,66806	0,0334	1,36756	0,15355	-0,24759	0,50073
268	1	0	0	0	-0,08414	-0,31635	0,96233	0,86885	1,12925	-0,573	0,46087	-0,24759	0,39441
269	0	0	0	0	0,05873	-0,31635	-0,41917	0,5017	0,0334	1,36756	0,1877	-0,24759	0,4582
270	0	0	0	0	-0,56036	-0,31635	1,37399	0,63938	1,12925	-1,51094	0,23892	-0,24759	0,54326
271	0	0	0	0	-0,17938	-0,31635	-0,23776	0,42138	0,32502	-1,15517	-0,56352	-0,24759	0,58579
272	0	0	0	0	-0,75085	-0,31635	0,45997	0,35827	0,42492	-1,09048	-0,18791	-0,24759	0,90477
273	0	0	0	0	-0,98896	-0,31635	0,84372	0,3468	1,12925	-0,18489	0,35843	-0,24759	0,46884
274	0	0	0	0	0,05873	-0,31635	0,7251	0,35254	0,97688	-0,3466	-0,05132	-0,24759	0,50073

275	0	0	0	0	-0,40559	-0,31635	0,30647	0,63938	0,32502	-0,63768	-0,47815	-0,24759	-0,12659
276	0	0	0	0	-0,88181	-0,31635	0,90651	1,12701	1,12925	-1,41391	-0,23913	-0,24759	0,38378
277	0	0	0	0	-0,84609	-0,31635	1,67401	0,1288	0,98294	0,36494	0,46087	-0,24759	-0,04153
278	0	0	0	0	0,32065	-0,31635	-0,74012	0,71396	0,0334	1,30288	-0,3245	-0,24759	0,87287
279	0	0	0	0	0,18969	-0,31635	-0,36335	0,74264	0,03542	0,42963	0,52917	-0,24759	0,86224
280	1	0	0	0	0,8564	-0,31635	-0,75408	-1,10463	0,03441	-1,28454	-0,3245	-0,24759	0,23492
282	0	0	0	0	-0,72704	-0,31635	-0,51685	0,13454	0,03441	0,91477	-1,31474	-0,24759	-0,54125
283	0	0	0	0	-0,0008	-0,31635	-0,81687	0,85164	0,03946	0,52665	-0,13669	-0,24759	-0,04153
284	0	0	0	0	0,17778	-0,31635	-0,60755	0,35254	0,0334	0,46197	0,32429	-0,24759	1,10679
285	0	0	0	0	0,36827	-0,31635	-0,48196	0,62791	0,0334	0,97945	0,01697	-0,24759	1,26628
286	1	0	0	0	0,68972	-0,31635	-0,43312	0,59922	0,0334	0,559	0,97307	-0,24759	0,9792
287	0	0	0	0	0,6302	-0,31635	-0,62151	0,28369	0,0334	0,81774	1,10966	-0,24759	1,04299
288	0	0	0	0	-0,92943	-0,31635	-0,51685	0,07143	0,03441	0,91477	-1,38304	-0,24759	-0,09469
289	0	0	0	0	-1,09611	-0,31635	-0,20287	0,63364	0,81442	0,52665	-1,26352	-0,24759	1,05363
290	0	0	0	0	-0,54845	-0,31635	2,17637	0,34106	1,12925	-1,28454	-0,3245	-0,24759	-0,10532
291	0	0	0	0	-0,83419	-0,31635	-0,44708	-2,21184	0,03744	-0,05552	-2,03182	-0,24759	0,37314
292	0	0	0	0	-1,54852	-0,31635	-0,40521	-0,77762	0,04652	1,01179	-2,03182	-0,24759	0,10733
293	0	0	0	0	0,82068	-0,31635	-0,14705	0,73117	0,05157	-0,83174	-1,34889	-0,24759	0,9473
294	1	1	1	1	1,74932	4,0071	-0,51685	1,07538	0,03845	0,59134	0,97307	-0,24759	-0,13722
295	0	0	0	0	-0,17938	-0,31635	1,01117	0,30664	0,97789	-1,60796	-0,52937	-0,24759	-0,08406
296	0	0	0	0	0,08254	-0,31635	-0,65639	-1,63815	0,0334	-0,28191	-0,92206	-0,24759	0,70275
297	0	1	0	0	-0,03652	-0,31635	-1,04712	-0,01462	0,05258	0,20323	-1,16108	-0,24759	1,06426
298	0	0	0	0	0,53495	-0,31635	-0,73314	-0,73173	0,05157	-1,76968	-1,40011	-0,24759	0,31998
299	0	1	1	1	1,0588	-0,31635	-0,80989	1,20159	0,05157	0,62368	-0,7684	-0,24759	0,59643
300	0	0	0	0	0,46352	-0,31635	-0,43312	-0,6801	0,04047	-1,51094	-0,35864	-0,24759	0,08607
301	0	0	0	0	0,01111	-0,31635	-0,27962	-1,6611	0,04047	-1,31688	-0,44401	-0,24759	-0,27544
302	0	0	0	0	-0,20319	-0,31635	-0,16101	-0,36457	0,0556	-0,67003	-0,81962	-0,24759	-0,00963
303	0	0	0	0	0,18969	-0,31635	-0,93548	-0,81778	0,04249	-1,54328	-0,58059	-0,24759	0,60706
304	0	0	0	0	-0,81038	-0,31635	0,29251	-0,7432	0,97688	-0,05552	0,17063	-0,24759	0,56453
305	0	1	0	0	0,88021	-0,31635	-1,13085	1,28191	0,03441	1,01179	0,27307	-0,24759	0,78781

CLASS_ID	Exp	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	X0Av	F02[C-P]	ATSC1e	MATS3v	SpMax1_Bh(s)	SpMAD_X	RBF	B05[C-P]	MATS1s
306	1	1	1	1	0,61829	2,92624	-0,0424	-0,23836	0,05258	1,20585	0,63161	4,02661	0,0329
307	1	1	1	1	0,61829	2,92624	0,13901	-0,21541	0,03441	1,27054	0,63161	4,02661	-0,04153
308	0	0	0	0	-0,17938	-0,31635	-0,41917	0,86312	0,32704	-1,28454	-1,45133	-0,24759	-0,3605
309	0	0	0	0	0,11826	-0,31635	-0,46801	0,47875	0,0334	0,00917	-0,39279	-0,24759	0,68149
310	0	0	0	0	-1,14373	-0,31635	0,01342	0,21485	1,12925	-0,02317	-0,75133	-0,24759	0,1286
311	0	0	0	0	0,13016	-0,31635	-0,67035	-1,72421	0,05157	-1,38157	-1,48548	-0,24759	0,80908
312	0	0	0	0	-0,47702	-0,31635	0,61347	-0,15231	1,12925	-0,41128	-1,4684	-0,24759	0,58579
313	0	0	0	0	0,57067	-0,31635	-0,90758	-0,15804	0,03441	-1,15517	-0,88791	-0,24759	0,11796
314	0	0	0	0	-0,91753	-0,31635	-0,49592	-1,12184	0,03643	0,97945	-0,81962	-0,24759	0,86224
315	1	1	1	1	0,97546	2,92624	0,06924	1,59744	0,03946	0,00917	1,19502	4,02661	-0,40303
316	1	0	0	0	0,07063	-0,31635	-0,87269	-0,49652	0,03239	-0,50831	-2,03182	-0,24759	0,31998
317	0	0	0	0	-0,83419	-0,31635	0,25065	-0,15231	0,81442	-0,05552	-1,07572	-0,24759	0,522
318	1	0	0	0	-0,09604	-0,31635	-0,6843	0,09438	0,03845	-0,08786	-1,84401	-0,24759	-0,12659
319	1	0	0	0	0,26112	-0,31635	-0,56569	0,22059	0,03946	0,23557	-0,83669	-0,24759	0,9473
322	1	1	1	1	1,60645	2,92624	0,03435	-0,66862	0,06771	2,46721	1,38283	4,02661	0,20302
323	0	1	1	1	1,01117	-0,31635	0,81581	-0,04331	0,06267	-0,83174	-0,13669	-0,24759	0,79845
324	0	0	0	0	-1,2866	-0,31635	-0,75408	-0,51373	0,05964	2,20847	-0,47815	-0,24759	0,87287
325	1	1	1	1	2,67795	1,84538	-0,24474	-2,28068	0,00212	1,56162	0,40965	-0,24759	-2,78472
326	1	1	1	1	1,96362	2,92624	1,31119	-0,87515	-0,0009	0,13854	0,66575	-0,24759	-2,08297
327	1	1	1	1	3,76136	2,92624	1,39492	-1,11036	0,00313	-0,44363	1,57063	-0,24759	-2,1893
328	1	1	1	1	1,54692	2,92624	1,10885	-2,80847	0,00313	1,43225	1,98039	-0,24759	-1,95538
329	1	1	1	1	2,77319	2,92624	-0,53778	1,06964	-0,00696	0,81774	1,656	-0,24759	-1,29617
331	1	1	1	1	1,20166	2,92624	-0,37033	0,7828	0,0001	-0,47597	0,8877	4,02661	-1,74273
332	1	1	1	1	1,96362	2,92624	-0,16799	1,48843	-0,00191	0,559	1,58771	4,02661	-0,99845
333	1	1	1	1	2,90416	2,92624	-0,61453	1,28191	-0,012	1,46459	2,73161	4,02661	-0,72201
334	0	0	0	0	-0,75085	-0,31635	2,37871	-1,11036	1,12925	-2,12545	-0,3245	-0,24759	0,1286
335	0	1	0	0	1,08261	-0,31635	-0,0703	-1,09315	0,80938	-1,76968	0,10234	-0,24759	-0,59442
336	1	0	0	0	0,93974	-0,31635	-0,02146	-0,63994	0,80938	-1,41391	0,32429	-0,24759	-0,59442
337	0	0	0	0	-1,39375	-0,31635	3,53694	0,31238	1,13329	-1,60796	-0,27328	-0,24759	-0,72201
338	0	0	0	0	-0,09604	-0,31635	0,96233	-0,26704	0,34722	-0,83174	1,10966	-0,24759	-1,35996

339	0	0	0	0	-0,59608	-0,31635	0,96931	0,75985	0,32603	-0,1202	0,40965	-0,24759	-1,29617
340	1	1	1	1	1,40406	2,92624	-0,09821	0,84017	0,32603	-0,96111	1,60478	-0,24759	-1,3387
341	0	0	0	0	-1,16754	-0,31635	1,35306	0,11159	0,82451	-0,21723	-0,01718	-0,24759	0,04354
343	0	0	0	0	-0,06033	-0,31635	0,62742	0,09438	0,32603	0,30025	0,05112	-0,24759	-0,79644
344	0	0	0	0	-1,11992	-0,31635	3,01364	0,80001	1,12925	-0,67003	0,46087	-0,24759	-0,19038
345	0	0	0	0	-0,94134	-0,31635	2,06474	0,40417	1,12925	-0,37894	0,29014	-0,24759	-0,49872
346	0	0	0	0	-0,20319	-0,31635	-0,14705	-0,9612	0,3775	-1,12282	-0,0684	-0,24759	-1,53008
347	0	0	0	0	-0,59608	-0,31635	0,92744	1,13275	0,3331	0,26791	1,31453	-0,24759	-1,32806
348	0	0	0	0	-0,03652	-0,31635	1,03908	0,5017	1,01018	-0,50831	1,2121	-0,24759	0,68149
349	0	0	0	0	0,32065	-0,31635	0,43206	-0,18099	0,81442	-1,15517	-0,35864	-0,24759	-0,31797
350	1	1	1	1	2,20173	2,92624	-0,05635	-0,10641	0,00313	2,24082	2,39015	4,02661	-0,54125
351	1	0	0	0	0,2016	-0,31635	-0,53778	-0,5711	0,01322	0,04151	0,30721	-0,24759	0,01164
352	1	1	1	1	1,52311	-0,31635	1,08792	1,0008	0,01322	-0,50831	-2,03182	-0,24759	0,57516
353	0	1	0	0	1,33262	-0,31635	-0,75408	-6,41122	0,42492	-3,06339	-2,03182	-0,24759	-0,89213
354	1	1	1	0	-0,63179	-0,31635	0,18088	-0,65715	0,43098	0,78539	1,02429	-0,24759	1,4045
355	1	0	0	0	-0,07223	-0,31635	0,20879	-0,32441	0,43098	0,49431	1,67307	-0,24759	0,80908
356	1	0	0	0	-0,73894	-0,31635	0,20181	-0,50225	0,43098	-0,05552	-0,20498	-0,24759	1,30881
357	0	0	0	0	-0,25082	-0,31635	0,01342	-1,25378	0,42492	-0,21723	-0,97328	-0,24759	-0,10532
358	0	0	0	0	-0,96515	-0,31635	1,45772	1,56875	1,12925	-0,89643	0,8877	-0,24759	0,63896
359	1	0	0	0	-0,858	-0,31635	0,62044	0,69101	0,42896	0,04151	0,32429	-0,24759	0,9154
361	0	0	0	0	-0,10795	-0,31635	0,43206	-0,42194	0,42896	-1,09048	0,40965	-0,24759	1,17058
362	1	0	0	0	-0,4294	-0,31635	0,89953	1,30486	0,42997	-0,70237	1,29746	-0,24759	0,47947
363	1	0	0	0	-0,8699	-0,31635	1,65308	1,18438	1,12925	-0,70237	0,81941	-0,24759	0,96857
365	0	0	0	0	-1,15564	-0,31635	0,23669	-0,65715	0,42492	0,39728	-0,23913	-0,24759	0,05417
369	0	0	0	0	-1,47709	-0,31635	-0,05635	-0,47357	0,42492	-0,02317	-0,88791	-0,24759	0,02227
370	0	0	1	0	-1,40565	-0,31635	3,70439	-1,37999	1,13329	-1,93139	-0,01718	-0,24759	0,54326
371	1	1	1	1	-0,06033	4,0071	0,24367	1,20159	0,42492	0,3326	0,81941	4,02661	0,29872
372	0	0	0	0	-1,00086	-0,31635	1,57633	1,05817	1,12925	-0,73471	0,34136	-0,24759	0,35188
373	0	0	0	0	-0,89371	-0,31635	0,1111	0,25501	0,42492	0,3326	-0,20498	-0,24759	0,14986
374	0	0	0	0	-0,97705	-0,31635	-0,11915	-0,61126	0,06267	0,20323	-1,34889	-0,24759	0,69212

CLASS_ID	Exp	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	X0Av	F02[C-P]	ATSC1e	MATS3v	SpMax1_Bh(s)	SpMAD_X	RBF	B05[C-P]	MATS1s
375	1	0	0	0	0,26112	-0,31635	-0,64244	-1,51194	-0,95346	-0,83174	-0,27328	-0,24759	-0,53062
376	0	1	1	1	-0,04842	-0,31635	-0,61453	-1,68978	-0,95346	-0,50831	0,15355	-0,24759	-0,55189
377	0	0	0	0	-0,45321	-0,31635	-0,72617	-1,12184	-0,95346	0,00917	0,01697	-0,24759	-0,45619
378	0	0	0	0	0,22541	-0,31635	-0,46103	-0,22689	-0,95649	-0,08786	0,22185	-0,24759	-0,13722
382	0	0	0	0	-0,4413	-0,31635	-0,73314	-0,14657	-0,95447	0,52665	-0,61474	-0,24759	-0,79644
383	0	0	0	0	-0,09604	-0,31635	-0,78198	-1,21936	-0,95649	-0,41128	-0,87084	-0,24759	-0,04153
386	1	1	1	1	1,51121	-0,31635	-0,78896	0,42712	-0,9464	2,79064	2,62917	-0,24759	0,23492
387	1	1	1	1	1,59455	-0,31635	0,54369	0,40417	-0,93227	1,30288	-2,03182	-0,24759	-1,83843
388	1	0	0	0	0,72544	-0,31635	0,06924	-0,91531	-0,92622	-0,28191	-2,03182	-0,24759	-2,56144
389	1	1	1	1	0,91593	-0,31635	-0,64942	-0,42194	-2,25718	-0,7994	-0,85377	-0,24759	-0,4243
390	0	0	0	0	-0,61989	-0,31635	-0,95642	0,58775	-2,52256	1,01179	-2,03182	-0,24759	2,23384
391	1	1	1	1	0,86831	2,92624	0,01342	1,67201	-2,26222	1,39991	1,38283	4,02661	-1,91285
394	1	1	1	1	1,16595	2,92624	-0,46801	1,36222	-2,28543	-0,54066	0,956	4,02661	-2,48701
395	0	0	0	0	-0,41749	-0,31635	-0,95642	-0,41047	-2,41459	-0,31426	-0,85377	-0,24759	-0,67948
397	0	0	0	0	0,02301	-0,31635	-0,18892	0,68527	-2,15829	-0,28191	-0,47815	-0,24759	-1,20047
398	0	0	0	0	-1,01277	-0,31635	-0,5308	1,73512	-2,07554	-0,7994	-0,54645	-0,24759	-2,48701
401	1	0	0	0	0,28493	-0,31635	-0,98433	0,16322	-2,27938	-0,67003	0,40965	-0,24759	-0,56252
402	1	0	0	0	0,89212	-0,31635	-0,88664	0,41564	-2,34597	-0,67003	0,40965	-0,24759	-1,16857
403	0	1	0	0	0,76116	-0,31635	-0,94246	1,08112	-2,34597	0,26791	0,63161	-0,24759	-0,7858
404	0	0	0	0	-1,16754	-0,31635	-0,95642	0,08864	-2,24305	0,52665	-1,31474	-0,24759	-0,22228
405	0	0	0	0	-1,40565	-0,31635	-0,58662	-1,816	-2,10783	1,14116	-2,03182	-0,24759	-4,48593
406	1	0	0	0	-0,04842	-0,31635	-0,9006	-0,12362	-2,38432	0,62368	0,97307	-0,24759	0,1286
407	1	1	1	1	0,29684	2,92624	-0,00053	1,70644	-2,21782	0,559	1,19502	4,02661	-2,16803
408	1	0	1	0	-0,25082	-0,31635	-0,87269	-0,82352	-2,30763	-0,21723	-0,22206	-0,24759	0,4901
410	0	0	0	0	-0,39368	-0,31635	-0,24474	0,22059	-2,31268	0,04151	-0,22206	-0,24759	0,08607
411	0	0	0	0	-0,02461	-0,31635	-0,50987	0,66233	-2,31368	0,07386	0,05112	-0,24759	0,71339
412	0	0	0	0	0,04682	-0,31635	-0,55174	1,17864	-2,13407	1,36756	0,93892	-0,24759	1,29818
413	0	0	0	0	-0,23891	-0,31635	-0,89362	-0,20968	-2,22892	-0,28191	-2,03182	-0,24759	2,32953
14	0				-0,51274	-0,31635	-0,62849	0,05996	0,97688	-0,573	-1,04157	-0,24759	0,39441
15	0				-0,82228	-0,31635	-0,48196	0,88033	0,97688	-0,21723	-0,54645	-0,24759	0,47947

17	0				-1,36994	-0,31635	1,32515	-0,03183	0,97789	-0,15254	-0,78547	-0,24759	0,23492
28	1				-0,69132	-0,31635	0,50183	-1,49473	0,86588	0,30025	0,51209	-0,24759	1,05363
35	1				0,45161	-0,31635	-0,93548	-0,74894	0,03845	-0,18489	0,40965	-0,24759	1,00046
36	1				-0,75085	-0,31635	0,16692	-0,82926	1,12925	-0,573	0,27307	-0,24759	0,47947
37	1				-0,36987	-0,31635	-0,40521	-0,81205	0,81442	-0,15254	0,40965	-0,24759	0,92604
38	1				-0,25082	-0,31635	-0,80989	-0,7891	0,03845	-0,18489	0,40965	-0,24759	0,9473
39	1				-1,21516	-0,31635	2,1694	-0,56536	1,14237	-0,92877	0,76819	-0,24759	0,39441
56	1				3,05893	2,92624	-0,57267	1,7007	-2,46605	1,46459	2,76576	4,02661	-2,6146
57	1				3,22561	2,92624	-0,53778	-0,33589	-2,46605	-1,15517	1,87795	4,02661	-3,41204
77	0				0,4278	-0,31635	-1,10992	0,09438	0,0334	1,36756	-0,08547	-0,24759	0,68149
79	0				0,58257	-0,31635	-1,02619	0,70249	0,0334	1,46459	0,05112	-0,24759	0,63896
81	0				0,45161	-0,31635	-1,08898	0,83443	0,0334	1,88505	0,22185	-0,24759	0,5539
88	0				-1,04849	-0,31635	2,02985	0,11733	0,82754	-0,15254	-0,92206	-0,24759	0,63896
89	0				-1,32231	-0,31635	4,92541	0,33533	1,13631	-1,38157	0,05112	-0,24759	0,27745
92	0				-1,56043	-0,31635	5,06496	0,15748	1,13329	-1,51094	0,40965	-0,24759	0,10733
93	0				-1,0723	-0,31635	0,34833	-0,14083	0,82451	-0,15254	-0,92206	-0,24759	0,46884
94	0				-1,38184	-0,31635	2,86014	-0,05478	1,12925	-0,83174	-0,54645	-0,24759	0,34125
95	0				-1,20326	-0,31635	1,27631	0,10585	1,13228	-0,86408	-0,56352	-0,24759	-0,07342
96	0				0,34446	-0,31635	-0,86571	-1,50621	0,03542	-0,89643	-0,7684	-0,24759	-0,05216
101	0				-0,29844	-0,31635	-0,69128	-0,15804	0,03542	1,33522	0,15355	-0,24759	0,86224
112	0				0,01111	-0,31635	-0,69128	-0,56536	0,03946	-0,7994	-0,90499	-0,24759	0,44757
113	0				-0,15557	-0,31635	-0,8378	-0,92105	0,04047	-0,24957	-0,44401	-0,24759	0,33061
120	0				0,43971	-0,31635	-0,75408	0,96638	0,03542	0,94711	-0,47815	-0,24759	0,41567
122	0				0,29684	-0,31635	-0,12612	0,13454	0,03643	1,01179	-0,66596	-0,24759	0,23492
125	0				0,14207	-0,31635	-0,89362	-1,50047	0,03643	0,26791	-1,31474	-0,24759	0,20302
128	0				-0,90562	-0,31635	0,18785	-1,48899	1,12925	-1,31688	-1,40011	-0,24759	0,001
129	0				-0,40559	-0,31635	-0,38428	0,25501	0,32401	-0,67003	-0,97328	-0,24759	-0,13722
132	0				-0,25082	-0,31635	-0,82385	-0,04904	0,03744	0,68837	-1,12694	-0,24759	0,93667
137	0				-1,03658	-0,31635	1,89728	0,3468	1,12925	-0,63768	0,10234	-0,24759	-0,46683
139	0				-0,83419	-0,31635	1,2484	0,49596	1,12925	-1,09048	0,23892	-0,24759	-0,17975

CLASS_ID	Exp	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	X0Av	F02[C-P]	ATSC1e	MATS3v	SpMax1_Bh(s)	SpMAD_X	RBF	B05[C-P]	MATS1s
142	1				-0,29844	-0,31635	-0,8378	-0,86368	0,03643	0,20323	-0,61474	-0,24759	0,11796
147	0				-0,04842	-0,31635	-0,5308	-0,3072	0,08083	-0,21723	0,01697	-0,24759	-0,3605
154	0				-0,27463	-0,31635	-0,39126	-0,98415	0,08083	-0,50831	0,29014	-0,24759	-0,34987
158	0				0,11826	-0,31635	0,77394	-0,00315	0,08083	-0,24957	0,81941	-0,24759	-0,97719
160	0				-0,33415	-0,31635	-0,10519	0,32385	0,81442	-0,60534	-0,0001	-0,24759	-0,15848
163	0				-0,40559	-0,31635	-0,12612	-0,5711	0,08184	-0,21723	-1,2123	-0,24759	-0,49872
168	0				-0,28653	-0,31635	0,1739	-0,27852	0,08184	-0,44363	-0,97328	-0,24759	-2,38068
171	0				-1,36994	-0,31635	2,38569	-0,38752	1,12925	-1,12282	0,10234	-0,24759	0,02227
172	0				-0,36987	-0,31635	-0,62849	-0,55389	0,08184	-0,86408	-0,68303	-0,24759	0,01164
173	0				-0,31034	-0,31635	-0,71919	0,21485	0,08184	-0,21723	-0,35864	-0,24759	0,10733
186	1				0,05873	-0,31635	-0,80989	0,11733	0,04148	-0,99345	-0,0684	-0,24759	0,05417
189	0				0,16588	-0,31635	-0,18194	1,20733	0,81644	0,97945	1,84381	-0,24759	0,53263
190	1				0,07063	-0,31635	-1,12387	-0,84647	0,04249	0,04151	0,4438	-0,24759	0,74528
191	1				-0,08414	-0,31635	-1,13085	-0,87515	0,04249	0,23557	0,71697	-0,24759	0,62832
192	1				0,09445	-0,31635	-1,07503	-0,60552	0,04249	0,23557	0,61453	-0,24759	0,76655
197	0				-0,227	-0,31635	-1,04014	-0,01462	0,04148	0,78539	0,10234	-0,24759	0,63896
198	0				-0,13176	-0,31635	-0,30055	0,27796	0,03845	0,59134	0,73404	-0,24759	0,1286
210	0				-0,50083	-0,31635	0,54369	0,1288	0,32704	0,36494	0,23892	-0,24759	-0,3605
212	0				-1,2985	-0,31635	3,20203	0,29517	0,98294	-0,83174	0,68283	-0,24759	-0,27544
213	0				-0,52464	-0,31635	0,15297	0,0829	0,32704	-0,54066	-0,0001	-0,24759	-0,3605
217	0				-0,57226	-0,31635	0,73906	-0,31294	0,42492	0,30025	0,63161	-0,24759	0,71339
221	1				0,24922	-0,31635	-0,91455	-1,1792	0,03946	-0,73471	0,05112	-0,24759	0,50073
222	1				0,36827	-0,31635	-0,87269	-1,05873	0,03946	-0,76705	-0,0684	-0,24759	0,60706
223	1				0,17778	-0,31635	-0,87967	-1,08168	0,03946	-0,54066	0,22185	-0,24759	0,50073
224	1				0,39208	-0,31635	-0,8378	-0,81778	0,03946	-0,54066	0,13648	-0,24759	0,61769
227	0				0,61829	-0,31635	-1,06805	0,79428	0,03845	1,01179	1,67307	-0,24759	0,77718
230	0				-0,56036	-0,31635	-0,45405	1,05243	0,03845	0,39728	0,63161	-0,24759	0,11796
241	0				-0,48893	-0,31635	-0,03542	0,40991	0,03845	0,17088	1,07551	-0,24759	0,51137
248	1				-0,19129	-0,31635	-0,90758	-1,28247	0,03845	-0,41128	0,1877	-0,24759	0,4901

CLASS_ID	Exp	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	X0Av	F02[C-P]	ATSC1e	MATS3v	SpMax1_Bh(s)	SpMAD_X	RBF	B05[C-P]	MATS1s
252	1				0,86831	-0,31635	-0,45405	-0,72026	0,04148	2,7583	0,66575	-0,24759	0,70275
260	0				-0,94134	-0,31635	-0,50289	0,55906	0,04349	0,26791	0,59746	-0,24759	0,19239
265	1				0,57067	-0,31635	-0,69128	0,06569	0,03946	0,30025	0,85356	-0,24759	0,73465
281	0				-0,92943	-0,31635	0,73906	0,47875	1,12925	-0,15254	-0,88791	-0,24759	0,08607
330	1				2,74938	2,92624	-0,57964	1,63759	-0,00696	1,98207	2,37308	-0,24759	-1,03035
342	0				-0,97705	-0,31635	0,4181	0,20338	0,82451	-0,60534	-0,47815	-0,24759	-0,08406
360	0				-0,25082	-0,31635	0,43904	0,84017	0,42896	0,42963	0,59746	-0,24759	1,30881
364	0				-1,47709	-0,31635	-0,05635	-0,94973	0,42492	-0,1202	-0,88791	-0,24759	0,01164
366	1				0,26112	2,92624	0,84372	-0,31868	0,42492	0,85008	1,24624	4,02661	-0,08406
367	1				0,45161	2,92624	0,75999	-0,54815	0,42492	0,3326	0,90478	4,02661	-0,02026
368	1				0,33256	2,92624	0,68324	1,97033	0,42492	0,00917	1,70722	4,02661	0,05417
379	0				0,2016	-0,31635	-0,72617	-1,50621	-0,95649	-0,76705	0,29014	-0,24759	-0,11595
380	0				-0,2151	-0,31635	-0,80989	0,15175	-0,95649	0,26791	0,10234	-0,24759	0,33061
381	0				-0,69132	-0,31635	-0,97037	0,65659	-0,95548	1,27054	-0,18791	-0,24759	0,1286
384	0				0,02301	-0,31635	-0,6843	-0,54815	-0,9575	-0,15254	-0,29035	-0,24759	0,47947
385	0				0,01111	-0,31635	-0,73314	-1,15626	-0,9575	-0,41128	-0,54645	-0,24759	0,28808
392	1				1,64217	2,92624	0,94838	-0,85794	-2,20168	0,65602	0,93892	4,02661	-2,31689
393	1				1,55883	2,92624	0,6344	2,62433	-2,20571	-0,1202	1,50234	4,02661	-1,77463
396	0				-0,56036	-0,31635	-0,92153	0,64512	-2,41762	0,07386	-0,52937	-0,24759	0,95793
399	1				0,58257	-0,31635	-0,91455	0,49596	-2,22892	-1,67265	0,17063	-0,24759	-0,87086
400	0				0,4159	-0,31635	-0,82385	1,30486	-2,22892	-0,73471	0,40965	-0,24759	-0,62632
409	0				0,61829	-0,31635	-0,4401	-0,74894	-2,32377	-0,31426	-0,52937	-0,24759	1,12805

Table S6.6. Decision tree classification prediction and descriptors.

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
1	1	0	0	0	0	-0,84563	-0,31635	-5,0103	1,65815	-3,15001	-3,37102	-0,4866
2	1	0	0	0	0	-0,84563	-0,31635	0,14997	-1,6746	-0,43231	-1,85575	-0,4866
3	0	0	0	0	0	-0,84563	-0,31635	-1,19685	2,13426	-2,36695	-0,56906	-0,4866
4	1	1	1	1	1	-0,84563	-0,31635	3,18408	-2,62682	0,55398	-0,85663	-0,4866
5	0	0	0	0	0	-0,84563	-0,31635	-5,87287	1,18204	-2,03096	-2,61154	-0,4866
6	0	0	0	0	0	-0,84563	-0,31635	-1,19685	3,00713	-2,67584	-1,535	-0,4866
7	0	0	0	0	0	1,17894	-0,31635	0,58125	-0,00823	-0,08006	0,19411	-0,4866
8	0	0	0	0	0	1,17894	-0,31635	0,46775	-1,83331	1,33162	0,88354	-0,4866
9	0	0	0	0	0	1,17894	-0,31635	0,00621	-0,48433	0,48895	0,82455	-0,4866
10	0	0	0	0	0	1,17894	-0,31635	0,61152	-0,56368	0,62172	0,76925	-0,4866
11	0	0	0	0	0	1,17894	-0,31635	-0,31915	-0,16693	1,07963	0,6439	-0,4866
12	1	0	0	0	0	-0,84563	-0,31635	-0,44021	-0,00823	1,13112	0,99414	-0,4866
13	1	0	0	0	0	-0,84563	-0,31635	-0,16782	-0,24628	0,38056	-0,03816	-0,4866
16	0	0	0	0	0	-0,84563	-0,31635	0,57368	0,15048	-0,06923	0,08719	-0,4866
18	1	0	0	0	0	-0,84563	-0,31635	0,89147	-0,40498	0,00122	0,60334	-0,4866
19	0	1	1	1	1	-0,84563	-0,31635	1,84483	-0,48433	0,25592	0,7545	-0,4866
20	0	0	0	0	0	1,17894	-0,31635	0,49802	0,22983	0,35889	0,87616	-0,4866
21	0	0	0	0	0	-0,84563	-0,31635	0,32399	-0,08758	0,35076	0,90566	-0,4866
22	0	0	0	0	0	-0,84563	-0,31635	0,27859	0,38853	-0,03671	0,59597	-0,4866

23	0	0	0	0	0	1,17894	-0,31635	0,1651	0,07113	1,11757	0,84298	-0,4866
24	0	0	0	0	0	-0,84563	-0,31635	0,30886	-0,88109	0,69487	0,68076	-0,4866
25	0	0	0	0	0	-0,84563	-0,31635	-0,03163	0,38853	0,22341	0,86879	-0,4866
26	0	0	0	0	0	-0,84563	-0,31635	0,34669	-0,96044	-0,03671	0,73975	-0,4866
27	0	0	0	0	0	1,17894	-0,31635	0,39209	-0,48433	0,9848	0,88354	-0,4866
29	0	0	0	0	0	-0,84563	-0,31635	-1,46167	1,65815	-1,43214	-0,94142	-0,4866
30	0	0	0	0	0	1,17894	-0,31635	1,05037	0,07113	0,8141	0,93515	-0,4866
31	0	1	1	1	0	-0,84563	-0,31635	-0,50074	0,38853	-0,81165	-0,41421	-0,4866
32	1	0	0	0	0	1,17894	-0,31635	-0,02406	-0,08758	1,1826	0,97571	-0,4866
33	1	0	0	0	0	1,17894	-0,31635	-0,14512	0,15048	0,72468	0,99045	-0,4866
34	1	1	1	1	1	1,17894	-0,31635	1,9205	-1,27785	1,87354	1,03469	-0,4866
40	0	0	0	0	0	-0,84563	-0,31635	0,9974	-1,43655	0,12857	0,49274	-0,4866
41	0	0	0	0	0	-0,84563	-0,31635	-0,02406	0,07113	-1,00132	-0,93036	-0,4866
42	1	1	1	0	1	-0,84563	-0,31635	-1,06822	0,38853	-1,03113	-0,52482	-0,4866
43	1	0	0	0	0	-0,84563	-0,31635	-0,03919	-1,03979	0,06354	0,42637	0,61373
44	0	0	0	0	0	-0,84563	-0,31635	0,33156	-2,15071	-0,66262	-1,60873	-0,4866
45	0	0	0	0	0	-0,84563	-0,31635	-0,8034	-0,96044	-1,78168	-2,07327	-0,4866
46	0	0	0	0	0	-0,84563	-0,31635	0,83851	-0,72239	-0,50818	0,23466	-0,4866
47	0	0	0	0	0	-0,84563	-0,31635	-0,28132	0,22983	0,35618	0,13143	-0,4866
48	1	0	0	0	0	1,17894	-0,31635	1,01253	-1,91266	1,00648	0,72132	-0,4866

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
49	1	0	0	0	0	-0,84563	-0,31635	1,18656	-0,80174	0,57836	0,48905	-0,4866
50	0	0	0	0	0	1,17894	-0,31635	-0,44778	0,78529	0,2776	0,65864	-0,4866
51	0	0	0	0	0	-0,84563	-0,31635	1,0428	-1,27785	0,57565	0,81717	1,00071
52	0	0	0	0	0	-0,84563	-0,31635	-0,06946	0,22983	0,69758	0,90197	-0,4866
53	0	0	0	0	0	-0,84563	-0,31635	-0,7353	0,46788	-0,23451	1,00889	-0,4866
54	0	0	0	0	0	-0,84563	-0,31635	-2,28641	0,22983	-1,12867	-0,51744	-0,4866
55	1	0	0	0	1	-0,84563	-0,31635	-0,50831	-1,1985	-1,31563	-1,17	-0,4866
58	1	1	1	1	1	-0,84563	2,92624	-1,14388	1,10269	0,5892	-0,45477	1,48518
59	1	1	1	1	1	-0,84563	6,16883	-0,8034	0,38853	1,8627	1,72781	3,45669
60	1	1	1	1	1	-0,84563	2,92624	-0,49318	-0,80174	0,6732	-0,56537	2,62743
61	0	1	1	1	1	-0,84563	2,92624	-0,08459	-1,03979	0,53501	0,85404	2,62743
62	1	1	1	1	1	-0,84563	2,92624	-0,41751	-0,32563	0,51875	0,90566	2,62743
63	1	0	0	0	0	1,17894	-0,31635	0,2559	-0,24628	0,16651	0,57753	-0,4866
64	1	0	0	0	0	1,17894	-0,31635	-0,28132	0,86464	0,48353	0,96465	-0,4866
65	0	0	0	0	0	1,17894	-0,31635	-0,42508	1,34075	-0,15322	0,3084	-0,4866
66	0	0	0	0	0	-0,84563	-0,31635	-1,19685	2,21361	-1,49988	-1,41333	-0,4866
67	1	1	1	1	1	-0,84563	-0,31635	3,517	-2,30941	0,61088	0,49274	-0,4866
68	1	0	0	0	0	-0,84563	-0,31635	0,49802	2,21361	-0,16948	0,05401	-0,4866
69	1	1	1	1	1	-0,84563	2,92624	-0,53101	0,78529	0,3074	-0,03816	3,93688

70	0	0	0	0	0	1,17894	-0,31635	0,08944	0,62659	0,58107	0,61809	-0,4866
71	0	0	0	0	0	-0,84563	-0,31635	-0,90933	0,07113	-0,20741	-0,15982	-0,4866
72	0	0	0	0	0	-0,84563	-0,31635	-0,17539	1,4201	-0,31851	-0,12664	-0,4866
73	0	0	0	0	0	-0,84563	-0,31635	-0,07702	0,62659	-0,7683	0,57385	-0,4866
74	1	1	0	0	0	-0,84563	-0,31635	-1,30278	2,45167	-1,04467	-0,41421	-0,4866
75	0	0	0	0	0	-0,84563	-0,31635	-0,03919	0,94399	-0,42689	-0,09715	-0,4866
76	0	0	0	1	0	-0,84563	-0,31635	-1,19685	0,86464	-0,5136	-0,21881	-0,4866
78	0	0	0	0	0	-0,84563	-0,31635	0,0516	0,22983	-0,34831	0,30471	-0,4866
80	0	1	1	0	1	-0,84563	-0,31635	0,02134	1,18204	-0,42147	-0,48058	-0,4866
82	1	1	1	1	1	-0,84563	-0,31635	0,08944	1,4201	-0,55153	-0,76814	-0,4866
83	0	0	0	0	0	-0,84563	-0,31635	-0,15269	-0,24628	0,45372	0,51117	-0,4866
84	0	0	0	0	0	-0,84563	-0,31635	-0,13756	0,15048	-1,42943	-0,52482	-0,4866
85	0	0	0	0	0	-0,84563	0,76451	-1,19685	-0,00823	-1,13409	-1,41702	1,63874
86	0	0	0	0	0	1,17894	-0,31635	0,42992	-0,32563	1,59174	0,89828	-0,4866
87	0	0	0	0	0	1,17894	-0,31635	0,52829	0,15048	0,68404	0,90934	-0,4866
90	0	0	0	0	0	1,17894	-0,31635	0,08187	-0,80174	0,89267	0,87248	-0,4866
91	0	0	0	0	0	-0,84563	-0,31635	0,52072	-0,96044	0,78971	0,89828	-0,4866
97	0	0	0	0	0	1,17894	-0,31635	-0,00136	-0,32563	-0,03129	0,28259	-0,4866
98	0	0	0	0	1	1,17894	-0,31635	0,2332	-0,08758	0,38869	1,08262	-0,4866
99	0	0	0	0	0	-0,84563	-0,31635	0,38452	0,46788	-0,77643	-0,1119	-0,4866

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
100	0	0	0	0	0	-0,84563	-0,31635	0,62665	-0,40498	0,06896	-0,10821	-0,4866
102	0	0	0	0	0	1,17894	-0,31635	-0,20565	0,86464	-0,05839	0,91672	-0,4866
103	1	1	1	1	1	-0,84563	2,92624	-1,24981	0,30918	0,26947	-0,90087	1,48518
104	0	0	0	0	0	-0,84563	-0,31635	-0,03919	0,15048	-0,73307	-0,61699	-0,4866
105	1	0	0	0	0	-0,84563	-0,31635	0,1424	-0,24628	-0,70598	-0,70178	-0,4866
106	0	0	0	0	0	-0,84563	-0,31635	0,56612	0,46788	0,42392	0,56279	-0,4866
107	0	0	0	0	0	1,17894	-0,31635	-0,47048	0,78529	0,64339	0,92778	-0,4866
108	1	0	0	0	0	1,17894	-0,31635	0,24833	0,15048	0,08522	0,79137	-0,4866
109	0	0	0	0	0	-0,84563	-0,31635	-0,11486	-0,96044	0,40224	0,78399	-0,4866
110	0	0	0	0	0	1,17894	-0,31635	0,11213	0,30918	0,20444	1,02363	-0,4866
111	0	0	0	0	0	1,17894	-0,31635	0,20293	0,22983	0,34263	1,0052	-0,4866
114	0	0	0	0	0	1,17894	-0,31635	0,49045	-0,80174	1,02273	-0,09346	1,06427
115	0	0	0	0	0	-0,84563	-0,31635	-0,05433	-0,40498	0,57294	0,81717	-0,4866
116	0	0	0	0	0	1,17894	-0,31635	-0,24349	-0,32563	0,45914	0,82455	-0,4866
117	0	0	0	0	0	1,17894	-0,31635	0,05917	0,30918	-0,20741	0,68814	-0,4866
118	1	1	1	1	1	-0,84563	-0,31635	-0,6672	0,54723	-0,94171	-0,90456	-0,4866
119	1	0	1	0	0	-0,84563	-0,31635	-0,74287	-0,32563	-0,31309	-1,16632	-0,4866
121	0	0	1	0	1	-0,84563	-0,31635	-0,77313	-0,16693	-0,93629	-0,89718	-0,4866
123	0	0	0	0	0	-0,84563	-0,31635	-1,09092	0,22983	-0,62469	-1,17369	-0,4866

124	0	1	1	1	1	-0,84563	-0,31635	-0,87149	1,10269	-1,24247	-0,66491	-0,4866
126	0	0	0	0	0	1,17894	-0,31635	1,16386	-1,3572	2,14721	1,5066	1,1281
127	0	0	0	0	0	1,17894	-0,31635	0,05917	0,15048	-0,01233	0,76187	-0,4866
130	0	0	0	0	0	-0,84563	-0,31635	-0,61424	0,54723	-0,58946	0,38213	-0,4866
131	0	0	0	0	0	-0,84563	-0,31635	-0,26618	0,78529	0,12044	0,81349	-0,4866
133	0	0	0	0	0	-0,84563	-0,31635	-1,13632	0,07113	-0,67888	0,33789	-0,4866
134	0	0	0	0	0	-0,84563	-0,31635	0,44506	-0,16693	0,40224	0,64021	-0,4866
135	0	0	0	0	0	1,17894	-0,31635	-0,10729	-0,24628	1,19073	0,53329	1,00071
136	1	0	0	0	0	1,17894	-0,31635	0,28616	-0,32563	1,10402	0,19779	1,06427
138	0	0	0	0	0	1,17894	-0,31635	0,49045	-0,88109	1,52942	0,83192	2,67898
140	0	0	0	0	0	1,17894	-0,31635	-0,01649	-0,80174	0,77616	0,00608	1,06427
141	0	0	0	0	0	1,17894	-0,31635	0,2105	0,78529	0,1367	0,72132	-0,4866
143	0	0	0	0	0	-0,84563	-0,31635	-0,06946	-0,48433	0,20715	0,6992	-0,4866
144	0	0	0	0	0	-0,84563	-0,31635	-2,98251	0,62659	-3,69735	-5,11118	-0,4866
145	0	0	0	0	0	-0,84563	0,76451	-1,6584	0,15048	-0,93358	-0,6133	1,13264
146	0	0	0	0	0	-0,84563	-0,31635	-2,98251	1,97556	-3,69735	-3,55904	-0,4866
148	0	1	1	1	1	-0,84563	-0,31635	-0,02406	0,70594	-1,23705	-0,64648	-0,4866
149	0	0	0	0	0	-0,84563	-0,31635	0,62665	-0,16693	-0,28057	0,32683	-0,4866
150	0	0	0	0	0	-0,84563	-0,31635	-2,28641	2,05491	-2,04451	-1,30642	-0,4866
151	0	1	1	0	1	1,17894	-0,31635	-0,87906	0,94399	-1,00403	-0,34048	-0,4866

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
152	1	1	0	0	1	1,17894	-0,31635	-0,87906	0,46788	-0,78726	-0,34048	-0,4866
153	0	0	0	0	0	-0,84563	-0,31635	-2,00645	0,07113	-0,97152	-1,15157	-0,1533
155	0	0	1	1	0	-0,84563	-0,31635	-0,85636	-0,24628	-1,09074	-1,26586	-0,4866
156	0	0	0	0	0	1,17894	-0,31635	-0,34941	0,22983	-1,16931	-0,68335	-0,4866
157	0	0	1	1	1	-0,84563	-0,31635	-0,07702	-0,64304	-0,70598	-1,26217	-0,4866
159	0	0	0	0	0	-0,84563	-0,31635	-0,13756	0,07113	-1,41589	-0,558	-0,4866
161	0	0	0	0	0	-0,84563	-0,31635	-0,06946	-0,64304	-1,09074	-2,0364	-0,4866
162	0	0	0	0	0	-0,84563	-0,31635	0,32399	-0,88109	-0,97152	-0,42159	-0,4866
164	1	0	0	0	0	-0,84563	-0,31635	-0,31158	-1,03979	-1,31292	-2,00322	-0,4866
165	0	0	0	0	0	1,17894	-0,31635	0,1878	-0,64304	1,29098	0,20148	1,06427
166	0	0	0	0	0	-0,84563	-0,31635	-1,19685	-0,32563	-1,96593	-2,0917	-0,4866
167	0	0	0	0	0	-0,84563	-0,31635	-0,19052	-0,72239	-1,59743	-2,19124	-0,4866
169	0	0	0	0	0	-0,84563	-0,31635	-0,67477	-1,03979	-1,47008	-2,0364	-0,4866
170	0	0	0	0	0	-0,84563	-0,31635	0,30886	-0,32563	-1,03384	-0,03447	-0,4866
174	0	0	0	0	0	1,17894	-0,31635	0,93687	-0,40498	-0,00962	0,55173	-0,4866
175	0	0	0	0	0	-0,84563	-0,31635	0,77798	-0,08758	-0,53256	0,29365	-0,4866
176	0	0	0	0	0	1,17894	-0,31635	0,17267	0,22983	0,91977	0,09825	-0,4866
177	0	0	0	0	0	1,17894	-0,31635	0,08187	-0,24628	-0,19929	1,0052	-0,4866
178	0	0	0	0	0	1,17894	-0,31635	-1,24981	1,4201	-0,87939	-0,31467	-0,4866

179	0	0	0	0	0	1,17894	-0,31635	0,13483	-0,40498	0,94416	0,81349	-0,4866
180	0	0	0	0	0	1,17894	-0,31635	0,74014	-0,40498	1,00648	0,14987	-0,4866
181	1	0	0	0	0	1,17894	-0,31635	0,18023	0,22983	0,80868	0,88722	-0,4866
182	0	0	0	0	0	1,17894	-0,31635	-0,40995	0,94399	0,4754	0,17936	-0,4866
183	0	0	0	1	0	1,17894	-0,31635	0,24076	0,22983	0,55669	0,21991	-0,4866
184	1	0	0	0	0	1,17894	-0,31635	-0,22079	0,86464	-0,13968	0,3084	-0,4866
185	0	1	1	1	1	1,17894	-0,31635	-0,49318	1,7375	-0,51631	-0,55062	-0,4866
187	1	0	0	0	0	-0,84563	-0,31635	-1,45411	1,34075	-1,5893	-0,55062	-0,4866
188	1	1	1	1	1	-0,84563	2,92624	-0,09972	0,54723	0,79242	0,77662	1,87269
193	0	0	0	0	0	1,17894	-0,31635	-0,15269	0,30918	0,35889	0,57753	-0,4866
194	0	0	0	0	0	1,17894	-0,31635	-0,19809	0,38853	0,64068	0,99414	-0,4866
195	0	0	0	0	0	1,17894	-0,31635	-0,20565	0,78529	-0,02316	0,8393	-0,4866
196	0	1	1	1	0	-0,84563	-0,31635	0,2332	1,02334	-0,42689	-0,36997	-0,4866
199	0	0	0	0	0	1,17894	-0,31635	-0,11486	-0,16693	0,41308	0,56279	-0,4866
200	1	1	1	1	1	-0,84563	2,92624	-1,00012	0,70594	-0,26432	-0,46951	3,93688
201	0	0	0	0	0	1,17894	-0,31635	0,52829	-0,16693	0,89809	0,34895	-0,4866
202	1	1	1	1	1	1,17894	-0,31635	1,56488	-0,48433	1,43188	0,8651	-0,4866
203	0	0	0	0	0	1,17894	-0,31635	-0,11486	0,15048	0,62172	0,91672	-0,4866
204	0	1	1	1	1	-0,84563	-0,31635	-0,62937	0,46788	0,12044	-0,65017	-0,4866
205	0	0	0	0	0	-0,84563	-0,31635	-0,28132	-0,56368	-0,23451	-0,12296	1,06427

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
206	0	0	1	0	0	-0,84563	-0,31635	0,65691	-1,1985	-0,02316	-1,33591	-0,4866
207	0	0	0	0	0	1,17894	-0,31635	0,83851	-0,72239	1,24492	0,19411	1,06427
208	0	0	0	0	0	1,17894	-0,31635	0,33913	-1,03979	1,53755	0,6697	1,06427
209	0	0	0	0	0	1,17894	-0,31635	0,46775	-0,48433	1,49149	0,03558	1,06427
211	0	0	0	0	0	1,17894	-0,31635	0,51315	-0,16693	1,03357	-0,03079	1,06427
214	1	0	0	0	0	1,17894	-0,31635	0,21806	-0,08758	1,09589	-0,02341	1,06427
215	0	0	0	0	0	1,17894	-0,31635	0,73258	-0,96044	1,09589	0,02451	1,06427
216	0	0	0	0	0	-0,84563	-0,31635	1,09576	-0,00823	0,63255	-0,0271	-0,4866
218	0	0	0	0	0	1,17894	-0,31635	1,25466	0,62659	0,65965	0,26416	-0,4866
219	0	0	0	0	0	1,17894	-0,31635	1,34545	0,78529	0,47811	0,66233	-0,4866
220	1	1	1	1	1	1,17894	-0,31635	1,67837	0,30918	-0,21012	0,8946	-0,4866
225	0	0	0	0	0	1,17894	-0,31635	0,28616	-0,40498	1,11486	0,5038	1,06427
226	0	0	0	0	0	1,17894	-0,31635	0,72501	1,7375	0,40766	0,39319	-0,4866
228	0	0	0	0	0	1,17894	-0,31635	0,9747	-0,08758	0,57023	0,78768	-0,4866
229	0	0	0	1	0	1,17894	-0,31635	0,72501	-0,24628	0,6461	-0,23356	-0,4866
231	1	1	1	1	1	-0,84563	2,92624	-0,25862	0,22983	0,91164	0,87248	1,48518
232	0	0	0	0	0	-0,84563	-0,31635	1,0882	0,07113	0,39953	0,2789	-0,4866
233	1	1	1	1	1	-0,84563	2,92624	-0,60667	0,94399	0,92248	-0,29255	1,48518
234	0	0	0	0	0	1,17894	-0,31635	0,65691	-0,24628	0,67591	0,10562	-0,4866

235	1	1	1	0	0	1,17894	-0,31635	0,38452	0,38853	0,05812	-0,35891	-0,4866
236	0	0	0	0	0	1,17894	-0,31635	-0,90933	1,34075	-0,12071	0,58122	-0,4866
237	1	0	0	0	0	1,17894	-0,31635	0,22563	-0,48433	1,09589	0,26784	1,06427
238	0	0	0	0	0	1,17894	-0,31635	-0,09216	0,22983	0,91164	0,78399	-0,4866
239	0	0	0	0	0	-0,84563	-0,31635	-2,5815	3,80064	-1,57575	-0,60593	-0,4866
240	0	0	0	0	0	1,17894	-0,31635	-0,47804	1,02334	0,20173	-0,21881	-0,4866
242	1	1	1	1	1	1,17894	-0,31635	1,65568	0,38853	0,15025	1,00889	-0,4866
243	1	1	1	1	1	1,17894	-0,31635	1,64811	0,38853	-0,0042	0,10931	-0,4866
244	0	0	0	0	0	1,17894	-0,31635	0,1197	0,70594	1,75703	0,6697	-0,4866
245	1	1	1	1	1	-0,84563	-0,31635	1,69351	-0,72239	0,62172	0,98308	-0,4866
246	0	0	0	0	0	1,17894	-0,31635	0,68718	-0,48433	0,84932	0,89828	-0,4866
247	1	1	1	1	1	1,17894	-0,31635	1,64054	0,38853	0,20173	1,08262	-0,4866
249	1	1	1	1	1	1,17894	-0,31635	1,64054	-0,08758	0,55127	1,00151	-0,4866
250	1	1	1	1	1	-0,84563	-0,31635	-1,31791	1,10269	-0,92816	-1,39859	-0,4866
251	1	1	0	0	0	1,17894	-0,31635	-1,22711	0,38853	0,59733	-0,39578	-0,4866
253	1	1	1	1	1	-0,84563	-0,31635	-0,96229	1,34075	-0,97964	-0,97092	-0,4866
254	1	1	1	1	1	-0,84563	-0,31635	-1,62813	1,34075	-1,3075	-1,40965	-0,4866
255	1	1	0	1	1	-0,84563	-0,31635	-0,33428	0,46788	-0,80623	-1,27692	-0,4866
256	0	0	0	0	0	1,17894	-0,31635	-0,12999	1,10269	0,52688	0,3084	-0,4866
257	1	1	0	1	1	-0,84563	-0,31635	-0,50831	1,02334	-1,17202	-1,28798	-0,4866

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
258	1	1	1	1	1	1,17894	-0,31635	-1,00769	1,18204	-1,00132	-0,95248	-0,4866
259	0	0	0	0	0	1,17894	-0,31635	-0,77313	0,54723	-0,14509	0,80243	-0,4866
261	1	1	1	1	1	-0,84563	-0,31635	0,57368	0,30918	-0,83062	-0,78289	-0,4866
262	1	1	1	1	1	-0,84563	-0,31635	-0,12242	0,30918	-1,16119	-0,76446	-0,4866
263	1	1	1	1	1	-0,84563	-0,31635	-0,14512	0,62659	-0,77914	-0,98567	-0,4866
264	1	0	0	0	0	1,17894	-0,31635	0,59638	0,78529	0,82493	0,04295	-0,4866
266	1	1	0	0	1	-0,84563	-0,31635	0,57368	0,38853	-0,89294	-0,54325	-0,4866
267	0	0	0	0	0	1,17894	-0,31635	0,76284	-0,00823	0,63797	0,78031	-0,4866
268	1	0	0	0	0	-0,84563	-0,31635	1,01253	-0,48433	0,67591	-0,34416	-0,4866
269	0	0	0	0	0	1,17894	-0,31635	0,76284	0,15048	0,37514	0,60334	-0,4866
270	0	0	0	0	0	1,17894	-0,31635	0,62665	-0,72239	1,14737	0,72869	1,1281
271	0	0	0	0	0	1,17894	-0,31635	0,30886	-1,03979	0,49437	0,725	1,1281
272	0	0	0	0	0	1,17894	-0,31635	0,46019	-1,03979	0,59462	0,73238	1,1281
273	0	0	0	0	0	1,17894	-0,31635	0,8839	0,38853	0,6163	0,72132	-0,4866
274	0	0	0	0	0	-0,84563	-0,31635	0,30886	-0,48433	0,73552	0,69551	-0,4866
275	0	0	0	0	0	1,17894	-0,31635	0,83851	-0,88109	0,85203	0,71026	1,1281
276	0	0	0	0	0	-0,84563	-0,31635	0,80067	-1,1985	0,79242	1,01626	1,1281
277	0	0	0	0	0	1,17894	-0,31635	0,61152	-0,64304	1,06067	0,68445	1,06427
278	0	0	1	0	1	-0,84563	-0,31635	0,1197	-1,1985	-0,62198	-0,83082	-0,4866

279	0	0	0	0	0	1,17894	-0,31635	-0,37211	0,15048	0,42121	0,40794	-0,4866
280	1	0	0	0	0	1,17894	-0,31635	0,20293	-0,32563	0,42392	0,5591	-0,4866
282	0	0	0	0	0	-0,84563	-0,31635	-0,02406	-0,24628	-0,92545	0,06876	-0,4866
283	0	0	0	0	0	-0,84563	-0,31635	1,10333	0,15048	0,21257	0,14618	-0,4866
284	0	0	0	0	0	1,17894	-0,31635	0,20293	1,02334	1,10131	0,85404	-0,4866
285	0	0	0	0	0	1,17894	-0,31635	0,19536	1,7375	0,15838	0,28259	-0,4866
286	1	0	0	0	0	1,17894	-0,31635	0,03647	1,97556	0,08793	0,0577	-0,4866
287	0	0	0	0	0	1,17894	-0,31635	0,02134	1,49945	0,31553	0,0282	-0,4866
288	0	0	0	0	0	-0,84563	-0,31635	-0,02406	0,15048	-1,11512	-0,38841	-0,4866
289	0	0	0	0	0	1,17894	-0,31635	1,17899	-0,08758	0,4185	0,725	-0,4866
290	0	0	0	0	0	1,17894	-0,31635	0,86121	-1,11914	1,50775	0,83192	1,00071
291	0	0	0	0	0	-0,84563	-0,31635	-1,4768	0,94399	-2,98744	-2,80694	-0,4866
292	0	0	0	0	0	-0,84563	-0,31635	-1,69623	0,78529	-2,749	-3,11663	-0,4866
293	0	0	1	0	0	-0,84563	-0,31635	0,77798	-0,96044	-0,02316	-0,53588	-0,4866
294	1	1	1	1	1	-0,84563	4,0071	0,21806	0,38853	0,62984	-0,14139	2,04041
295	0	0	0	0	0	1,17894	-0,31635	0,73258	-1,5159	0,93061	0,70657	1,06427
296	0	0	0	0	1	-0,84563	-0,31635	0,26346	0,15048	-0,71682	0,28996	-0,4866
297	0	0	0	0	1	-0,84563	-0,31635	0,42992	1,2614	-0,54611	-0,28518	-0,4866
298	0	0	0	0	0	-0,84563	-0,31635	0,39966	-0,32563	-1,12054	-0,40684	-0,4866
299	0	0	0	0	0	-0,84563	-0,31635	0,12727	0,07113	-0,64908	-0,40684	-0,4866

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
300	0	0	0	0	0	-0,84563	-0,31635	0,59638	-0,40498	0,20986	0,20517	-0,4866
301	0	0	0	0	0	1,17894	-0,31635	0,49802	-0,24628	0,18005	0,20517	-0,4866
302	0	0	0	0	0	-0,84563	-0,31635	0,89904	-0,96044	-0,29683	0,14618	-0,4866
303	0	0	0	1	0	1,17894	-0,31635	1,42868	0,22983	0,4185	0,67708	-0,4866
304	0	0	0	0	0	1,17894	-0,31635	-0,37968	0,30918	-0,2589	0,14987	-0,4866
305	0	0	0	0	0	-0,84563	-0,31635	0,48289	0,46788	0,36431	-0,0271	-0,4866
306	1	1	1	1	1	-0,84563	2,92624	0,72501	-0,80174	0,96041	1,16742	1,48518
307	1	1	1	1	1	1,17894	2,92624	0,10457	-0,56368	0,9848	0,99414	1,48518
308	0	0	0	0	0	-0,84563	-0,31635	0,81581	-0,96044	-0,034	0,32314	1,00071
309	0	0	0	0	0	1,17894	-0,31635	0,47532	0,86464	0,04187	0,54067	-0,4866
310	0	0	0	0	0	-0,84563	-0,31635	0,77041	-0,16693	0,15296	0,85036	-0,4866
311	0	1	1	1	1	-0,84563	-0,31635	2,7528	-1,3572	-0,24535	0,81349	-0,4866
312	0	0	0	0	0	-0,84563	-0,31635	0,46775	-0,48433	-0,11529	0,30471	-0,4866
313	0	0	0	0	0	-0,84563	-0,31635	0,38452	0,30918	0,03374	0,28259	-0,4866
314	0	0	0	0	0	-0,84563	-0,31635	-0,76556	0,62659	-0,90649	-0,17088	-0,4866
315	1	1	1	1	1	-0,84563	2,92624	0,49802	-0,40498	1,23679	0,97939	1,48518
316	1	0	0	0	0	-0,84563	-0,31635	0,47532	-1,5159	-0,32935	0,13143	-0,4866
317	0	0	0	0	0	-0,84563	-0,31635	0,9747	-1,11914	0,57294	0,78399	-0,4866
318	1	1	0	1	0	1,17894	-0,31635	1,50435	0,86464	1,50775	0,78399	-0,4866
319	1	0	0	0	0	1,17894	-0,31635	1,28492	1,49945	2,32604	0,67708	-0,4866

322	1	1	1	1	1	-0,84563	2,92624	-1,11362	-0,00823	0,46998	-0,65754	1,48518
323	0	0	0	0	0	1,17894	-0,31635	0,08944	-0,16693	0,69758	-0,3331	-0,4866
324	0	0	0	0	0	-0,84563	-0,31635	-1,83242	0,78529	-3,12292	-2,6779	-0,4866
325	1	1	1	1	1	-0,84563	1,84538	-1,95349	-0,80174	-0,74391	-2,34609	2,04041
326	1	1	1	1	1	-0,84563	2,92624	-0,10729	-1,3572	-0,10445	-1,3949	3,43078
327	1	1	1	1	1	-0,84563	2,92624	-0,58397	-2,38877	0,787	-0,34048	3,93688
328	1	1	1	1	1	-0,84563	2,92624	-1,28764	-0,56368	-0,32393	-0,99304	3,93688
329	1	1	1	1	1	-0,84563	2,92624	-0,72773	1,34075	0,62713	-0,18563	1,65263
331	1	1	1	1	1	-0,84563	2,92624	-0,28888	0,30918	0,48353	0,75081	3,18241
332	1	1	1	1	1	-0,84563	2,92624	-0,31915	-0,48433	1,03899	0,79137	2,79489
333	1	1	1	1	1	-0,84563	2,92624	-1,05309	2,21361	1,10131	-0,24462	0,51037
334	0	0	0	0	0	-0,84563	-0,31635	0,9066	-1,03979	0,96041	0,40425	1,1281
335	0	0	0	0	1	-0,84563	-0,31635	0,26346	-1,43655	0,76532	0,59597	1,00071
336	1	0	0	0	0	-0,84563	-0,31635	0,22563	-1,43655	0,68404	0,48168	1,00071
337	0	0	0	0	0	1,17894	-0,31635	0,42992	-1,03979	1,23679	-0,01973	1,06427
338	0	0	0	0	0	1,17894	-0,31635	-0,04676	-1,11914	1,22866	0,28996	2,55159
339	0	0	0	0	0	1,17894	-0,31635	0,86877	-1,03979	1,17718	0,40425	1,06427
340	1	1	1	1	1	1,17894	2,92624	-0,44021	-0,32563	1,6351	1,65776	3,03605
341	0	0	0	0	0	1,17894	-0,31635	0,84607	-1,3572	0,86558	0,73238	1,06427
343	0	0	0	0	0	1,17894	-0,31635	0,82337	-1,3572	1,19344	0,90566	1,06427

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
344	0	0	0	0	0	1,17894	-0,31635	0,92174	-0,80174	1,51317	0,8098	1,06427
345	0	0	0	0	0	1,17894	-0,31635	0,87634	-0,96044	1,29911	0,73238	1,06427
346	0	0	0	0	0	1,17894	-0,31635	-0,91689	-0,64304	-0,16948	0,53698	-0,19283
347	0	0	0	0	0	1,17894	-0,31635	-0,30402	0,07113	1,08776	0,44112	1,51936
348	0	0	0	0	0	-0,84563	-0,31635	-0,15269	-1,6746	0,31824	0,02451	1,1281
349	0	0	0	0	0	1,17894	-0,31635	1,39085	-2,38877	1,74348	1,30383	2,55159
350	1	1	1	1	1	-0,84563	2,92624	-1,30278	0,30918	0,51062	0,44481	2,79489
351	1	0	0	0	0	1,17894	-0,31635	-0,28132	0,86464	0,62172	0,5038	-0,4866
352	1	1	1	1	1	-0,84563	-0,31635	2,91926	-2,38877	0,98751	-0,62436	-0,4866
353	0	0	0	0	0	-0,84563	-0,31635	-1,6584	-3,10293	-2,00115	-4,41806	-0,4866
354	1	1	1	1	1	-0,84563	-0,31635	-1,42384	1,97556	-1,17473	-0,93405	-0,4866
355	1	0	0	1	0	-0,84563	-0,31635	-1,06065	-0,24628	-0,49463	-0,87506	-0,4866
356	1	0	0	0	0	1,17894	-0,31635	-0,43264	0,15048	-0,49734	-0,19301	-0,4866
357	0	0	0	1	0	-0,84563	-0,31635	-0,38725	-1,5159	-1,23976	-1,86312	-0,4866
358	0	0	0	0	0	-0,84563	-0,31635	0,92174	0,07113	0,1638	-0,28149	-0,4866
359	1	0	0	0	0	-0,84563	-0,31635	0,17267	-0,24628	-0,81165	-0,42159	-0,4866
361	0	0	0	0	0	-0,84563	-0,31635	0,48289	-0,00823	-0,28057	-0,12296	-0,4866
362	1	0	0	0	0	1,17894	-0,31635	0,64178	-0,48433	0,64339	0,27153	1,06427
363	1	0	0	0	0	-0,84563	-0,31635	0,64178	-0,24628	0,20173	-0,36997	-0,4866

365	0	0	0	0	0	-0,84563	-0,31635	-0,65207	-0,00823	-1,68142	-1,86681	-0,4866
369	0	0	0	0	0	-0,84563	-0,31635	-0,96229	-0,24628	-2,21521	-2,07327	-0,4866
370	0	0	0	0	0	1,17894	-0,31635	0,85364	-1,6746	0,94957	0,65127	-0,4866
371	1	1	1	1	1	1,17894	4,0071	0,02134	-0,48433	0,71113	0,83561	2,12107
372	0	0	0	0	0	1,17894	-0,31635	0,00621	-0,48433	0,31553	0,75081	-0,4866
373	0	0	0	0	0	1,17894	-0,31635	-0,03919	-0,56368	-0,47024	0,66233	-0,4866
374	0	0	0	0	0	-0,84563	-0,31635	0,31643	-0,80174	-0,46753	0,21991	-0,4866
375	1	0	0	0	0	-0,84563	-0,31635	0,0743	-0,40498	0,02832	0,39688	-0,4866
376	0	0	0	0	0	-0,84563	-0,31635	-0,01649	0,07113	-0,17761	0,39319	-0,4866
377	0	0	0	0	0	1,17894	-0,31635	0,03647	0,22983	0,22612	0,49642	-0,4866
378	0	0	0	0	0	-0,84563	-0,31635	1,31519	-0,40498	0,24779	0,71763	-0,4866
382	0	0	0	0	0	-0,84563	-0,31635	0,89147	-0,56368	0,12315	0,64758	-0,4866
383	0	0	0	0	0	-0,84563	-0,31635	1,15629	-0,08758	0,05541	0,68814	-0,4866
386	1	0	0	0	0	1,17894	-0,31635	-3,42136	4,91156	-1,41318	-0,28149	-0,4866
387	1	0	0	0	0	-0,84563	-0,31635	0,14997	-2,62682	-0,6572	-2,44195	-0,4866
388	1	0	0	0	0	-0,84563	-0,31635	-0,6899	-1,75396	-1,38879	-2,57098	-0,4866
389	1	0	0	0	0	-0,84563	-0,31635	0,22563	-1,27785	0,17193	0,65127	-0,4866
390	0	0	0	0	0	-0,84563	-0,31635	0,30886	0,15048	-1,19641	0,113	-0,4866
391	1	1	1	1	1	-0,84563	2,92624	-0,34185	0,62659	0,52959	0,24204	2,62743
394	1	1	1	1	1	-0,84563	2,92624	-0,25862	0,54723	0,58107	0,45956	2,62743

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
395	0	0	0	0	0	-0,84563	-0,31635	-0,7353	0,46788	-1,18286	0,39319	-0,4866
397	0	0	0	0	0	-0,84563	-0,31635	-0,55371	-0,96044	-0,38896	-0,20407	-0,4866
398	0	0	0	0	0	-0,84563	-0,31635	-0,46291	0,70594	-1,56762	-0,36628	-0,4866
401	1	1	1	1	0	-0,84563	-0,31635	-0,5991	1,4201	-0,93358	-0,5027	-0,4866
402	1	1	0	1	0	-0,84563	-0,31635	-0,5991	1,02334	-0,60572	-0,5027	-0,4866
403	0	1	1	1	1	-0,84563	-0,31635	-0,65964	1,18204	-0,74391	-0,60593	-0,4866
404	0	0	0	0	0	-0,84563	-0,31635	0,20293	-0,56368	-1,19641	0,31577	-0,4866
405	0	0	0	0	0	-0,84563	-0,31635	-2,11238	0,70594	-3,35594	-3,99777	-0,4866
406	1	0	0	0	0	-0,84563	-0,31635	-0,28132	0,22983	-0,18845	0,30102	-0,4866
407	1	1	1	1	1	1,17894	2,92624	-0,12242	0,15048	0,73281	0,92778	2,62743
408	1	0	0	0	0	-0,84563	-0,31635	-0,15269	0,22983	-0,034	1,16005	-0,4866
410	0	0	0	0	0	1,17894	-0,31635	0,83851	-0,24628	0,787	0,7545	-0,4866
411	0	0	0	0	0	-0,84563	-0,31635	0,85364	0,15048	0,28573	0,09088	-0,4866
412	0	0	0	0	0	-0,84563	-0,31635	0,47532	1,65815	-0,36728	0,16461	-0,4866
413	0	0	0	0	0	-0,84563	-0,31635	0,85364	-1,43655	-0,9119	-0,42159	-0,4866
14	0	1				-0,84563	-0,31635	1,54975	-0,00823	0,42934	0,6697	-0,4866
15	0	0				-0,84563	-0,31635	1,18656	0,07113	0,24237	0,76556	-0,4866
17	0	0				-0,84563	-0,31635	0,30886	-0,56368	0,49437	0,65864	-0,4866
28	1	1				1,17894	-0,31635	1,65568	-0,00823	0,43746	0,98677	-0,4866
35	1	1				1,17894	-0,31635	1,66324	-0,64304	1,1826	1,02732	-0,4866

36	1	1				1,17894	-0,31635	1,70864	-0,32563	0,89538	1,02363	-0,4866
37	1	1				1,17894	-0,31635	1,66324	-0,40498	0,80868	1,01995	-0,4866
38	1	1				1,17894	-0,31635	1,66324	-0,24628	0,703	1,02363	-0,4866
39	1	1				1,17894	-0,31635	1,64811	-0,24628	1,37227	1,02363	-0,4866
56	1	1				-0,84563	2,92624	-1,14388	1,4201	0,6732	-0,22619	1,48518
57	1	1				-0,84563	2,92624	-1,02282	0,46788	0,75719	0,21991	1,48518
77	0	1				-0,84563	-0,31635	-0,07702	1,10269	-0,6843	-0,60224	-0,4866
79	0	0				-0,84563	-0,31635	0,19536	1,2614	-0,42147	-0,3331	-0,4866
81	0	1				-0,84563	-0,31635	0,00621	1,4201	-0,55153	-0,71653	-0,4866
88	0	0				-0,84563	-0,31635	0,03647	-0,96044	0,38598	0,86879	-0,4866
89	0	0				1,17894	-0,31635	0,2559	-0,88109	1,15279	0,87248	-0,4866
92	0	0				1,17894	-0,31635	-0,03919	-0,56368	1,07963	0,84667	-0,4866
93	0	0				-0,84563	-0,31635	-0,28888	-0,32563	-0,12613	0,8393	-0,4866
94	0	0				1,17894	-0,31635	-0,03163	-0,72239	1,10131	0,89828	-0,4866
95	0	0				1,17894	-0,31635	-0,56884	-0,32563	0,26134	0,81717	-0,4866
96	0	0				-0,84563	-0,31635	-0,45534	-0,64304	-0,73849	0,0577	-0,4866
101	0	0				1,17894	-0,31635	-0,17539	0,78529	0,06083	0,91672	-0,4866
112	0	0				1,17894	-0,31635	0,39209	0,22983	0,58649	1,10474	-0,4866
113	0	0				-0,84563	-0,31635	-0,01649	-0,00823	0,04729	0,88354	-0,4866
120	0	0				-0,84563	-0,31635	-0,84123	-0,00823	-1,15035	-0,60224	-0,4866

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
122	0	0				-0,84563	-0,31635	-0,74287	0,07113	-0,72223	-1,18844	-0,4866
125	0	1				-0,84563	-0,31635	-1,14388	0,38853	-1,33731	-0,90456	-0,4866
128	0	0				-0,84563	-0,31635	-0,39481	0,22983	-0,94713	-0,8935	-0,4866
129	0	0				-0,84563	-0,31635	0,26346	-0,40498	0,12044	0,8098	1,1281
132	0	0				-0,84563	-0,31635	-0,88663	1,7375	-0,83604	0,64758	-0,4866
137	0	0				1,17894	-0,31635	0,42236	-0,64304	1,0688	0,19779	1,06427
139	0	0				1,17894	-0,31635	0,11213	-0,48433	1,10944	0,05032	1,06427
142	1	0				1,17894	-0,31635	-0,7353	0,22983	-0,23993	0,64758	-0,4866
147	0	0				-0,84563	-0,31635	-0,70503	0,22983	-1,1639	-0,69441	-0,4866
154	0	1				-0,84563	-0,31635	-0,85636	0,30918	-1,3346	-1,26586	-0,4866
158	0	0				-0,84563	-0,31635	-0,49318	-0,88109	-0,69514	-1,25849	-0,4866
160	0	0				1,17894	-0,31635	0,2332	-0,16693	0,25592	0,20148	-0,4866
163	0	0				-0,84563	-0,31635	0,32399	-1,27785	-0,75746	-0,59855	-0,4866
168	0	0				-0,84563	-0,31635	-0,31158	-1,11914	-1,27499	-2,01059	-0,4866
171	0	0				1,17894	-0,31635	0,06674	-1,03979	0,29115	0,8098	-0,4866
172	0	0				-0,84563	-0,31635	0,81581	-0,08758	-0,40521	0,36001	-0,4866
173	0	0				1,17894	-0,31635	0,82337	0,15048	-0,28328	0,38213	-0,4866
186	1	1				-0,84563	-0,31635	-1,15145	0,62659	-1,15577	-0,54694	-0,4866
189	0	0				1,17894	-0,31635	-0,06189	1,02334	0,2505	0,23835	-0,4866

190	1	1				1,17894	-0,31635	1,70107	0,62659	0,27218	0,9204	-0,4866
191	1	1				1,17894	-0,31635	1,70107	0,54723	0,35347	0,92778	-0,4866
192	1	1				1,17894	-0,31635	1,70107	0,78529	0,36702	0,92778	-0,4866
197	0	0				1,17894	-0,31635	0,28616	0,62659	0,01748	0,89091	-0,4866
198	0	0				1,17894	-0,31635	-0,49318	-0,00823	0,1638	0,65127	-0,4866
210	0	0				1,17894	-0,31635	0,29373	-0,48433	1,0146	0,07244	1,06427
212	0	0				1,17894	-0,31635	0,22563	-0,48433	1,37498	0,20148	1,06427
213	0	0				-0,84563	-0,31635	0,2105	-0,56368	0,80326	0,50011	1,06427
217	0	0				1,17894	-0,31635	-0,01649	-0,80174	0,18005	0,77293	-0,4866
221	1	1				1,17894	-0,31635	1,67837	0,46788	-0,19387	0,90566	-0,4866
222	1	1				1,17894	-0,31635	1,67837	0,62659	-0,08277	0,9204	-0,4866
223	1	1				1,17894	-0,31635	1,67837	0,62659	0,01206	0,92778	-0,4866
224	1	1				1,17894	-0,31635	1,67837	0,86464	0,02561	0,92778	-0,4866
227	0	0				1,17894	-0,31635	-0,6445	1,4201	0,09064	0,36001	-0,4866
230	0	0				1,17894	-0,31635	-0,09216	-0,00823	0,48624	0,78399	-0,4866
241	0	0				1,17894	-0,31635	-0,09216	0,22983	0,94145	0,78399	-0,4866
248	1	1				1,17894	-0,31635	1,64054	0,38853	0,28573	1,00151	-0,4866
252	1	0				-0,84563	-0,31635	-2,52096	1,5788	-1,67871	-1,47232	-0,4866
260	0	0				1,17894	-0,31635	-0,81853	0,78529	-0,14509	0,75819	-0,4866
265	1	0				1,17894	-0,31635	0,59638	1,34075	0,62713	0,12037	-0,4866

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	B09[C-O]	F02[C-P]	SpMax_A	ChiA_B(m)	ATS1m	SpMax2_Bh(p)	P_VSA_s_1
281	0	0				-0,84563	-0,31635	0,2559	-0,48433	-0,1478	0,15355	-0,4866
330	1	1				-0,84563	2,92624	-1,12118	1,7375	0,45101	-0,52113	1,65263
342	0	0				-0,84563	-0,31635	0,59638	-1,43655	0,6461	0,74344	1,06427
360	0	0				-0,84563	-0,31635	0,48289	0,30918	-0,4025	-0,06028	-0,4866
364	0	0				-0,84563	-0,31635	-1,19685	-0,24628	-2,21521	-2,06221	-0,4866
366	1	1				1,17894	2,92624	-0,56127	-0,64304	0,32366	0,74713	2,62743
367	1	1				1,17894	2,92624	-0,28132	-0,64304	0,41579	0,86142	2,62743
368	1	1				1,17894	2,92624	-0,48561	-0,00823	0,5052	0,75081	2,62743
379	0	0				-0,84563	-0,31635	0,21806	0,22983	-0,11258	0,54435	-0,4866
380	0	0				-0,84563	-0,31635	0,74771	0,15048	-0,14509	0,65127	-0,4866
381	0	0				1,17894	-0,31635	1,08063	0,46788	-0,45128	0,41163	-0,4866
384	0	0				-0,84563	-0,31635	0,8839	0,46788	0,17464	0,72869	-0,4866
385	0	0				-0,84563	-0,31635	1,15629	0,15048	0,07167	0,73606	-0,4866
392	1	1				-0,84563	2,92624	-0,17539	-1,59525	0,67862	0,56647	2,62743
393	1	1				-0,84563	2,92624	-0,12999	-0,80174	0,83848	0,57016	2,62743
396	0	0				-0,84563	-0,31635	0,06674	0,46788	-0,7683	0,76925	-0,4866
399	1	1				-0,84563	-0,31635	-0,6899	0,86464	-0,92816	-0,5027	-0,4866
400	0	1				-0,84563	-0,31635	-0,758	1,10269	-1,08261	-0,66491	-0,4866
409	0	0				-0,84563	-0,31635	0,80824	-0,80174	0,48895	0,19042	-0,4866

Table S6.7. k-NN classification prediction, neighbors and descriptors.

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
1	1	1	0	0	0	Train	1	324	45	74	357	55	292	413	169	-1,83202	0,2994	-2,69332	0,21125	-0,11442	-0,86795	-0,49967	-1,15628	0,11581	-0,85339
2	1	1	1	1	1	Train	2	387	352	44	67	389	206	388	45	-2,95934	-1,11141	-0,85 34	0,21125	0,01883	-0,86795	-0,49967	-0,4866	5,00243	-0,85339
3	0	0	0	0	0	Train	3	66	395	74	123	371	187	45	73	-0,86235	-0,35945	-1,08899	2,23 45	1,49075	-0,86795	-0,49967	0,4063	-0,25299	-0,85339
4	1	1	0	1	1	Train	4	323	67	352	389	2	11	387	206	-1,42385	-0,55691	-0,46352	0,21125	-0,96818	2,67642	-0,49967	0,62952	10,35005	-0,85339
5	0	0	0	1	1	Train	5	291	254	353	6	405	325	328	66	-2,95934	-0,72113	-2,84596	0,21125	6,63604	-0,86795	-0,49967	-1,82595	-0,25299	-0,85339
6	0	0	0	0	0	Train	6	353	291	324	45	66	292	405	55	-3,05868	-0,96551	-1,85383	0,21125	1,79972	1,0843	-0,49967	-1,15628	-0,25299	-0,85339
7	0	0	0	0	0	Train	7	275	273	114	290	18	375	383	376	0,40103	-0,26952	0,98651	0,96875	-1,03255	1,06535	0,7562	0,18307	-0,25299	1,00009
8	0	0	0	0	0	Train	8	48	90	275	10	301	376	295	215	0,58676	-0,37704	1,08938	-0,02183	1,22469	0,57255	0,12827	-0,4866	-0,25299	1,00009
9	0	0	0	0	0	Train	9	91	10	116	90	275	301	127	87	0,64939	-0,74264	1,66 01	-0,67898	-0,92526	0,76208	-0,49967	0,4063	-0,25299	1,00009
10	0	0	0	0	0	Train	10	91	9	90	275	108	116	377	301	1,23032	-1,04567	0,93674	-0,54 95	0,20335	0,64836	0,12827	0,62952	-0,25299	1,00009
11	0	0	0	0	0	Train	11	109	97	206	274	389	279	63	410	1,83718	0,12149	0,47552	-0,45885	-0,71499	-0,20456	-0,49967	0,62952	3,25062	-0,85339
12	1	1	0	1	0	Train	12	32	111	116	110	64	33	202	9	1,42037	-0,97724	2,81 15	0,10442	-0,47038	0,93267	-0,49967	2,41532	-0,25299	0,71084
13	1	0	0	0	0	Train	13	232	300	49	272	411	312	302	406	0,09436	0,23489	0,5452	-0,43619	-0,07558	0,02288	-0,49967	-0,26338	0,11581	-0,85339
16	0	0	0	0	0	Train	16	22	174	373	156	285	375	175	24	0,08357	-0,93423	-0,37559	-0,47828	0,25485	-0,86795	0,7562	0,62952	-0,25299	1,35726

18	1	0	0	0	0	Train	18	382	383	275	375	376	149	24	7	-0,33354	-1,51293	1,07761	1,17917	-0,51759	0,02288	0,7562	0,85275	-0,25299	1,00009
19	0	0	0	0	0	Train	19	317	271	70	27	411	26	410	267	0,04685	-1,69279	0,91186	0,38282	-0,01121	-0,01502	-0,49967	0,62952	-0,25299	-0,85339
20	0	0	0	0	0	Train	20	313	219	70	27	63	203	218	195	0,69042	-0,45916	0,74263	0,62561	0,01454	-0,09084	-0,49967	1,2992	-0,25299	-0,85339
21	0	0	0	0	0	Train	21	26	296	411	242	243	272	408	372	0,61051	-1,43277	-0,94299	-0,31318	0,23768	0,00393	-0,49967	0,4063	-0,25299	-0,85339
22	0	0	0	0	0	Train	22	283	285	309	24	16	375	184	127	0,29953	-0,11702	0,40418	-0,04449	-0,26869	-0,86795	0,7562	1,07597	-0,25299	1,00009
23	0	0	0	0	0	Train	23	27	70	203	52	269	116	289	267	1,15042	-1,13365	1,68333	0,66446	-0,74074	-0,86795	-0,49967	0,4063	-0,25299	-0,85339
24	0	0	0	0	0	Train	24	246	22	106	115	90	108	273	310	0,70338	-1,00071	1,09435	-0,05097	-0,56908	-0,86795	0,7562	0,62952	0,02361	1,00009
25	0	0	0	0	0	Train	25	310	275	108	375	301	383	219	106	0,83943	-0,77783	0,01927	0,94285	-0,76648	-0,01502	-0,49967	0,85275	-0,25299	1,00009
26	0	0	0	0	0	Train	26	411	130	317	21	19	302	406	400	0,12446	-1,45037	0,04582	-0,28404	0,37071	-0,86795	-0,49967	0,4063	0,11581	-0,85339
27	0	0	0	0	0	Train	27	70	203	23	313	267	20	269	193	0,6969	-0,73482	1,32165	0,70978	-0,35881	-0,16666	-0,49967	0,4063	-0,25299	-0,85339
29	0	0	0	0	0	Train	29	250	54	239	157	358	355	354	104	-0,21446	1,64448	-0,76605	-2,01918	-0,72357	-0,86795	1,38414	-1,60273	-0,25299	0,71084
30	0	0	0	0	0	Train	30	194	218	180	63	202	34	219	270	0,66019	0,61026	1,71651	-0,08981	-0,81369	1,08043	-0,49967	1,96887	-0,25299	-0,85339
31	0	0	0	0	0	Train	31	41	403	402	346	344	134	299	253	-0,25981	0,79208	-0,59791	0,90077	-1,054	0,38301	0,7562	-0,93305	-0,25299	0,71084
32	1	1	1	1	1	Train	32	249	33	116	64	34	110	244	111	1,84115	-0,84043	1,29013	0,10118	-0,60771	-0,28038	-0,49967	2,1921	-0,25299	-0,07128
33	1	1	1	1	1	Train	33	64	32	319	110	111	53	249	181	1,51108	-1,00462	2,02178	-0,40706	-0,067	-0,86795	-0,49967	2,63855	-0,25299	-0,85339

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
34	1	1	1	1	1	Train	34	249	32	63	87	310	30	194	228	1,2756	-0,53345	0,89527	-0,16751	-1,83931	0,19347	-0,49967	2,41532	-0,25299	-0,85339
40	0	0	0	0	0	Train	40	289	78	269	267	297	411	312	406	-0,26845	-0,69572	0,51368	-0,16751	-0,69782	-0,86795	-0,49967	-0,04015	-0,25299	-0,85339
41	0	0	0	0	0	Train	41	31	346	298	299	344	119	403	308	-0,41099	1,03646	-0,24452	1,80071	-1,04542	0,51568	0,12827	-1,15628	-0,25299	-0,07128
42	1	1	1	1	1	Train	42	258	119	185	196	257	235	118	255	0,15051	1,37273	-0,05539	0,22 42	-0,0198	-0,86795	-0,49967	-0,70983	-0,25299	-0,85339
43	1	1	0	1	1	Train	43	334	59	335	69	407	231	391	61	0,19587	0,8605	0,43404	-0,01859	0,43938	2,20258	-0,49967	-0,93305	-0,25299	-0,85339
44	0	0	0	0	0	Train	44	45	206	55	357	388	406	169	413	-1,75643	-0,35939	-0,43532	0,21125	0,56812	-0,86795	-0,49967	-0,4866	1,86762	-0,85339
45	0	0	0	0	0	Train	45	324	55	357	413	66	74	44	169	-1,91192	-0,35945	-1,18522	0,21125	0,68398	-0,86795	-0,49967	-0,4866	0,11581	-0,85339
46	0	0	0	0	0	Train	46	341	343	312	216	316	272	296	302	-0,50817	-0,92837	-0,20636	-1,19046	-1,19991	0,30719	0,12827	-0,4866	-0,25299	-0,12447
47	0	0	0	0	0	Train	47	72	75	256	80	179	236	240	42	0,74441	2,04331	-0,36563	-0,55597	-0,73645	-0,86795	-0,49967	0,62952	-0,25299	-0,85339
48	1	0	0	0	0	Train	48	295	215	275	8	301	91	211	375	0,22178	-0,24019	0,52197	-0,35526	0,14757	0,72418	0,12827	-0,70983	0,11581	0,57514
49	1	0	0	0	0	Train	49	300	232	216	13	272	363	312	302	0,0123	0,24857	0,60161	-1,27463	-0,45751	-0,05293	-0,49967	0,18307	-0,25299	-0,85339

50	0	0	0	0	0	Train	50	179	229	199	107	228	234	244	203	1,9170 8	1,2495 6	0,56345	1,04321	-0,29444	-0,86795	-0,49967	0,85275	-0,25299	-0,85339
51	0	0	0	0	0	Train	51	302	272	243	61	271	13	266	411	-0,2727 7	-0,1932 7	0,1885	0,08176	0,98008	0,66731	-0,49967	-0,04015	-0,25299	-0,85339
52	0	0	0	0	0	Train	52	194	23	203	202	27	70	193	267	1,1137 1	-0,1209 3	2,04501	0,1724	-0,68924	-0,86795	-0,49967	0,85275	-0,25299	-0,85339
53	0	0	0	0	0	Train	53	70	410	111	98	110	181	116	20	0,9517 3	-1,3311 1	1,23 04	-0,23549	0,88567	-0,86795	-0,49967	1,2992	-0,25299	-0,85339
54	0	0	0	0	0	Train	54	250	29	365	121	359	157	104	239	-0,9379 3	0,9582 5	-1,05083	-2,59864	-0,58625	-0,86795	-0,49967	-1,60273	-0,25299	0,71084
55	1	0	0	0	0	Train	55	357	169	161	388	292	45	166	84	-1,7953	-0,4083 3	-1,0724	0,21125	0,68827	-0,86795	-0,49967	-0,4866	0,11581	0,57514
58	1	1	1	1	1	Train	58	294	332	62	231	61	394	333	315	-0,3677 9	1,6249 3	-1,22006	0,54144	0,61532	1,59605	-0,49967	-0,93305	-0,25299	-0,85339
59	1	1	1	1	1	Train	59	43	69	348	391	231	394	335	333	0,0187 8	1,9748 8	0,5452	-0,611	0,49087	2,52479	-0,49967	-0,93305	-0,25299	-0,85339
60	1	1	1	1	1	Train	60	335	327	394	69	348	391	59	336	-0,9444 1	0,8996	-0,47016	-1,48181	-0,70211	2,41107	-0,49967	-0,4866	1,13002	-0,85339
61	0	1	1	1	1	Train	61	62	231	391	294	394	407	315	51	-0,4109 9	0,5613 8	-0,69413	-0,12866	1,32768	1,29279	-0,49967	-0,4866	0,11581	-0,85339
62	1	1	1	1	1	Train	62	61	294	329	231	331	332	58	407	-0,1280 8	0,7001 9	-0,72731	0,56086	1,47358	1,31175	-0,49967	-0,4866	-0,25299	-0,85339
63	1	0	0	0	0	Train	63	218	20	127	87	108	310	249	219	0,4161 5	-0,5197 6	0,80402	-0,12542	-0,68924	0,00393	-0,49967	1,2992	0,30021	-0,85339

64	1	1	1	1	1	Train	64	181	33	110	111	319	32	53	20	1,78319	-0,99484	1,85255	0,31484	0,70544	-0,16666	-0,49967	2,41532	-0,25299	-0,85339
65	0	1	1	1	1	Train	65	220	408	247	243	242	351	20	68	0,60835	-0,8834	-0,48841	0,61913	1,75252	0,02288	-0,49967	1,52242	-0,25299	-0,85339
66	0	0	0	0	0	Train	66	45	395	413	3	374	324	55	292	-1,69596	-1,05936	-0,69579	1,18564	1,50362	-0,86795	-0,49967	-0,93305	-0,25299	-0,85339
67	1	1	1	1	1	Train	67	352	389	2	387	206	44	97	60	-1,8169	-1,21185	-0,62279	0,21125	-1,55518	1,80455	-0,49967	0,62952	4,44923	-0,85339
68	1	1	1	1	1	Train	68	243	188	331	351	220	51	407	231	0,76601	0,70801	0,00434	-0,07686	1,37488	1,00848	-0,49967	0,85275	-0,25299	-0,85339
69	1	1	1	1	1	Train	69	391	348	394	315	307	231	407	306	-0,23606	1,34926	-0,38223	-0,94443	0,56812	1,29279	-0,49967	-0,93305	-0,25299	-0,85339
70	0	0	0	0	0	Train	70	27	203	267	317	23	20	131	19	0,58244	-0,99093	1,22221	0,73244	-0,04125	-0,86795	-0,49967	0,85275	-0,25299	-0,85339
71	0	0	1	1	1	Train	71	363	250	268	216	121	80	72	49	-0,56648	1,14203	-0,78372	-1,98681	-1,87794	-0,86795	-0,49967	-0,04015	0,48461	-0,85339
72	0	0	0	0	0	Train	72	47	80	75	196	305	82	42	256	0,04253	1,8654	0,05245	-0,33584	-0,98963	-0,86795	-0,49967	0,62952	-0,25299	-0,85339
73	0	0	0	0	0	Train	73	267	195	313	78	70	20	102	317	-0,64638	-0,4748	0,42243	0,71625	0,19048	-0,86795	-0,49967	1,07597	-0,25299	-0,85339
74	1	1	1	1	1	Train	74	187	124	255	119	257	235	123	262	-0,96385	0,99345	-1,15702	0,69683	0,38788	-0,86795	-0,49967	-0,04015	-0,25299	-0,85339
75	0	0	0	0	0	Train	75	80	305	72	195	196	47	232	406	0,1138	0,99345	-0,23954	-0,10924	-0,28586	-0,86795	-0,49967	0,85275	-0,25299	-0,85339

76	0	0	0	1	1	Train	76	305	119	235	229	255	42	74	234	0,24338	1,4783	-0,76005	1,57087	-0,48326	-0,86795	-0,49967	0,18307	-0,25299	-0,85339
78	0	0	0	0	0	Train	78	40	262	297	289	406	312	305	411	-0,57944	-0,43961	0,04582	0,23067	-0,92526	-0,86795	-0,49967	0,18307	-0,25299	-0,85339
80	0	0	0	0	0	Train	80	72	75	82	196	305	262	195	47	-0,47793	1,1518	-0,17484	-0,10924	-0,99822	-0,86795	-0,49967	0,62952	-0,25299	-0,85339
82	1	0	0	0	0	Train	82	80	78	262	305	72	196	75	255	-1,06319	0,86441	-0,16323	0,32131	-1,57754	-0,86795	-0,49967	0,4063	-0,25299	-0,85339
83	0	0	0	0	0	Train	83	372	201	312	411	245	297	289	199	0,73145	-0,70549	-0,00023	0,32455	-1,13554	0,00393	-0,49967	-0,26338	-0,25299	-0,85339
84	0	0	0	0	0	Train	84	159	156	55	262	124	395	406	375	-0,96385	-0,29689	-0,55311	-0,28404	0,55524	-0,86795	-0,49967	0,62952	-0,25299	0,57514
85	0	0	1	0	1	Train	85	326	145	103	167	325	153	205	328	-1,81474	0,13127	-1,62653	0,21125	1,34484	3,60516	-0,49967	-1,60273	-0,25299	2,18758
86	0	0	0	0	0	Train	86	238	21	408	372	83	245	242	280	2,14816	-1,23727	-0,89654	0,24038	-0,45322	-0,29933	-0,49967	0,62952	-0,25299	-0,85339
87	0	0	0	0	0	Train	87	310	108	127	63	106	98	289	249	0,42695	-1,10237	0,52695	-0,63366	-0,97676	-0,12875	-0,49967	1,2992	-0,25299	-0,07128
90	0	0	0	0	0	Train	90	91	115	301	116	10	8	24	275	1,15258	-0,33012	1,71319	-0,00565	0,07891	-0,12875	0,12827	-0,04015	0,11581	0,71084
91	0	0	0	0	0	Train	91	9	10	90	275	115	127	48	116	0,4831	-0,79151	1,57715	-0,70812	-0,45322	0,64836	0,12827	0,18307	-0,25299	0,71084
97	0	0	0	0	0	Train	97	206	232	229	80	228	75	234	13	0,04253	1,11857	0,06075	0,19182	-0,93385	0,06079	-0,49967	0,62952	1,77542	-0,07128

98	0	0	1	0	0	Train	98	108	127	219	242	218	20	87	63	0,70986	-0,55886	0,53027	-0,42324	0,2291	-0,09084	-0,49967	1,2992	-0,25299	-0,07128
99	0	0	0	0	0	Train	99	175	166	104	48	156	402	31	401	-0,44986	0,44408	-0,64436	-0,09629	0,52949	0,36405	0,7562	-0,4866	-0,25299	1,35192
100	0	0	0	0	0	Train	100	134	105	185	124	255	257	119	229	0,17859	-0,00754	0,2615	0,78423	0,09178	-0,86795	0,7562	-0,26338	0,11581	-0,07128
102	0	0	0	0	0	Train	102	219	267	195	218	98	70	20	203	0,27145	0,01201	1,2221	0,18859	0,0832	-0,86795	-0,49967	1,07597	-0,25299	-0,07128
103	1	1	1	1	1	Train	103	322	328	294	329	254	62	58	61	-1,0416	1,01495	-1,83724	0,84898	2,46488	2,01304	-0,49967	-1,82595	-0,25299	-0,07128
104	0	1	1	1	1	Train	104	253	261	266	394	315	407	121	391	-0,35484	0,64349	-1,17195	-0,91206	-0,02409	0,43987	0,12827	-0,70983	-0,25299	-0,07128
105	1	1	1	1	1	Train	105	118	121	255	257	263	262	119	258	-0,35484	0,26812	-0,49338	-0,1006	-0,18716	-0,86795	0,12827	-0,70983	-0,25299	-0,07128
106	0	0	0	0	0	Train	106	108	310	87	246	131	24	143	289	0,83511	-1,43277	0,54686	0,06881	-0,64633	-0,86795	0,12827	1,07597	-0,25299	-0,07128
107	0	0	0	0	0	Train	107	244	351	219	236	98	20	203	50	1,56507	0,20752	0,20675	0,39253	0,06603	-0,86795	-0,49967	1,52242	-0,25299	-0,07128
108	1	0	0	0	0	Train	108	127	310	98	87	106	219	141	63	0,74225	-0,67226	0,43238	-0,19 34	-0,38885	-0,03398	0,12827	1,07597	-0,25299	-0,07128
109	0	0	0	0	0	Train	109	274	410	246	411	63	90	143	279	0,63859	-0,88536	0,99481	-0,29376	-0,10991	-0,01502	-0,49967	0,18307	1,59102	-0,07128
110	0	0	0	0	0	Train	110	111	98	64	377	303	53	218	32	0,72065	-0,74264	1,24035	-0,08658	0,69256	-0,10979	-0,49967	2,41532	-0,25299	-0,07662

111	0	0	0	0	0	Train	111	110	377	64	20	102	181	98	53	0,56948	-0,75437	1,71319	0,33426	0,66682	-0,12875	-0,49967	1,96887	-0,25299	-0,07662
114	0	0	0	0	0	Train	114	345	215	295	211	7	344	48	370	0,61915	-0,37704	0,87204	0,47022	-0,57767	1,36861	1,38414	-0,93305	-0,25299	-0,14102
115	0	0	0	0	0	Train	115	116	90	24	91	108	106	246	202	1,13746	-1,07304	1,92224	-0,09305	-0,39743	-0,10979	0,7562	0,62952	-0,25299	-0,07128
116	0	0	0	0	0	Train	116	115	27	410	98	108	202	23	32	1,24976	-1,10628	1,60535	-0,17074	-0,24723	-0,09084	-0,49967	0,85275	-0,25299	-0,07128
117	0	0	0	0	0	Train	117	174	175	177	25	87	9	310	378	0,01878	-1,249	-0,16323	-0,56892	-1,61617	0,13661	-0,49967	0,62952	0,02361	1,78221
118	1	1	1	1	1	Train	118	258	121	257	263	105	255	119	42	-0,10432	0,64154	-0,32748	-0,38116	0,03599	-0,86795	-0,49967	-0,93305	-0,25299	-0,07128
119	1	1	1	1	1	Train	119	257	255	187	124	123	42	118	185	-0,2447	0,90351	-0,26111	0,80689	-0,07129	-0,86795	-0,49967	-0,70983	-0,25299	-0,07128
121	0	1	1	1	1	Train	121	118	105	263	258	255	257	304	406	-0,33108	0,48709	-0,43532	-0,74049	-0,40172	-0,86795	-0,49967	-0,70983	0,02361	-0,07128
123	0	1	1	1	1	Train	123	187	119	257	124	255	74	185	42	-0,21446	0,99149	-0,37227	1,51584	0,71831	-0,86795	-0,49967	-0,70983	-0,25299	-0,07128
124	0	1	1	1	1	Train	124	187	257	255	119	185	123	258	262	-0,17559	0,64154	-0,27273	0,48317	0,5767	-0,86795	-0,49967	-0,04015	-0,25299	-0,07128
126	0	0	0	0	0	Train	126	7	290	30	91	9	12	10	273	1,25408	0,4343	1,86914	0,02349	-0,67637	2,08885	0,7562	1,2992	-0,25299	0,71084

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
127	0	0	0	0	0	Train	127	108	310	141	219	98	134	87	63	0,30601	-0,32035	0,56843	-0,23225	-0,45322	0,04184	0,12827	0,85275	-0,25299	-0,07128
130	0	0	0	0	0	Train	130	314	395	133	374	26	288	317	413	-0,30948	-1,45428	-0,05705	0,20477	0,48658	-0,86795	0,12827	-0,26338	-0,25299	-0,07128
131	0	0	0	0	0	Train	131	70	106	19	20	317	111	23	27	0,63859	-1,69279	0,97158	1,01731	-0,16557	-0,86795	-0,49967	1,52242	-0,25299	-0,07128
133	0	0	0	0	0	Train	133	314	130	395	166	374	156	288	26	-0,17559	-1,6576	-0,47679	-0,38444	0,53378	-0,86795	-0,49967	-0,70983	-0,25299	0,71084
134	0	0	0	0	0	Train	134	141	127	100	211	295	283	108	372	0,29521	0,06089	-0,07198	0,20477	-0,47038	-0,01502	0,7562	0,18307	-0,25299	-0,07128
135	0	0	0	0	0	Train	135	214	347	339	207	209	136	208	140	0,99925	-0,35749	1,28847	0,2242	-0,30302	0,32615	5,15174	-0,4866	-0,25299	1,35348
136	1	0	0	0	0	Train	136	207	208	140	237	209	138	339	347	0,7984	-0,32035	1,48459	0,57381	-0,40172	0,36405	3,26794	-0,93305	-0,25299	0,6411
138	0	0	0	0	0	Train	138	136	209	208	140	207	237	337	347	0,72929	-0,32035	1,36976	0,22743	0,20335	1,55815	3,26794	-0,93305	-0,25299	0,6411
140	0	0	0	0	0	Train	140	136	208	138	347	207	237	339	214	0,53061	-0,91859	0,58317	0,60295	-0,07558	0,59315	3,26794	-0,4866	-0,25299	0,6411
141	0	0	0	0	0	Train	141	134	127	108	283	361	343	300	211	0,2153	-0,1092	-0,10848	-0,63366	-0,43176	-0,01502	0,7562	0,62952	-0,25299	-0,07128
143	0	0	0	0	0	Train	143	246	106	87	116	24	410	63	108	0,67314	-1,54225	1,30838	-0,64985	-0,57337	-0,86795	0,12827	1,52242	0,76122	-0,07128

144	0	0	0	0	0	Train	144	146	164	405	292	169	388	161	369	-3,13643	-0,88145	-2,22546	0,21125	0,04028	-0,86795	-0,49967	-1,82595	-0,25299	2,85362
145	0	0	0	0	0	Train	145	153	85	167	205	326	346	99	164	-1,25324	0,62003	-1,73272	1,41225	0,38359	2,75224	-0,49967	-1,60273	-0,25299	4,04106
146	0	0	0	0	0	Train	146	144	405	164	292	169	388	161	357	-3,13643	-0,88145	-2,3416	0,21125	0,41363	-0,86795	-0,49967	-1,60273	-0,25299	2,85362
148	0	0	0	0	0	Train	148	157	175	174	16	99	159	204	156	-0,29869	0,13322	-0,34573	-1,51418	-0,27298	-0,86795	0,7562	-0,4866	-0,25299	2,85362
149	0	0	0	0	0	Train	149	376	18	288	259	382	100	375	299	-0,5384	-0,58623	0,41745	2,22802	-0,35881	-0,86795	0,7562	0,4063	-0,25299	1,00009
150	0	0	0	0	0	Train	150	365	54	162	261	157	148	359	29	-0,93793	0,95825	-0,98281	-3,81259	1,5637	-0,86795	-0,49967	-1,3795	-0,25299	1,77691
151	0	0	0	1	1	Train	151	152	386	183	178	156	185	187	258	0,221878	1,37273	-1,03258	-0,14161	0,28918	-0,86795	-0,49967	-0,04015	-0,25299	1,00009
152	1	1	1	0	1	Train	152	151	386	187	183	255	119	185	257	0,221878	1,37273	-1,03424	0,53582	0,04887	-0,86795	-0,49967	-0,26338	0,11581	1,00009
153	0	0	0	0	0	Train	153	145	85	205	326	167	346	398	170	-1,31587	0,76666	-1,63483	2,07264	0,27201	3,28295	-0,49967	-1,82595	-0,25299	5,00708
155	0	0	0	0	0	Train	155	123	3	149	205	278	41	76	376	-0,36779	0,57311	-0,80363	4,53291	0,95433	-0,86795	-0,49967	-0,70983	0,11581	1,00009
156	0	0	0	0	0	Train	156	84	124	151	159	257	255	262	105	-0,10432	0,11758	-0,55975	-0,15456	0,52091	-0,86795	-0,49967	0,18307	-0,25299	1,00009
157	0	0	0	0	0	Train	157	148	175	29	99	166	54	402	204	-0,35484	0,3952	-0,68086	-1,36851	-0,47038	-0,86795	0,7562	-1,3795	0,11581	2,42867

159	0	0	0	0	0	Train	159	84	156	16	175	266	174	55	121	-0,96385	-0,29689	-0,53652	-1,26492	0,35355	-0,86795	-0,49967	0,62952	-0,25299	1,00009
161	0	0	0	0	0	Train	161	169	357	55	166	388	369	164	292	-2,0523	-0,27147	-0,66427	0,21125	0,11753	-0,86795	0,12827	-0,70983	0,20801	1,00009
162	0	0	0	0	0	Train	162	365	359	397	150	54	293	159	363	-0,87314	-0,66444	-0,23954	-4,69959	-0,35023	-0,86795	0,12827	-0,04015	-0,25299	1,00009
164	1	1	0	0	0	Train	164	169	161	388	357	55	146	144	292	-2,25746	-0,47871	-0,78206	0,21125	0,34497	-0,86795	-0,49967	-0,70983	0,20801	2,42867
165	0	0	0	0	0	Train	165	339	214	347	136	207	208	209	237	0,72713	-0,17958	2,03505	1,30218	0,08749	0,36405	3,26794	-0,4866	-0,25299	2,49459
166	0	0	0	0	0	Train	166	357	369	161	402	169	55	105	403	-0,86235	-0,35945	-1,11554	0,21125	-0,02409	-0,86795	0,12827	-0,93305	-0,25299	1,00009
167	0	0	0	0	0	Train	167	99	31	175	369	166	346	402	403	-1,2662	-0,22064	-1,15204	0,21125	-0,56908	1,10325	0,7562	-1,60273	-0,25299	1,72367
169	0	0	0	0	0	Train	169	357	55	388	161	292	166	164	45	-1,89033	-0,40833	-0,99994	0,21125	0,67969	-0,86795	-0,49967	-0,93305	0,20801	1,00009
170	0	0	0	0	0	Train	170	398	403	369	149	288	299	164	402	-1,71324	-0,60383	-0,68584	2,27982	-1,42306	-0,86795	0,12827	-0,93305	-0,25299	2,42867
174	0	0	0	0	0	Train	174	175	117	16	375	22	275	301	177	0,12028	-0,70354	0,19679	-0,68222	-0,47897	0,02288	0,12827	0,62952	-0,25299	1,78221
175	0	0	0	0	0	Train	175	174	99	117	375	275	16	48	378	-0,47793	-0,54322	-0,30757	-0,57863	-0,47038	0,21242	0,12827	-0,26338	-0,25299	1,78221

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
176	0	0	0	0	0	Train	176	259	90	149	290	24	18	7	115	1,0510 8	-0,6370 7	2,27064	2,08882	-0,80081	-0,86795	0,7562	-0,4866	-0,25299	1,00009
177	0	0	0	0	0	Train	177	174	24	22	377	375	25	117	246	0,0835 7	-1,1180 1	0,84549	-0,03478	-0,56 05	-0,86795	-0,49967	1,52242	-0,25299	1,78221
178	0	0	0	0	1	Train	178	386	151	183	185	258	152	236	118	0,3708	2,0452 6	0,05577	-0,39087	-0,03267	-0,86795	-0,49967	-0,26338	-0,25299	0,71084
179	0	0	0	0	0	Train	179	199	50	229	372	236	201	305	228	1,4894 8	0,8644 1	0,02425	0,15945	-0,74503	-0,18561	-0,49967	0,62952	-0,25299	-0,85339
180	0	0	0	0	0	Train	180	202	193	199	194	270	179	232	30	1,4268 5	0,9875 8	1,46931	-0,18045	-0,65491	0,45882	-0,49967	0,4063	-0,25299	-0,85339
181	1	1	0	0	0	Train	181	64	111	20	53	110	203	32	107	1,8263 8	-0,5295 4	1,37308	0,71949	0,97 15	-0,22352	-0,49967	1,74565	-0,25299	-0,85339
182	0	0	0	0	0	Train	182	287	309	184	286	178	285	185	273	0,7271 3	2,1977 6	1,32165	0,17 24	0,14328	-0,86795	0,7562	0,85275	-0,25299	0,71084
183	0	0	0	0	0	Train	183	386	151	178	236	152	287	185	184	1,1180 2	1,7305	-0,797	-0,26138	0,32351	-0,86795	-0,49967	0,4063	-0,25299	0,71084
184	1	0	0	0	0	Train	184	285	236	22	283	301	309	183	141	0,7552 1	0,7940 3	0,39422	-0,79229	0,12611	-0,86795	0,12827	0,85275	-0,25299	0,70016
185	0	1	1	1	1	Train	185	258	42	257	124	255	119	187	118	0,4442 2	1,2710 6	-0,14498	0,21448	0,23768	-0,86795	0,12827	-0,26338	-0,25299	-0,07128
187	1	1	0	0	0	Train	187	124	257	123	119	255	74	185	258	-0,3677 9	1,0599 2	-0,64104	0,6774	0,7226	-0,86795	-0,49967	-0,26338	-0,25299	-0,07128
188	1	1	1	1	1	Train	188	331	332	333	231	407	394	68	235	0,1310 8	1,2925 7	-0,22461	0,3116	0,46512	0,8379	-0,49967	0,4063	-0,25299	-0,27411

193	0	0	0	0	0	Train	193	194	199	27	195	269	228	270	313	0,6148	0,2446	1,07279	0,59324	-1,18703	-0,05293	-0,49967	0,4063	-0,25299	-0,85339
194	0	0	0	0	0	Train	194	52	202	193	27	203	180	228	23	1,02084	0,12931	1,92721	0,54792	-0,92955	-0,14 77	-0,49967	1,07597	-0,25299	-0,85339
195	0	0	0	0	0	Train	195	305	267	228	289	313	102	219	75	0,27145	0,14886	0,53359	0,31807	-0,49184	-0,86795	-0,49967	0,85275	-0,25299	-0,85339
196	0	0	1	1	1	Train	196	42	80	75	305	263	255	406	118	-0,60967	1,15376	0,26316	-0,29376	-0,04554	-0,86795	-0,49967	-0,04015	-0,25299	-0,85339
199	0	0	0	0	0	Train	199	229	179	193	201	270	234	313	203	1,30159	0,35219	0,55535	0,70 33	-0,46 18	-0,03398	-0,49967	-0,04015	-0,25299	-0,85339
200	1	1	1	1	1	Train	200	350	233	254	306	307	348	328	391	-0,46282	1,37859	-0,65431	-3,28493	3,68361	2,05095	-0,49967	-1,3795	-0,25299	-0,85339
201	0	0	0	0	0	Train	201	305	229	234	83	372	199	297	195	0,79624	0,0472	-0,01889	0,44109	-0,84802	-0,86795	-0,49967	-0,26338	-0,25299	-0,85339
202	1	0	0	0	0	Train	202	194	52	116	180	27	23	193	115	1,49596	-0,24997	2,30548	0,06881	-0,65062	0,30719	-0,49967	0,62952	-0,25299	-0,85339
203	0	0	0	0	0	Train	203	70	27	23	20	267	52	195	313	1,19145	-0,3008	1,17067	0,80689	-0,05842	-0,86795	-0,49967	0,62952	-0,25299	-0,85339
204	0	0	0	0	0	Train	204	148	157	29	287	286	285	99	184	0,78112	1,79306	0,42741	-1,38469	0,63678	-0,86795	2,01207	-0,93305	-0,25299	2,27503
205	0	0	0	0	0	Train	205	346	331	41	376	31	99	123	308	-0,29869	0,90351	-0,28766	1,84 28	0,65823	1,40652	-0,49967	-0,93305	-0,25299	1,28752

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
206	0	0	0	0	0	Train	206	279	97	44	406	312	121	78	234	-0,37427	0,17037	0,03586	0,07528	-1,01967	-0,86795	-0,49967	-0,4866	1,95982	-0,85339
207	0	0	1	0	1	Train	207	136	214	209	339	208	347	140	138	0,74873	-0,3008	1,59871	0,63208	-0,50042	0,36405	3,89587	-1,15628	-0,25299	0,6411
208	0	0	1	1	1	Train	208	136	237	209	138	140	207	225	214	0,66882	-0,42788	1,67337	0,13032	0,64965	0,30719	3,26794	-0,4866	-0,25299	0,6411
209	0	0	0	0	0	Train	209	207	136	208	138	214	237	135	347	1,16986	0,04916	2,06658	0,07852	-0,11142	0,85685	3,89587	-0,4866	-0,25299	0,6411
211	0	0	0	0	0	Train	211	215	134	273	295	141	127	48	114	1,06835	0,11563	0,70281	-0,47828	-0,77936	0,42091	0,7562	-0,26338	-0,25299	-0,14102
214	1	0	0	0	0	Train	214	135	347	339	207	209	136	208	140	0,98413	-0,15612	1,44 94	0,55439	-0,44034	0,36405	4,52381	-0,4866	-0,25299	1,42318
215	0	0	0	0	0	Train	215	295	48	114	211	345	344	134	275	0,48094	-0,34185	0,60659	0,02996	-0,12708	0,49673	0,7562	-1,15628	-0,25299	-0,14102
216	0	0	1	1	1	Train	216	268	49	363	300	232	312	13	40	-0,63775	0,08239	0,70613	-1,23254	-1,23853	-0,05293	-0,49967	-0,04015	-0,25299	-0,85339
218	0	0	0	1	1	Train	218	63	219	20	98	242	195	303	102	0,07277	-0,02905	0,85877	-0,26462	-0,18287	-0,16666	-0,49967	1,74565	-0,25299	-0,85339
219	0	0	0	0	0	Train	219	102	127	218	20	98	195	313	108	0,32328	0,18796	0,64474	0,26304	-0,16141	-0,10979	-0,49967	1,2992	-0,25299	-0,07128
220	1	1	1	1	1	Train	220	351	65	247	243	242	408	303	20	0,58676	-0,01341	-0,08359	0,30836	1,11311	0,04184	-0,49967	1,74565	-0,25299	-0,85339
225	0	0	0	0	0	Train	225	237	208	115	370	140	136	10	24	1,16554	-1,4	1,57549	-0,17722	0,30205	0,3451	2,01207	-0,04015	-0,25299	0,6411

																	0931								
226	0	0	0	0	0	Train	226	256	47	72	50	244	179	351	228	1,43981	2,61223	0,21172	0,01702	-0,87806	-0,86795	-0,49967	2,63855	-0,25299	-0,85339
228	0	0	0	0	0	Train	228	195	193	63	179	194	219	203	229	0,79408	0,39911	0,63811	0,63856	-1,07975	-0,86795	-0,49967	1,2992	0,20801	-0,85339
229	0	0	0	0	0	Train	229	234	305	201	199	195	42	179	279	0,81784	0,68259	0,26979	0,68064	-0,24294	-0,86795	-0,49967	-0,04015	0,20801	-0,85339
231	1	1	1	1	1	Train	231	332	61	62	394	294	331	188	69	-0,47793	1,34731	-0,27936	0,02673	1,04016	1,02744	-0,49967	-0,26338	-0,25299	-0,85339
232	0	0	0	1	1	Train	232	13	300	49	195	372	411	305	312	0,26282	0,46363	0,58668	-0,40382	-0,44893	-0,05293	-0,49967	0,18307	0,11581	-0,85339
233	1	1	1	1	1	Train	233	348	307	306	391	69	350	261	315	-0,26197	1,59951	-0,39384	-2,02242	1,59803	1,08043	-0,49967	-1,15628	-0,25299	-0,85339
234	0	0	0	0	0	Train	234	229	201	199	305	42	179	279	185	1,11802	0,87419	0,11716	0,76157	-0,62487	-0,86795	-0,49967	-0,4866	0,39241	-0,85339
235	1	1	1	1	1	Train	235	332	42	253	407	255	315	188	119	0,05981	1,41183	-0,76547	0,35692	-0,11885	0,07974	-0,49967	-0,4866	-0,25299	-0,85339
236	0	0	0	0	0	Train	236	184	183	179	185	107	229	47	75	1,54131	1,10879	0,152	-0,43296	0,1819	-0,86795	-0,49967	0,4063	-0,25299	-0,07128
237	1	0	0	0	0	Train	237	208	136	209	225	140	138	290	207	1,023	-0,11311	1,60867	-0,03478	-0,26011	0,38301	2,64001	-0,04015	-0,25299	0,6411

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
238	0	0	0	0	0	Train	238	86	410	106	408	70	203	131	21	1,94948	-1,70648	-0,05207	0,68064	0,17331	-0,86795	-0,49967	0,62952	-0,25299	-0,85339
239	0	0	1	1	1	Train	239	258	104	151	250	29	118	386	178	0,29521	2,15279	-1,72276	-1,28111	0,30205	-0,86795	0,12827	-1,15628	-0,25299	-0,07128
240	0	0	0	0	0	Train	240	47	234	72	42	76	235	179	229	1,25408	2,78623	-0,74059	0,2242	-1,13554	-0,86795	-0,49967	-0,26338	0,85342	-0,85339
242	1	1	1	1	1	Train	242	280	98	408	243	220	218	108	20	0,53925	-0,54127	-0,27439	-0,30023	0,06174	-0,01502	-0,49967	1,52242	-0,25299	-0,85339
243	1	1	1	1	1	Train	243	242	220	351	68	51	300	98	411	0,68178	-0,05837	-0,25116	-0,22578	0,71402	0,02288	-0,49967	0,62952	-0,25299	-0,85339
244	0	0	0	0	0	Train	244	107	351	284	179	50	32	228	236	1,88685	0,3385	0,24491	0,22096	-0,31316	-0,86795	0,12827	1,74565	-0,25299	-0,07128
245	1	0	0	0	0	Train	245	372	272	83	312	280	300	343	411	0,80272	-0,2832	-0,42038	-0,36497	-0,94243	0,72418	-0,49967	0,4063	-0,25299	-0,85339
246	0	0	0	0	0	Train	246	24	106	143	310	108	116	115	87	1,14826	-1,05349	0,98032	-0,07686	-1,04113	-0,86795	0,12827	0,62952	0,57681	0,57514
247	1	1	1	1	1	Train	247	408	220	65	242	351	243	280	303	1,2692	-0,67812	-0,55809	-0,1066	1,16461	-0,07189	-0,49967	2,1921	-0,25299	-0,85339
249	1	1	1	1	1	Train	249	34	32	63	242	280	20	218	87	1,26272	-0,72113	0,39092	-0,10276	-0,71107	-0,12875	-0,49967	2,1921	-0,06859	-0,85339
250	1	0	0	0	0	Train	250	54	121	71	118	29	363	263	239	-0,46282	1,37859	-0,67588	-1,72246	-1,22136	-0,86795	-0,49967	-1,60273	-0,25299	-0,07128

251	1	1	1	1	1	Train	251	264	254	329	258	185	42	187	123	1,076999	2,55944	0,46722	0,42249	3,03562	-0,86795	-0,49967	-0,93305	-0,25299	-0,85339
253	1	1	1	1	1	Train	253	235	104	394	332	407	315	69	188	-0,63991	1,378459	-0,80695	-0,12542	-0,44464	0,53464	-0,49967	-0,70983	-0,25299	-0,07128
254	1	1	1	1	1	Train	254	322	329	350	328	103	294	61	62	-0,54704	1,4626	-1,22504	-0,45885	3,52483	0,76208	-0,49967	-1,15628	-0,25299	-0,07128
255	1	1	1	1	1	Train	255	119	257	262	124	105	305	118	187	-0,17559	0,64154	-0,48011	0,40548	-0,27727	-0,86795	-0,49967	-0,26338	-0,25299	-0,07128
256	0	0	0	0	0	Train	256	226	47	72	179	75	236	244	50	1,465772	2,03549	-0,05539	-0,69517	-1,12266	-0,86795	-0,49967	1,74565	-0,25299	-0,85339
257	1	1	1	1	1	Train	257	119	124	255	118	258	187	105	185	-0,10432	0,64154	-0,38057	0,34397	0,36642	-0,86795	-0,49967	-0,70983	-0,25299	-0,07128
258	1	1	1	1	1	Train	258	118	257	185	42	263	121	119	124	0,243338	0,98954	-0,22129	-0,28728	0,40075	-0,86795	-0,49967	-0,93305	-0,25299	-0,07128
259	0	0	0	0	0	Train	259	376	100	377	203	102	90	149	375	0,878331	-0,21673	1,04079	1,50937	0,44367	-0,86795	0,12827	0,4063	-0,25299	0,71084
261	1	1	1	1	1	Train	261	266	307	306	391	104	348	315	394	-0,41099	0,25639	-0,58131	-1,38793	0,95433	0,43987	-0,49967	-0,70983	-0,25299	-0,07128
262	1	1	1	1	0	Train	262	255	78	305	406	105	124	119	257	-0,47793	0,11758	-0,38072	0,36663	-0,58196	-0,86795	-0,49967	0,18307	-0,25299	-0,07128
263	1	1	1	1	1	Train	263	118	121	105	258	257	304	42	406	-0,40235	0,47536	-0,30591	-0,36821	0,10036	-0,86795	-0,49967	-1,15628	-0,25299	-0,85339

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
264	1	1	1	1	1	Train	264	68	251	333	351	188	331	332	236	1,26272	2,22122	-0,10335	0,42814	1,87697	-0,14777	-0,49967	0,62952	-0,25299	-0,85339
266	1	1	1	1	1	Train	266	261	104	307	391	306	394	51	61	-0,41099	0,25639	-0,51163	-0,93472	0,93288	0,42091	-0,49967	-0,26338	-0,25299	-0,07128
267	0	0	0	0	0	Train	267	269	289	70	195	40	102	27	317	0,1246	-0,4572	1,10431	0,16593	-0,02409	-0,86795	-0,49967	0,4063	-0,25299	-0,85339
268	1	1	0	0	0	Train	268	216	363	49	71	362	46	232	312	-0,61183	0,33068	0,60493	-1,38146	-2,12683	-0,05293	-0,49967	0,18307	-0,25299	-0,85339
269	0	0	0	0	0	Train	269	289	40	267	27	195	193	78	23	0,1246	-0,4572	1,20219	0,20801	-0,85231	-0,86795	-0,49967	-0,04015	-0,25299	-0,85339
270	0	0	0	0	0	Train	270	199	313	193	271	276	277	180	27	0,70338	0,23489	0,93384	0,62885	-0,36739	1,08043	-0,49967	0,18307	-0,25299	-0,85339
271	0	0	0	0	0	Train	271	272	276	411	296	312	19	313	302	0,17859	-0,82475	0,70613	-0,05542	-0,23007	0,68627	-0,49967	0,18307	-0,25299	-0,85339
272	0	0	0	0	0	Train	272	271	302	300	411	296	312	13	245	0,18291	-0,65662	0,13375	-0,65309	-0,10133	0,62941	-0,49967	-0,04015	-0,25299	-0,85339
273	0	0	0	0	0	Train	273	211	301	22	283	90	134	285	24	0,6753	0,51055	1,06117	-0,21283	-0,98534	-0,10979	0,7562	0,18307	-0,25299	1,00009
274	0	0	0	0	0	Train	274	109	410	300	267	289	26	13	40	0,66019	-0,73091	1,16072	-1,02536	-0,18287	-0,86795	-0,49967	0,18307	0,85342	-0,85339
275	0	0	0	0	0	Train	275	375	376	383	48	91	378	7	295	0,1246	-0,94401	0,77249	0,42949	-0,17857	0,47778	0,12827	-0,04015	-0,25299	1,00009

276	0	0	0	0	0	Train	276	271	272	270	245	312	411	335	295	0,15051	-0,70549	0,73102	-0,11571	-0,80081	1,44442	-0,49967	0,18307	-0,25299	-0,85339
277	0	0	0	0	0	Train	277	270	295	199	193	27	269	13	271	0,29953	0,07848	1,46101	0,92667	-0,41446	0,45882	-0,49967	-0,93305	-0,25299	-0,85339
278	0	0	0	0	0	Train	278	299	298	41	297	76	288	78	262	-1,17549	-0,19522	-0,87331	2,50966	-1,47455	-0,86795	-0,49967	-0,70983	-0,25299	-0,85339
279	0	0	0	0	0	Train	279	201	297	411	229	406	305	83	234	0,48094	-0,58623	-0,07861	0,75 51	-0,45751	-0,86795	-0,49967	-0,26338	0,85342	-0,85339
280	1	1	1	1	1	Train	280	242	245	408	98	249	108	63	243	0,65587	-0,68594	-0,29595	-0,32289	-0,25153	0,68627	-0,49967	1,52242	-0,25299	-0,85339
282	0	0	0	0	0	Train	282	288	373	130	166	402	299	356	403	-0,47793	-1,25877	-0,26111	0,22 42	-1,08404	-0,86795	0,7562	-0,4866	-0,25299	0,57514
283	0	0	0	0	0	Train	283	285	22	141	134	184	412	127	309	0,27145	0,48122	0,00434	-0,00888	-0,74503	-0,86795	0,7562	1,07597	-0,25299	0,57514
284	0	0	0	0	0	Train	284	244	22	377	309	106	107	24	108	1,68384	-0,90882	0,16859	0,56 41	-0,08416	-0,86795	0,7562	1,96887	-0,25299	1,00009
285	0	0	0	0	0	Train	285	283	22	184	287	412	309	286	134	0,74657	0,61612	-0,17152	-0,12866	-0,12708	-0,86795	0,7562	0,85275	-0,25299	1,00009
286	1	0	0	0	0	Train	286	287	285	183	184	185	151	182	386	0,82864	1,48221	-0,28102	0,03644	1,37488	-0,86795	0,7562	0,85275	-0,25299	1,00009
287	0	0	0	0	0	Train	287	286	183	285	182	412	386	184	185	1,21089	2,01203	-0,26111	0,04291	0,53378	-0,86795	0,7562	0,85275	-0,25299	1,00009
288	0	0	0	0	0	Train	288	282	130	299	166	297	262	149	395	-0,47793	-1,25877	-0,32416	1,04645	-0,56908	-0,86795	0,12827	-0,26338	-0,25299	0,56984

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
289	0	0	0	0	0	Train	289	40	269	267	195	78	411	27	70	0,24554	-0,72113	0,69949	-0,01859	-0,83085	-0,86795	-0,49967	0,4063	-0,25299	-0,85339
290	0	0	0	0	0	Train	290	237	7	90	114	273	225	8	136	1,100075	0,360001	2,01514	0,76481	0,03599	0,87581	1,38414	-0,26338	-0,25299	1,000009
291	0	0	0	0	0	Train	291	405	292	324	66	55	357	6	45	-2,14084	-0,8834	-2,26 86	0,21125	2,38334	-0,86795	-0,49967	-1,3795	-0,25299	0,04224
292	0	0	0	0	0	Train	292	357	169	55	388	161	166	324	291	-1,87953	-0,96551	-1,50044	0,21125	0,88138	-0,86795	-0,49967	-1,3795	-0,25299	0,70016
293	0	0	0	0	0	Train	293	71	363	397	359	316	268	365	162	-0,89906	-0,18545	-1,08899	-4,09747	-2,586	0,25033	-0,49967	0,18307	0,30021	-0,85339
294	1	1	1	1	1	Train	294	62	329	58	231	61	332	322	311	-0,96385	1,28084	-0,93138	0,63856	1,43067	1,34965	-0,49967	-0,70983	-0,25299	-0,85339
295	0	0	0	0	0	Train	295	215	48	13	344	134	271	411	275	0,18291	-0,27343	0,52031	0,19182	-0,21 72	0,57255	0,12827	-0,70983	0,11581	-0,14102
296	0	0	0	0	0	Train	296	302	411	272	312	271	21	300	26	-0,56648	-1,2099	-0,17318	-0,26138	-0,07987	0,3451	-0,49967	0,18307	-0,25299	-0,85339
297	0	0	0	0	0	Train	297	78	40	406	201	374	83	279	312	-0,17775	-0,7661	-0,30093	0,41843	-0,69782	-0,86795	-0,49967	-0,70983	-0,25299	-0,85339
298	0	0	0	0	0	Train	298	308	299	288	41	376	31	346	295	-1,15605	-0,63511	-0,51993	1,57411	-0,43176	0,55359	0,12827	-0,70983	-0,25299	-0,07128
299	0	0	0	0	0	Train	299	403	402	262	288	278	298	255	297	-0,96385	-0,05642	-0,67754	1,27629	-1,03255	-0,86795	0,12827	-0,70983	-0,25299	-0,07128

300	0	0	0	0	0	Train	300	49	232	272	13	312	411	302	372	0,15915	-0,23042	0,13873	-0,87645	-0,33306	0,02288	-0,49967	0,4063	0,11581	-0,85339
301	0	0	1	0	0	Train	301	48	90	275	375	273	127	184	174	0,78976	0,11172	0,48879	-0,21606	-0,13995	0,00393	-0,49967	-0,04015	-0,25299	1,00009
302	0	0	0	0	0	Train	302	296	272	411	300	312	51	13	271	-0,46498	-0,73677	-0,02884	-0,63366	0,25056	0,21242	-0,49967	-0,04015	0,11581	-0,85339
303	0	0	0	1	0	Train	303	218	220	242	110	280	98	351	219	0,10084	0,06284	0,49377	-0,43619	0,6754	0,61045	-0,49967	2,41532	-0,25299	-0,85339
304	0	0	1	1	0	Train	304	406	263	121	118	302	105	400	300	0,05117	-0,43961	-0,37559	-0,98328	0,01024	-0,86795	-0,49967	-0,70983	-0,25299	-0,85339
305	0	0	0	0	0	Train	305	201	229	195	75	255	406	262	372	0,23474	0,45972	-0,11511	0,399	-0,49184	-0,86795	-0,49967	0,18307	-0,25299	-0,85339
306	1	1	1	1	1	Train	306	307	391	348	261	233	315	69	394	-0,26197	0,73147	-1,10889	-1,75373	1,13886	1,04639	-0,49967	-0,4866	-0,25299	-0,85339
307	1	1	1	1	1	Train	307	391	306	348	69	315	261	233	266	0,20451	0,665	-0,51661	-1,42678	1,26331	1,04639	-0,49967	-0,70983	-0,25299	-0,85339
308	0	0	0	0	0	Train	308	298	336	295	376	344	275	346	296	-0,89906	-0,67421	-0,04046	1,33132	-0,39743	1,19802	0,12827	-0,26338	-0,25299	-0,14102
309	0	0	0	0	0	Train	309	22	285	377	283	184	102	284	24	0,68826	0,41475	1,12009	0,23391	0,1304	-0,86795	0,7562	1,74565	-0,25299	1,00009
310	0	0	0	0	0	Train	310	108	127	87	106	63	289	219	141	0,43127	-0,79738	0,54686	0,08176	-1,16987	0,00393	0,12827	1,07597	-0,25299	-0,07128

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
311	0	0	0	0	0	Train	311	349	302	272	296	46	300	49	216	-1,03512	-1,19035	0,02259	-2,35909	-0,02838	1,23593	-0,49967	0,4063	0,11581	-0,85339
312	0	0	0	0	0	Train	312	411	372	300	272	296	302	83	40	-0,22094	-0,54322	-0,11014	-0,22901	-0,80081	0,17451	-0,49967	-0,04015	0,11581	-0,85339
313	0	0	0	0	0	Train	313	20	27	411	195	219	267	271	199	0,34272	-0,21673	0,48713	0,83279	-0,03267	0,04184	-0,49967	0,62952	-0,25299	-0,85339
314	0	0	0	0	0	Train	314	130	133	374	395	26	297	413	304	-0,30948	-1,45428	0,06407	0,02673	0,58957	-0,86795	-0,49967	-0,93305	-0,25299	-0,07662
315	1	1	1	1	1	Train	315	394	407	391	69	307	332	231	348	0,30601	1,187	-0,93047	-0,69517	0,59815	0,8379	-0,49967	-0,4866	-0,25299	-0,85339
316	1	0	0	0	0	Train	316	46	304	312	40	216	341	78	268	-0,42394	-1,40931	-0,69745	-1,62101	-1,74061	-0,86795	-0,49967	-0,26338	-0,25299	-0,85339
317	0	0	0	0	0	Train	317	70	19	267	27	26	289	269	40	0,04901	-1,48556	1,05786	0,70 33	0,17331	-0,86795	-0,49967	0,18307	0,11581	-0,85339
318	1	0	0	0	0	Train	318	377	177	309	284	383	382	22	174	0,22178	-1,22945	1,40294	-0,08 01	0,40505	0,17451	1,38414	3,30822	-0,25299	2,85362
319	1	1	1	1	1	Train	319	33	64	32	181	110	111	249	53	2,42028	-0,89904	1,87246	0,21448	0,44367	-0,86795	-0,49967	3,75467	-0,25299	-0,85339
322	1	1	1	1	1	Train	322	103	328	294	329	62	58	254	61	-0,46282	1,37859	-1,54 69	0,52202	2,33614	1,69082	-0,49967	-1,82595	-0,25299	-0,85339
323	0	1	1	1	1	Train	323	4	389	352	67	2	11	387	97	-0,00282	0,86246	0,96495	-0,20635	-1,26428	0,78104	-0,49967	-0,4866	8,32164	0,71084

324	0	0	0	0	0	Train	324	1	45	292	291	357	55	66	413	-1,98967	-0,32426	-2,21219	0,21125	0,81701	-0,86795	-0,49967	-1,3795	-0,25299	-0,85339
325	1	1	1	1	1	Train	325	328	103	85	353	322	326	254	350	-2,14084	0,01983	-2,12758	0,21125	4,42172	3,98424	-0,49967	-1,82595	-0,25299	0,372235
326	1	1	0	0	0	Train	326	85	327	336	167	103	334	205	328	-1,91624	0,02961	-0,87331	0,21125	0,60674	3,52935	-0,49967	-1,15628	0,30021	1,000009
327	1	1	1	1	1	Train	327	336	335	58	60	43	334	394	253	-1,31803	0,5868	-0,80031	0,21125	-0,48755	2,50584	-0,49967	-0,93305	-0,06859	-0,85339
328	1	1	1	1	1	Train	328	103	322	329	294	254	62	61	350	-1,18197	0,82336	-1,07074	0,21125	3,00129	2,43002	-0,49967	-1,3795	-0,06859	-0,85339
329	1	1	1	1	1	Train	329	294	62	231	61	322	331	58	332	-0,47793	1,441115	-0,64436	0,61266	2,23744	1,40652	-0,49967	-0,26338	-0,25299	-0,85339
331	1	1	1	1	1	Train	331	188	332	231	62	333	407	340	68	0,20451	1,1518	-0,42536	0,65798	0,76122	1,08043	-0,49967	0,18307	-0,25299	-0,27411
332	1	1	1	1	1	Train	332	333	331	188	231	235	407	58	394	0,05117	1,55144	-0,72731	0,46051	0,46942	0,95162	-0,49967	-0,04015	-0,25299	-0,85339
333	1	1	1	1	1	Train	333	332	188	331	231	340	235	394	315	0,27577	2,20753	-0,47845	0,21448	0,42221	1,06535	-0,49967	0,18307	-0,25299	-0,85339
334	0	0	0	0	0	Train	334	43	335	276	349	51	336	270	295	0,00582	-0,34576	0,60099	0,15621	0,30635	2,35421	-0,49967	-0,93305	-0,25299	-0,85339
335	0	1	1	1	1	Train	335	336	394	407	43	327	231	276	188	-0,47793	0,56138	-0,03888	-0,17722	-0,27298	1,69082	-0,49967	-0,04015	-0,06859	-0,85339

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
336	1	1	0	1	1	Train	336	335	327	308	332	58	231	43	311	-1,03512	0,64349	-0,25282	0,87811	-0,25153	1,84245	-0,49967	-0,26338	-0,06859	-0,85339
337	0	0	0	0	0	Train	337	138	237	209	339	136	208	347	225	0,46582	-0,65075	2,04169	-0,58834	-0,57337	1,93722	2,64001	-0,70983	-0,25299	1,782221
338	0	0	0	0	0	Train	338	138	214	135	209	207	208	165	347	0,29953	1,41574	0,98154	1,00112	1,88984	2,14572	5,15174	-1,60273	-0,25299	1,712447
339	0	0	0	0	0	Train	339	347	214	207	136	135	165	140	209	0,78112	-0,84626	1,52903	0,59647	-1,03255	0,38301	3,89587	-0,70983	-0,25299	1,712447
340	1	1	1	1	1	Train	340	333	331	332	188	58	231	329	62	0,59108	2,2349	-0,87663	0,53497	0,94575	1,97513	-0,49967	0,18307	-0,25299	-0,14102
341	0	0	0	0	0	Train	341	46	343	272	104	312	304	295	48	0,07061	-0,95183	-0,72068	-1,09982	-0,78365	0,61045	0,12827	-1,15628	-0,25299	-0,14102
343	0	0	0	0	0	Train	343	341	245	272	372	46	312	141	300	0,62995	-0,77001	-0,651	-0,71783	-0,59054	0,38301	0,12827	-0,26338	0,11581	-0,14102
344	0	0	0	0	0	Train	344	345	295	215	134	83	31	343	346	0,76817	-0,02123	-0,49 67	0,66122	-0,67208	0,95162	0,12827	-0,93305	-0,25299	-0,14102
345	0	0	0	0	0	Train	345	114	344	215	295	211	134	343	31	0,62563	-0,57646	-0,30923	0,43785	-0,87806	1,12221	1,38414	-1,15628	-0,25299	-0,14102
346	0	0	0	0	0	Train	346	31	41	344	205	253	308	336	298	-0,29869	0,64154	-0,27273	1,05939	-1,10 55	1,36861	-0,49967	-0,93305	-0,25299	0,84433
347	0	0	0	0	0	Train	347	214	339	135	136	207	140	209	237	1,023	-0,35554	0,77083	0,32779	-0,62058	0,42091	3,89587	-0,4866	-0,25299	1,712447

348	0	1	1	1	1	Train	348	69	391	233	307	306	394	315	261	-0,19286	1,2945 2	-0,51495	-1,64691	0,68827	1,16012	-0,49967	-1,15628	-0,25299	-0,85339
349	0	0	0	0	0	Train	349	272	276	271	334	302	49	13	300	0,32112	-0,58037	0,7194	-1,45915	0,2291	1,53919	-0,49967	-0,4866	-0,25299	-0,85339
350	1	1	1	1	1	Train	350	200	233	254	306	307	328	348	391	-0,51033	1,4958 9	-1,29472	-2,11954	3,01417	1,84245	-0,49967	-1,60273	-0,25299	-0,85339
351	1	1	1	1	1	Train	351	220	243	107	242	68	219	179	20	1,11586	0,6962 8	-0,11677	0,14974	0,55524	-0,05293	-0,49967	1,52242	-0,25299	-0,85339
352	1	1	1	1	1	Train	352	67	389	2	387	206	44	11	97	-1,13014	-1,2020 8	-0,71238	0,21125	-0,67208	1,80455	-0,49967	-0,04015	4,44923	-0,85339
353	0	0	0	0	0	Train	353	6	291	324	328	103	405	45	66	-3,25521	-0,9889 8	-2,34658	0,21125	2,1087	1,69082	-0,49967	-1,15628	0,30021	-0,85339
354	1	1	1	1	1	Train	354	355	401	402	29	403	31	356	239	-0,40667	1,6972 6	-1,95503	0,04939	-0,34594	-0,86795	2,64001	-1,3795	-0,25299	0,71084
355	1	1	1	1	1	Train	355	354	401	29	402	403	356	239	31	-0,2447	1,6210 2	-1,92351	-0,10924	-0,15712	-0,86795	2,64001	-1,3795	-0,25299	0,71084
356	1	1	0	0	0	Train	356	105	373	166	402	282	343	403	104	0,10516	-0,4024 6	-1,64313	-0,27757	-0,56908	-0,86795	0,7562	-0,93305	-0,25299	-0,07128
357	0	0	0	0	0	Train	357	55	169	292	388	161	166	45	369	-1,76075	-0,4787 1	-1,08402	0,21125	0,46942	-0,86795	-0,49967	-0,93305	0,11581	0,57514
358	0	0	0	0	0	Train	358	273	362	361	211	46	141	268	216	-0,32244	0,6063 5	0,53359	-1,34585	-1,98951	0,06079	1,38414	-0,4866	-0,25299	0,57514

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
359	1	0	0	0	0	Train	359	365	162	54	358	397	29	363	148	-1,21221	0,15864	-0,11014	-3,22999	-1,39731	-0,86795	0,7562	-0,26338	-0,25299	1,000009
361	0	0	0	0	0	Train	361	141	134	370	283	211	127	273	22	-0,2231	0,46558	0,11882	-0,55273	-0,31316	0,11765	2,01207	0,4063	-0,25299	-0,07128
362	1	1	1	1	1	Train	362	363	216	49	46	48	268	121	104	-0,26845	0,66304	0,43902	-1,54332	-0,96388	0,64836	-0,49967	-0,4866	-0,25299	0,5054
363	1	1	1	0	1	Train	363	216	268	49	362	71	250	300	46	-0,26845	0,66304	0,36934	-2,29758	-1,36298	0,06079	-0,49967	-0,4866	-0,25299	-0,85339
365	0	0	0	0	0	Train	365	162	359	54	150	397	29	293	250	-0,88826	-0,21673	-1,13711	-4,34026	-0,55192	-0,86795	0,12827	-1,15628	-0,25299	1,000009
369	0	0	0	0	0	Train	369	166	402	161	403	357	299	169	282	-1,72619	-0,35945	-1,45067	0,21125	-0,77507	-0,86795	0,7562	-0,93305	-0,25299	1,000009
370	0	0	0	0	0	Train	370	361	211	215	237	225	114	48	141	0,39671	-0,54909	0,84881	-1,14119	-0,07558	0,64836	2,01207	-0,4866	0,39241	-0,07128
371	1	1	1	1	1	Train	371	331	62	313	51	408	332	243	65	0,34272	-0,33599	-1,24329	1,54174	0,59815	0,93267	-0,49967	0,4063	-0,25299	-0,85339
372	0	0	0	0	0	Train	372	245	83	312	411	201	300	305	272	0,58676	-0,18154	-0,43863	-0,06068	-0,71928	0,00393	-0,49967	0,18307	-0,25299	-0,85339
373	0	0	0	0	0	Train	373	282	134	356	16	141	288	283	22	0,25634	-0,82671	-0,63938	0,02349	-0,89952	-0,86795	1,38414	-0,04015	-0,25299	0,57514
374	0	0	0	0	0	Train	374	314	130	297	395	413	317	390	26	-0,13671	-1,49533	-0,10516	0,76157	0,40934	-0,86795	-0,49967	-0,93305	-0,25299	-0,85339

375	1	0	0	0	0	Train	375	376	275	22	127	174	377	378	383	-0,31164	-0,43961	0,517	0,32779	0,03599	0,04184	0,12827	0,62952	0,11581	1,00009
376	0	0	0	0	0	Train	376	375	275	383	259	378	149	18	100	-0,30732	-0,43179	0,46556	1,17 27	0,41363	0,11765	0,12827	0,18307	-0,25299	1,00009
377	0	0	0	0	0	Train	377	111	383	375	110	309	98	10	108	0,61051	-0,76023	1,08 44	0,42 49	0,66252	-0,07189	0,12827	1,52242	-0,25299	1,00009
378	0	0	0	0	0	Train	378	383	275	375	376	174	175	382	18	-0,27277	-1,37607	-0,02221	0,58676	0,14328	0,89476	0,12827	0,4063	0,30021	1,77691
382	0	0	1	0	1	Train	382	18	383	375	275	376	24	378	377	-0,3354	-1,51293	1,01 14	1,12738	-0,37168	-0,01502	0,7562	1,07597	0,48461	1,00009
383	0	0	0	0	0	Train	383	275	377	18	378	382	375	376	10	0,25202	-1,4054	0,79406	0,87487	0,21194	0,89476	0,12827	1,07597	-0,25299	1,00009
386	1	0	0	0	0	Train	386	178	151	183	152	185	287	412	187	0,46366	2,37762	-0,73893	0,09795	0,19477	-0,86795	-0,49967	-0,04015	-0,25299	1,00009
387	1	1	0	1	1	Train	387	2	388	44	161	169	357	164	55	-2,95934	-1,1141	-0,92972	0,21125	-0,52617	-0,86795	-0,49967	-0,70983	3,25062	1,00009
388	1	0	0	0	0	Train	388	169	357	55	292	161	164	166	45	-1,83202	-0,99289	-1,28311	0,21125	0,52949	-0,86795	-0,49967	-0,93305	0,85342	1,00009
389	1	1	1	0	0	Train	389	11	352	67	109	2	206	97	44	-0,08272	-0,81107	0,52695	-0,802	-0,25582	0,02288	-0,49967	1,74565	5,09463	-0,85339
390	0	0	0	0	0	Train	390	404	413	374	395	130	314	288	66	-1,14094	-2,67618	-0,2866	1,24391	0,04028	-0,86795	-0,49967	-0,4866	-0,25299	-0,85339

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
391	1	1	1	1	1	Train	391	69	307	348	394	315	306	61	407	-0,23606	0,86441	-0,66427	-0,95415	0,91571	1,25488	-0,49967	-0,93305	-0,25299	-0,85339
394	1	1	1	1	1	Train	394	315	407	69	391	231	332	335	348	-0,26197	1,1518	-0,69247	-0,70812	0,29776	1,10325	-0,49967	-0,26338	-0,25299	-0,85339
395	0	0	0	0	0	Train	395	130	314	374	133	26	413	84	288	-0,42394	-1,40931	-0,54115	0,65474	0,98008	-0,86795	-0,49967	-0,04015	-0,25299	-0,07128
397	0	0	0	1	0	Train	397	316	46	359	162	365	311	304	341	-0,47793	-1,30178	-0,42702	-2,90618	-0,90381	-0,86795	0,12827	-0,26338	1,40662	-0,07128
398	0	0	0	0	0	Train	398	170	369	149	166	288	167	164	401	-1,17765	-1,249	-1,22172	2,44815	-0,25582	-0,86795	1,38414	-1,3795	-0,25299	2,785779
401	1	1	1	1	1	Train	401	402	403	355	354	99	166	369	100	-0,84939	0,45581	-0,76215	0,61759	0,17761	-0,86795	2,01207	-0,93305	-0,25299	0,7108 4
402	1	0	0	0	0	Train	402	403	299	166	369	105	31	401	255	-0,84939	0,45581	-0,83535	0,55115	-0,75361	-0,86795	0,7562	-0,70983	-0,25299	0,7108 4
403	0	0	0	0	0	Train	403	402	299	31	369	166	105	401	282	-0,87314	0,49295	-0,93304	0,74862	-1,16558	-0,86795	0,7562	-0,93305	-0,25299	0,7108 4
404	0	0	0	0	0	Train	404	390	413	395	374	288	130	133	314	-0,83211	-2,67618	-1,61824	1,06587	0,10036	-0,86795	-0,49967	-0,4866	-0,25299	-0,12447
405	0	0	0	0	0	Train	405	291	292	146	144	169	388	357	324	-2,95934	-0,90903	-2,49423	0,21125	1,70102	-0,86795	-0,49967	-1,82595	-0,25299	1,29872
406	1	0	0	0	0	Train	406	305	304	262	411	78	40	255	263	-0,22094	-0,10529	-0,30259	-0,11247	0,04887	-0,86795	-0,49967	-0,04015	0,11581	-0,85339

407	1	1	1	1	1	Train	407	315	394	332	391	188	69	335	231	0,34056	0,69237	-0,70409	-0,30671	0,24627	1,02744	-0,49967	-0,26338	-0,25299	-0,85339
408	1	1	1	1	1	Train	408	247	242	65	220	280	20	21	243	0,82432	-1,15711	-0,58463	0,51554	0,49945	-0,01502	-0,49967	1,74565	-0,25299	-0,85339
410	0	0	0	0	0	Train	410	116	53	19	274	109	70	26	27	1,19145	-1,64392	1,02799	-0,14808	0,23768	-0,14 77	-0,49967	0,62952	0,57681	-0,85339
411	0	0	0	0	0	Train	411	312	271	272	296	302	300	372	313	0,08357	-0,72113	0,10 72	0,05262	-0,2644	0,00393	-0,49967	0,18307	0,11581	-0,85339
412	0	0	0	0	0	Train	412	285	283	287	183	185	386	151	47	0,63859	1,73245	-0,85 34	-0,00888	-0,96818	-0,86795	0,7562	0,62952	-0,25299	0,57514
413	0	0	0	0	0	Train	413	45	390	374	297	395	404	130	314	-1,31803	-1,56376	-0,97452	0,21125	-0,00263	-0,86795	-0,49967	-0,4866	-0,25299	-0,85339
14	0	0				Test	97	406	279	262	255	124	305	75	229	0,29089	-0,81107	0,43404	-0,01536	-0,57767	-0,86795	0,7562	1,2992	-0,25299	1,00009
15	0	0				Test	406	262	255	75	124	305	187	156	279	0,32112	-1,08673	0,26813	0,01054	-1,05829	-0,86795	0,7562	0,85275	-0,25299	1,00009
17	0	0				Test	124	255	262	187	76	305	74	75	119	0,67962	-1,00071	0,96495	0,02996	-0,8094	-0,86795	0,7562	0,18307	-0,25299	1,00009
28	1	0				Test	185	234	229	258	236	42	232	118	13	1,00788	0,13713	-0,26443	-0,25491	-0,59483	-0,07189	-0,49967	0,4063	-0,25299	-0,85339
35	1	1				Test	234	206	229	240	279	97	152	185	42	1,26272	-0,54127	0,89029	-0,24196	-1,89081	-0,24247	-0,49967	2,1921	-0,25299	-0,85339

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
36	1	1				Test	234	97	229	185	240	100	206	152	42	1,57737	-0,25388	1,01472	0,0591	-0,44893	0,43987	-0,49967	1,96887	-0,25299	-0,85339
37	1	1				Test	234	97	229	152	185	206	279	240	257	1,26704	-0,53345	0,70115	0,15298	-0,82227	-0,20456	-0,49967	1,96887	-0,06859	-0,85339
38	1	1				Test	97	234	206	229	240	279	152	185	119	1,26272	-0,54127	0,89527	-0,22578	-0,77078	-0,18561	-0,49967	2,1921	-0,06859	-0,85339
39	1	0				Test	234	97	229	100	41	76	185	123	240	1,84815	0,52228	2,27894	-0,06068	-0,10133	0,81895	-0,49967	1,96887	-0,25299	-0,85339
56	1	1				Test	211	370	114	91	237	215	225	115	290	-0,21446	1,76373	-1,14872	-0,45238	1,41351	1,44442	-0,49967	-0,93305	-0,25299	-0,85339
57	1	1				Test	138	114	237	370	225	208	209	337	140	-0,47793	1,44115	-0,18314	0,21772	2,15161	2,41107	-0,49967	-0,4866	-0,25299	-0,85339
77	0	0				Test	40	269	267	289	274	13	317	314	406	-0,59455	1,16994	-0,36895	-0,17722	-1,04971	-0,86795	-0,49967	-0,04015	-0,25299	-0,85339
79	0	0				Test	269	267	40	289	274	317	13	406	279	-0,55136	1,15118	-0,09552	-0,20635	-1,18703	-0,86795	-0,49967	0,4063	-0,25299	-0,85339
81	0	0				Test	267	269	274	40	317	289	406	13	314	-0,50817	1,21632	-0,23789	-0,10766	-1,36298	-0,86795	-0,49967	0,4063	-0,25299	-0,85339
88	0	0				Test	255	187	262	124	119	76	152	257	123	0,46115	-1,27105	1,24533	-0,11247	-0,36331	-0,86795	0,12827	-0,26338	0,11581	0,71084
89	0	0				Test	41	31	76	123	331	119	188	134	205	1,45277	-0,11116	2,08648	-0,35526	0,34068	0,47778	0,12827	-0,04015	0,20801	0,71084
92	0	0				Test	185	235	188	119	134	76	42	255	100	1,61474	0,19188	0,95499	-0,25491	-0,067	0,51568	0,12827	-0,04015	-0,25299	0,71084
93	0	0				Test	407	235	253	372	188	332	394	255	104	0,62779	-1,30374	0,03586	-0,36497	0,42221	-0,86795	0,12827	0,18307	-0,25299	0,71084
94	0	0				Test	76	41	123	235	152	119	187	74	332	1,88253	-1,07891	2,42066	-0,42324	-0,13566	-0,22352	0,12827	0,85275	-0,25299	0,71084
95	0	0				Test	76	119	255	235	41	305	229	123	262	1,15906	-0,57841	1,48258	-0,61424	-0,16999	0,00393	0,12827	0,4063	-0,25299	0,71084

96	0	1				Test	295	211	46	312	310	7	83	276	91	-0,41099	0,80576	-0,21964	-1,86638	0,56812	0,28824	-0,49967	0,4063	0,11581	-0,07128
101	0	0				Test	279	229	124	313	406	411	97	13	255	0,29521	0,16059	0,85379	0,25009	-0,16999	-0,86795	0,12827	1,07597	-0,25299	-0,07128
112	0	0				Test	97	206	279	41	44	235	371	234	331	0,69906	-1,01183	1,69 66	-0,05744	0,53808	-0,18561	-0,49967	2,41532	-0,25299	-0,07662
113	0	0				Test	279	187	235	123	257	331	97	119	124	0,48094	-0,91859	1,08938	0,42167	0,22481	-0,01502	-0,49967	1,52242	-0,25299	-0,07662
120	0	0				Test	40	216	269	289	121	312	78	49	46	-0,36779	0,57311	-0,75718	-1,37822	-0,67637	-0,86795	-0,49967	-0,26338	0,02361	-0,07128
122	0	1				Test	51	8	13	314	266	267	48	301	302	-0,24 47	0,90351	-0,25945	0,99465	-0,16999	-0,86795	-0,49967	-0,70983	0,02361	-0,07128
125	0	1				Test	51	266	272	302	261	13	48	271	349	-0,28789	0,57311	-0,52325	1,05292	0,56812	-0,86795	-0,49967	-0,70983	-0,25299	-0,07128
128	0	1				Test	100	130	215	105	134	295	267	314	40	-0,41099	0,56138	-0,09686	0,32779	-0,59054	0,49673	-0,49967	-0,70983	-0,25299	-0,07128
129	0	0				Test	141	134	242	280	108	127	408	21	343	-0,48225	-1,28028	-0,06866	-0,07363	0,33638	0,97058	1,38414	-0,4866	-0,25299	-0,07128
132	0	0				Test	104	266	407	315	121	261	406	118	343	0,10516	-1,49533	-0,36232	0,23714	0,1819	-0,86795	-0,49967	0,62952	-0,25299	0,71084
137	0	0				Test	283	309	244	220	284	351	219	22	107	0,56084	-0,62729	1,36147	-0,05 42	-0,56 05	1,25488	3,26794	-0,93305	-0,25299	0,6411
139	0	0				Test	283	303	309	219	218	351	244	226	220	1,10075	-0,12875	0,72 77	-0,57863	0,31064	1,14116	3,26794	-0,04015	-0,25299	0,6411
142	1	0				Test	331	188	199	235	332	407	229	234	270	0,91718	-0,07792	0,75093	0,27275	0,5252	-0,86795	-0,49967	0,18307	-0,25299	-0,07128
147	0	0				Test	51	266	261	8	307	61	68	62	243	-0,33108	0,48709	-0,52988	2,38665	0,42221	-0,86795	-0,49967	-0,26338	-0,25299	1,00009
154	0	0				Test	254	329	328	307	51	61	8	62	266	-0,36779	0,57311	-0,80529	4,26422	0,82989	-0,86795	-0,49967	-0,4866	-0,25299	1,00009

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
158	0	0				Test	8	266	261	51	156	307	254	133	314	-0,33108	0,48709	-0,81359	2,80101	0,16044	-0,86795	-0,49967	-0,93305	0,85342	1,00009
160	0	0				Test	13	232	49	362	295	216	48	300	196	0,71849	1,25151	-0,61947	-0,14808	0,2806	-0,86795	-0,49967	-0,04015	-0,25299	3,63569
163	0	0				Test	117	268	316	46	83	310	245	312	87	-0,87314	-0,66444	-0,24995	-4,25609	0,3278	-0,86795	0,12827	-0,26338	0,39241	1,00009
168	0	0				Test	133	21	296	314	130	26	395	272	302	-2,25746	-0,47871	-0,81027	0,21125	-0,09275	-0,86795	0,12827	-1,15628	0,20801	2,42867
171	0	0				Test	255	105	406	134	121	262	305	185	75	0,64507	-0,53736	-0,22793	-0,26138	-0,32877	0,02288	0,12827	0,18307	-0,25299	1,00009
172	0	0				Test	141	134	372	343	406	312	105	411	262	-0,30732	-0,56864	-0,28102	-0,35526	-0,29444	0,15556	0,12827	-0,04015	-0,25299	1,78221
173	0	0				Test	262	134	105	406	255	305	141	372	121	0,17859	-0,44938	-0,11677	-0,52683	-0,54763	0,11765	0,12827	0,18307	-0,25299	1,78221
186	1	1				Test	272	13	271	48	349	8	295	301	51	-0,40667	0,99149	-0,52657	0,64827	0,51662	-0,86795	-0,49967	-0,26338	-0,25299	-0,07128
189	0	0				Test	102	52	182	267	184	194	196	195	232	1,24544	2,57508	-0,61645	-0,08981	-1,4488	-0,86795	0,7562	0,18307	0,11581	0,57514
190	1	1				Test	97	13	109	232	234	229	295	48	199	1,16777	0,73538	0,05411	0,30189	0,39217	-0,09084	-0,49967	1,74565	-0,25299	-0,85339
191	1	1				Test	97	234	109	13	232	229	295	11	206	1,41173	0,87614	0,11055	0,21125	0,39217	-0,12875	-0,49967	1,96887	-0,25299	-0,85339
192	1	1				Test	97	234	109	229	232	13	11	206	295	1,41173	0,87614	0,06407	0,15945	0,26772	-0,12875	-0,49967	1,96887	-0,25299	-0,85339
197	0	0				Test	279	206	411	97	109	295	406	13	312	-0,05681	-0,205	0,65968	0,01378	0,11324	-0,86795	-0,49967	1,52242	-0,25299	-0,85339
198	0	0				Test	331	188	124	187	123	257	231	119	258	1,18713	0,25444	0,72106	1,09177	0,18619	-0,86795	-0,49967	0,4063	0,11581	-0,85339
210	0	0				Test	124	187	185	100	188	123	331	134	313	0,71202	-0,42788	1,29676	0,67093	-0,25153	0,42091	0,7562	-0,4866	-0,25299	-0,14102

212	0	0				Test	220	351	65	303	247	219	110	309	107	1,04646	-0,07597	2,09478	1,57087	-0,17857	0,30719	3,26794	-0,4866	-0,25299	0,64011
213	0	0				Test	219	351	220	313	283	134	20	188	242	0,62779	-0,58232	1,41456	-0,16751	0,24198	0,49673	2,01207	-0,04015	-0,25299	0,64011
217	0	0				Test	266	118	13	121	257	258	255	407	261	0,35352	-0,26952	-0,30923	0,14327	0,17331	-0,86795	-0,49967	0,18307	0,11581	-0,85339
221	1	1				Test	295	97	48	109	407	13	344	232	343	0,58676	-0,01341	-0,11677	0,21772	0,98866	0,04184	-0,49967	1,74565	-0,25299	-0,85339
222	1	1				Test	97	109	295	48	13	232	11	407	188	0,78976	0,32677	-0,03888	0,17888	0,95863	-0,01502	-0,49967	1,96887	-0,25299	-0,85339
223	1	1				Test	97	109	111	295	13	232	48	206	234	1,15906	0,64545	0,05577	0,07528	0,93717	-0,05293	-0,49967	2,1921	-0,25299	-0,85339
224	1	1				Test	97	109	111	295	13	232	206	48	234	1,15906	0,64545	-0,00023	0,00073	0,80414	-0,05293	-0,49967	2,1921	-0,25299	-0,85339
227	0	0				Test	90	13	102	232	274	109	267	196	277	1,06403	2,50047	-0,30923	0,399	-0,69782	-0,86795	-0,49967	1,07597	0,11581	-0,85339
230	0	0				Test	231	188	258	331	332	42	235	187	124	1,62554	0,25639	-0,05539	1,05616	0,14328	-0,86795	-0,49967	0,4063	-0,25299	-0,85339
241	0	0				Test	231	188	333	178	42	59	331	69	13	2,32093	1,27888	-0,05373	0,68064	0,50804	-0,86795	-0,49967	0,62952	-0,25299	-0,85339
248	1	1				Test	97	111	206	407	332	331	188	235	234	1,26272	-0,72113	0,40252	0,02349	0,96292	-0,10979	-0,49967	2,41532	-0,25299	-0,85339
252	1	1				Test	349	272	8	307	10	271	51	334	91	-0,96817	1,08142	-1,54192	1,33132	1,41351	-0,86795	-0,49967	-1,3795	-0,25299	-0,07128
260	0	0				Test	331	188	124	187	231	62	332	123	68	1,14826	0,07066	0,77581	1,52555	0,12611	-0,86795	0,12827	0,4063	-0,25299	0,71084
265	1	1				Test	180	270	277	295	211	202	90	48	43	0,7768	2,06872	-0,17484	-0,45562	1,57228	-0,10979	-0,49967	1,07597	-0,25299	-0,85339
281	0	0				Test	124	187	255	262	305	100	406	185	156	0,20667	-0,58623	0,62981	0,65798	-1,41447	0,17451	0,7562	-0,4866	-0,25299	-0,07128

CLASS_ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	6	7	8	9	ICR	LOC	VE1_B(p)	MATS8e	GATS3m	Eta_sh_x	CATS2D_04_DA	CATS2D_00_LL	T(Cl..Cl)	SAdon
330	1	1				Test	370	225	237	211	215	114	208	91	138	-0,36779	1,62493	-1,28311	0,62561	1,14315	1,70978	-0,49967	-0,70983	-0,25299	-0,85339
342	0	0				Test	356	141	104	105	134	343	121	373	372	0,20451	-1,33111	-0,92474	-0,53007	-0,50901	0,70522	0,12827	-0,70983	-0,25299	-0,14102
360	0	0				Test	87	63	218	195	228	249	289	102	219	-0,65718	0,53205	0,04084	-1,00918	-1,17845	-0,86795	2,01207	0,18307	-0,25299	-0,07128
364	0	0				Test	304	272	261	302	266	341	296	300	46	-0,86235	-0,35945	-1,51538	0,21125	0,01883	-0,86795	-0,49967	-0,93305	-0,25299	1,00009
366	1	1				Test	8	307	68	370	349	233	51	391	261	-0,19286	0,70019	-1,34615	3,04056	1,75681	1,53919	-0,49967	-0,93305	-0,25299	-0,85339
367	1	1				Test	8	307	68	370	349	48	51	215	266	-0,13671	0,63372	-1,09895	2,61001	1,44784	1,38756	-0,49967	-0,70983	-0,25299	-0,85339
368	1	1				Test	8	370	307	48	215	225	68	90	266	0,25634	1,21632	-1,33786	2,15681	0,38788	1,25488	-0,49967	-0,4866	-0,25299	-0,85339
379	0	0				Test	100	134	331	371	62	61	99	188	51	0,18939	-0,89513	0,90688	0,96552	0,50375	0,98953	0,12827	0,85275	-0,25299	1,00009
380	0	0				Test	406	343	105	356	262	372	134	255	21	-0,27493	-1,45037	0,19347	0,15621	-0,2129	0,0987	0,12827	0,62952	-0,25299	1,00009
381	0	0				Test	262	255	134	305	406	75	372	124	283	0,10732	-1,13365	0,41413	-0,29699	-0,68924	0,19347	0,7562	0,18307	-0,25299	1,00009
384	0	0				Test	124	406	279	187	255	97	257	100	331	0,27145	-1,04958	0,69286	0,43138	-0,01121	-0,03398	0,12827	1,2992	-0,25299	1,00009
385	0	0				Test	371	100	134	187	331	124	99	62	61	0,25202	-1,40554	0,91517	0,87811	0,17331	0,89476	0,12827	1,07597	-0,25299	1,00009
392	1	1				Test	114	211	48	215	370	10	91	345	225	-0,41099	0,56138	-0,78372	-0,03478	1,54654	1,27384	-0,49967	-0,93305	0,85342	-0,85339
393	1	1				Test	90	211	273	91	215	48	225	10	8	-0,31596	1,09315	-0,75054	0,05586	-0,08416	1,04639	-0,49967	-0,4866	0,85342	-0,85339
396	0	0				Test	104	266	406	261	304	343	121	407	21	-0,24747	-1,73758	-0,40711	0,30513	-0,00263	-0,86795	-0,49967	0,4063	-0,25299	-0,07128

399	1	0				Test	98	53	243	26	102	242	236	267	107	-0,47793	0,38151	-0,90152	0,87811	-1,04542	-0,86795	0,7562	-0,70983	-0,25299	0,71084
400	0	0				Test	26	98	1022	84	267	53	243	159	406	-0,46282	0,43235	-0,99774	1,01407	-1,43164	-0,86795	0,7562	-0,93305	-0,25299	0,71084
409	0	0				Test	104	343	407	344	266	341	61	99	48	-0,30732	-1,45037	-0,01059	0,16269	1,13886	0,00393	-0,49967	-0,04015	0,11581	-0,85339

Table S6.8. Random forest regression prediction and descriptors.

Set	CATS2D_07_LL	MATS8v	X3v	CATS2D_03_DL	TI2_L	IVDE	SpMax2_Bh(p)	CATS2D_09_LL	SaaN	SpPosA_B(i)	GATS5m	MATS4v	N-074
Train	-0,63121	0,39293	-1,97677	-0,49238	-0,96649	-3,02359	-3,37649	-0,58266	-0,51995	-1,53039	-3,08995	0,29861	-0,45766
Train	-0,63121	0,39293	0,64256	-0,49238	-1,85317	-2,88691	-1,8248	-0,58266	-0,51995	-0,7772	0,70419	0,21134	-0,45766
Train	-0,63121	0,39293	2,54874	-0,49238	-1,74279	0,71865	-0,80167	-0,58266	-0,51995	-0,71443	1,09534	-0,97218	-0,45766
Train	2,25433	0,6416	0,27147	3,73487	0,97098	0,50656	1,09358	3,82799	-0,51995	1,4196	-0,55027	1,21488	3,56978
Train	-0,15029	0,78666	-0,3531	2,67805	-1,00571	-0,39366	0,69338	-0,58266	1,80559	0,81286	-0,19265	-0,00136	-0,45766
Train	2,25433	0,11525	2,532	-0,49238	1,47713	0,91189	1,13511	2,44967	-0,51995	0,03875	0,1957	1,38396	1,55606
Train	2,73525	-0,30749	0,59334	-0,49238	2,99923	1,01558	1,12378	1,07134	-0,51995	1,58697	-0,56704	1,14398	1,55606
Train	-0,63121	-0,24533	-1,15927	-0,49238	0,03565	-0,38894	-0,46188	-0,58266	0,6459	0,83378	-1,62035	-0,41042	3,56978
Train	-0,63121	-0,0008	-0,04845	-0,49238	-0,15679	1,07685	0,51217	-0,58266	0,4401	2,21463	-0,5838	-1,35396	1,55606
Train	-0,63121	-0,95404	0,0746	-0,49238	-0,13936	0,67152	0,8142	-0,58266	0,48825	2,96781	0,1482	0,4295	1,55606
Train	0,09017	-0,58932	-0,19708	-0,49238	-0,88661	0,74693	0,57635	-0,58266	-0,51995	1,56605	-0,29603	-0,03954	1,55606
Train	-0,15029	0,25616	1,68596	-0,49238	0,05308	-1,95371	-0,15608	-0,03133	-0,51995	-2,15804	1,16798	-1,02127	-0,45766
Train	-0,63121	0,00749	1,76865	-0,49238	-0,35431	-1,76047	-0,39015	-0,58266	-0,51995	-2,22081	1,8497	-0,90128	-0,45766
Train	0,81156	-1,20685	0,23652	-0,49238	-0,39425	0,36045	-0,50341	-0,58266	-0,51995	-0,88181	-0,73467	-0,54677	-0,45766
Train	-0,63121	-1,05765	0,40583	-0,49238	-0,36666	0,64324	1,00297	-0,58266	-0,51995	-1,07011	-0,61453	-1,39759	-0,45766
Train	-0,63121	1,02289	-0,33046	-0,49238	-0,45743	0,43586	0,66696	-0,58266	-0,51995	-0,25415	0,86345	-0,08317	-0,45766
Train	2,97571	-0,02152	0,00225	-0,49238	2,02831	-1,28444	1,06337	1,07134	-0,51995	-0,42153	-0,04458	0,33679	-0,45766
Train	0,33063	-0,99548	0,38565	-0,49238	0,13804	0,44057	0,13085	0,52001	-0,51995	-1,67684	-0,46087	-0,09954	-0,45766
Train	-0,63121	0,76593	-0,24039	-0,49238	0,29635	0,63381	0,03647	-0,58266	-0,51995	-0,71443	0,27672	-0,75947	-0,45766
Train	0,33063	1,05191	-1,01408	-0,49238	-0,78059	-0,26169	-0,34862	-0,58266	-0,51995	-1,53039	0,88301	1,77665	-0,45766
Train	0,09017	0,98974	-0,69614	-0,49238	-0,88806	-0,30882	-0,71106	-0,58266	-0,51995	-0,79812	-0,30441	0,05318	-0,45766
Train	-0,63121	-1,14054	1,08453	-0,49238	-0,26499	0,17664	-0,84697	-0,58266	-0,51995	-1,53039	-0,65924	-2,31932	-0,45766
Train	0,09017	0,97316	-0,13556	1,62124	1,02109	0,74693	0,88593	0,79567	0,73281	0,14336	1,00594	-0,12681	-0,45766

Train	-0,63121	0,58772	-0,78178	0,56443	-0,17785	-0,12972	-1,1188	-0,58266	-0,51995	0,28982	-1,89694	-1,49031	-0,45766
Train	-0,63121	1,74404	-0,02482	0,56443	1,29704	0,57254	0,27809	-0,30699	2,8358	1,08485	-0,31559	0,2168	-0,45766
Train	-0,39075	0,13182	-1,14894	0,56443	0,42416	-0,38894	-0,27312	-0,58266	-0,51995	0,20613	-0,16192	0,35315	-0,45766
Train	-0,63121	0,39293	-1,28724	2,67805	-1,38841	-0,61046	-1,97582	-0,58266	-0,51995	1,06392	0,54215	-0,84128	-0,45766
Train	1,29248	0,95244	-0,14934	-0,49238	2,53592	0,01168	0,98409	0,52001	0,77097	0,45719	0,28231	0,72402	1,55606
Train	-0,15029	-0,61419	-0,78867	2,67805	0,44086	-0,83669	0,39135	0,52001	-0,51995	0,14336	0,20408	-0,19771	-0,45766
Train	-0,15029	0,43023	-1,42849	0,56443	-0,0711	-0,26169	-0,48454	-0,58266	-0,51995	0,3735	0,89418	0,70766	-0,45766
Train	1,29248	0,12353	0,27343	-0,49238	0,16418	0,61025	0,87083	-0,58266	-0,51995	-0,296	2,19335	1,05126	-0,45766
Train	2,49479	-0,67636	0,44422	-0,49238	1,24258	0,66209	1,02562	1,89833	-0,51995	-0,19139	0,06439	0,0859	-0,45766
Train	-0,63121	-0,45255	-0,94518	-0,49238	-0,01083	0,53955	-0,40525	-0,58266	-0,51995	-0,37968	-0,11722	-0,6613	-0,45766
Train	-0,15029	1,15137	0,53871	-0,49238	0,71028	0,8082	0,96144	2,174	-0,51995	1,2313	-0,12281	-0,37769	1,55606
Train	-0,63121	1,1348	-0,00809	1,62124	1,31447	0,34631	0,05157	-0,58266	2,89863	1,08485	-0,06693	-0,33951	-0,45766
Train	2,01386	-0,36552	0,08001	-0,49238	0,48807	0,64795	0,99164	2,174	-0,51995	-0,06586	0,11188	0,65857	-0,45766
Train	2,01386	-0,44012	0,11446	-0,49238	0,48807	0,64795	1,00297	2,174	-0,51995	-0,31692	0,24879	0,56585	-0,45766
Train	2,49479	-0,3821	0,29952	-0,49238	0,80977	0,72336	1,02562	2,44967	-0,51995	-0,63075	0,12027	0,45132	-0,45766
Train	-0,15029	-0,55202	1,73813	-0,49238	-0,60049	0,12008	0,96899	-0,58266	-0,51995	-0,90273	0,02248	-0,89582	-0,45766
Train	-0,63121	-1,34777	1,58802	-0,49238	-0,33398	0,41229	-0,22404	-0,30699	-0,51995	-1,13287	-0,42734	-1,37578	-0,45766
Train	-0,63121	-0,7178	-0,59672	-0,49238	-0,02971	0,75164	-0,29199	-0,30699	2,01815	0,12244	0,75728	1,96754	-0,45766
Train	-0,39075	1,12651	0,018	1,62124	1,22297	0,44057	0,34982	0,52001	1,69197	1,10577	-0,32117	-0,63403	-0,45766
Train	0,81156	-0,78411	0,37876	-0,49238	0,91216	0,69037	1,10868	2,72533	-0,51995	-0,31692	0,58964	0,60949	-0,45766
Train	-0,63121	-0,75925	-0,14491	-0,49238	0,9521	0,70451	0,18748	0,24434	-0,51995	0,28982	0,14541	1,12217	-0,45766
Train	-0,39075	-0,99548	0,15531	-0,49238	1,32028	0,32274	1,08225	0,79567	-0,51995	1,85896	0,23202	0,78402	-0,45766
Train	0,81156	-0,75096	0,13267	-0,49238	1,62673	0,28504	1,18419	1,89833	-0,51995	-0,1077	0,15659	1,10035	-0,45766

Set	CATS2D_07_LL	MATS8v	X3v	CATS2D_03_DL	TI2_L	IVDE	SpMax2_Bh(p)	CATS2D_09_LL	SaaN	SpPosA_B(i)	GATS5m	MATS4v	N-074
Train	2,49479	-0,33236	0,12923	-0,49238	1,58388	0,16721	1,10113	1,89833	-0,51995	-0,04493	-0,10325	1,13853	-0,45766
Train	-0,63121	0,4178	-0,28469	-0,49238	1,53087	-0,17685	-0,32975	-0,58266	-0,51995	-0,46337	-1,51697	-0,4595	3,56978
Train	-0,63121	1,30058	-1,54021	-0,49238	-0,21271	-0,42665	-1,43216	-0,58266	-0,51995	-0,5889	-1,30184	-1,14671	1,55606
Train	-0,63121	-1,05351	-1,18093	-0,49238	-0,21343	0,53955	-0,91871	-0,58266	-0,51995	-0,75628	2,11512	1,10035	1,55606
Train	-0,63121	-1,80781	-1,06527	-0,49238	0,16055	0,17664	-1,36798	-0,58266	-0,51995	-0,7772	-1,36889	-0,47586	1,55606
Train	-0,63121	-1,43895	-0,71189	-0,49238	-0,58089	-0,12972	-1,23206	-0,58266	-0,51995	-0,33784	-0,15633	-1,17943	-0,45766
Train	-0,63121	-0,83799	-1,22965	-0,49238	-0,66149	0,75164	-0,72616	-0,58266	-0,51995	0,47811	-0,00546	0,34224	-0,45766
Train	-0,63121	-0,02567	-1,17896	-0,49238	-0,67456	-1,29387	-0,70729	-0,58266	-0,51995	0,49903	0,00292	-0,52495	-0,45766
Train	-0,63121	1,07263	-1,11695	-0,49238	-0,44581	-0,54919	-0,93381	-0,58266	1,97411	0,18521	-0,27368	0,69675	-0,45766
Train	-0,63121	-0,74681	0,39894	-0,49238	1,66231	0,57254	0,11953	1,07134	-0,51995	-0,31692	-0,51395	0,25498	-0,45766
Train	-0,63121	-0,94575	-0,1016	-0,49238	-1,37606	0,22848	-0,27689	-0,58266	0,36259	1,00116	-0,32117	1,72211	-0,45766
Train	0,5711	0,25201	0,37433	-0,49238	0,87004	1,14283	0,64808	1,347	-0,51995	-0,25415	0,02527	1,13853	1,55606
Train	0,5711	0,65403	0,12923	0,56443	0,04364	-0,34652	0,13463	0,24434	-0,51995	-1,00734	0,517	0,25498	1,55606
Train	-0,15029	1,68602	1,9473	-0,49238	-0,75009	0,64324	-0,06924	-0,58266	-0,51995	-0,81904	0,16776	-0,42678	-0,45766
Train	-0,63121	-0,19974	1,70614	-0,49238	0,06397	0,20963	1,09358	-0,58266	-0,51995	-0,1077	-0,91069	-1,6812	-0,45766
Train	-0,63121	0,01163	1,76225	-0,49238	0,24624	0,43586	1,07847	0,52001	-0,51995	-0,12862	1,26018	-1,27215	-0,45766
Train	-0,39075	-3,95467	-0,25368	-0,49238	-1,31797	-0,86497	0,21013	-0,58266	2,07305	1,10577	0,13703	1,66757	-0,45766
Train	-0,63121	-3,41174	1,20757	-0,49238	0,07632	0,53955	-0,5978	-0,58266	-0,51995	-0,92365	-0,49719	-1,9539	-0,45766
Train	-0,63121	0,39293	-0,51699	-0,49238	-1,40075	-1,10063	-2,32693	-0,58266	-0,51995	-1,11195	-2,61498	-2,79382	-0,45766
Train	-0,63121	0,39293	-0,18182	2,67805	-0,99046	0,00696	-1,35288	-0,58266	-0,51995	-0,69351	-1,35213	-1,2067	-0,45766
Train	-0,63121	0,39293	0,42059	-0,49238	-0,48503	0,26619	-0,27312	-0,58266	-0,51995	-0,90273	-0,73188	1,25852	-0,45766
Train	1,29248	0,38049	0,99397	-0,49238	-0,23449	0,29918	0,88593	-0,03133	-0,51995	-0,48429	-1,01965	0,32042	-0,45766
Train	-0,15029	1,54925	0,30149	-0,49238	-0,99772	1,07685	0,5688	-0,58266	-0,51995	0,7501	0,50303	1,82028	-0,45766

Train	-0,63121	2,58538	2,38337	1,62124	1,40452	0,51127	1,77315	-0,58266	-0,51995	-0,27508	2,79684	0,30406	-0,45766
Train	-0,63121	0,39293	2,16337	-0,49238	-1,56197	1,50103	-0,56382	-0,58266	-0,51995	-0,00309	1,82735	-0,57404	-0,45766
Train	-0,63121	0,61259	-1,25525	0,56443	0,11335	-0,93567	-0,88095	-0,58266	-0,51995	0,7501	-0,33514	0,89855	1,55606
Train	-0,63121	0,4178	-1,01507	0,56443	0,23971	-0,38894	-0,82055	-0,58266	0,6048	1,69158	-0,47484	-0,28497	1,55606
Train	-0,63121	0,46339	-0,93337	2,67805	-1,16982	-0,54919	-0,35617	-0,58266	-0,51995	1,21038	-0,70673	-0,57404	-0,45766
Train	-0,63121	0,31418	-0,29059	1,62124	-0,91275	0,6668	0,3536	-0,58266	-0,51995	1,29407	-1,10347	0,7622	-0,45766
Train	1,53294	1,05605	0,61303	-0,49238	0,00442	-0,14386	0,93123	-0,58266	-0,51995	0,43627	1,81897	-0,21407	-0,45766
Train	-0,39075	0,41365	-0,79999	0,56443	1,06466	-4,53651	-0,21271	-0,58266	-0,51995	-2,72293	-0,65924	0,10226	-0,45766
Train	-0,63121	0,39293	-0,54554	1,62124	-1,85317	-2,88691	-2,42509	-0,58266	-0,51995	0,58272	1,02549	-1,31578	-0,45766
Train	-0,63121	0,39293	-1,34482	1,62124	-1,72681	-0,42665	-2,55723	-0,58266	0,48707	0,98024	0,02248	-0,90128	-0,45766
Train	1,7734	0,94829	-0,32406	-0,49238	-0,71233	0,70922	0,74246	-0,58266	-0,51995	0,22705	0,62317	0,47859	-0,45766
Train	-0,39075	-0,80484	-0,06961	-0,49238	-0,36375	0,29918	0,3234	-0,58266	1,94915	-0,50522	-0,51954	-0,70493	-0,45766
Train	-0,63121	0,23958	0,14842	-0,49238	-0,50464	0,90717	0,65563	-0,58266	0,60568	0,01783	-1,57564	-0,045	-0,45766
Train	0,33063	-0,19559	0,20699	-0,49238	-0,31655	0,75164	0,65941	-0,03133	0,62565	0,05967	-0,34911	-0,39405	-0,45766
Train	-0,63121	-0,7178	0,00422	-0,49238	-0,16623	0,63381	0,54614	0,52001	2,03753	-0,40061	1,3915	0,39133	-0,45766
Train	-0,63121	0,27688	-1,44621	-0,49238	-0,55765	-0,12972	-0,43923	-0,58266	3,0419	0,62457	1,01711	1,83119	-0,45766
Train	-0,63121	-1,27731	-1,43637	-0,49238	-0,67819	-0,158	-0,43923	-0,58266	3,1338	0,33166	1,21548	1,53667	-0,45766
Train	0,09017	-0,14586	-0,65381	-0,49238	-0,53949	-1,07706	0,3838	-0,30699	0,66	0,64549	-0,74306	0,23316	-0,45766
Train	-0,63121	0,6789	0,049	-0,49238	0,24043	-0,80841	1,02562	0,52001	1,91304	0,47811	1,08137	-0,52495	-0,45766
Test	-0,63121	-0,12928	-0,48303	-0,49238	-0,11903	0,5584	0,03647	-0,58266	1,70694	1,81711	-0,19265	-0,19771	-0,45766
Test	-0,15029	-1,02864	0,07164	-0,49238	1,14744	0,50656	1,08603	0,24434	-0,51995	1,16853	0,10909	0,90946	1,55606
Test	-0,63121	0,39293	-1,23458	-0,49238	-1,42835	-0,18628	-1,12258	-0,58266	-0,51995	0,58272	3,01756	-0,47041	-0,45766
Test	-0,63121	-1,18613	1,99849	-0,49238	0,12134	-0,86497	0,30074	-0,58266	-0,51995	-1,96975	1,95587	0,45132	-0,45766

Set	CATS2D_07_LL	MATS8v	X3v	CATS2D_03_DL	TI2_L	IVDE	SpMax2_Bh(p)	CATS2D_09_LL	SaaN	SpPosA_B(i)	GATS5m	MATS4v	N-074
Test	-0,63121	0,15669	4,4367	-0,49238	0,6137	-1,35985	1,84488	-0,58266	-0,51995	-1,67684	1,60384	-1,16852	-0,45766
Test	-0,63121	0,38878	-0,66464	-0,49238	-0,6201	-0,50206	-0,6431	-0,58266	0,76158	0,83378	0,08394	-0,77038	-0,45766
Test	-0,63121	-0,72195	-1,20406	0,56443	-0,32454	-0,12972	-0,85075	-0,58266	-0,51995	0,0806	1,06461	0,31497	-0,45766
Test	-0,15029	1,37103	-0,8556	1,62124	0,54906	-0,60103	0,73869	-0,58266	-0,51995	0,35258	-1,49462	0,66403	-0,45766
Test	0,33063	-0,12928	-1,56925	1,62124	-0,11467	-0,733	-0,48831	-0,58266	-0,51995	0,16428	1,20151	0,49495	-0,45766
Test	2,01386	0,0738	0,42601	-0,49238	1,07991	0,64795	1,01807	1,89833	-0,51995	-0,31692	0,00851	0,08045	-0,45766
Test	2,49479	-0,55617	0,5082	-0,49238	1,24258	0,66209	1,02562	1,89833	-0,51995	-0,44245	0,03365	0,0859	-0,45766
Test	2,01386	0,33076	0,21733	-0,49238	0,62967	0,70451	1,01807	2,44967	-0,51995	-0,48429	0,08953	0,47314	-0,45766
Test	2,49479	-0,49814	0,23554	-0,49238	0,80977	0,72336	1,02562	2,44967	-0,51995	-0,31692	0,13982	0,48404	-0,45766
Test	2,73525	-0,51058	0,10708	-0,49238	1,58388	0,16721	1,10113	2,174	-0,51995	-0,25415	0,33819	1,19307	-0,45766
Test	-0,63121	-1,0618	-1,22375	-0,49238	-0,4647	-0,12972	-1,24339	-0,58266	-0,51995	-0,02401	0,40804	-0,24679	-0,45766
Test	-0,15029	-0,17487	0,34874	-0,49238	1,24548	0,20963	0,19881	1,07134	-0,51995	-0,71443	0,40245	0,91491	-0,45766
Test	-0,63121	0,39293	-1,17994	-0,49238	-0,50536	0,30389	-0,94136	-0,58266	-0,51995	-0,37968	-1,28228	0,86037	-0,45766
Test	0,81156	0,94415	2,88981	-0,49238	-0,53441	0,44057	-0,11455	-0,58266	-0,51995	-2,03251	2,29673	-0,48677	-0,45766
Test	0,33063	0,81152	2,24113	-0,49238	-0,63462	-1,76047	-0,45811	-0,58266	-0,51995	-2,01159	2,78566	-0,31224	-0,45766
Test	0,81156	-0,23704	5,03715	-0,49238	-0,37828	-2,48158	-0,17496	1,07134	-0,51995	-2,63925	3,1349	0,23316	-0,45766
Test	-0,15029	0,38049	-0,70254	-0,49238	-0,91275	0,6668	-0,30332	-0,58266	-0,51995	1,73342	-0,76541	0,60949	-0,45766
Test	-0,63121	-1,08252	-0,03713	-0,49238	-0,29331	0,8459	0,95766	-0,58266	-0,51995	0,35258	0,32702	-0,83038	-0,45766
Test	0,33063	1,42491	-0,13408	-0,49238	-0,07255	0,29918	0,8444	-0,58266	-0,51995	0,33166	2,25482	-0,74311	-0,45766
Test	1,53294	-0,3448	-0,46236	3,73487	-0,28315	0,75164	0,48196	0,24434	1,83876	0,66641	-1,26831	1,70575	-0,45766
Test	-0,63121	-1,11568	-1,16666	-0,49238	-0,67819	-0,158	-0,43923	-0,58266	3,26209	0,0806	0,92491	1,61848	-0,45766

Table S6.9. Multi linear regression predictions and descriptors.

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	CATS2D_07_LL	MATS4m	GATS8s	GGI8	JGI4	CATS2D_00_DD	CATS2D_03_DA	CATS2D_03_DL	F03[C-N]	Psychotic-50
1	-1,22173	-0,53715	-0,28701	-0,04084	-0,27225	Train	-0,63121	0,36102	-1,33019	-0,77192	-2,95462	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
2	-2,90345	-2,16897	-1,98303	-2,15274	-2,08393	Train	-0,63121	0,33932	-1,33019	-0,77192	2,10828	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
3	-0,90803	-1,58704	-1,69597	-1,6584	-1,68993	Train	-0,63121	0,40009	-1,33019	-0,77192	0,37624	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
4	-0,86782	-1,42556	-2,17403	-1,9833	-1,46617	Train	2,25433	0,60844	-0,54013	2,64878	-0,1567	2,06287	2,85874	3,73487	2,14405	-0,26896
6	-1,16256	0,012005	0,365853	0,788368	0,334216	Train	-0,15029	0,00942	1,09336	-0,34433	-0,55641	0,68762	-0,3665	2,67805	0,13952	-0,26896
8	-3,70781	-4,30962	-4,76083	-4,45193	-4,33631	Train	2,25433	0,43047	0,47901	1,8745	0,30962	-0,68762	-0,3665	-0,49238	-0,61218	-0,26896
9	-3,80902	-4,6079	-5,45729	-5,26124	-5,04932	Train	2,73525	1,92369	-0,11842	4,5267	-0,42317	-0,68762	-0,3665	-0,49238	-0,61218	-0,26896
10	-1,43854	-2,25442	-2,28416	-2,31258	-2,09994	Train	-0,63121	-0,21196	1,90686	-0,44256	0,57609	-0,68762	-0,3665	-0,49238	1,14178	-0,26896
11	-2,86392	-2,49312	-2,47013	-2,46322	-2,4886	Train	-0,63121	-0,3335	1,86391	-0,22876	-0,35655	-0,68762	-0,3665	-0,49238	0,13952	-0,26896
12	-4,93172	-3,93891	-3,83111	-3,76135	-3,56162	Train	-0,63121	-0,07305	0,66905	1,24468	1,24226	0,68762	1,24612	-0,49238	1,14178	-0,26896
13	-3,52931	-2,52907	-2,52349	-2,60676	-2,52421	Train	0,09017	0,8385	1,46432	0,34328	0,70932	-0,68762	-0,3665	-0,49238	0,64065	-0,26896
15	-2,45552	-2,56037	-2,56011	-2,49465	-2,644	Train	-0,15029	-1,43605	0,19397	-0,48878	-1,55566	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
17	-1,41197	-1,86586	-1,88731	-2,00522	-1,86573	Train	-0,63121	-1,32753	-1,03994	-0,77192	-1,42243	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
19	-3,74545	-3,33097	-3,32099	-3,36983	-3,16522	Train	0,81156	-0,38125	1,21311	-0,77192	0,57609	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
20	-2,95553	-2,73185	-2,6777	-2,77353	-2,62121	Train	-0,63121	-1,7182	-0,72496	-0,48878	0,04315	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896

21	-1,16422	-1,86193	-1,82637	-2,03228	-1,85094	Train	-0,63121	0,57806	-0,30454	-0,48878	0,10977	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
22	-4,39933	-3,24576	-3,08105	-3,04557	-2,76089	Train	2,97571	0,23948	-0,35661	0,08326	-1,15596	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
23	-1,84887	-2,33801	-2,32662	-2,33726	-2,27044	Train	0,33063	0,04849	-0,36442	0,03703	-0,35655	-0,68762	-0,3665	-0,49238	-0,36161	-0,26896
24	-2,31142	-2,76057	-2,75475	-2,70698	-2,68221	Train	-0,63121	-0,61131	1,02698	-0,19987	-0,09008	-0,68762	-0,3665	-0,49238	-0,61218	-0,26896
25	-0,87436	-1,32242	-1,39807	-1,47664	-1,39708	Train	0,33063	0,6779	-1,16359	-0,77192	0,17638	-0,68762	-0,3665	-0,49238	0,39009	-0,26896
26	-1,22582	-1,52872	-1,59173	-1,66619	-1,62043	Train	0,09017	0,29591	-0,66638	-0,77192	0,10977	-0,68762	-0,3665	-0,49238	0,39009	-0,26896
27	-3,1815	-3,79642	-3,81365	-3,83113	-3,99884	Train	-0,63121	-2,63409	-1,11153	-0,77192	0,57609	0,68762	-0,3665	-0,49238	-0,86275	-0,26896
29	-1,69658	-1,12279	-0,89401	-1,14408	-1,02891	Train	0,09017	0,55202	-0,47245	0,67841	-0,22332	0,68762	1,24612	1,62124	0,64065	-0,26896
31	-1,26121	-2,06114	-2,07142	-2,03133	-2,15485	Train	-0,63121	-0,26839	-0,74969	-0,5119	0,70932	0,68762	1,24612	0,56443	0,13952	-0,26896
32	-0,73208	-2,0339	-2,28914	-2,38864	-2,18391	Train	-0,63121	0,53899	0,13539	0,93843	-0,28994	2,06287	1,24612	0,56443	2,64518	-0,26896
34	-1,23308	-2,05887	-2,18122	-2,25867	-2,20509	Train	-0,39075	0,68658	0,73933	-0,36744	0,17638	0,68762	-0,3665	0,56443	-0,86275	-0,26896
35	-1,18498	-0,73651	-0,53725	-0,04047	-0,56595	Train	-0,63121	-0,40729	-1,33019	-0,77192	1,7752	2,06287	-0,3665	2,67805	-0,36161	-0,26896
36	-1,42583	-2,72523	-3,13945	-2,98489	-2,99541	Train	1,29248	0,84285	-0,83689	0,73042	-0,48979	-0,68762	-0,3665	-0,49238	0,89122	3,67578
37	-1,56658	-2,01266	-2,24363	-2,09507	-1,5651	Train	-0,15029	0,17437	0,96841	0,48773	-0,95611	2,06287	4,47136	2,67805	0,39009	-0,26896
38	-0,34577	-0,6994	-0,5913	-0,69281	-0,50112	Train	-0,15029	1,10329	-0,99959	-0,53501	-0,95611	0,68762	-0,3665	0,56443	-0,11105	-0,26896

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	CATS2D_07_LL	MATS4m	GATS8s	GGI8	JGI4	CATS2D_00_DD	CATS2D_03_DA	CATS2D_03_DL	F03[C-N]	Psychotic-50
40	-3,84777	-3,33158	-3,21969	-3,26899	-3,37297	Train	1,29248	0,24816	-0,1887	0,1526	-0,75626	0,68762	-0,3665	-0,49238	-0,36161	-0,26896
42	-4,45683	-4,47686	-4,96877	-4,45042	-4,349	Train	2,49479	0,05283	-0,3488	0,73042	-0,22332	-0,68762	-0,3665	-0,49238	-0,86275	3,67578
44	-3,50524	-3,31224	-3,39964	-3,2721	-3,23164	Train	-0,63121	-1,61836	2,65787	-0,77192	-0,82287	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
45	-3,60767	-3,20448	-2,95698	-3,45484	-2,94609	Train	-0,15029	-0,12948	-0,33188	4,128	0,04315	-0,68762	-0,3665	-0,49238	2,14405	-0,26896
46	-1,94825	-1,279	-1,02534	-0,87269	-1,12516	Train	-0,63121	0,07019	0,01174	0,93843	-0,48979	3,43812	-0,3665	1,62124	3,14631	-0,26896
47	-4,04631	-3,14248	-3,16884	-3,28278	-2,92096	Train	2,01386	0,88191	-0,61432	0,44729	0,30962	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
48	-1,9488	-3,22937	-3,26787	-3,36903	-3,00104	Train	2,01386	0,82982	-0,40476	0,44729	0,30962	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
51	-4,40486	-4,30668	-3,87225	-4,16655	-4,33636	Train	2,49479	0,74735	-0,03251	0,58596	0,10977	-0,68762	-0,3665	-0,49238	-0,86275	3,67578
52	-3,01947	-2,74568	-2,75917	-2,65899	-2,73082	Train	-0,15029	-1,28412	-0,69111	-0,41945	-0,02347	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
53	-3,31444	-3,37479	-3,44109	-3,34418	-3,34668	Train	-0,63121	-1,67913	1,44219	-0,34433	-0,1567	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
54	-1,06576	-0,82383	-1,00996	-0,68211	-0,76199	Train	-0,63121	1,11197	-0,53883	0,26238	0,84256	-0,68762	-0,3665	-0,49238	2,39462	-0,26896
55	-1,52013	-1,20754	-0,98797	-1,43126	-1,24858	Train	-0,39075	0,35234	0,44778	0,79976	-0,42317	2,06287	-0,3665	1,62124	1,64292	-0,26896
56	-3,32973	-3,00751	-2,95517	-3,01283	-2,89452	Train	0,81156	0,93834	0,74454	1,27357	-0,09008	-0,68762	-0,3665	-0,49238	-0,11105	-0,26896
57	-2,90058	-1,98793	-2,08469	-1,81781	-1,88766	Train	-0,63121	1,15972	0,15362	0,80553	0,30962	-0,68762	-0,3665	-0,49238	0,13952	-0,26896
58	-3,17408	-3,1342	-3,17705	-2,88257	-3,34984	Train	-0,39075	0,68658	-0,00257	1,51625	0,84256	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896

59	-3,72982	-3,57987	-3,20268	-3,38242	-3,77439	Train	0,81156	1,21181	0,14841	0,85176	-0,02347	-0,68762	-0,3665	-0,49238	-0,86275	3,67578
61	-4,21094	-3,56759	-3,4632	-3,61181	-3,27328	Train	2,49479	0,85153	-0,1887	0,89799	-0,02347	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
62	-2,0577	-2,00835	-2,03664	-2,00077	-2,00389	Train	-0,63121	-0,57224	-0,86162	0,51085	-0,02347	-0,68762	-0,3665	-0,49238	0,64065	-0,26896
63	-3,00552	-3,47322	-3,52199	-3,67487	-3,70641	Train	-0,63121	-1,26242	-1,22216	-0,77192	0,57609	0,68762	1,24612	-0,49238	-0,61218	-0,26896
64	-3,57525	-2,51032	-2,40601	-2,30616	-2,26637	Train	-0,63121	1,03818	0,11847	-0,77192	0,10977	0,68762	1,24612	-0,49238	-0,11105	-0,26896
65	-2,82409	-2,72282	-2,77542	-2,63334	-2,59459	Train	-0,63121	0,23514	0,43867	-0,77192	-0,95611	0,68762	1,24612	-0,49238	-0,36161	-0,26896
66	-2,88446	-3,15452	-3,08962	-3,12598	-3,30083	Train	-0,63121	-1,47511	-0,64296	-0,62746	-0,1567	0,68762	-0,3665	-0,49238	-0,61218	-0,26896
68	-2,71703	-3,42947	-3,57763	-3,45274	-3,64553	Train	-0,63121	0,16569	2,09559	-0,77192	0,57609	0,68762	-0,3665	-0,49238	-0,61218	-0,26896
69	-3,15724	-2,75282	-2,65538	-2,74419	-2,87998	Train	-0,63121	-0,7068	-0,44902	-0,65635	-0,48979	0,68762	-0,3665	-0,49238	-0,61218	-0,26896
70	-1,10453	-1,58267	-1,81705	-1,84469	-1,26571	Train	-0,63121	-0,43334	0,40743	-0,77192	1,30888	-0,68762	-0,3665	-0,49238	2,14405	-0,26896
71	-3,33423	-3,13753	-3,17367	-3,41846	-3,00896	Train	-0,63121	-0,15553	0,48813	0,67841	0,243	-0,68762	-0,3665	-0,49238	0,64065	3,67578
73	-0,69472	-1,67671	-1,65049	-1,91944	-1,72462	Train	-0,63121	1,45489	0,92936	-0,77192	0,90917	-0,68762	-0,3665	-0,49238	-0,11105	-0,26896
74	-4,18105	-3,30644	-2,95116	-2,99583	-3,02376	Train	0,5711	1,12499	0,25254	1,31402	0,10977	-0,68762	-0,3665	-0,49238	0,39009	3,67578
75	-1,5148	-1,79692	-1,74799	-1,5777	-1,87017	Train	0,5711	0,68224	0,20828	0,03703	-0,62302	0,68762	-0,3665	0,56443	0,13952	-0,26896
76	-3,04362	-1,97061	-1,89808	-1,88064	-1,96988	Train	-0,15029	-0,75021	-0,71975	-0,77192	-0,55641	-0,68762	-0,3665	-0,49238	-0,36161	-0,26896

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	CATS2D_07_LL	MATS4m	GATS8s	GGI8	JGI4	CATS2D_00_DD	CATS2D_03_DA	CATS2D_03_DL	F03[C-N]	Psychotic-50
77	-2,87782	-3,15093	-3,28654	-3,08274	-3,23483	Train	-0,63121	-2,14793	1,74025	-0,00919	0,243	-0,68762	-0,3665	-0,49238	1,14178	-0,26896
78	-1,9212	-3,15276	-3,22334	-3,04934	-3,22032	Train	-0,63121	-1,73122	0,66384	0,51085	-0,75626	-0,68762	-0,3665	-0,49238	-0,36161	-0,26896
79	-0,54658	-1,73939	-1,91953	-1,93135	-1,87127	Train	-0,39075	1,33769	1,78581	-0,77192	0,77594	-0,68762	-0,3665	-0,49238	0,64065	-0,26896
80	-3,23383	-2,75907	-2,66483	-2,73354	-2,65232	Train	-0,63121	-2,22172	-0,6013	-0,77192	-0,28994	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
81	-2,0125	-2,12088	-2,16863	-1,94728	-2,02456	Train	-0,63121	-2,00903	-1,33019	-0,77192	-2,95462	0,68762	-0,3665	-0,49238	-0,36161	-0,26896
82	-0,63141	-0,54806	-0,25759	-0,41996	-0,15227	Train	-0,63121	-0,12514	-1,33019	-0,77192	-0,68964	0,68762	4,47136	2,67805	-0,86275	-0,26896
83	-2,896	-1,89301	-1,76965	-1,90798	-1,80734	Train	-0,63121	0,8168	-1,33019	-0,77192	1,90843	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
87	-3,59274	-2,91211	-2,83551	-2,90694	-2,90516	Train	1,29248	0,08755	-0,83038	-0,17676	0,44285	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
89	-1,31697	-1,56416	-1,7121	-1,46751	-1,59003	Train	-0,15029	0,90362	-0,7601	-0,77192	1,44211	-0,68762	-0,3665	-0,49238	0,39009	-0,26896
90	-1,21865	-1,22237	-1,53674	-1,46664	-1,4345	Train	-0,63121	-0,6851	-0,72235	1,12911	-1,08934	0,68762	-0,3665	1,62124	-0,36161	-0,26896
91	-1,95297	-1,68753	-1,78786	-1,47101	-1,79912	Train	-0,63121	0,26987	-1,33019	-0,77192	0,50947	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
92	-3,63348	-2,94466	-2,68626	-2,64858	-2,53094	Train	-0,63121	0,79076	1,47994	-0,41367	-0,28994	2,06287	2,85874	0,56443	0,64065	-0,26896
93	-3,82889	-3,52183	-3,40721	-3,52853	-3,65892	Train	-0,63121	-0,36388	1,05562	-0,12476	-0,09008	2,06287	2,85874	0,56443	0,64065	-0,26896
94	-0,87168	-0,82311	-0,73024	-0,16567	-0,81329	Train	-0,63121	0,2308	2,83359	-0,77192	0,57609	0,68762	1,24612	2,67805	0,13952	-0,26896
95	-0,92853	-0,40574	-0,22774	-0,44082	-0,3122	Train	-0,63121	0,8168	1,2079	0,08326	0,57609	0,68762	-0,3665	1,62124	1,64292	-0,26896

99	-3,12009	-3,36566	-3,44741	-3,24934	-3,45186	Train	1,53294	-1,3666	-0,75229	0,3895	-0,68964	-0,68762	-0,3665	-0,49238	-0,36161	-0,26896
101	-0,52396	-1,18357	-1,32294	-1,41121	-1,4041	Train	-0,39075	0,26553	0,22781	-0,62746	-2,28845	0,68762	-0,3665	0,56443	-0,86275	-0,26896
102	-0,74011	-1,02309	-0,84771	-1,42832	-1,4591	Train	-0,63121	0,18305	-1,33019	-0,77192	2,10828	0,68762	-0,3665	1,62124	-0,86275	-0,26896
103	-0,71565	-1,74169	-2,12583	-2,15044	-2,30699	Train	-0,63121	-0,45938	-1,33019	-0,77192	3,77371	0,68762	-0,3665	1,62124	-0,61218	-0,26896
104	-2,81404	-3,2344	-3,38696	-3,43807	-3,31113	Train	1,7734	0,33932	0,51416	-0,48878	0,37624	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
105	-3,46548	-2,42621	-2,49974	-2,26725	-2,41286	Train	-0,39075	-1,34923	0,7094	-0,48878	-0,48979	-0,68762	-0,3665	-0,49238	0,39009	-0,26896
106	-3,33027	-2,8409	-2,97956	-2,96416	-2,75763	Train	-0,63121	-0,76757	0,2239	-0,77192	1,50873	-0,68762	-0,3665	-0,49238	-0,61218	-0,26896
107	-3,77188	-3,17872	-3,08918	-3,11817	-3,11775	Train	0,33063	-0,91082	0,04428	-0,25188	0,6427	-0,68762	-0,3665	-0,49238	-0,61218	-0,26896
108	-3,14195	-2,11175	-2,13688	-2,12823	-2,10425	Train	-0,63121	-0,57224	0,45819	0,03703	-0,28994	-0,68762	-0,3665	-0,49238	0,64065	-0,26896
109	-1,15592	-0,96378	-0,83086	-1,02495	-0,93319	Train	-0,63121	2,05391	-1,17531	-0,77192	-0,55641	2,06287	-0,3665	-0,49238	1,89348	-0,26896
110	-0,79619	-0,98119	-1,18678	-0,94863	-0,96321	Train	-0,63121	1,78479	-0,78223	-0,77192	-0,62302	2,06287	-0,3665	-0,49238	2,39462	-0,26896
112	-0,8803	-1,94175	-1,83978	-2,0013	-1,80261	Train	0,09017	0,90796	0,42695	0,00814	-0,82287	-0,68762	-0,3665	-0,49238	-0,36161	-0,26896
113	-0,71893	-1,82144	-1,87454	-1,75127	-1,87413	Train	-0,63121	-1,10181	-0,05724	-0,2461	-1,35581	-0,68762	-0,3665	-0,49238	0,39009	-0,26896
5	-0,77006	-2,50647				Test	-0,63121	0,41745	1,0387	1,24468	-0,55641	-0,68762	-0,3665	-0,49238	0,13952	-0,26896
7	-3,28458	-2,86631				Test	-0,15029	1,27258	0,18355	1,51625	0,97579	-0,68762	-0,3665	-0,49238	-0,36161	-0,26896

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	CATS2D_07_LL	MATS4m	GATS8s	GGI8	JGI4	CATS2D_00_DD	CATS2D_03_DA	CATS2D_03_DL	F03[C-N]	Psychotic-50
14	-0,42347	-1,16078				Test	-0,63121	1,78913	-1,33019	-0,77192	-0,68964	0,68762	-0,3665	-0,49238	-0,36161	-0,26896
16	-1,84668	-1,17217				Test	-0,63121	0,15267	-0,95013	-0,77192	-1,62228	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
18	-1,27151	-2,73243				Test	-0,63121	-1,79199	0,77708	-0,19987	-1,88875	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
28	-2,44142	-2,68169				Test	-0,63121	-1,54457	0,28378	-0,59279	-0,1567	0,68762	-0,3665	-0,49238	1,39235	-0,26896
30	-1,15394	-1,55496				Test	-0,63121	0,16569	-0,03641	-0,48878	0,6427	0,68762	-0,3665	0,56443	0,39009	-0,26896
33	-1,2584	-0,88938				Test	-0,15029	-0,238	-0,51149	-0,19987	0,50947	0,68762	-0,3665	1,62124	0,39009	-0,26896
39	-1,04896	0,244756				Test	0,33063	1,18142	-1,12064	-0,53501	-1,22258	0,68762	-0,3665	1,62124	-0,11105	-0,26896
41	-4,44259	-4,24181				Test	2,01386	0,06151	-0,56616	0,73042	-0,1567	-0,68762	-0,3665	-0,49238	-0,86275	3,67578
43	-4,45917	-4,47209				Test	2,49479	0,08321	-0,32016	0,73042	-0,22332	-0,68762	-0,3665	-0,49238	-0,86275	3,67578
49	-4,38602	-3,27384				Test	2,01386	0,76037	-0,42429	0,58596	0,17638	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
50	-4,4022	-4,28501				Test	2,49479	0,75603	-0,09108	0,58596	0,10977	-0,68762	-0,3665	-0,49238	-0,86275	3,67578
60	-3,71823	-3,48617				Test	2,73525	1,20747	-0,24337	0,89799	-0,02347	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
67	-3,23553	-2,51153				Test	-0,63121	-0,32916	-0,23165	-0,48878	-0,22332	0,68762	-0,3665	-0,49238	0,13952	-0,26896
72	-3,32475	-2,11269				Test	-0,15029	0,63015	1,47863	0,10637	0,04315	-0,68762	-0,3665	-0,49238	0,89122	-0,26896

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	CATS2D_07_LL	MATS4m	GATS8s	GGI8	JGI4	CATS2D_00_DD	CATS2D_03_DA	CATS2D_03_DL	F03[C-N]	Psychotic-50
84	-2,64335	-1,83709				Test	-0,63121	0,31761	-1,33019	-0,77192	1,04241	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
85	-2,39808	-2,51512				Test	0,81156	-1,18863	-1,1727	-0,77192	-0,82287	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
86	-1,63884	-2,51498				Test	0,33063	-1,76595	-1,19223	-0,77192	-1,02273	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
88	-0,85164	-2,3201				Test	0,81156	-0,54185	-1,15058	-0,34433	-1,15596	-0,68762	-0,3665	-0,49238	-0,86275	-0,26896
96	-0,56802	-2,42499				Test	-0,15029	0,45652	1,72464	0,08326	0,57609	-0,68762	-0,3665	-0,49238	0,89122	-0,26896
97	-3,04885	-2,4239				Test	-0,63121	-1,55759	-0,66118	-0,19987	-0,22332	-0,68762	-0,3665	-0,49238	-0,11105	-0,26896
98	-3,2209	-3,06506				Test	0,33063	-2,11755	-0,9241	0,03703	-0,55641	-0,68762	-0,3665	-0,49238	-0,36161	-0,26896
100	-1,21095	0,294375				Test	1,53294	1,23785	-0,11191	0,34906	0,44285	0,68762	1,24612	3,73487	0,39009	-0,26896
111	-0,39175	-1,03898				Test	-0,63121	1,75006	-0,64296	-0,77192	-0,62302	2,06287	-0,3665	-0,49238	2,39462	-0,26896

Table S6.10. Plartial least squares regression prediction and descriptors.

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	CATS2D_07_LL	F06[N-S]	MATS4m	GATS4e	GATS8s	JGI8	JGT	CATS2D_03_DL	F02[N-S]	F04[O-O]
1	-1,22173	-0,84201	-0,59692	-0,73958	-1,01084	Train	-0,63121	-0,26896	0,36102	0,49912	-1,33019	-1,22528	-2,97967	-0,49238	-0,36927	-0,49103
2	-2,90345	-1,416	-1,42507	-1,44454	-1,42027	Train	-0,63121	-0,26896	0,33932	-0,31269	-1,33019	-1,22528	-0,22825	-0,49238	-0,36927	-0,49103
3	-0,90803	-1,91024	-2,05503	-2,0997	-1,96886	Train	-0,63121	-0,26896	0,40009	-0,62742	-1,33019	-1,22528	1,93982	-0,49238	-0,36927	-0,49103
4	-0,86782	-1,50589	-1,60429	-1,68175	-1,5333	Train	2,25433	-0,26896	0,60844	-0,75231	-0,54013	0,73791	-0,09214	3,73487	-0,36927	-0,49103
6	-1,16256	-0,97049	-1,02372	-1,01153	-0,80214	Train	-0,15029	-0,26896	0,00942	-0,31269	1,09336	0,13385	-0,98659	2,67805	-0,36927	-0,49103
8	-3,70781	-3,78908	-3,84978	-3,62047	-3,74854	Train	2,25433	-0,26896	0,43047	-0,08288	0,47901	0,73791	0,19953	-0,49238	-0,36927	-0,49103
9	-3,80902	-3,69356	-3,70361	-3,76975	-3,52857	Train	2,73525	-0,26896	1,92369	-1,65654	-0,11842	1,644	0,48148	-0,49238	-0,36927	-0,49103
10	-1,43854	-2,47055	-2,50315	-2,40821	-2,44778	Train	-0,63121	-0,26896	-0,21196	-0,80477	1,90686	0,43588	-0,56853	-0,49238	-0,36927	-0,49103
11	-2,86392	-2,87971	-2,98029	-2,77131	-2,86326	Train	-0,63121	-0,26896	-0,3335	-1,0046	1,86391	0,13385	1,46343	-0,49238	-0,36927	-0,49103
12	-4,93172	-2,87667	-2,77992	-2,6139	-2,70076	Train	-0,63121	-0,26896	-0,07305	-0,4176	0,66905	0,73791	1,32732	-0,49238	-0,36927	-0,49103
13	-3,52931	-2,93079	-2,87717	-2,81303	-2,84097	Train	0,09017	-0,26896	0,8385	-0,93466	1,46432	1,49299	-0,1602	-0,49238	-0,36927	-0,49103
15	-2,45552	-2,91384	-2,93921	-2,97806	-2,84343	Train	-0,15029	-0,26896	-1,43605	1,55072	0,19397	0,5869	-1,19076	-0,49238	-0,36927	-0,49103
17	-1,41197	-1,90386	-1,97614	-1,97336	-1,89062	Train	-0,63121	-0,26896	-1,32753	1,39585	-1,03994	-1,22528	-0,95743	-0,49238	-0,36927	-0,49103
19	-3,74545	-3,2331	-3,34335	-3,10278	-3,10813	Train	0,81156	-0,26896	-0,38125	0,54658	1,21311	-1,22528	1,42454	-0,49238	-0,36927	-0,49103
20	-2,95553	-2,81599	-2,8801	-2,85036	-2,73057	Train	-0,63121	-0,26896	-1,7182	1,11609	-0,72496	0,28487	0,35509	-0,49238	-0,36927	-0,49103

21	-1,16422	-1,58094	-1,55625	-1,58882	-1,50744	Train	-0,63121	-0,26896	0,57806	-1,12699	-0,30454	-0,77224	-0,14075	-0,49238	-0,36927	-0,49103
22	-4,39933	-3,23231	-3,24238	-2,78778	-3,02467	Train	2,97571	-0,26896	0,23948	0,36673	-0,35661	-0,47021	-1,54077	-0,49238	-0,36927	-0,49103
23	-1,84887	-2,82122	-2,77251	-2,89633	-2,9063	Train	0,33063	-0,26896	0,04849	-0,46506	-0,36442	0,5869	0,50092	-0,49238	-0,36927	-0,49103
24	-2,31142	-3,16544	-3,31928	-3,21557	-3,20432	Train	-0,63121	-0,26896	-0,61131	1,00868	1,02698	0,28487	0,74398	-0,49238	-0,36927	0,74355
25	-0,87436	-1,78509	-1,98036	-1,97807	-1,88872	Train	0,33063	-0,26896	0,6779	1,52325	-1,16359	-1,22528	-0,18936	-0,49238	1,66174	-0,49103
26	-1,22582	-1,61491	-1,74853	-1,71752	-1,66149	Train	0,09017	-0,26896	0,29591	-0,91467	-0,66638	-1,22528	-1,4241	-0,49238	-0,36927	0,74355
27	-3,1815	-2,51006	-2,22729	-2,72756	-2,68703	Train	-0,63121	-0,26896	-2,63409	1,29594	-1,11153	-1,22528	0,52037	-0,49238	-0,36927	-0,49103
29	-1,69658	-1,64075	-1,6646	-1,62588	-1,60275	Train	0,09017	-0,26896	0,55202	-0,82475	-0,47245	0,88893	0,28703	1,62124	-0,36927	-0,49103
31	-1,26121	-1,80182	-1,76106	-1,77258	-1,73844	Train	-0,63121	-0,26896	-0,26839	-0,08288	-0,74969	0,43588	-0,02408	0,56443	-0,36927	-0,49103
32	-0,73208	-1,79752	-2,01089	-2,08439	-2,01417	Train	-0,63121	-0,26896	0,53899	-0,93965	0,13539	0,43588	-0,03381	0,56443	1,66174	3,21271
34	-1,23308	-1,75905	-1,75314	-1,71079	-1,79939	Train	-0,39075	-0,26896	0,68658	0,02452	0,73933	0,13385	-0,61715	0,56443	-0,36927	-0,49103
35	-1,18498	-0,21125	0,124766	-0,08165	-0,1339	Train	-0,63121	-0,26896	-0,40729	-0,20778	-1,33019	-1,22528	-0,09214	2,67805	-0,36927	-0,49103
36	-1,42583	-2,59894	-2,64055	-2,70744	-2,72209	Train	1,29248	-0,26896	0,84285	0,12194	-0,83689	0,13385	-0,90882	-0,49238	-0,36927	-0,49103
37	-1,56658	-0,85532	-0,91172	-0,66631	-0,68318	Train	-0,15029	-0,26896	0,17437	-0,59744	0,96841	0,43588	-1,40465	2,67805	-0,36927	-0,49103
38	-0,34577	-1,42024	-1,41337	-1,41843	-1,46362	Train	-0,15029	-0,26896	1,10329	-0,81226	-0,99959	0,28487	-0,43242	0,56443	-0,36927	-0,49103

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	CATS2D_07_LL	F06[N-S]	MATS4m	GATS4e	GATS8s	JGI8	JGT	CATS2D_03_DL	F02[N-S]	F04[O-O]
40	-3,84777	-3,08208	-3,00664	-2,8364	-2,9875	Train	1,29248	-0,26896	0,24816	0,13193	-0,1887	0,13385	0,13147	-0,49238	1,66174	1,97813
42	-4,45683	-4,05078	-4,16068	-4,03644	-3,96126	Train	2,49479	-0,26896	0,05283	0,19688	-0,3488	0,43588	0,63703	-0,49238	-0,36927	0,74355
44	-3,50524	-3,34585	-3,21163	-3,02804	-3,2143	Train	-0,63121	-0,26896	-1,61836	0,90627	2,65787	-1,22528	0,86065	-0,49238	-0,36927	1,97813
45	-3,60767	-3,23396	-3,2277	-3,34163	-3,26239	Train	-0,15029	-0,26896	-0,12948	-0,75231	-0,33188	1,34197	0,08286	-0,49238	-0,36927	3,21271
46	-1,94825	-1,60568	-1,4524	-1,53426	-1,65575	Train	-0,63121	-0,26896	0,07019	-0,4101	0,01174	0,43588	-0,31575	1,62124	-0,36927	1,97813
47	-4,04631	-3,73755	-3,92242	-3,90729	-3,69862	Train	2,01386	-0,26896	0,88191	-0,23525	-0,61432	0,5869	1,16204	-0,49238	-0,36927	0,74355
48	-1,9488	-3,78463	-3,97964	-3,9511	-3,74475	Train	2,01386	-0,26896	0,82982	-0,24025	-0,40476	0,5869	1,16204	-0,49238	-0,36927	0,74355
51	-4,40486	-3,98949	-3,79181	-3,75658	-4,06469	Train	2,49479	-0,26896	0,74735	-0,17281	-0,03251	0,5869	0,76342	-0,49238	-0,36927	0,74355
52	-3,01947	-2,68997	-2,71338	-2,63984	-2,68616	Train	-0,15029	-0,26896	-1,28412	0,60903	-0,69111	0,5869	-0,6852	-0,49238	-0,36927	-0,49103
53	-3,31444	-3,03143	-3,07865	-2,72368	-3,0778	Train	-0,63121	-0,26896	-1,67913	1,05864	1,44219	-0,01716	0,12175	-0,49238	-0,36927	-0,49103
54	-1,06576	-1,51524	-1,36801	-1,60753	-1,87253	Train	-0,63121	-0,26896	1,11197	0,12944	-0,53883	0,43588	0,53981	-0,49238	3,69274	-0,49103
55	-1,52013	-1,68201	-1,57301	-1,85784	-1,51451	Train	-0,39075	-0,26896	0,35234	-0,69736	0,44778	0,28487	-0,14075	1,62124	-0,36927	1,97813
56	-3,32973	-3,70842	-3,82262	-3,60541	-3,77723	Train	0,81156	-0,26896	0,93834	-0,36015	0,74454	1,03994	0,43286	-0,49238	-0,36927	3,21271
57	-2,90058	-2,92943	-2,87814	-3,00394	-2,98277	Train	-0,63121	-0,26896	1,15972	-0,57996	0,15362	0,73791	0,99676	-0,49238	-0,36927	3,21271
58	-3,17408	-2,83715	-2,79731	-2,79604	-2,9539	Train	-0,39075	-0,26896	0,68658	-0,78978	-0,00257	1,03994	1,67732	-0,49238	-0,36927	-0,49103

59	-3,72982	-3,04201	-2,96225	-2,97973	-3,05711	Train	0,81156	-0,26896	1,21181	-0,42509	0,14841	1,03994	-0,18936	-0,49238	-0,36927	0,74355
61	-4,21094	-3,53028	-3,31953	-3,35984	-3,44144	Train	2,49479	-0,26896	0,85153	-0,09288	-0,1887	0,73791	-0,4227	-0,49238	-0,36927	-0,49103
62	-2,0577	-2,15685	-2,36969	-2,41594	-2,62292	Train	-0,63121	-0,26896	-0,57224	1,96787	-0,86162	1,03994	-0,02408	-0,49238	3,69274	-0,49103
63	-3,00552	-1,83853	-1,95399	-1,5136	-1,79678	Train	-0,63121	-0,26896	-1,26242	2,7522	-1,22216	-1,22528	-0,58798	-0,49238	1,66174	-0,49103
64	-3,57525	-2,65753	-2,45019	-2,52916	-2,62789	Train	-0,63121	3,67578	1,03818	0,12444	0,11847	-1,22528	0,52037	-0,49238	-0,36927	-0,49103
65	-2,82409	-2,86896	-3,0211	-2,84777	-2,83597	Train	-0,63121	3,67578	0,23514	0,12444	0,43867	-1,22528	0,51064	-0,49238	-0,36927	-0,49103
66	-2,88446	-2,4219	-2,32875	-2,44498	-2,34747	Train	-0,63121	-0,26896	-1,47511	0,24434	-0,64296	-0,01716	-0,0727	-0,49238	-0,36927	-0,49103
68	-2,71703	-2,36383	-2,41036	-2,26427	-2,4906	Train	-0,63121	-0,26896	0,16569	-0,12285	2,09559	-1,22528	0,14119	-0,49238	-0,36927	0,74355
69	-3,15724	-2,62951	-2,64992	-2,41458	-2,55517	Train	-0,63121	-0,26896	-0,7068	1,21851	-0,44902	1,79502	-1,75466	-0,49238	-0,36927	-0,49103
70	-1,10453	-2,356	-2,5101	-2,29863	-2,47966	Train	-0,63121	-0,26896	-0,43334	1,68061	0,40743	-1,22528	0,25786	-0,49238	-0,36927	-0,49103
71	-3,33423	-2,3376	-2,29555	-2,18932	-2,34091	Train	-0,63121	-0,26896	-0,15553	-0,33017	0,48813	0,13385	0,00508	-0,49238	-0,36927	-0,49103
73	-0,69472	-1,58052	-1,57741	-1,60477	-1,51499	Train	-0,63121	-0,26896	1,45489	-1,19444	0,92936	-1,22528	0,26758	-0,49238	-0,36927	-0,49103
74	-4,18105	-3,20842	-3,1021	-3,16935	-3,21056	Train	0,5711	-0,26896	1,12499	-1,37178	0,25254	1,94603	0,88009	-0,49238	-0,36927	-0,49103
75	-1,5148	-2,15101	-2,08759	-2,1856	-2,14167	Train	0,5711	-0,26896	0,68224	-0,70236	0,20828	0,43588	-0,90882	0,56443	-0,36927	0,74355
76	-3,04362	-1,89951	-1,44735	-1,34952	-1,89302	Train	-0,15029	-0,26896	-0,75021	1,12359	-0,71975	-1,22528	0,04397	-0,49238	3,69274	1,97813

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	CATS2D_07_LL	F06[N-S]	MATS4m	GATS4e	GATS8s	JGI8	JGT	CATS2D_03_DL	F02[N-S]	F04[O-O]
77	-2,87782	-3,69347	-4,16764	-3,88722	-3,70327	Train	-0,63121	3,67578	-2,14793	0,95123	1,74025	0,5869	-0,46159	-0,49238	1,66174	-0,49103
78	-1,9212	-2,65261	-2,80484	-2,52261	-2,85092	Train	-0,63121	-0,26896	-1,73122	0,36174	0,66384	0,88893	-0,13103	-0,49238	1,66174	-0,49103
79	-0,54658	-0,7741	-1,16437	-1,21794	-1,23339	Train	-0,39075	-0,26896	1,33769	-1,88385	1,78581	-1,22528	-0,74354	-0,49238	3,69274	-0,49103
80	-3,23383	-2,66025	-2,57503	-2,41649	-2,65266	Train	-0,63121	-0,26896	-2,22172	0,89129	-0,6013	-1,22528	0,68564	-0,49238	-0,36927	0,74355
81	-2,0125	-2,1454	-2,15576	-2,06182	-2,16481	Train	-0,63121	-0,26896	-2,00903	1,60568	-1,33019	-1,22528	0,88981	-0,49238	1,66174	-0,49103
82	-0,63141	-0,71735	-0,3994	-0,97585	-0,70539	Train	-0,63121	-0,26896	-0,12514	-0,09537	-1,33019	-1,22528	1,99816	2,67805	-0,36927	-0,49103
83	-2,896	-2,37347	-2,12761	-2,14592	-2,62261	Train	-0,63121	-0,26896	0,8168	2,21266	-1,33019	-1,22528	2,11483	-0,49238	-0,36927	-0,49103
87	-3,59274	-3,54801	-3,39485	-3,49394	-3,763	Train	1,29248	-0,26896	0,08755	1,78053	-0,83038	1,03994	-0,0727	-0,49238	-0,36927	-0,49103
89	-1,31697	-1,99192	-1,98065	-1,92177	-2,03707	Train	-0,15029	-0,26896	0,90362	0,02952	-0,7601	-1,22528	0,96759	-0,49238	-0,36927	-0,49103
90	-1,21865	-1,96444	-2,01171	-2,16076	-1,99113	Train	-0,63121	-0,26896	-0,6851	0,63151	-0,72235	1,34197	0,46203	1,62124	-0,36927	-0,49103
91	-1,95297	-1,91578	-1,93667	-1,9141	-1,87948	Train	-0,63121	-0,26896	0,26987	-0,03792	-1,33019	-1,22528	1,46343	-0,49238	-0,36927	-0,49103
92	-3,63348	-1,62829	-1,59093	-1,51605	-1,35941	Train	-0,63121	-0,26896	0,79076	-1,41675	1,47994	1,19096	-1,49216	0,56443	-0,36927	-0,49103
93	-3,82889	-2,73181	-2,41556	-2,82406	-2,64762	Train	-0,63121	3,67578	-0,36388	-0,8847	1,05562	1,644	-0,62687	0,56443	1,66174	-0,49103
94	-0,87168	-1,03791	-0,86783	-1,07967	-1,18825	Train	-0,63121	-0,26896	0,2308	-0,84224	2,83359	-1,22528	-0,17964	2,67805	-0,36927	1,97813
95	-0,92853	-1,7669	-1,79211	-1,64219	-1,82633	Train	-0,63121	-0,26896	0,8168	-0,91218	1,2079	1,644	0,15092	1,62124	-0,36927	-0,49103

99	-3,12009	-3,57453	-3,64719	-3,51047	-3,72829	Train	1,53294	-0,26896	-1,3666	0,8738	-0,75229	0,88893	-0,72409	-0,49238	-0,36927	-0,49103
101	-0,52396	-0,27752	-0,23141	0,110813	-0,28262	Train	-0,39075	-0,26896	0,26553	-2,19858	0,22781	-0,31919	-4,22413	0,56443	-0,36927	-0,49103
102	-0,74011	-0,43869	-0,16655	-0,52101	-0,43113	Train	-0,63121	-0,26896	0,18305	-0,72234	-1,33019	-1,22528	-0,22825	1,62124	-0,36927	-0,49103
103	-0,71565	-0,86572	-0,8859	-0,84065	-0,88315	Train	-0,63121	-0,26896	-0,45938	0,42418	-1,33019	-1,22528	0,16064	1,62124	-0,36927	-0,49103
104	-2,81404	-3,32278	-3,35906	-3,41116	-3,33126	Train	1,7734	-0,26896	0,33932	-0,94215	0,51416	0,28487	0,27731	-0,49238	-0,36927	-0,49103
105	-3,46548	-3,75635	-4,21505	-3,8565	-3,62392	Train	-0,39075	3,67578	-1,34923	0,63151	0,7094	-0,16818	-0,1602	-0,49238	-0,36927	0,74355
106	-3,33027	-2,43554	-2,58686	-2,23075	-2,54076	Train	-0,63121	-0,26896	-0,76757	1,06364	0,2239	-1,22528	0,86065	-0,49238	-0,36927	-0,49103
107	-3,77188	-3,41709	-3,22368	-3,44661	-3,38622	Train	0,33063	-0,26896	-0,91082	0,7539	0,04428	1,03994	0,4037	-0,49238	-0,36927	-0,49103
108	-3,14195	-3,74488	-3,78412	-4,01633	-3,91613	Train	-0,63121	3,67578	-0,57224	0,7614	0,45819	0,88893	0,19953	-0,49238	-0,36927	-0,49103
109	-1,15592	-0,89567	-0,83422	-0,81278	-0,93442	Train	-0,63121	-0,26896	2,05391	-2,06619	-1,17531	-1,22528	0,13147	-0,49238	-0,36927	-0,49103
110	-0,79619	-0,99158	-1,25018	-1,0041	-0,95765	Train	-0,63121	-0,26896	1,78479	-1,5916	-0,78223	-1,22528	-0,29631	-0,49238	-0,36927	-0,49103
112	-0,8803	-2,40488	-2,57388	-2,36216	-2,47427	Train	0,09017	-0,26896	0,90796	-0,4151	0,42695	1,34197	-1,57966	-0,49238	-0,36927	-0,49103
113	-0,71893	-2,18716	-2,3112	-2,25366	-2,15915	Train	-0,63121	-0,26896	-1,10181	0,67397	-0,05724	-0,16818	-1,21021	-0,49238	-0,36927	-0,49103
5	-0,77006	-2,40342				Test	-0,63121	-0,26896	0,41745	-1,26937	1,0387	1,03994	-0,18936	-0,49238	-0,36927	-0,49103
7	-3,28458	-2,6282				Test	-0,15029	-0,26896	1,27258	-0,80976	0,18355	1,03994	0,78287	-0,49238	-0,36927	-0,49103

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	CATS2D_07_LL	F06[N-S]	MATS4m	GATS4e	GATS8s	JGI8	JGT	CATS2D_03_DL	F02[N-S]	F04[O-O]
14	-0,42347	-0,65513				Test	-0,63121	-0,26896	1,78913	-0,72484	-1,33019	-1,22528	-0,43242	-0,49238	1,66174	-0,49103
16	-1,84668	-1,86396				Test	-0,63121	-0,26896	0,15267	0,0495	-0,95013	-1,22528	0,85092	-0,49238	-0,36927	-0,49103
18	-1,27151	-2,84278				Test	-0,63121	-0,26896	-1,79199	1,28595	0,77708	-0,31919	-0,00464	-0,49238	-0,36927	-0,49103
28	-2,44142	-2,31488				Test	-0,63121	-0,26896	-1,54457	1,84048	0,28378	1,19096	-1,04493	-0,49238	3,69274	-0,49103
30	-1,15394	-1,95684				Test	-0,63121	-0,26896	0,16569	0,31178	-0,03641	0,5869	-0,05325	0,56443	-0,36927	-0,49103
33	-1,2584	-1,46529				Test	-0,15029	-0,26896	-0,238	1,00119	-0,51149	0,13385	-0,84076	1,62124	-0,36927	-0,49103
39	-1,04896	-1,44548				Test	0,33063	-0,26896	1,18142	-0,62742	-1,12064	1,79502	-1,21993	1,62124	-0,36927	-0,49103
41	-4,44259	-3,89006				Test	2,01386	-0,26896	0,06151	0,17689	-0,56616	0,5869	0,82176	-0,49238	-0,36927	0,74355
43	-4,45917	-4,05049				Test	2,49479	-0,26896	0,08321	0,19938	-0,32016	0,43588	0,63703	-0,49238	-0,36927	0,74355
49	-4,38602	-3,80178				Test	2,01386	-0,26896	0,76037	-0,20528	-0,42429	0,73791	0,97731	-0,49238	-0,36927	0,74355
50	-4,4022	-3,97503				Test	2,49479	-0,26896	0,75603	-0,1853	-0,09108	0,5869	0,76342	-0,49238	-0,36927	0,74355
60	-3,71823	-3,59543				Test	2,73525	-0,26896	1,20747	0,12194	-0,24337	0,73791	-0,4227	-0,49238	-0,36927	-0,49103
67	-3,23553	-2,34702				Test	-0,63121	-0,26896	-0,32916	0,06449	-0,23165	0,28487	-0,04353	-0,49238	-0,36927	-0,49103
72	-3,32475	-2,31799				Test	-0,15029	-0,26896	0,63015	-0,0654	1,47863	-0,31919	-0,60742	-0,49238	-0,36927	-0,49103
84	-2,64335	-2,13883				Test	-0,63121	-0,26896	0,31761	2,28759	-1,33019	-1,22528	0,78287	-0,49238	-0,36927	-0,49103

85	-2,39808	-2,8017				Test	0,81156	-0,26896	-1,18863	0,9962	-1,1727	-1,22528	0,51064	-0,49238	-0,36927	-0,49103
86	-1,63884	-2,5314				Test	0,33063	-0,26896	-1,76595	2,25762	-1,19223	-1,22528	-0,98659	-0,49238	-0,36927	-0,49103
88	-0,85164	-2,60798				Test	0,81156	-0,26896	-0,54185	1,80051	-1,15058	-0,01716	-1,82271	-0,49238	-0,36927	-0,49103
96	-0,56802	-3,10527				Test	-0,15029	-0,26896	0,45652	-0,7623	1,72464	1,644	0,15092	-0,49238	-0,36927	-0,49103
97	-3,04885	-2,90379				Test	-0,63121	-0,26896	-1,55759	0,69895	-0,66118	0,5869	0,67592	-0,49238	-0,36927	-0,49103
98	-3,2209	-3,36942				Test	0,33063	-0,26896	-2,11755	1,15106	-0,9241	0,88893	-0,09214	-0,49238	-0,36927	-0,49103
100	-1,21095	-1,70338				Test	1,53294	-0,26896	1,23785	-0,1853	-0,11191	2,09705	-0,00464	3,73487	-0,36927	-0,49103
111	-0,39175	-0,42489				Test	-0,63121	-0,26896	1,75006	-0,97213	-0,64296	-1,22528	-0,29631	-0,49238	3,69274	-0,49103

Table S6.11. Decision tree regression prediction and descriptors.

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	piPC06	ChiA_B(p)	MATS8v	SpMax2_Bh(p)	P_VSA_m_2	SpMaxA_EA(ed)	Eig03_EA(bo)	Eig05_EA(dm)	F04[N-P]	F04[O-S]
1	-1,22173	-2,22646	-2,31019	-2,32747	-2,50923	Train	-3,28461	2,7301	0,39293	-3,37649	-1,53243	0,13917	-4,21954	-0,39841	-0,24404	-0,3623
2	-2,90345	-2,22646	-2,48029	-2,32747	-1,12593	Train	-1,16302	-0,50889	0,39293	-1,8248	-1,98087	1,17878	-0,48417	-0,39841	-0,24404	-0,3623
3	-0,90803	-1,14496	-1,1568	-1,18166	-1,15702	Train	0,04839	-1,84674	0,39293	-0,80167	-1,48103	1,99952	0,39644	-0,39841	-0,24404	-0,3623
4	-0,86782	-2,11804	-2,27432	-1,84151	-1,80632	Train	1,12648	-0,29765	0,6416	1,09358	2,84154	-1,53879	0,96714	-0,39841	-0,24404	-0,3623
6	-1,16256	-1,14496	-1,09639	-1,9337	-2,06773	Train	1,00397	-1,14261	0,78666	0,69338	1,22824	-0,01586	0,92431	-0,39841	-0,24404	-0,3623
8	-3,70781	-4,10456	-3,68364	-4,0469	-3,80411	Train	0,73517	-1,42426	0,11525	1,13511	1,49563	-0,69981	0,93817	1,42673	-0,24404	-0,3623
9	-3,80902	-4,10456	-3,68364	-3,742	-4,14335	Train	0,84615	-0,29765	-0,30749	1,12378	2,34038	-1,20137	1,10824	4,76252	-0,24404	-0,3623
10	-1,43854	-1,34048	-2,38109	-2,48969	-2,37056	Train	-0,06403	-0,22724	-0,24533	-0,46188	0,85449	-0,47638	0,00716	-0,39841	-0,24404	-0,3623
11	-2,86392	-3,01573	-4,07254	-4,0469	-3,45541	Train	0,44331	-0,64972	-0,0008	0,51217	0,78639	0,19845	0,61313	-0,39841	-0,24404	-0,3623
12	-4,93172	-4,10456	-3,95861	-4,0142	-3,52555	Train	1,25043	-0,86096	-0,95404	0,8142	1,08914	-0,85483	1,04777	-0,39841	-0,24404	-0,3623
13	-3,52931	-4,10456	-3,95861	-3,55888	-4,18777	Train	1,16323	-0,57931	-0,58932	0,57635	0,68024	-0,41255	0,68368	-0,39841	-0,24404	-0,3623
15	-2,45552	-3,01573	-3,00697	-3,0735	-2,16673	Train	-1,5565	1,46267	0,25616	-0,15608	-1,20311	1,14686	-1,81454	-0,39841	-0,24404	-0,3623
17	-1,41197	-3,01573	-2,48029	-2,48969	-2,16673	Train	-1,62496	1,32184	0,00749	-0,39015	-1,23959	1,4934	-1,81454	-0,39841	-0,24404	2,99373
19	-3,74545	-3,01573	-2,96361	-3,10438	-2,16673	Train	-1,4938	0,47689	-1,20685	-0,50341	-1,43677	0,95991	-1,55376	-0,39841	-0,24404	-0,3623
20	-2,95553	-2,94707	-3,70535	-3,13027	-3,80411	Train	0,22495	-0,36807	-1,05765	1,00297	-0,43606	0,36716	0,10794	-0,39841	-0,24404	-0,3623

21	-1,16422	-1,14496	-1,2942	-1,9337	-1,15702	Train	0,73445	-0,72013	1,02289	0,66696	0,33937	-0,30767	0,70006	-0,39841	-0,24404	-0,3623
22	-4,39933	-4,10456	-3,70535	-3,55888	-3,52555	Train	0,8094	-0,36807	-0,02152	1,06337	1,7812	-1,09194	0,90667	-0,39841	-0,24404	-0,3623
23	-1,84887	-1,34048	-1,284	-1,284	-2,37056	Train	-0,27662	0,96978	-0,99548	0,13085	-0,5447	-0,32135	-0,15788	-0,39841	-0,24404	-0,3623
24	-2,31142	-2,22646	-2,36375	-2,32747	-2,23819	Train	-0,5411	1,04019	0,76593	0,03647	-0,58131	0,17565	-0,03442	-0,39841	-0,24404	-0,3623
25	-0,87436	-2,22646	-2,33259	-2,27856	-2,50923	Train	-1,5565	1,32184	1,05191	-0,34862	-0,85696	0,09358	-0,72102	-0,39841	-0,24404	-0,3623
26	-1,22582	-1,14496	-1,1568	-1,66261	-1,1376	Train	0,32296	0,68813	0,98974	-0,71106	-0,23713	-0,07057	-0,10497	-0,39841	-0,24404	-0,3623
27	-3,1815	-3,01573	-2,48029	-3,0735	-2,16673	Train	-1,5565	1,18102	-1,14054	-0,84697	-0,89048	1,82625	-2,33107	-0,39841	4,05109	-0,3623
29	-1,69658	-2,11804	-1,2942	-1,84151	-2,15649	Train	0,61482	-0,86096	0,97316	0,88593	0,91416	-0,75908	0,60683	-0,39841	-0,24404	-0,3623
31	-1,26121	-1,14496	-1,2942	-1,00251	-1,12593	Train	0,08586	0,05441	0,58772	-1,1188	-0,2565	-0,39431	-0,00292	-0,39841	-0,24404	-0,3623
32	-0,73208	-1,14496	-1,90748	-1,68683	-2,15649	Train	0,88218	-0,22724	1,74404	0,27809	1,84614	-0,92323	1,22289	0,50491	-0,24404	1,31572
34	-1,23308	-1,34048	-2,38109	-1,30679	-1,10469	Train	-0,06403	0,40647	0,13182	-0,27312	-0,94184	-0,33047	0,0248	-0,39841	-0,24404	-0,3623
35	-1,18498	-1,14496	-1,11961	-1,07138	-1,15702	Train	0,08514	-0,86096	0,39293	-1,97582	-0,64869	0,86416	0,04369	-0,39841	-0,24404	-0,3623
36	-1,42583	-2,11804	-2,1175	-2,06185	-2,06773	Train	0,94488	-0,57931	0,95244	0,98409	2,27755	-1,48407	0,94069	-0,39841	-0,24404	-0,3623
37	-1,56658	-1,34048	-4,07254	-1,30679	-1,70709	Train	0,79138	0,12482	-0,61419	0,39135	1,28868	-1,08282	0,22889	-0,39841	-0,24404	-0,3623
38	-0,34577	-1,14496	-1,1568	-1,07138	-1,15702	Train	-0,0907	0,12482	0,43023	-0,48454	-0,22527	-0,45814	-0,31158	-0,39841	-0,24404	-0,3623

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	piPC06	ChiA_B(p)	MATS8v	SpMax2_Bh(p)	P_VSA_m_2	SpMaxA_EA(ed)	Eig03_EA(bo)	Eig05_EA(dm)	F04[N-P]	F04[O-S]
40	-3,84777	-4,10456	-3,70535	-1,45657	-4,18777	Train	0,3734	0,05441	0,12353	0,87083	0,36268	-0,47638	0,0248	-0,39841	-0,24404	-0,3623
42	-4,45683	-4,10456	-4,07254	-3,55888	-3,45541	Train	0,40367	-0,36807	-0,67636	1,02562	0,73604	-0,44446	0,84746	1,69803	-0,24404	-0,3623
44	-3,50524	-3,01573	-2,38109	-3,10438	-2,9613	Train	-0,8099	1,53308	-0,45255	-0,40525	-0,69569	1,13318	-0,49677	-0,39841	-0,24404	-0,3623
45	-3,60767	-2,11804	-1,2942	-1,68683	-2,06773	Train	1,28502	-0,29765	1,15137	0,96144	1,39024	-1,17401	1,1158	3,99178	-0,24404	-0,3623
46	-1,94825	-1,14496	-1,2942	-1,03508	-1,07803	Train	0,87497	-0,08642	1,1348	0,05157	2,02913	-0,92779	1,17879	0,03938	-0,24404	2,99373
47	-4,04631	-4,10456	-3,00697	-3,742	-2,90372	Train	0,11541	-0,86096	-0,36552	0,99164	0,47057	0,10726	0,22385	1,69803	-0,24404	-0,3623
48	-1,9488	-2,94707	-3,68364	-3,742	-3,52555	Train	0,08298	-0,72013	-0,44012	1,00297	0,16422	0,10726	0,22259	1,69803	-0,24404	-0,3623
51	-4,40486	-4,10456	-3,00697	-3,742	-3,45541	Train	0,20621	-0,43848	-0,3821	1,02562	0,267	-0,13897	0,22385	1,69803	-0,24404	-0,3623
52	-3,01947	-2,94707	-2,99167	-3,55888	-3,38545	Train	0,07577	0,12482	-0,55202	0,96899	-0,0964	-0,13441	0,22889	-0,39841	-0,24404	-0,3623
53	-3,31444	-3,01573	-2,38109	-3,10438	-2,37056	Train	-0,70181	1,39226	-1,34777	-0,22404	-0,65244	0,00694	0,05881	-0,39841	-0,24404	2,99373
54	-1,06576	-1,34048	-1,84465	-1,45657	-2,37056	Train	0,51249	0,12482	-0,7178	-0,29199	0,25804	-0,40343	0,20243	0,45558	-0,24404	1,31572
55	-1,52013	-1,14496	-1,11961	-1,14613	-2,15649	Train	0,82885	-0,43848	1,12651	0,34982	1,37797	-0,85483	1,18005	-0,39841	-0,24404	2,99373
56	-3,32973	-2,94707	-1,84465	-3,55888	-3,80411	Train	0,26603	-0,50889	-0,78411	1,10868	-0,1893	-0,2484	0,22007	1,44831	-0,24404	-0,3623
57	-2,90058	-3,01573	-3,0036	-2,48969	-3,38545	Train	0,00443	-0,36807	-0,75925	0,18748	0,43418	-0,02041	0,21755	1,73194	-0,24404	-0,3623
58	-3,17408	-2,94707	-3,70535	-3,742	-4,14335	Train	0,69193	-0,08642	-0,99548	1,08225	0,08424	-0,44446	1,04021	-0,39841	-0,24404	-0,3623

59	-3,72982	-4,10456	-4,13439	-3,55888	-3,80411	Train	0,5024	-0,86096	-0,75096	1,18419	0,91244	-0,34871	0,91171	-0,39841	-0,24404	-0,3623
61	-4,21094	-4,10456	-3,68364	-3,99231	-3,45541	Train	0,60329	-0,93137	-0,33236	1,10113	0,8919	-0,44446	0,90793	-0,39841	-0,24404	-0,3623
62	-2,0577	-2,22646	-2,17098	-2,32747	-2,24053	Train	-0,52525	1,04019	0,4178	-0,32975	0,17708	-1,0509	0,27046	3,58482	-0,24404	-0,3623
63	-3,00552	-2,22646	-2,17098	-2,27856	-1,12593	Train	-1,5565	2,1668	1,30058	-1,43216	-0,56751	0,17109	-2,34367	-0,39841	-0,24404	-0,3623
64	-3,57525	-3,01573	-3,0036	-2,48969	-2,37056	Train	-0,8099	1,25143	-1,05351	-0,91871	-0,15824	1,11038	-0,76133	-0,1641	-0,24404	1,31572
65	-2,82409	-3,01573	-3,00697	-2,99588	-2,16673	Train	-1,02538	1,53308	-1,80781	-1,36798	-0,22787	1,48428	-1,36353	-0,39841	-0,24404	1,31572
66	-2,88446	-3,01573	-1,1568	-3,13027	-2,99124	Train	0,32152	-0,08642	-1,43895	-1,23206	-0,15346	-0,00674	-0,18056	-0,39841	-0,24404	-0,3623
68	-2,71703	-3,01573	-3,0036	-3,0735	-2,37056	Train	0,46493	-0,016	-0,83799	-0,72616	-0,0551	0,16653	0,1533	-0,39841	-0,24404	-0,3623
69	-3,15724	-3,01573	-2,38109	-2,48969	-2,16673	Train	0,85263	-0,57931	-0,02567	-0,70729	0,24729	-0,06145	-0,09237	-0,39841	-0,24404	-0,3623
70	-1,10453	-1,14496	-1,73533	-1,18166	-1,12593	Train	0,42097	0,19524	1,07263	-0,93381	0,25907	-0,34415	0,34479	-0,39841	-0,24404	-0,3623
71	-3,33423	-1,34048	-2,38109	-1,45657	-1,10469	Train	0,58816	0,19524	-0,74681	0,11953	0,52079	-1,13297	0,41534	0,72997	-0,24404	-0,3623
73	-0,69472	-1,34048	-1,84465	-4,0469	-1,61189	Train	1,04577	0,05441	-0,94575	-0,27689	-0,17039	-0,59949	0,60053	-0,39841	-0,24404	-0,3623
74	-4,18105	-4,10456	-4,13439	-4,0142	-3,52555	Train	0,62563	-1,21302	0,25201	0,64808	0,7887	-0,73172	0,70888	-0,39841	-0,24404	1,31572
75	-1,5148	-1,14496	-1,11961	-1,91455	-1,09789	Train	0,08947	1,04019	0,65403	0,13463	-0,2789	-0,7226	-0,28764	1,24175	-0,24404	-0,3623
76	-3,04362	-2,22646	-2,33259	-1,66261	-2,0724	Train	-0,78684	0,26565	1,68602	-0,06924	-1,17637	0,89152	-0,61519	-0,39841	-0,24404	4,67175

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	piPC06	ChiA_B(p)	MATS8v	SpMax2_Bh(p)	P_VSA_m_2	SpMaxA_EA(ed)	Eig03_EA(bo)	Eig05_EA(dm)	F04[N-P]	F04[O-S]
77	-2,87782	-2,94707	-3,68364	-3,55888	-3,80411	Train	0,55861	-0,50889	-0,19974	1,09358	0,18852	-0,33503	0,22007	-0,39841	4,05109	1,31572
78	-1,9212	-2,94707	-3,70535	-3,742	-3,80411	Train	0,55573	-0,36807	0,01163	1,07847	-0,41716	-0,53566	0,49219	-0,39841	-0,24404	2,99373
79	-0,54658	-1,34048	-2,38109	-2,48969	-1,61189	Train	1,14306	-2,05798	-3,95467	0,21013	-0,31002	-0,28944	0,23897	-0,39841	-0,24404	-0,3623
80	-3,23383	-3,01573	-3,00697	-3,13027	-2,9613	Train	-1,24301	0,82895	-3,41174	-0,5978	-0,96939	1,08759	-1,18715	-0,39841	4,05109	-0,3623
81	-2,0125	-2,22646	-2,33259	-2,39092	-2,23819	Train	-3,28461	0,68813	0,39293	-2,32693	-1,0421	5,11834	-3,0794	-0,39841	-0,24404	-0,3623
82	-0,63141	-2,22646	-2,36375	-2,39092	-2,50923	Train	-3,28461	0,5473	0,39293	-1,35288	-1,53955	2,76098	-2,17107	-0,39841	-0,24404	-0,3623
83	-2,896	-2,22646	-2,48029	-1,66261	-2,0724	Train	-1,88223	0,19524	0,39293	-0,27312	-1,53955	1,57091	-1,5777	-0,39841	-0,24404	-0,3623
87	-3,59274	-2,11804	-1,04718	-2,06185	-1,80632	Train	0,14424	-0,29765	0,38049	0,88593	-0,6446	0,04798	0,06889	-0,39841	-0,24404	-0,3623
89	-1,31697	-1,14496	-1,73533	-1,9337	-1,07803	Train	0,07217	-0,93137	1,54925	0,5688	-0,59301	0,49027	0,50101	-0,39841	-0,24404	-0,3623
90	-1,21865	-2,11804	-1,98235	-2,06185	-1,9071	Train	0,0347	-0,57931	2,58538	1,77315	0,12057	-0,51742	0,46069	-0,39841	4,05109	-0,3623
91	-1,95297	-1,14496	-1,09639	-1,03508	-1,15702	Train	0,44979	-1,49467	0,39293	-0,56382	-1,65269	1,27453	0,32841	1,74119	-0,24404	-0,3623
92	-3,63348	-2,22646	-2,48029	-2,32747	-2,50923	Train	-0,96484	2,09639	0,61259	-0,88095	-0,3111	-0,60405	-0,26622	-0,39841	-0,24404	-0,3623
93	-3,82889	-2,22646	-2,33259	-1,91455	-2,0724	Train	-0,40994	0,26565	0,4178	-0,82055	0,02783	-0,65877	-0,03568	-0,39841	-0,24404	-0,3623
94	-0,87168	-1,14496	-1,2942	-1,18166	-1,09789	Train	0,79714	-0,43848	0,46339	-0,35617	-0,78101	-0,09337	0,77565	-0,39841	-0,24404	-0,3623
95	-0,92853	-1,14496	-1,40096	-1,30679	-2,37056	Train	1,1272	-0,50889	0,31418	0,3536	0,07631	-0,58581	1,19895	-0,39841	-0,24404	-0,3623

99	-3,12009	-2,11804	-1,73533	-1,84151	-1,80632	Train	0,69697	-1,00178	1,05605	0,93123	0,30482	-0,2256	0,67864	-0,39841	-0,24404	-0,3623
101	-0,52396	-2,22646	-2,33259	-2,32747	-2,0724	Train	-1,88223	3,64547	0,41365	-0,21271	-1,45711	-1,38832	-1,60415	-0,39841	-0,24404	-0,3623
102	-0,74011	-1,14496	-1,11961	-1,14613	-1,12593	Train	0,18027	-1,56509	0,39293	-2,42509	-1,62598	1,17878	0,39392	-0,39841	-0,24404	-0,3623
103	-0,71565	-1,14496	-1,04718	-1,66261	-1,12593	Train	-0,20384	-0,93137	0,39293	-2,55723	-1,24415	1,05111	-0,63157	-0,39841	-0,24404	-0,3623
104	-2,81404	-2,11804	-1,09639	-1,84151	-1,09789	Train	0,63572	-1,70591	0,94829	0,74246	-0,17264	-0,03865	0,65849	-0,39841	-0,24404	-0,3623
105	-3,46548	-3,01573	-1,90748	-2,99588	-2,99124	Train	0,39142	0,68813	-0,80484	0,3234	0,07809	-0,00218	-0,08985	-0,39841	-0,24404	-0,3623
106	-3,33027	-2,94707	-2,97332	-2,97332	-3,52555	Train	0,25954	-0,57931	0,23958	0,65563	-0,70522	0,38084	0,38636	-0,39841	-0,24404	-0,3623
107	-3,77188	-2,94707	-3,00697	-3,13027	-3,38545	Train	0,29557	-0,22724	-0,19559	0,65941	-0,60242	0,04798	0,38762	-0,39841	-0,24404	-0,3623
108	-3,14195	-2,94707	-3,70535	-3,742	-3,80411	Train	0,36187	0,19524	-0,7178	0,54614	0,17357	-0,12073	0,08401	-0,39841	-0,24404	-0,3623
109	-1,15592	-1,14496	-1,40096	-2,48969	-1,70709	Train	0,21918	0,40647	0,27688	-0,43923	0,61686	-0,45358	-0,15158	-0,39841	-0,24404	-0,3623
110	-0,79619	-1,34048	-1,40096	-2,48969	-2,37056	Train	0,29629	0,96978	-1,27731	-0,43923	0,87286	-0,56758	-0,05835	-0,39841	-0,24404	-0,3623
112	-0,8803	-1,34048	-4,07254	-1,91455	-1,61189	Train	0,4397	-0,36807	-0,14586	0,3838	0,37021	-0,69069	0,23771	-0,39841	-0,24404	-0,3623
113	-0,71893	-2,11804	-1,98235	-2,06185	-2,06773	Train	0,635	-0,15683	0,6789	1,02562	0,55772	-0,38975	0,66604	-0,39841	4,05109	-0,3623
5	-0,77006	-1,34048				Test	0,47141	0,33606	-0,12928	0,03647	0,24486	-0,70893	0,81092	-0,39841	-0,24404	-0,3623
7	-3,28458	-4,10456				Test	0,76904	-0,016	-1,02864	1,08603	1,20066	-0,44446	0,54384	-0,39841	-0,24404	-0,3623

row ID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	piPC06	ChiA_B(p)	MATS8v	SpMax2_Bh(p)	P_VSA_m_2	SpMaxA_EA(ed)	Eig03_EA(bo)	Eig05_EA(dm)	F04[N-P]	F04[O-S]
14	-0,42347	-1,14496				Test	-0,10871	-1,42426	0,39293	-1,12258	-0,51643	1,33837	-0,25992	-0,39841	-0,24404	-0,3623
16	-1,84668	-3,01573				Test	-1,20194	0,40647	-1,18613	0,30074	-0,89995	0,85504	-1,75281	-0,39841	-0,24404	2,99373
18	-1,27151	-2,94707				Test	-0,90935	0,82895	0,15669	1,84488	-0,89267	0,0115	-1,67596	-0,39841	-0,24404	6,34977
28	-2,44142	-1,14496				Test	0,84471	-0,86096	0,38878	-0,6431	0,05249	-0,13441	-0,15914	-0,39841	-0,24404	1,31572
30	-1,15394	-1,34048				Test	0,17378	0,26565	-0,72195	-0,85075	-0,0949	-0,31223	0,00338	-0,39841	-0,24404	-0,3623
33	-1,2584	-2,11804				Test	0,54203	-0,50889	1,37103	0,73869	0,75145	-1,23785	0,47581	-0,39841	-0,24404	-0,3623
39	-1,04896	-1,34048				Test	-0,37031	0,47689	-0,12928	-0,48831	-0,1041	-0,35783	-0,76637	-0,39841	-0,24404	-0,3623
41	-4,44259	-4,10456				Test	0,34962	-0,29765	0,0738	1,01807	0,40315	-0,34871	0,84746	1,69803	-0,24404	-0,3623
43	-4,45917	-4,10456				Test	0,37701	-0,22724	-0,55617	1,02562	0,45453	-0,44446	0,84746	1,69803	-0,24404	-0,3623
49	-4,38602	-2,11804				Test	0,17162	-0,57931	0,33076	1,01807	0,21562	-0,02041	0,22385	1,69803	-0,24404	-0,3623
50	-4,4022	-4,10456				Test	0,23936	-0,64972	-0,49814	1,02562	0,54851	-0,13897	0,22385	1,69803	-0,24404	-0,3623
60	-3,71823	-4,10456				Test	0,60329	-0,86096	-0,51058	1,10113	1,23699	-0,44446	0,90793	-0,39841	-0,24404	-0,3623
67	-3,23553	-3,01573				Test	0,24369	0,33606	-1,0618	-1,24339	0,07765	-0,08881	-0,12764	-0,39841	-0,24404	-0,3623
72	-3,32475	-1,34048				Test	0,56077	0,40647	-0,17487	0,19881	0,29979	-0,99618	0,40148	-0,39841	-0,24404	-0,3623
84	-2,64335	-2,22646				Test	-1,70135	1,25143	0,39293	-0,94136	-1,22831	2,30502	-2,51122	-0,39841	-0,24404	-0,3623

85	-2,39808	-2,22646				Test	-1,43615	0,75854	0,94415	-0,11455	-0,8903	0,93256	-0,67818	-0,39841	-0,24404	-0,3623
86	-1,63884	-2,22646				Test	-1,78638	1,39226	0,81152	-0,45811	-1,21169	1,4934	-1,55502	-0,39841	-0,24404	-0,3623
88	-0,85164	-3,01573				Test	-1,05709	1,25143	-0,23704	-0,17496	-1,08099	0,37628	-1,26526	-0,39841	-0,24404	-0,3623
96	-0,56802	-1,14496				Test	1,00685	0,26565	0,38049	-0,30332	-0,54022	-0,58581	1,07045	-0,39841	-0,24404	-0,3623
97	-3,04885	-2,94707				Test	0,46132	-0,43848	-1,08252	0,95766	-0,44243	0,17109	0,64841	-0,39841	-0,24404	-0,3623
98	-3,2209	-2,11804				Test	0,36692	0,05441	1,42491	0,8444	-0,39105	0,03886	0,59927	-0,39841	-0,24404	-0,3623
100	-1,21095	-1,34048				Test	0,79714	-1,00178	-0,3448	0,48196	0,65064	-0,47638	0,70636	-0,39841	-0,24404	-0,3623
111	-0,39175	-1,34048				Test	0,29629	0,40647	-1,11568	-0,43923	0,72313	-0,56758	-0,05835	-0,39841	-0,24404	-0,3623

Table S6.12. k-NN regression prediction and descriptors.

rowID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	X4v	GGI10	SpMin2_Bh(s)	Eig02_AEA(dm)	SsOH	NssO	CATS2D_02_DD	CATS2D_07_LL	F01[N-O]	F05[C-S]	F06[N-S]	F07[C-N]
1	-1,22173	-1,8477	-2,32028	-2,31839	-2,31839	Train	1	81	69	2	103	-1,58573	-0,59637	-3,12377	-3,09956	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
2	-2,90345	-2,99672	-2,83949	-2,91315	-2,88649	Train	2	69	66	27	80	0,23488	-0,59637	-1,08248	-0,94348	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
3	-0,90803	-1,74205	-2,44834	-2,54962	-2,49695	Train	3	76	15	89	17	2,69845	-0,59637	0,30821	0,41535	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
4	-0,86782	-1,21159	-1,58938	-2,18257	-2,03816	Train	4	55	46	32	29	0,01012	3,78952	0,93036	1,8547	-0,33376	-1,22839	3,67578	2,25433	-0,33483	-0,66179	-0,26896	4,03861
6	-1,16256	-1,04791	-0,97958	-0,90488	-0,97958	Train	6	34	101	82	75	-0,17857	-0,59637	0,32447	-1,03577	3,64397	-1,22839	-0,26896	-0,15029	-0,33483	-0,66179	-0,26896	1,21051
8	-3,70781	-3,83083	-3,89349	-3,57927	-3,64277	Train	8	9	42	56	61	1,29425	2,43913	0,8653	1,38146	-0,33376	0,44755	-0,26896	2,25433	-0,33483	-0,66179	-0,26896	1,21051
9	-3,80902	-3,69856	-4,0176	-3,49516	-3,58801	Train	9	8	42	36	61	0,36324	2,39296	0,71078	1,39521	-0,33376	1,28552	-0,26896	2,73525	-0,33483	-0,66179	-0,26896	1,21051
10	-1,43854	-1,40971	-1,51064	-1,44441	-1,5785	Train	10	70	112	69	37	-1,08001	-0,59637	-0,379	-0,53505	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	1,21051
11	-2,86392	-3,35705	-3,68647	-3,68647	-3,75143	Train	11	13	12	57	66	-0,46899	-0,59637	-0,0171	1,8547	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	0,03987	-0,26896	0,26781
12	-4,93172	-3,99065	-3,34237	-2,9578	-3,13794	Train	12	11	57	74	13	0,1917	1,12336	-0,36274	1,85863	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	0,74152	-0,26896	0,73916
13	-3,52931	-3,20371	-2,92361	-2,7395	-2,94814	Train	13	11	57	66	56	0,02609	-0,59637	0,20655	1,12619	-0,33376	-0,39042	-0,26896	0,09017	-0,33483	-0,66179	-0,26896	0,26781
15	-2,45552	-2,34854	-2,35741	-1,4467	-2,31237	Train	15	76	17	3	87	2,16433	-0,59637	0,31228	0,4507	-0,33376	0,44755	-0,26896	-0,15029	-0,33483	0,03987	-0,26896	-0,67489
17	-1,41197	-1,91904	-2,44188	-2,63503	-2,85759	Train	17	78	76	15	52	2,76411	-0,59637	0,26755	0,01281	-0,33376	0,44755	-0,26896	-0,63121	-0,33483	1,44317	-0,26896	-0,67489
19	-3,74545	-3,56368	-3,16219	-3,26705	-2,78478	Train	19	107	80	27	21	0,28634	-0,59637	0,04796	-0,36028	-0,33376	1,28552	-0,26896	0,81156	-0,33483	-0,66179	-0,26896	-0,67489

20	-2,95553	-3,17844	-3,27188	-3,36268	-3,3004	Train	20	106	107	83	27	-0,16851	-0,59637	-0,74497	-0,27192	-0,33376	1,28552	-0,26896	-0,63121	-0,33483	0,74152	-0,26896	-0,67489
21	-1,16422	-1,25554	-1,53797	-0,93824	-1,14594	Train	21	26	113	27	38	-0,03839	-0,12316	0,37327	-1,0397	-0,33376	0,44755	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
22	-4,39933	-4,29942	-4,11585	-3,90079	-4,21422	Train	22	61	42	51	48	0,00952	1,06565	0,88969	-0,81388	-0,33376	1,28552	-0,26896	2,97571	-0,33483	-0,66179	-0,26896	-0,67489
23	-1,84887	-2,52132	-3,08531	-3,15861	-2,38128	Train	23	59	56	37	21	0,52708	0,69631	1,43865	-0,42312	-0,33376	0,44755	-0,26896	0,33063	-0,33483	-0,66179	-0,26896	-0,20354
24	-2,31142	-2,39904	-2,852	-2,50622	-2,14979	Train	24	68	83	113	80	-0,52637	-0,59637	-0,20008	-0,27781	-0,33376	1,28552	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	0,26781
25	-0,87436	-0,9622	-1,30487	-1,32597	-1,11275	Train	25	89	79	54	26	-0,64171	-0,59637	1,05641	-0,57039	-0,33376	-1,22839	-0,26896	0,33063	-0,33483	2,14482	-0,26896	-0,67489
26	-1,22582	-0,95089	-0,88783	-1,5377	-0,68949	Train	26	38	21	112	110	-0,48201	-0,59637	0,3814	-0,67054	-0,33376	-0,39042	-0,26896	0,09017	-0,33483	-0,66179	-0,26896	-0,67489
27	-3,18115	-3,12625	-3,08655	-2,90168	-2,95665	Train	27	80	83	66	69	0,06986	-0,59637	-0,85069	-0,27977	-0,33376	0,44755	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
29	-1,69658	-2,47039	-3,09662	-3,0862	-1,8863	Train	29	74	57	70	37	-0,11232	1,70045	0,62945	0,04423	-0,33376	-1,22839	-0,26896	0,09017	-0,33483	-0,66179	-0,26896	2,62456
31	-1,26121	-2,18204	-2,92496	-2,82453	-2,79559	Train	31	63	69	66	27	-1,00429	-0,59637	-0,85476	-0,37796	-0,33376	-0,39042	-0,26896	-0,63121	1,00 45	-0,66179	-0,26896	-0,67489
32	-0,73208	-1,45035	-2,20272	-2,72847	-2,33596	Train	32	46	55	71	92	0,01603	1,16953	-0,39527	0,25237	-0,33376	0,44755	3,67578	-0,63121	-0,33483	0,74152	-0,26896	1,68186
34	-1,23308	-0,95191	-0,65995	-0,70266	-0,89474	Train	34	101	82	6	26	-0,90078	-0,59637	0,43833	-0,42705	2,59567	-0,39042	-0,26896	-0,39075	-0,33483	-0,66179	-0,26896	-0,67489
35	-1,18498	-0,95174	-1,36793	-0,78323	-0,8274	Train	35	103	102	82	34	-1,06522	-0,59637	-2,92046	-0,74908	2,66993	-1,22839	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
36	-1,42583	-2,54977	-2,7092	-3,36911	-3,5223	Train	36	9	8	71	37	-0,10937	1,31957	0,78804	0,4939	-0,33376	1,28552	-0,26896	1,29248	1,00 45	-0,66179	-0,26896	2,15321
37	-1,56658	-1,42547	-1,28803	-1,77698	-1,38219	Train	37	112	24	113	21	-0,59498	0,16539	0,54812	-0,7589	-0,33376	0,44755	-0,26896	-0,15029	-0,33483	-0,66179	-0,26896	0,73916

rowID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	X4v	GGI10	SpMin2_Bh(s)	Eig02_AEA(dm)	SsOH	NssO	CATS2D_02_DD	CATS2D_07_LL	F01[N-O]	F05[C-S]	F06[N-S]	F07[C-N]
38	-0,34577	-0,728	-1,26014	-1,45381	-1,25439	Train	38	26	110	109	21	-1,10367	-0,59637	0,52372	-1,07505	-0,33376	-0,39042	-0,26896	-0,15029	-0,33483	-0,66179	-0,26896	-0,67489
40	-3,84777	-3,68975	-3,4866	-3,53177	-3,12503	Train	40	87	107	52	53	0,87192	-0,59637	0,65791	-0,25032	-0,33376	1,28552	-0,26896	1,29248	-0,33483	1,44317	-0,26896	-0,20354
42	-4,45683	-4,36281	-4,22889	-4,19469	-4,22889	Train	42	61	51	22	59	0,2615	1,61966	0,7799	0,53514	-0,33376	0,44755	-0,26896	2,49479	-0,33483	-0,66179	-0,26896	-0,67489
44	-3,50524	-3,13672	-2,98713	-2,77341	-3,11707	Train	44	83	24	106	20	-1,00784	-0,59637	-1,04994	-0,29352	-0,33376	2,12349	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
45	-3,60767	-3,55831	-3,45372	-3,26574	-3,27637	Train	45	71	8	9	58	0,64065	2,73922	0,63352	1,68779	-0,33376	2,12349	-0,26896	-0,15029	-0,33483	-0,66179	-0,26896	0,73916
46	-1,94825	-1,75742	-1,609	-1,64491	-1,45786	Train	46	55	32	71	29	0,07163	1,16953	0,09269	0,22292	-0,33376	0,44755	3,67578	-0,63121	-0,33483	0,03987	-0,26896	2,62456
47	-4,04631	-3,47237	-3,96564	-4,09846	-3,9153	Train	47	48	51	56	104	0,05093	-0,21549	0,83277	0,54888	-0,33376	-0,39042	-0,26896	2,01386	-0,33483	-0,66179	-0,26896	-0,67489
48	-1,94988	-3,00094	-3,96564	-4,09846	-3,92027	Train	48	47	51	56	104	0,09766	-0,21549	0,8653	0,53317	-0,33376	-0,39042	-0,26896	2,01386	-0,33483	-0,66179	-0,26896	-0,67489
51	-4,40486	-3,71649	-3,00893	-3,09346	-3,09346	Train	51	48	47	42	61	0,23015	0,39623	0,8653	0,53317	-0,33376	-0,39042	-0,26896	2,49479	-0,33483	-0,66179	-0,26896	-0,67489
52	-3,01947	-2,86376	-3,37436	-3,0223	-2,14475	Train	52	78	53	107	40	1,23805	-0,59637	0,08456	-0,14821	-0,33376	1,28552	-0,26896	-0,15029	-0,33483	2,14482	-0,26896	-0,67489
53	-3,31444	-3,05239	-3,44138	-3,23106	-2,02457	Train	53	52	78	107	20	1,16412	-0,59637	0,08456	0,27594	-0,33376	2,12349	-0,26896	-0,63121	-0,33483	1,44317	-0,26896	-0,67489
54	-1,06576	-0,92308	-0,90388	-0,88891	-0,88891	Train	54	73	79	25	52	-0,53938	-0,59637	0,12522	0,52728	-0,33376	-0,39042	-0,26896	-0,63121	-0,33483	3,54813	-0,26896	0,73916
55	-1,52013	-1,56841	-1,36129	-2,4435	-1,57533	Train	55	46	32	29	71	0,13788	1,16953	0,07236	0,24059	-0,33376	0,44755	3,67578	-0,39075	-0,33483	0,03987	-0,26896	2,62456
56	-3,32973	-3,11017	-3,53291	-3,35422	-3,1181	Train	56	59	23	48	47	0,39873	0,93869	0,8653	0,92786	-0,33376	-0,39042	-0,26896	0,81156	-0,33483	-0,66179	-0,26896	-0,20354

57	-2,90058	-3,20114	-3,24694	-3,59096	-3,61622	Train	57	13	56	58	12	-0,16792	0,8579	0,45866	0,97695	-0,33376	-0,39042	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	0,73916
58	-3,17408	-3,28609	-3,1286	-3,34025	-3,33578	Train	58	59	56	57	23	-0,16615	2,21983	0,71484	1,1203	-0,33376	-0,39042	-0,26896	-0,39075	-0,33483	-0,66179	-0,26896	-0,67489
59	-3,72982	-3,36268	-3,02707	-3,04274	-3,65115	Train	59	56	23	61	58	0,10653	1,71199	0,8653	-0,24443	-0,33376	-0,39042	-0,26896	0,81156	-0,33483	-0,66179	-0,26896	-0,67489
61	-4,21094	-4,30714	-3,99467	-4,28014	-4,21422	Train	61	42	22	51	59	0,09411	1,67737	0,69858	-0,23068	-0,33376	0,44755	-0,26896	2,49479	-0,33483	-0,66179	-0,26896	-0,67489
62	-2,05777	-2,3975	-2,61004	-3,29226	-2,79093	Train	62	63	11	31	13	-0,91439	-0,21549	-0,4156	1,99608	-0,33376	0,44755	-0,26896	-0,63121	2,34383	0,74152	-0,26896	0,26781
63	-3,00552	-2,55561	-2,1793	-3,12959	-3,17911	Train	63	31	66	69	80	-1,58751	-0,59637	-0,96455	0,90823	-0,33376	-0,39042	-0,26896	-0,63121	1,0045	-0,66179	-0,26896	-0,67489
64	-3,57525	-3,23643	-2,94357	-2,94357	-2,94357	Train	64	65	77	108	105	-0,92622	-0,59637	-0,32207	1,06728	-0,33376	-0,39042	-0,26896	-0,63121	1,0045	0,03987	3,67578	-0,20354
65	-2,82409	-3,0676	-3,30587	-3,27981	-3,27502	Train	65	64	77	108	105	-1,49997	-0,59637	-0,96455	1,13601	-0,33376	-0,39042	-0,26896	-0,63121	1,0045	0,03987	3,67578	-0,20354
66	-2,88446	-3,00218	-3,11782	-3,13194	-3,14931	Train	66	80	27	69	31	-0,45716	-0,59637	-0,81003	0,28379	-0,33376	-0,39042	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
68	-2,71703	-2,46958	-2,6814	-2,4674	-2,32959	Train	68	24	37	83	44	-0,91338	-0,59637	0,23908	0,27004	-0,33376	1,28552	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	0,26781
69	-3,15724	-2,8161	-2,26429	-2,44384	-2,28698	Train	69	81	66	27	2	-0,91557	-0,59637	-0,90762	-0,8885	-0,33376	-0,39042	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,20354
70	-1,10453	-1,24451	-1,48606	-1,47602	-1,56952	Train	70	10	112	37	69	-1,13679	-0,59637	-0,33427	-0,57432	-0,33376	-0,39042	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	2,15321
71	-3,33423	-3,2035	-2,84818	-3,29905	-3,06489	Train	71	45	57	36	74	0,37921	2,15058	1,13367	0,33092	-0,33376	1,28552	-0,26896	-0,63121	-0,33483	0,74152	-0,26896	1,68186
73	-0,69472	-0,81327	-0,98104	-1,22185	-1,23618	Train	73	54	79	25	91	0,09115	-0,59637	0,42613	2,1355	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	3,54813	-0,26896	-0,67489
74	-4,18105	-3,49372	-3,76628	-3,19316	-3,6883	Train	74	56	29	57	12	0,24671	1,21757	1,04421	0,61368	-0,33376	-1,22839	-0,26896	0,5711	-0,33483	0,74152	-0,26896	1,21051

rowID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	X4v	GGI10	SpMin2_Bh(s)	Eig02_AEA(dm)	SsOH	NssO	CATS2D_02_DD	CATS2D_07_LL	F01[N-O]	F05[C-S]	F06[N-S]	F07[C-N]
75	-1,51 48	-1,2854 7	-1,108 71	-1,175 28	-1,522 51	Train	75	34	6	101	82	0,51 939	-0,01 928	1,15401	0,99463	3,27 966	-0,39 042	-0,26896	0,5711	1,00 45	0,74 152	-0,26 896	-0,20 354
76	-3,04 362	-2,4993 2	-1,970 15	-1,510 43	-1,970 15	Train	76	15	17	91	3	2,73 808	-0,59 637	0,64165	1,05943	-0,33 376	-0,39 042	-0,26896	-0,15029	-0,33 483	0,74 152	-0,26 896	-0,67 489
77	-2,87 782	-3,0917 5	-3,210 26	-3,366 31	-3,210 26	Train	77	108	105	64	65	0,60 22	-0,30 783	0,0683	0,17383	-0,33 376	0,44 755	-0,26896	-0,63121	-0,33 483	1,44 317	3,67 578	0,26 781
78	-1,92 12	-2,3778 4	-2,158 77	-2,527 36	-3,070 51	Train	78	52	53	17	91	1,87 036	0,07 305	0,35701	0,33288	-0,33 376	1,28 552	-0,26896	-0,63121	-0,33 483	2,14 482	-0,26 896	-0,67 489
79	-0,54 658	-0,8414	-1,324 62	-1,352 27	-1,106 43	Train	79	89	25	54	73	0,13 019	-0,59 637	-0,92795	0,52335	-0,33 376	-1,22 839	-0,26896	-0,39075	-0,33 483	2,14 482	-0,26 896	-0,67 489
80	-3,23 383	-3,1162 3	-2,809 19	-2,984 61	-2,989 61	Train	80	27	66	83	24	0,13 906	-0,59 637	-0,73684	-0,06181	-0,33 376	0,44 755	-0,26896	-0,63121	-0,33 483	-0,66 179	-0,26 896	-0,67 489
81	-2,01 25	-2,2229 9	-1,133 57	-2,152 41	-2,073 81	Train	81	69	31	2	38	-1,36 806	-0,59 637	-1,37118	-1,64647	-0,33 376	-0,39 042	-0,26896	-0,63121	-0,33 483	-0,66 179	-0,26 896	-0,67 489
82	-0,63 141	-0,8825 2	-1,299 52	-1,150 94	-1,150 94	Train	82	34	35	102	103	-0,32 408	-0,59 637	-1,44031	-0,57432	2,87 213	0,44 755	-0,26896	-0,63121	-0,33 483	-0,66 179	-0,26 896	-0,67 489
83	-2,89 6	-3,1099	-3,132 08	-3,278 54	-3,165 11	Train	83	44	27	80	24	-0,51 986	-0,59 637	-1,12314	-0,3485	-0,33 376	1,28 552	-0,26896	-0,63121	-0,33 483	-0,66 179	-0,26 896	-0,67 489
87	-3,59 274	-3,5971 9	-3,358 95	-3,544 23	-3,585 1	Train	87	40	107	15	19	1,39 776	-0,59 637	0,53186	-0,44865	-0,33 376	0,44 755	-0,26896	1,29248	-0,33 483	0,74 152	-0,26 896	-0,67 489
89	-1,31 697	-1,2714 3	-1,348 5	-1,335 26	-1,381 61	Train	89	79	26	66	2	0,52 353	-0,59 637	0,0439	0,09724	-0,33 376	-1,22 839	-0,26896	-0,15029	-0,33 483	0,74 152	-0,26 896	-0,67 489
90	-1,21 865	-1,8261 6	-2,312 87	-2,507 91	-2,392 23	Train	90	15	76	17	3	3,24 381	0,78 865	0,78804	0,55477	-0,33 376	0,44 755	-0,26896	-0,63121	-0,33 483	0,03 987	-0,26 896	1,21 051
91	-1,95 297	-2,1604 8	-2,409 42	-2,593 24	-2,878 97	Train	91	76	17	78	15	2,13 831	-0,59 637	0,25941	2,1571	-0,33 376	0,44 755	-0,26896	-0,63121	-0,33 483	1,44 317	-0,26 896	-0,67 489
92	-3,63 348	-3,0669	-1,582 46	-2,577 28	-2,466 28	Train	92	93	31	32	63	-1,03 091	-0,59 637	-0,24888	-1,18894	-0,33 376	-0,39 042	3,67578	-0,63121	2,34 383	-0,66 179	-0,26 896	-0,20 354

93	-3,82889	-3,64069	-3,24549	-3,32449	-3,29815	Train	93	92	64	65	108	-0,75409	-0,59637	-0,91576	-0,88654	-0,33376	-1,22839	3,67578	-0,63121	2,34383	-0,66179	3,67578	0,26781
94	-0,87168	-1,09152	-1,28636	-1,43734	-1,49397	Train	94	95	75	62	31	-0,89783	-0,59637	0,28381	0,46248	3,04888	-1,22839	-0,26896	-0,63121	5,02225	-0,66179	-0,26896	-0,67489
95	-0,92853	-1,27437	-1,7529	-1,60778	-1,81505	Train	95	94	62	99	31	-0,12474	-0,59637	0,86936	0,69026	-0,33376	-1,22839	-0,26896	-0,63121	5,02225	-0,66179	-0,26896	0,26781
99	-3,12009	-3,43084	-3,38495	-3,32464	-3,44764	Train	99	40	87	107	52	0,7341	-0,16932	0,55219	-0,02057	-0,33376	0,44755	-0,26896	1,53294	2,34383	1,44317	-0,26896	-0,20354
101	-0,52396	-0,8631	-1,07525	-1,1455	-0,97467	Train	101	34	6	109	110	-0,64526	-0,40016	1,14587	-2,04508	2,63576	-1,22839	-0,26896	-0,39075	-0,33483	0,03987	-0,26896	-0,67489
102	-0,74011	-0,80536	-1,59803	-1,18986	-1,15799	Train	102	103	35	82	1	-0,74764	-0,59637	-3,73372	-0,94348	2,87387	-1,22839	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
103	-0,71565	-0,79986	-0,90869	-1,18986	-1,15799	Train	103	102	35	82	1	-1,21605	-0,59637	-3,73372	-1,10254	2,74732	-1,22839	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
104	-2,81404	-2,69091	-2,53807	-3,10389	-2,2208	Train	104	48	47	26	51	-0,05613	-0,59637	0,35294	-0,97686	-0,33376	-1,22839	-0,26896	1,7734	-0,33483	-0,66179	-0,26896	-0,67489
105	-3,46548	-3,28475	-3,20505	-3,11938	-3,012	Train	105	108	77	64	65	-0,58788	-0,59637	0,09676	-0,98276	-0,33376	2,12349	-0,26896	-0,39075	-0,33483	0,74152	3,67578	0,26781
106	-3,33027	-3,19918	-3,18865	-3,15908	-3,00378	Train	106	20	83	107	27	-0,46899	-0,59637	-0,81003	-0,44865	-0,33376	1,28552	-0,26896	-0,63121	-0,33483	0,74152	-0,26896	-0,67489
107	-3,77188	-3,55279	-3,27144	-3,32841	-2,87353	Train	107	20	19	40	106	0,36146	-0,59637	0,04796	-0,42116	-0,33376	1,28552	-0,26896	0,33063	-0,33483	0,74152	-0,26896	-0,67489
108	-3,14195	-3,21204	-3,21077	-3,17898	-3,11496	Train	108	105	77	64	65	0,23666	-0,59637	0,25941	-0,89046	-0,33376	1,28552	-0,26896	-0,63121	-0,33483	0,74152	3,67578	0,26781
109	-1,15592	-0,99383	-0,76466	-0,94766	-1,12116	Train	109	110	38	26	69	-1,08119	-0,59637	0,71484	-1,79767	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
110	-0,79619	-0,86362	-1,10466	-1,46485	-1,11414	Train	110	109	38	26	21	-1,11845	-0,59637	0,71484	-1,83497	-0,33376	-0,39042	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
112	-0,88803	-1,21208	-1,58044	-1,50321	-1,39909	Train	112	37	26	10	69	-0,40215	-0,59637	0,02763	-1,12414	-0,33376	-0,39042	-0,26896	0,09017	-0,33483	-0,66179	-0,26896	0,73916

rowID	Exp	Prediction	Prediction (CV) Rep1	Prediction (CV) Rep2	Prediction (CV) Rep3	Set	1	2	3	4	5	X4v	GGI10	SpMin2_Bh(s)	Eig02_AEA(dm)	SsOH	NssO	CATS2D_02_DD	CATS2D_07_LL	F01[N-O]	F05[C-S]	F06[N-S]	F07[C-N]
113	-0,71893	-1,34227	-1,85351	-1,97897	-2,03616	Train	113	24	21	37	107	0,26387	-0,12316	0,25128	-0,9749	-0,33376	1,28552	-0,26896	-0,63121	-0,33483	0,03987	-0,26896	0,26781
5	-0,77006	-1,25372				Test	66	69	80	27	26	-0,30752	-0,59637	0,03577	-0,27977	-0,33376	-0,39042	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	1,68186
7	-3,28458	-3,24462				Test	34	38	68	63	26	-0,09162	1,68891	0,75144	1,04175	-0,33376	0,44755	-0,26896	-0,15029	-0,33483	-0,66179	-0,26896	-0,67489
14	-0,42347	-2,47056				Test	81	110	69	112	109	-0,77953	-0,59637	-2,58702	-0,81781	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	0,03987	-0,26896	-0,67489
16	-1,84668	-2,01308				Test	58	59	57	12	56	3,08706	-0,59637	0,37327	0,69223	-0,33376	0,44755	-0,26896	-0,63121	-0,33483	1,44317	-0,26896	-0,67489
18	-1,27151	-1,67208				Test	58	59	45	32	71	6,345	-0,59637	0,29194	-0,21105	-0,33376	2,12349	-0,26896	-0,63121	-0,33483	2,14482	-0,26896	-0,67489
28	-2,44142	-2,44226				Test	69	81	38	66	10	-0,47373	-0,59637	-0,52132	-1,01614	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	0,03987	-0,26896	0,26781
30	-1,15394	-2,45005				Test	69	66	26	38	27	-1,03328	-0,59637	-0,32614	-0,51541	-0,33376	-0,39042	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
33	-1,25 84	-1,16421				Test	26	66	38	13	68	-0,81502	0,30389	0,40173	-0,81781	-0,33376	-0,39042	-0,26896	-0,15029	-0,33483	-0,66179	-0,26896	0,26781
39	-1,04896	-0,67137				Test	66	63	31	69	13	-1,24267	-0,59637	0,52779	-1,85461	-0,33376	-0,39042	-0,26896	0,33063	-0,33483	-0,66179	-0,26896	-0,67489
41	-4,44259	-4,29839				Test	63	62	31	38	68	0,23429	1,61966	0,7799	0,53514	-0,33376	0,44755	-0,26896	2,01386	-0,33483	-0,66179	-0,26896	-0,67489
43	-4,45917	-4,36281				Test	63	62	31	38	68	0,31237	1,61966	0,7799	0,53514	-0,33376	0,44755	-0,26896	2,49479	-0,33483	-0,66179	-0,26896	-0,67489
49	-4,38602	-3,63081				Test	63	62	31	38	68	0,15207	0,39623	0,8653	0,53317	-0,33376	-0,39042	-0,26896	2,01386	-0,33483	-0,66179	-0,26896	-0,67489
50	-4,40 22	-3,71649				Test	63	62	31	38	68	0,17928	0,39623	0,8653	0,53317	-0,33376	-0,39042	-0,26896	2,49479	-0,33483	-0,66179	-0,26896	-0,67489

60	-3,71823	-4,30714				Test	63	62	31	95	38	0,06572	1,67737	0,8653	-0,24443	-0,33376	0,44755	-0,26896	2,73525	-0,33483	-0,66179	-0,26896	-0,67489
67	-3,23553	-2,46865				Test	81	69	38	10	63	-0,96585	-0,59637	-0,53759	-0,68821	-0,33376	-0,39042	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
72	-3,32475	-3,16288				Test	65	64	63	68	57	0,4632	1,10028	1,25566	0,28575	-0,33376	0,44755	-0,26896	-0,15029	-0,33483	2,14482	-0,26896	1,21051
84	-2,64335	-3,03568				Test	81	69	38	110	109	-1,15335	-0,59637	-1,44031	-1,07112	-0,33376	1,28552	-0,26896	-0,63121	-0,33483	-0,66179	-0,26896	-0,67489
85	-2,39808	-2,31375				Test	58	59	57	12	45	4,315	-0,59637	1,17434	0,71972	-0,33376	-0,39042	-0,26896	0,81156	-0,33483	0,74152	-0,26896	-0,67489
86	-1,63884	-2,63284				Test	58	57	59	12	56	2,83745	-0,59637	0,93849	-0,05788	-0,33376	-0,39042	-0,26896	0,33063	-0,33483	0,74152	-0,26896	-0,67489
88	-0,85164	-1,94987				Test	58	45	59	64	71	5,98597	-0,59637	1,34919	-0,05396	-0,33376	-1,22839	-0,26896	0,81156	-0,33483	3,54813	-0,26896	-0,67489
96	-0,56802	-1,27437				Test	54	73	79	25	52	-0,41517	-0,59637	0,69858	0,70401	-0,33376	-1,22839	-0,26896	-0,15029	5,02225	-0,66179	-0,26896	-0,67489
97	-3,04885	-1,96596				Test	25	54	106	79	81	-0,52519	-0,59637	-0,7531	-0,27192	-0,33376	1,28552	-0,26896	-0,63121	2,34383	0,74152	-0,26896	-0,67489
98	-3,22209	-3,19162				Test	54	25	79	63	106	0,15562	-0,59637	0,22688	-0,23657	-0,33376	1,28552	-0,26896	0,33063	2,34383	0,74152	-0,26896	-0,67489
100	-1,21095	-1,09797				Test	44	68	24	83	106	-0,40689	-0,59637	0,69451	-0,61556	3,36089	-1,22839	-0,26896	1,53294	-0,33483	-0,66179	-0,26896	0,73916
111	-0,39175	-0,87069				Test	66	13	80	27	11	-0,90788	-0,59637	0,71484	-1,20857	-0,33376	-1,22839	-0,26896	-0,63121	-0,33483	0,74152	-0,26896	-0,67489

Table S6.13. Harmonised mode of action (MoA) classification scheme for pesticides active substances (n= 113) according to i) function (e.g. insecticide, fungicide, etc.), ii) chemical class (e.g. carbamates, organophosphate, etc.), and iii) site of action (e.g. sodium channel modulators). Qualitative information on MoA were extracted from three different sources: Pesticide Properties DataBase (PPDB), Resistance Action Committee classifications (i.e. IRAC, FRAC, HRAC), and the peer reviewed scientific literature (Sanchez-Bayo, 2012; Johnson et al. 2012, 2013; Simon-Delso et al. 2015). Available at: https://doi.org/10.5281/zenodo.3755675

Chemical Name	PPDB_Common Name	Other names	CAS Number	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Uniq ue conc ate nate	Original Unit_Uniq ue conc at enat e	Source dataset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical group	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo 2018	Mode of action_Sanchez_bayo2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees	Sanchez_value
2-[[Ethoxy[(1-methylethyl)amino]phosphinothioyl]oxy]benzoic acid 1-methylethyl ester	isofenphos		25311-71-1	O=C(OC(C)C)c1ccccc1(OP(OCC)(NC(C)C)=S)	-3,847694754	0,0001422	0.049	µg/org	Ecotox	Insecticide	Organophosphate	Selective with contact and stomach action. Acetylcholine esterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		isofenphos	1B	organophosphate	Organophosphate	Acetylcholin esterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
chlorpyrifos	chlorpyrifos		2921-88-2	n1c(OP(OCC)(OCC)=S)c(cc(c1Cl)Cl)Cl	-3,771881674	0,0001691	0.059	µg/org	PPDB	Insecticide	Organophosphate	Non-systemic with contact, inhalation and stomach action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		chlorpyrifos	1B	organophosphate	Organophosphate	Acetylcholin esterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee

chloret hoxyfos	c hl or et h o xy fo s	Phosp horoth ioic acid O,O-diethy l-O-(1,2,2 ,2-tetrac hloroe thyl) ester	5 4 5 9 3 - 8 3 - 8	O(C C)P (OC C)(OC(C (Cl)(Cl)Cl) Cl) =S	- 3 , 7 4 5 4 5 4 8 4 5	0 , 0 0 0 1 7 9 7	0.0 4| 0.0 9	µg/ org	P P D B | E c o t o x	In s e ct ici d e	Or g a n o p h os ph at e	Cholines terase inhibitor , vapour action	Or ga no ph os ph at e	**Ace tylc holi nest eras e (AC hE) inhi bito rs_ ACh E(-)**		ch lo re th ox yf os	1B	organ ophos phate	O rg a n o p h os ph at e	Acetylcholin esterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Org ano pho sph oru s	AC hE(-)	Ne ur ot ox ic	Or ga no ph os ph at e	A C h E(-)	N e u r o t o x ic	0 . 5 8	µ g / b e e
profeno fos	pr of e n of os	Phosp horoth ioic acid O-(4-bromo -2-chloro phenyl) O-ethyl S-propyl ester	4 1 1 9 8 - 0 8 - 7	O= P(O c1c cc(c c1Cl)Br)(O CC) SCC C	- 3 , 5 9 2 7 4 3 6 0 9	0 , 0 0 0 2 5 5 4	0.0 95	µg/ org	P P D B | E c o t o x	In s e ct ici d e, A c ar ici d e	Or g a n o p h os ph at e	Non-systemic with contact and stomach action. Acetylch olinester ase (AChE) inhibitor .	Or ga no ph os ph at e	**Ace tylc holi nest eras e (AC hE) inhi bito rs_ ACh E(-)**		pr of en of os	1B	organ ophos phate	O rg a n o p h os ph at e	Acetylcholin esterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Org ano pho sph oru s	AC hE(-)	Ne ur ot ox ic	Or ga no ph os ph at e	A C h E(-)	N e u r o t o x ic	0 . 5 8	µ g / b e e

Chemical Name	PPDB_Common Name	Other names	CAS Number	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Uniqueconcatenate	Original Unit_Uniqueconcatenate	Sourcedataset	PPDB_Pesticidetype	PPDB_Substancegroup	PPDB_Mode of action	Harmonised Chemicalgorup	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substancegroup	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo 2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees	Sanchez_value
thiofanox	thiofanox		39196-18-4	O=C(ON=C(CSC)C(C)(C)C)NC	-3,5752453922	0,000265 9	0.058	µg/org	PPDB	Insecticide, Acaricide	Carbamate	Systemic. Acetylcholinesterase (AChE) inhibitor.	Carbamate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		thiofanox	1A	carbamate	Carbamate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Carbamate	AChE(-)	Neurotoxic	Carbamate	AChE(-)	Neurotoxic	0.71	µg/bee
3-[(Dimethoxyphosphinyl)oxy]-2-butenoic acid, Methyl ester	mevinphos		7786-34-7	O=C(OC)C=C(OP(=O)(OC)OC)C	-3,5052371167	0,000312 4	0.07	µg/org	Ecotox	Insecticide, Acaricide	Organophosphate	Systemic with contact, stomach and respiratory action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		mevinphos	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee

																				insecticidal effects}								
etrimfos	etrimfos		38260-54-7	n1c(OCC)cc(nc1C)OP(OC)(OC)=S	-3,465479017	0,000342424	0.1	µg/org	PPDB	Insecticide	Organophosphate	Contact and stomach action, non-systemic. Acetylcholinesterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**	Not in Sanchez-bayo as compound NOR in IRAC. "Pyrimidine organothiophosphate insecticides" from AlanWood								Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
benfuracarb	benfuracarb	Benfuracarb	82560-54-1	O=C(Oc2ccccc1c2(OC(C)(C)C1))N(C)SN(CCC(=O)OCC)C(C)C	-3,334228866	0,000463632	0.19	µg/org	PPDB|OpenFoodTox	Insecticide	Carbamate	Systemic, stomach and contact action. Acetylcholinesterase (AChE) inhibitor.	Carbamate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		benfuracarb	1A	carbamate	Carbamate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Carbamate	AChE(-)	Neurotoxic	Carbamate	AChE(-)	Neurotoxic	0.71	µg/bee
chlorpyrifos-methyl	chlorpyrifos-methyl		55598-13-0	n1c(OP(OC)(OC)=S)c(cc(c1Cl)Cl)Cl	-3,330271669	0,000467 4	0.15	µg/org	PPDB	Insecticide, Acaricide	Organophosphate	Non-systemic with contact, stomach and respiratory action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		chlorpyrifos-methyl	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
carbosulfan	carb	Car	552	O=C(Oc2cccc	-3,	0,0	0.18	µg/org	PPD	Insectid	Carb	Systemic with contact	Carba	**Acetylcholi**		carbosulfan	1A	carbamate	Carb	Acetylcholinesterase (AChE)	1	Carba	AChE(-)	Neuro	Carba	ACh	Neu	0.	µg/

	osulfan	bosulfan	85-14-8	1c2(OC(C)(C)C1))N(C)SN(CCCC)CCCC	32475487 2	0047 34			BlOpenFoodTox	e, Nematicide	amate	and stomach action. Acetylcholine esterase inhibitor.	mate	**nesterase (AChE) inhibitors_AChE(-)**					amate	inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}		mate		toxic	mate	E(-)	rotoxic	71	bee
malathion	malathion	Malathion	121-75-5	O=C(OCC)CC(C(=O)OCC)SP(OC)(OC)=S	-3, 31444142	0,0004848	0.16	µg/org	PPDBlOpenFoodTox	Insecticide, Acaricide, Veterinary substance	Organophosphate	Broad-spectrum, non-systemic with contact, stomach and respiratory action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		malathion	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
aminocarb	aminocarb		2032-59-9	O=C(Oc1ccc(c(c1)C)N(C)C)NC	-3, 23553095	0,0005814	0.121	µg/org	PPDB	Insecticide	Carbamate	Non-systemic, acetylcholinesterase (AChE) inhibitor.	Carbamate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**	Classified as "phenyl methylcarbamate insecticides". OR "carbamate acaricides"	aminocarb	1A	carbamate	Carbamate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Carbamate	AChE(-)	Neurotoxic	Carbamate	AChE(-)	Neurotoxic	0.71	µg/bee
formothion	formothi		2540-	O=CN(C(=O)CSP(OC)(	-3, 23	0,000	0.15	µg/org	PPDB	Insecticide, Aca	Organo	Systemic with contact action. Cholinest	Organophos	**Acetylcholinester**	Classified as "Organothiophosphate insecticide" in AlanWood. Not		1B														

	on		82-1	OC)=S)C	3832632	5837				ricide	phosphate	erase inhibitor.	phate	**ase (AChE) inhibitors_AChE(-)**	present in IRAC nor Sanchez-bayo. However, it can be considered as "organophosphate"												
Phosphorothioic acid, O,O-Diethyl-O-(4-nitrophenyl)ester	parathion-ethyl		56-38-2	O=[N+]([O-])c1ccc(OP(OCC)(OC)=S)cc1	-3,220904232	0,000601 3	0.175	µg/org	Ecotox	Insecticide, Acaricide	Organophosphate	Non-systemic with contact, stomach and some respiratory action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		parathion	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee

Chemical Name	PPDB_Common Name	Other names	CAS Number	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Unique concatenate	Original Unit_Unique concatenate	Source dataset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical group	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo 2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees	Sanchez_value
dimethoate	dimethoate	Phosphorodithioic acid, O,O-Dimethyl S-[2-(methylamino)-2-oxoethyl] ester	60-51-5	O=C(NC)CSP(OC)(OC)=S	-3,1815041139	0,000658 4	0.1 | 0.16 | 0.17 | 0.19	µg/org	PPDB | Ecotox	Insecticide, Acaricide, Metabolite	Organophosphate	Systemic with contact and stomach action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		dimethoate	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
carbaryl	carbaryl		63-25-2	O=C(Oc1ccc2ccccc12)NC	-3,1572386 35	0,000696 2	0.14	µg/org	PPDB	Insecticide, Plant growth regulator	Carbamate	Stomach and contact activity with slight systemic properties. Cholinesterase inhibitor.	Carbamate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		carbaryl	1A	carbamate	Carbamate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Carbamates	AChE(-)	Neurotoxic	Carbamate	AChE(-)	Neurotoxic	0.71	µg/bee

diazinon	diazinon	O,O-Diethyl O-[6-methyl-2-(1-methylethyl)-4-pyrimidinyl] ester phosphorothioic acid	333-41-5	n1c(OP(OCC)(OCC)=S)cc(nc1C(C)C)C	-3	0,000721 2	0.13|0.37	µg/org	PPDB|Ecotox	Insecticide, Acaricide, Repellent, Veterinary substance	Organophosphate	Non-systemic with respiratory, contact and stomach action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		diazinon	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
P-Phenylphosphonothioic acid O-ethyl O-(4-nitrophenyl) ester	EPN		2104-64-5	O=[N+](O-)c2ccc(OP(OCC)(c1ccccc1)=S)cc2	-3	0,000758 4	0.245	µg/org	Ecotox	Insecticide, Acaricide	Organophosphate	Non-systemic with contact and stomach action, works by cholinesterase inhibition	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		epn	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
fenitrothion	fenitrothion	Phosphorothioic acid O,O-dimethyl O-(3-methyl-4-nitrophenyl)ester	122-14-5	O=[N+](O-)c1ccc(OP(OC)(OC)=S)cc1C	-3	0,000893 6	0.16|0.383	µg/org	PPDB|Ecotox	Insecticide	Organophosphate	Non-systemic with contact and stomach action. Cholinesterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		fenitrothion	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
fosthiazate	fosthiaz		98883	O=C1N(CCS1	-3	0,00	0.256	µg/org	PPDB	Insecticide, Nem	Organ	Systemic. Cholinesterase	Organoph	**Acetylcholines**	Not in Sa	fosthiaz	1B	organophosphate	Organ	Acetylcholinesterase (AChE) inhibitors	1	Organophos	AChE(-)	Neuroto	Organoph	ACh	Neur	0.58	µg/b

	zate		6-44-3	)P(=O)(OCC)SC(C)CC	436179334	090444				aticide	ophosphate	inhibitor.	osphate	**terase (AChE) inhibitors_ AChE(-)**	nchez-bayoascompound	ate			ophosphate	Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}		phorus		xic	osphate	E(-)	otoxic		ee
alpha-[(Dimethoxyphosphinothioyl)thio]benzeneacetic acid, Ethyl ester	phenthoate		2597-03-7	O=C(OCC)C(c1ccccc1)SP(OC)(OC)=S	-3,0194700063	0,000956 2	0.306	µg/org	Ecotox	Insecticide, Acaricide	Organophosphate	Non-systemic with contact and stomach action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_ AChE(-)**		phenthoate	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee

Chemical Name	PPDB_Common Name	Other names	CAS Number	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Unique concnatenate	Original Unit_Unique concnatenate	Sourced at aset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical group	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees	Sanchez_value
methomyl	methomyl	Methomyl| N-[[(Methylamino)carbonyl]oxy]ethanimidothioic acid methyl ester	16752-77-5	O=C(ON=C(C)SC)NC	-3,000551913 3	0,000098 74	0.16	µg/org	PPDB| OpenFood Tox | Ecotox	Insecticide, Acaricide, Metabolite	Carbamate	Systemic with contact and stomach action. Cholinesterase inhibitor.	Carbamate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		methomyl	1A	carbamate	Carbamate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Carbamates	AChE(-)	Neurotoxic	Carbamate	AChE(-)	Neurotoxic	0.71	µg/bee
O,O-Dimethyl O-[3-methyl-4-(methylthio)phenyl]ester phosphorothioic acid	fenthion		55-38-9	O(c1cc(c(c1)C)SC)P(OC)(OC)=S	-2,955525358	0,001107 8	0.308	µg/org	Ecotox	Insecticide, Veterinary substance, Avicide	Organophosphate	Contact, stomach and respiratory action. Cholinesterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		fenthion	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee

Phosphoric acid 1,2-dibromo-2,2-dichloroethyl dimethyl ester	naled		300-76-5	O=P(OC)(OC)OC(C(Cl)(Cl)Br)Br	-2,8960006850	0,0012706	0.48	µg/org	Ecotox	Insecticide, Acaricide	Organophosphate	Non-systemic with rapid contact and stomach action. Cholinesterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		naled	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
methiocarb	methiocarb	Methiocarb|3,5-Dimethyl-4-(methylthio)phenol, Methylcarbamate	2032-65-7	O=C(Oc1cc(c(c(c1)C)SC)C)NC	-2,8844618888	0,0013048	0.23|0.375	µg/org	PPDB|OpenFoodTox|Ecotox	Insecticide, Molluscicide, Bird repellent	Carbamate	Non-systemic with neurotoxic contact and stomach action. Acetylcholinesterase (AChE) inhibitor.	Carbamate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		methiocarb	1A	carbamate	Carbamate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Carbamates	AChE(-)	Neurotoxic	Carbamate	AChE(-)	Neurotoxic	0.71	µg/bee
azinphos-methyl	azinphos-methyl	O,O-Dimethyl S-[(4-oxo-1,2,3-benzotriazin-3(4H)-yl)methyl] ester, Phosphorodithioic acid	86-50-0	O=C1c2ccccc2(N=NN1CSP(OC)(OC)=S)	-2,8778178777	0,0013249	0.42	µg/org	PPDB|Ecotox	Insecticide, Acaricide, Molluscicide	Organophosphate	Non-systemic, contact and stomach action. Cholinesterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		azinphos-methyl	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
2-Methyl-2-(meth	aldic		116-	O=C(ON=C	-2,8	0,00	0.285	µg/org	Ecotox	Insecticide,	Carba	Systemic with contact and	Carbam	**Acetylcholines**		aldicarb	1A	carbamate	Carba	Acetylcholinesterase (AChE) inhibitors	1	Carbamate	AChE(-)	Neuroto	Carbam	AChE	Neur	0.71	µg/b

ylthio) propanol O-[(methylamino)carbonyl] oxime	arb		06-3	C(C)(C)SC)NC	240860777	149944				Acaricide, Nematicide	mate	stomach action absorbed through roots. Acetylcholinesterase (AChE) inhibitor.	ate	**terase (AChE) inhibitors_ AChE(-)**					mate	Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}				xic	ate	E(-)	otoxic		ee
2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate)	bendiocarb		2781-23-3	O=C(Oc1cccc2OC(Oc12)(C)C)NC	-2,71702554 5	0,001918 6	0.428	µg/org	Ecotox	Insecticide, Veterinary substance	Carbamate	Systemic, with contact and stomach action resulting in rapid knock-down. Acetylcholinesterase (AChE) inhibitor.	Carbamate	**Acetylcholinesterase (AChE) inhibitors_ AChE(-)**		bendiocarb	1A	carbamate	Carbamate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Carbamate	AChE(-)	Neurotoxic	Carbamate	AChE(-)	Neurotoxic	0.71	µg/bee

Chemical Name	PPDB_Common Name	Other names	CAS Number	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Uniq	Original Unit_Uniq	Source dataset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical goup	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Subst ance group	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo 2018	Mode of action_Sanchez_bayo2018	Group_Sanchez_bayo 2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees	Sanchez_value
Phosphoric acid 2,2-dichloroethenyl dimethyl ester	dichlorvos		62-73-7	O=P(OC=C(Cl)Cl)(OC)OC	-2,643458	0,0022733	0.5	µg/org	Ecotox	Insecticide, Acaricide, Metabolite	Organophosphate	Respiratory, contact and stomach action. Cholinesterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**	Not in Sanchez-bayo as compound	dichlorvos	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
disulfoton	disulfoton	Phosphorodithioic acid, O,O-Diethyl-S-[2-(ethylthio)ethyl] ester	298-04-4	O(CC)P(OCC)(=S)SCCSCC	-2,4555244688	0,0035033	0.96	µg/org	PDB|Ecotox	Insecticide, Acaricide	Organophosphate	Systemic absorbed via roots. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		disulfoton	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee

cadusafos	cadusafos		95465-99-9	O=P(OCC)(SC(C)CC)SC(C)CC	-2,398081065	0,003998 7	1.08	µg/org	PPDB	Insecticide, Nematicide	Organophosphate	Broad spectrum activity with contact and stomach action. Cholinesterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**	Not in Sanchez-bayo as compound	cadusafos	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
Phosphoric acid, 2-Chloro-3-(diethylamino)-1-methyl-3-oxo-1-propen-1-yl dimethyl ester	phosphamidon		13171-21-6	O=C(C(=C(OP(=O)(OC)OC)C)Cl)N(CC)CC	-2,311418451	0,004881 8	1.46	µg/org	Ecotox	Insecticide, Acaricide	Organophosphate	Strong stomach action and slight contact action. Cholinesterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		phosphamidon	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
thiodicarb	thiodicarb	Thiodicarb	59669-26-0	O=C(ON=C(C)SC)N(C)SN(C(=O)ON=C(C)SC)C	-2,057701 7	0,008755 58	3.1	µg/org	PPDB | OpenFoodTox	Insecticide, Molluscicide, Ovicide	Carbamate	Mainly stomach action but some contact effects. Cholinesterase inhibitor.	Carbamate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		thiodicarb	1A	carbamate	Carbamate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Carbamate	AChE(-)	Neurotoxic	Carbamate	AChE(-)	Neurotoxic	0.71	µg/bee
Phosphoramidothioic	meth		102	O=P(OC)(	-2,	0, 0	1.37	µg/org	Eco	Insecti	Orga	System ic with contact	Organo	**Acetylcholi**		meth	1B	organophosphate	Orga	Acetylcholinesterase (AChE)	1	Organopho	AChE(-)	Neurot	Organo	AChE	Neu	0.	µg/

acid, O,S-Dimethyl ester	amidophos		65-92-6	N)SC	012502661	0971622			tox	cide, Acaricide, Metabolite	nophosphate	and stomach action. Acetylcholinesterase (AChE) inhibitor.	phosphate	**nesterase (AChE) inhibitors_AChE(-)**		amidophos			nophosphate	inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}		sphorus		oxic	phosphate	E(-)	rotoxic	58	bee
phosalone	phosalone	Phosalone	2310-17-0	O=C2Oc1cc(ccc1N2CSP(OCC)(OCC)=S)Cl	-1,921197843	0,011989895	4.4	µg/org	PPDB | OpenFoodTox	Insecticide, Acaricide	Organophosphate	Non-systemic with contact and stomach action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**	Not in Sanchez-bayo as compound	phosalone	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
terbufos	terbufos		13071-79-9	O(CC)P(OCC)(=S)SCSC(C)(C)C	-1,846675172	0,014239	4.1	µg/org	PPDB	Insecticide, Nematicide	Organophosphate	Systemic. Cholinesterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		terbufos	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee

Chemical Name	PPDB_CommonName	Othernames	CASNumber	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Uniqueconcatenate	Original Unit_Uniqueconcatenate	Sourcedataset	PPDB_Pesticidetype	PPDB_Substancegroup	PPDB_Mode of action	HarmonisedChemicalgoup	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.ChemicalSub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substancegroup	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticideclass_Sanchez_bayo2018	Mode of action_Sanchez_bayo2018	Group_Sanchez_bayo2018	Sanchez_Insecticideclass	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees	Sanchez_value
ethoprophos	ethoprophos		13194-48-4	O=P(OCC)(SCCC)SCCC	-1,6384179	0,022969	5.56	µg/org	PPDB	Insecticide, Nematicide	Organophosphate	Non-systemic with contact action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		ethoprophos	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
O,O-Diethyl S-[(ethylthio)methyl]) ester, Phosphorodithioic acid	phorate		298-02-2	O(CC)P(OCC)(=S)SCSCC	-1,4119653 07	0,038728 9	10.07	µg/org	Ecotox	Insecticide, Acaricide, Nematicide	Organophosphate	Systemic with contact and stomach action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		phorate	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee

Phosphorodithioic acid, S,S'-Methylene O,O,O',O'-tetraethyl ester	ethion		563-12-2	O(CC)P(OCC)(=S)SCSP(OCC)(OCC)=S	-1,2715053922	0,053517	20.55	µg/org	Ecotox	Insecticide, Acaricide, Metabolite	Organophosphate	Non-systemic with a predominate contact action. Acetylcholine esterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		ethion	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
Dimethylcarbamic acid, 2-(Dimethylamino)-5,6-dimethyl-4-pyrimidinyl ester	pirimicarb		23103-98-2	O=C(Oc1nc(nc(c1C)C)N(C)C)N(C)C	-1,1045319332	0,078608	18.72	µg/org	Ecotox	Insecticide	Carbamate	Selective, systemic with contact, stomach and respiratory action. Acetylcholinesterase (AChE) inhibitor.	Carbamate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		pirimicarb	1A	carbamate	Carbamate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Carbamates	AChE(-)	Neurotoxic	Carbamate	AChE(-)	Neurotoxic	0.71	µg/bee
triazamate	triazamate		112143-82-5	O=C(OCC)CSc1nc(nn1(C(=O)N(C)C))C(C)(C)C	-1,0657612119	0,085948	27.0	µg/org	PPDB	Insecticide	Carbamoyltriazole	Systemic with contact and stomach action.	Carbamate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**	PPDB classification too specific. "Carbamate" fine for IRAC and Sanchez-Bayo	triazamate	1A	carbamate	Carbamoyltriazole	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Carbamates	AChE(-)	Neurotoxic	Carbamates	AChE(-)	Neurotoxic	0.71	µg/bee

Chemical Name	PPDB_Common Name	Other names	CASNumber	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Uniquconcatenate	Original Unit_Uniquconcatenate	Sourcedataset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical group	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No.Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees	Sanchez_value
triazophos	triazophos		24017-47-8	n1cn(nc1OP(OCC)(OCC)=S)c2ccccc2	-0,7189333332	0,191014646	59.8	µg/org	PPDB	Insecticide, Acaricide, Nematicide	Organophosphate	Non-systemic, broad spectrum with contact and stomach action. Cholinesterase inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		triazophos	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee
P-(2,2,2-Trichloro-1-hydroxyethyl)phosphonic acid dimethyl ester	trichlorfon		52-68-6	O=P(OC)(OC)C(O)C(Cl)(Cl)Cl	-0,6314074194	0,2336644	59.8	µg/org	Ecotox	Insecticide, Veterinary substance	Organophosphate	Non-systemic with contact and stomach action. Acetylcholinesterase (AChE) inhibitor.	Organophosphate	**Acetylcholinesterase (AChE) inhibitors_AChE(-)**		trichlorfon	1B	organophosphate	Organophosphate	Acetylcholinesterase (AChE) inhibitors Nerve action {Strong evidence that action at this protein is responsible for	1	Organophosphorus	AChE(-)	Neurotoxic	Organophosphate	AChE(-)	Neurotoxic	0.58	µg/bee

																				insecticidal effects}								
5-Amino-1-[2,6-dichloro-4-(trifluoromethyl)phenyl]-4-[(trifluoromethyl)sulfinyl]-1H-pyrazole-3-carbonitrile	fipronil	fipronil	120068-37-3	N#Cc1nn(c(N)c1S(=O)C(F)(F)F)c2c(cc(cc2Cl)C(F)(F)F)Cl	-4,9317228807	1,17E-05	0.00385999999999999997|0.000583|0.00059	µg/org	Ecotox|PPDB	Insecticide, Veterinary substance	Phenylpyrazole	Broad-spectrum with contact and stomach action. GABA-gated chloride channel antagonist.	Phenylpyrazole(azole)	**GABA-gated chloride channel blockers_GABA-R(-)**			2B	phenylpyrazoles (Fiproles)	phenylpyrazoles (Fiproles)	GABA-gated chloride channel blockers Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	2	Phenylpyrazole (Cyclodiene OC, fipronil)	GABA-R(-)	Neurotoxic	Phenylpyrazole	GABA-R(-)	Neurotoxic	2.24	µg/bee
lindane	lindane	(1alpha,2alpha,3beta,4alpha,5alpha,6beta)-1,2,3,4,5,6-Hexachlorocyclohexane	58-89-9	C1(C(C(C(C(C1Cl)Cl)Cl)Cl)Cl)Cl	-2,90344934 5	0,0012 49	0.23|0.562	µg/org	PPDB|Ecotox	Insecticide, Acaricide, Veterinary substance	Organochlorine	Acts as stimulant to the nervous system with contact and stomach action. GABA-gated chloride channel antagonist.	Organochlorine/cyclodiene organochlorine	**GABA-gated chloride channel blockers_GABA-R(-)**	Classified as "organochlorine insecticides". OR "organochlorine fungicides" from AlanWood. Not included in IRAC (maybe "cyclodiene organochlorine")														
6,7,8,9,10,10-Hexachloro-1,5,5a,6,9,9a-hexahydro-6,9-methano-2,4,3-benzodioxathiepin 3-oxide	endosulfan		115-29-7	O=S1OCC2C(CO1)C3(C(=C(C2(C3(Cl)Cl)Cl)Cl)Cl)Cl	-1,95297195 6	0,0111437	4.5	µg/org	Ecotox	Insecticide, Acaricide	Organochlorine	Non-systemic with contact and stomach action, acts as a non-competitive GABA antagonist	Organochlorine	**GABA-gated chloride channel blockers_GABA-R(-)**		endosulfan	2A	cyclodiene organochlorine	Organochlorine	GABA-gated chloride channel blockers Nerve action {Strong evidence that action at this protein is responsible for	2	Cyclodiene OC, fipronil	GABA-R(-)	Neurotoxic					

																				insecticidal effects}									
Toxaphene	camphechlor		8001-35-2	C=C1C(CCl)(CCl)C2(C(C(C1(C2(Cl)Cl)Cl)Cl)Cl)Cl	-0,908030978	0,123585599	50.4	µg/org	Ecotox	Insecticide, Acaricide, Veterinary substance	Chlorinated hydrocarbon	A non-systemic contact and stomach insecticide. It acts on the neurons, causing an imbalance in sodium and potassium ions, similar to that of the cyclodiene insecticides.	Organochlorine/cyclodiene organochlorine	**GABA-gated chloride channel blockers_GABA-R(-)**	Classified as "organochlorine insecticides" on AlanWood similar to that of the cyclodiene insecticides LINDANE							Cyclodiene OC, fiproni l	GABA-R(-)	Neurotoxic					
ethametsulfuron-methyl	ethametsulfuron-methyl		97780-06-8	O=C(OC)c1ccccc1S(=O)(=O)NC(=O)Nc2nc(nc(n2)NC)OCC	-1,948248697	0,011265 5	4.62	µg/org	PPDB	Herbicide	Triazinylsulfonylurea	Selective. Inhibits plant amino acid synthesis - acetohydroxy acid synthase AHAS	Sulfonylurea	**Inhibition of acetolactate synthase (ALS)/acetohydroxyacid synthase (AHAS)**			Group_B_(HRAC)	Group_2_(WSSA)	Inhibition of acetolactate synthase (ALS) also named acetohydroxyacid synthase (AHAS)										

Chemical Name	PPDB_Common Name	Other names	CAS Number	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Uniqueconcatenate	Original Unit_Uniqueconcatenate	Source dataset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical group	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo 2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action
chlorimuron-ethyl	chlorimuron-ethyl	2-[[[[(4-Chloro-6-methoxy-2-pyrimidinyl)amino]carbonyl]amino]sulfonyl]benzoic acid, Ethyl ester	90982-32-4	O=C(OCC)c1ccccc1S(=O)(=O)NC(=O)Nc2nc(OC)cc(n2)Cl	-1,5201323744	0,030190303	12.5	µg/org	PPDB | Ecotox	Herbicide	Sulfonylurea	Inhibits plant amino acid synthesis - acetohydroxyacid synthase AHAS	Sulfonylurea	**Inhibition of acetolactate synthase (ALS) /acetohydroxyacid synthase (AHAS)**			Group_B_(HRAC)	Group_2_(WSSA)	Inhibition of acetolactate synthase (ALS) also named acetohydroxyacid synthase (AHAS)							
nicosulfuron	nicosulfuron		111991-09-4	O=C(Nc1nc(OC)cc(n1)OC)NS(=O)(=O)c2nccc2(C(=O)N(C)C)	-0,732077081	0,185320303	76.0	µg/org	PPDB	Herbicide	Sulfonylurea	Selective, systemic absorbed by foliage and roots and translocated. Inhibits plant amino acid synthesis - acetohydroxyacid	Sulfonylurea	**Inhibition of acetolactate synthase (ALS) /acetohydroxyacid synthase**			Group_B_(HRAC)	Group_2_(WSSA)	Inhibition of acetolactate synthase (ALS) also named acetohydroxyacid synthase (AHAS)							

												synthase AHAS		**(AHAS)**												
2-[1-(Ethoxyimino)butyl]-5-[2-(ethylthio)propyl]-3-hydroxy-2-cyclohexen-1-one	sethoxydim		74051-80-2	O=C1C(=C(O)CC(C1)CC(C)SCC)C(=NOCC)CCC	-1,5147957594	0,030563636	10.0	µg/org	Ecotox	Herbicide	Cyclohexanedione	Selective, systemic absorbed mainly by foliage. An acetyl CoA carboxylase inhibitor (ACCase).	Cyclohexanedione (DIMs)	**Inhibition of acetyl-CoA carboxylase (ACCase)**			Group_A_(HRAC)	Group_1_(WSSA)	Inhibition of acetyl-CoA carboxylase (ACCase)								
chlorthiamid	chlorthiamid		1918-13-4	NC(c1c(ccc1Cl)Cl)=S	-0,423472615	0,377161615	77.3	µg/org	PPDB	Herbicide	Benzonitrile	Systemic, absorbed by roots and leaves, acting against germinating seeds. Inhibits cell wall synthesis	Nitrile	**Inhibition of cell wall (cellulose) synthesis**			Group_L_(HRAC)	Group_22_(WSSA)	Inhibition of cell wall (cellulose) synthesis								
bensulide	bensulide	Phosphorodithioic acid, O,O-Bis(1-methylethyl) S-[2-[(phenylsulfonyl)amino]ethyl] ester	741-58-2	O=S(=O)(NCCSP(OC(C)C)(OC(C)C)=S)c1ccccc1	-1,218645452	0,060442	24.0	µg/org	PPDB | Ecotox	Herbicide	Phosphorodithioate	Selective, acts by inhibiting germination, absorbed through leaves and roots	Phosphorodithionate	**Inhibition of lipid synthesis – not ACCase**			Group_N_(HRAC)	Group_8_(WSSA)	Inhibition of lipid synthesis – not ACCase								
N,N-Bis(2-methylpropyl)carbamothioic acid S-ethyl ester	butylate		2008-41-5	O=C(N(CC(C)C)CC(C)C)SCC	-0,8743619	0,133548 2	29.0	µg/org	Ecotox	Herbicide	Thiocarbamate	Selective, systemic absorbed by roots and coleoptiles. Lipid synthesis inhibitor.	Thiocarbamate	**Inhibition of lipid synthesis – not ACCase**			Group_N_(HRAC)	Group_8_(WSSA)	Inhibition of lipid synthesis – not ACCase								

					07																							
oryzalin	oryzalin		19044-88-3	O=[N+]([O-])c1cc(cc(c1N(CCC)CCC)[N+](=O)[O-])S(=O)(=O)N	-0,9285347922	0,1178868	40.8	µg/org	PPDB	Herbicide	Dinitroaniline	Selective affecting physiological growth process. Microtubule assembly inhibition.	Dinitroaniline	**Inhibition of microtubule assembly**			Group_K1_(HRAC)	Group_3_(WSSA)	Inhibition of microtubule assembly									
2-(Difluoromethyl)-4-(2-methylpropyl)-6-(trifluoromethyl)-3,5-pyridinedicarbothioic acid, S3,S5-Dimethyl ester	dithiopyr		97886-45-8	O=C(c1c(nc(c(C(=O)SC)c1CC(C)C)C(F)(F)F)C(F)F)SC	-0,6947181117	0,2019677	81.0	µg/org	Ecotox	Herbicide	Pyridine compound	Inhibits cell division - a mitotic inhibitor of normal cell division in susceptible plants. Thought to alter microtubule polymerization.	Pyridine	**Inhibition of microtubule assembly**			Group_K1_(HRAC)	Group_3_(WSSA)	Inhibition of microtubule assembly									
fluchloralin	fluchloralin		33245-39-5	O=[N+]([O-])c1cc(cc(c1N(CCC)CCCl)[N+](=O)[O-])C(F)(F)F	-0,5680239094	0,2703809	96.0	µg/org	PPDB	Herbicide	Chloroaniline	Inhibition of microtubule formation inhibiting cell division	Dinitroaniline	**Inhibition of microtubule assembly**			NA	NA	NA									

propham	propham		122-42-9	O=C(OC(C)C)Nc1ccccc1	-1,0489625778	0,089333822	16.0	µg/org	PPDB	Herbicide, Plant growth regulator	Carbamate	Selective, systemic, mitosis inhibitor	Carbamate	**Inhibition of mitosis/microtubule organization**			Group_K2_(HRAC)	Group_23_(WSSA)	Inhibition of mitosis/microtubule organization								
chlorpropham	chlorpropham		101-21-3	O=C(OC(C)C)Nc1cccc(c1)Cl	-0,345769681	0,4510558	96.1	µg/org	PPDB	Herbicide, Plant growth regulator	Carbamate	Mitosis inhibitor, absorbed predominately by roots	Carbamate	**Inhibition of mitosis/microtubule organization**			Group_K2_(HRAC)	Group_23_(WSSA)	Inhibition of mitosis/microtubule organization								

Chemical Name	PPDB_Common Name	Other names	CAS Number	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Unique concatenate	Original Unit_Unique concatenate	Sourced dataset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical group	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo 2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees
methabenzthiazuron	methabenztiazuron		18691-97-9	O=C(NC)N(c2nc1ccccc1s2)C	-2,44142464	0,00361889	0.8	µg/org	PPDB	Herbicide	Urea	Selective, absorbed through roots and leaves. Photosynthetic electron transport inhibitor at the photosystem II.	Urea/amide	**Inhibition of photosynthesis at PSII**			Group_C2_(HRAC)	Group_7_(WSSA)	(Inhibition of photosynthesis at photosystem II)										
chlorbromuron	chlorbromuron	chlorobromuron	13360-45-7	O=C(Nc1ccc(c(c1)Cl)Br)N(OC)C	-1,26120554	0,05480018	16.0	µg/org	PPDB	Herbicide	Urea	Photosynthetic electron transport inhibitor at the photosystem II. Sective, absorbed via roots and translocated	Urea/amide	**Inhibition of photosynthesis at PSII**			Group_C2_(HRAC)	Group_7_(WSSA)	(Inhibition of photosynthesis at photosystem II)										

	on				81																						
chloroxuron	chloroxuron		1982-47-4	O=C(Nc2ccc(Oc1ccc(cc1)Cl)cc2)N(C)C	-1,2584011106	0,055156	16.0	µg/org	PPDB	Herbicide	Dimethylurea	Photosynthetic electron transport inhibitor at the photosystem II receptor. Growth inhibition and chlorotic and necrotic effects on foliage	Urea/amide	**Inhibition of photosynthesis at PS II**			Group_C2_(HRAC)	Group_7_(WSSA)	(Inhibition of photosynthesis at photosystem II)									
propazine	propazine		139-40-2	n1c(nc(nc1NC(C)C)Cl)NC(C)C	-1,155922969	0,069835 56	16.0	µg/org	PPDB	Herbicide	Triazine	Selective, systemic, absorbed by roots and translocated	Triazine	**Inhibition of photosynthesis at PS II**			Group_C1_(HRAC)	Group_5_(WSSA)	Inhibition of photosynthesis at PS II									
metoxuron	metoxuron		19937-59-8	O=C(Nc1ccc(OC)c(c1)Cl)N(C)C	-1,153941 62	0,070155	16.0	µg/org	PPDB	Herbicide	Urea	Selective with contact and residual action. Photosynthetic electron transport inhibitor at the photosystem II.	Urea/amide	**Inhibition of photosynthesis at PS II**			Group_C2_(HRAC)	Group_7_(WSSA)	(Inhibition of photosynthesis at photosystem II)									
prometon	prometon	6-Methoxy-N2,N4-bis(1-methylethyl)-1,3,5-triazine-2,4-	1610-18-0	n1c(nc(nc1NC(C)C)NC(C)C)OC	-0,796186 87	0,159887	36.0	µg/org	PPDB|Ecotox	Herbicide	Methoxytriazine	Non-selective, systemic absorbed by foliage and roots, inhibits photosynthesis. Affects photosystem II competing with	Triazine	**Inhibition of photosynthesis at PS II**			Group_C1_(HRAC)	Group_5_(WSSA)	Inhibition of photosynthesis at PS II									

		diamine			34							plastoquinone and modifying electron transport processes.																	
N2,N4-Bis(1-methylethyl)-6-(methylthio)-1,3,5-triazine-2,4-diamine	prometryn	prometryn	7287-19-6	n1c(nc(nc1NC(C)C)SC)NC(C)C	-0,3917538385	0,405738384	96.69|99.0	µg/org	Ecotox|PPDB	Herbicide	Triazine	A selective, systemic, contact and residual triazine, a photosynthetic electron transport inhibitor at the photosystem II receptor site	Triazine	**Inhibition of photosynthesis at PS II**			Group_C1_(HRAC)	Group_5_(WSSA)	Inhibition of photosynthesis at PS II										
alachlor	alachlor		15972-60-8	O=C(N(c1c(cccc1CC)CC)COC)CCl	-1,2258231178	0,059453534	16.0	µg/org	PPDB	Herbicide	Chloroacetamide	Selective, systemic action absorbed by germinating shoots. Inhibition of VLCFA (inhibition of cell division)	Chloroacetamide (V1)	**Inhibition of very-long-chain fatty acid synthesis (VLCFAs)**		Chloroacetamide (V1)	Group_K3_(HRAC)	Group_15_(WSSA)	Inhibition of very-long-chain fatty acid synthesis (VLCFAs), inhibition of cell division									
pyridaben	pyridab	4-Chloro-2-(1,1-dimethylethyl)-5-	96489-	O=C1C(=C(C=NN1C(	-4,1811	6,5911E	0.024	µg/org	PPDB|E	Insectic	Pyridazin	Non-systemic with rapid knock down action and long residual activity.	Pyridazin	**Mitochondrial complex**	Classified as "Unclassified" on		21A	METI acaricides and insec	METI acaricides and insec	Mitochondrial complex I electron transport inhibitors Energy	21								

	en	[[[4-(1,1-dimethylethyl)phenyl]methyl]thio]-3(2H)pyridazinone	71-3	C)(C)C)SCc2ccc(cc2)C(C)(C)C)Cl	0542988	-05			cotox	ide, Acaricide	one	Inhibitor of mitochondrial electron transport at complex I.	one	**Electron transport inhibitors**	AlanWood		ticides	ticides	metabolism {Good evidence that action at this protein complex is responsible for insecticidal effects}										
Chemical Name	**PPDB_Common Name**	**Other names**	**CASNumber**	**SMILES (VEGA)**	**Exp_log**	**Exp_Value**	**Original Value_Unique concatenate**	**Original Unit_Unique concatenate**	**Source dataset**	**PPDB_Pesticide type**	**PPDB_Substance group**	**PPDB_Mode of action**	**Harmonised Chemical group**	**Harmonised MoA**	**NOTE**	**IRAC_Active Ingredient**	**IRAC_No.Chemical Sub-group or exemplifying Active Ingredient**	**IRAC_Chemical Sub-group or exemplifying Active Ingredient**	**IRAC_Substance group**	**IRAC_Main Group and Primary Site of Action**	**IRAC_No. Main Group and Primary Site of Action**	**Insecticide class_Sanchez_bayo2018**	**Mode of action_Sanchez_bayo2018**	**Group_Sanchez_bayo2018**	**Sanchez_Insecticide class**	**Sanchez_Mode of action**	**Sanchez_Group**	**Sanchez_Bees**

tebufenpyrad	tebufenpyrad		119168-77-3	O=C(NCc1ccc(cc1)C(C)(C)C)c2c(c(nn2C)CC)Cl	-1,6965790811	0,020110104	6.7	µg/org	PPDB	Acaricide	Pyrazolium	A mitochondrial electron transport inhibitor, non-systemic with contact and stomach action	Pyrazolium (azole)	**Mitochondrial complex I electron transport inhibitors**		Tebufenpyrad	21A	METI acaricides and insecticides	METI acaricides and insecticides	Mitochondrial complex I electron transport inhibitors Energy metabolism {Good evidence that action at this protein complex is responsible for insecticidal effects}	21								
fenpyroximate	fenpyroximate		134098-61-6	O=C(OC(C)(C)C)c1ccc(cc1)CON=Cc3c(nn(c3(Oc2ccccc2))C)C	-1,4258314437	0,037511519	15.8	µg/org	PPDB	Acaricide, Insecticide	Pyrazolium	Mitochondrial electron transport inhibitor with contact action	Pyrazolium (azole)	**Mitochondrial complex I electron transport inhibitors**		fenpyroximate	21A	METI acaricides and insecticides	METI acaricides and insecticides	Mitochondrial complex I electron transport inhibitors Energy metabolism {Good evidence that action at this protein complex is responsible for insecticidal effects}	21								
2-(4-Methoxy[1,1'-biphenyl]-3-yl)hydrazinecarboxylic acid, 1-Methylethyl ester	bifenazate	Bifenazate	149877-41-8	O=C(OC(C)C)NNc1cc(ccc1(OC))c2ccccc2	-1,5665778111	0,027128283	7.8|8.5	µg/org	Ecotox|OpenFoodTox	Insecticide, Acaricide	Hydrazinecarboxylate	Neuronal inhibitor, non-systemic having contact and residual action	Hydrazine carboxylate	**Mitochondrial complex III electron transport inhibitors**		bifenazate	20D	METI acaricides and insecticides	METI acaricides and insecticides	Mitochondrial complex III electron transport inhibitors Energy metabolism {Good evidence that action at this protein complex is responsible for insecticidal effects}	20								
Tetrahydro-5,5-dimethyl-2(1H)-pyrimidio[3-[4-trifluoromethyl)phenyl]-1-[2-[4-(trifluoromethyl)phenyl]ethenyl]-2-propenylidene]hydrazone	hydramethylnon		67485-29-4	FC(F)(F)c3ccc(C=CC(=NN=C1NCC(C)(C)CN1)C=Cc2ccc(cc2)C(F)(F)F)cc3	-0,867819604	0,135575 52	67.0	µg/org	Ecotox	Insecticide	Unclassified	Non-systemic, selective with stomach action. Mitochondrial complex III electron transport inhibitor.	Unclassified	**Mitochondrial complex III electron transport inhibitors**		hydramethylnon	20A	METI acaricides and insecticides	METI acaricides and insecticides	Mitochondrial complex III electron transport inhibitors Energy metabolism {Good evidence that action at this protein complex is responsible for insecticidal effects}	20								

dichlofluanid	dichlofluanid		1085-98-9	O=S(=O)(N(c1ccccc1)SC(F)(Cl)Cl)N(C)C	-1,3169687 86	0,04819 82	16.0	µg/org	PPDB	Fungicide	Sulphamide	Foliar with protective action. Multi-site activity. Potent inhibitor of fungal spore germination.	Sulphamide/electrophiles	**Multi-site contact activity**		dichlofluanid	M06_FRAC	sulfamides	sulfamides (electrophiles)	Chemicals with multi-site activity. TARGET: multi-site contact activity	undefined_FRAC							

Chemical Name	PPDB_Common Name	Other names	CAS Number	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Unique concatenate	Original Unit_Unique concatenate	Sourced dataset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical group	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees	Sanchez_value
1,3-dichloropropene	1,3-dichloropropene		542-75-6	C(=CCl)CCl	-1,2217263 62	0,0600169	6.6	µg/org	PPDB	Nematicide, Bactericide	Halogenated hydrocarbon	Soil fumigant that penetrates into the nematodes via mouth and cuticle	Organochlorine	**N/A**																
3-Amino-2,5-dichlorobenzoic acid ammonium salt (1:1)			1076-46-6	O=C(O)c1cc(cc(N)c1	-1,1849777	0,0653164	14.5	µg/org	Ecotox	Unclassified	Unclassified		Unclassified	**N/A**																

				Cl)Cl																									
tribufos	tribufos		78-48-8	O=P(SCCCC)(SCCCC)SCCCC	-0,851640365	0,140721212	44.2	µg/org	PPDB	Plant growth regulator, Herbicide	Organophosphate	Absorbed by leaves and induces early leaf abscission through changes in the levels of plant hormones	Organophosphate	**N/A**			NA	NA	NA										
pentachlorophenol	pentachlorophenol		87-86-5	Oc1c(c(c(c(c1Cl)Cl)Cl)Cl)Cl	-0,740110928	0,181923636	48.0	µg/org	PPDB	Insecticide, Herbicide, Fungicide, Molluscicide, Plant growth regulator, Wood preservative	Organochlorine	Accelerates aerobic metabolism and increases heat production	Organochlorine	**N/A**	Classified as "organochlorine insecticides". OR "aromatic fungicides". OR "unclassified herbicides" from AlanWood												
3,5,6-trichloro-2-pyridinol	3,5,6-trichloro-2-pyridinol		6515-38-4	Oc1nc(c(cc1Cl)Cl)Cl	-0,715651051	0,192463638	37.9	µg/org	PPDB	Herbicide, Metabolite	Unclassified	Not applicable	Unclassified	**N/A**	Metabolite of Chlorpyrifos and Triclopyr												
6-Methyl-1,3-dithiolo[4,5-	chinomethi		2439-01	O=C2Sc1nc3ccc	-0,546565	0,284006	66.47	µg/org	Ecotox	Fungicide, Acaricide, Miticide,	Carbamate	Selective, non-systemic contact action with protective	Unclassified	**N/A**	"Carbamate" from PPDB not appropriate. IRAC classifies it as "Compounds of unknown or uncertain MoA". FRAC classifies as "Chemicals with multi-site activity"												

b]quinoxalin-2-one	onat		-2	c(cc3(nc1S2))C	756188	94				Insecticide		and curative properties. Multi-site activity.																			
2-(Octylthio)ethanol	2-(octylthio)ethanol		3547-33-9	OCSCCCCCCCC	-0,523959225	0,29925445	56.9	µg/org	Ecotox	Insecticide, Insect repellent	Unclassified	Repellent	Unclassified	**N/A**	Tobechecked																	
Clothianidin	clothianidin	[C(E)]-N-[(2-Chloro-5-thiazolyl)methyl]-N'-methyl-N"-nitroguanidine|clothianidin	210880-92-5	O=[N+]([O-])N=C(NC)NCc1cnc(Cl)s1	-3,828885078	0,0001483	0.0275|0.035|0.0439|0.044	µg/org	OpenFoodTox|Ecotox|PPDB	Insecticide, Metabolite	Neonicotinoid	Translaminar and root systemic activity. Acetylcholine receptor (nAChR) agonist.	Neonicotinoid	**Nicotinic acetylcholine receptor (nAChR) competitive modulators_nACh-R(+)**		clothianidin	4A	neonicotinoid	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) competitive modulators Nerve action {Strong evidence that action at one or more of this class of protein is responsible for insecticidal effects}	4	Neonicotinoids, spinosad	nACh-R(+)	Neurotoxic	Neonicotinoid	nACh-R(+)	Neurotoxic	0.13	µg/bee
N"-Methyl-N-nitro-N'-[(tetrahy	dinotefuran		165252-7	O=[N+]([O-])N	-3,633348	0,000023	0.047	µg/org	Ecotox	Insecticide	Neonicotin	Systemic, with contact and stomach action, effects insects	Neonicotinoid	**Nicotinic acetylcholine receptor (nAC**		dinotefuran	4A	neonicotinoid	Neonicotin	Nicotinic acetylcholine receptor (nAChR) competitive modulators Nerve action {Strong	4	Neonicotinoids, spino	nACh-R(+)	Neurotoxic	Neonicotin	nACh-R(+)	Neurotox	0.13	µg/bee

dro-3-furanyl)methyl]guanidine			0-0	=C(NC)NCC1COCC1	2617	26					oid	nervous system. Nicotinic Acetylcholine receptor agonist /antagonist.		**hR) competitive modulators_nAChR-R(+)**					oid	evidence that action at one or more of this class of protein is responsible for insecticidal effects}		sad			oid		ic		

ChemicalName	PPDB_CommonName	Othernames	CASNumber	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Uniqueconcatenate	Original Unit_Uniqueconcatenate	Sourcedataset	PPDB_Pesticidetype	PPDB_Substancegroup	PPDB_Mode of action	Harmonised Chemicalgroup	Harmonised MoA	NOTE		IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees
sulfoxaflor	sulfoxaflor		946578-00-3	N#CN=S(=O)(C)C(c1cnc(cc1)C(F)(F)F)C	-2,863918423	0,001368	0.379	µg/org	PPDB	Insecticide	Sulfoximine	An agonist of n-acetylcholine receptors in insects. Systemic.	Sulfoximine	**Nicotinic acetylcholine receptor (nAChR) competitive modulators_nACh-R(+)**	Not in Sanchez-bayoas compound		4C	Sulfoximines	Sulfoximines	Nicotinic acetylcholine receptor (nAChR) competitive modulators Nerve action {Strong evidence that action at one or more of this class of protein is responsible for insecticidal effects}	4								
acetamiprid	acetamiprid		135410-20-7	N#CN=C(N(C)Cc1cnc(cc1)Cl)C	-1,438535942	0,036430304	8.09	µg/org	PPDB	Insecticide	Neonicotinoid	Systemic with translaminar activity having both contact and stomach action. Acetylcholine receptor (nAChR) agonist.	Neonicotinoid	**Nicotinic acetylcholine receptor (nAChR) competitive modulators_**		acetamiprid	4A	neonicotinoid	Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) competitive modulators Nerve action {Strong evidence that action at one or more of this class of protein is responsible	4	Neonicotinoids, spinosad	nACh-R(+)	Neurotoxic	Neonicotinoid	nACh-R(+)	Neurotoxic	0.13	µg/bee

													nACh-R(+)						for insecticidal effects}									
diniconazole	diniconazole		83657-24-3	OC(C(=Cc1ccc(cc1Cl)Cl)n2ncnc2)C(C)(C)C	-1,2109533939	0,061524343	20.0	µg/org	PPDB	Fungicide	Triazole	Systemic with curative and protective action. Inhibits the demethylation of steroids distrupting ergosterol biosynthesis.	Triazole	**SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)**		diniconazole	3_FRAC	triazoles	DMI-fungicides (DeMethylation Inhibitors) (SBI: Class I)	Sterol biosynthesis in membranes. TARGET: C14-demethylase in sterol biosynthesis (erg11/cyp51)	G1_FRAC								
nuarimol	nuarimol		63284-71-9	Fc1ccc(cc1)C(O)(c2cncnc2)c3ccccc3Cl	-1,162561944	0,068776762	21.6	µg/org	PPDB	Fungicide	Pyrimidine	Systemic with curative and protective action. Sterol demethylation (ergosterol biosynthesis) inhibitor.	Pyrimidine	**SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)**		nuarimol	3_FRAC	pyrimidines	DMI-fungicides (DeMethylation Inhibitors) (SBI: Class I)	Sterol biosynthesis in membranes. TARGET: C14-demethylase in sterol biosynthesis (erg11/cyp51)	G1_FRAC								
imazalil	imazalil		35554-44-0	n1ccn(c1)CC(OCC=C)c2ccc(cc2Cl)Cl	-0,880297992	0,131735353	39.0	µg/org	PPDB	Fungicide, Veterinary substance	Imidazole	Systemic with curative and protective properties. Disrupts membrane function.	Imidazole	**SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)**		imazalil	3_FRAC	imidazoles	DMI-fungicides (DeMethylation Inhibitors) (SBI: Class I)	Sterol biosynthesis in membranes. TARGET: C14-demethylase in sterol biosynthesis (erg11/cyp51)	G1_FRAC								
tet	tetra		112	FC(F)C(F)	-0,	0,1	63.0	µg/org	PP	Fung	Triaz	Systemic with protectant, eradicant and	Triaz	**SBI: Class I_C14**		tet	3_F	triazoles	DMI-fungicides	Sterol biosynthesis in membranes.	G1_FRAC								

raconazole	conazole		281-77-3	(F)OCC(c1ccc(cc1Cl)Cl)Cn2ncnc2	7700585563	6980015			DB	icide	ole	curative properties. Sterol biosynthesis inhibitor, acts mainly on the vegetative stages of fungi by blocking the mycelial growth either inside or on the surface of the host plant.	ole	**-demethylase in sterol biosynthesis (erg11/cyp51)**		raconazole	RAC		(DeMethylation Inhibitors) (SBI: Class I)	TARGET: C14-demethylase in sterol biosynthesis (erg11/cyp51)								

Chemical Name	PPDB_Common Name	Other names	CAS Number	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Unique concatenate	Original Unit_Unique concatenate	Source dataset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical group	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action		Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo 2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees
spiroxamine	spiroxamine	8-(1,1-Dimethylethyl)-N-ethyl-N-propyl-1,4-dioxaspiro[4.5]decane-2-methanamine	118134-30-8	O1CC(OC12(CCC(CC2)C(C)(C)C))CN(CC)CCC	-1,848865507	0,014162 3	4.2 | 4.22	µg/org	PPDB | Ecotox	Fungicide	Morpholine	Systemic with protective, curative and eradicative action. Disrupts membrane function. Inhibits sterol biosynthesis in membranes.	Morpholine/spiroketal-amines	**SBI: Class II_Δ14-reductase and Δ8 to Δ7-isomerase in sterol biosynthesis (erg24, erg2)**		spiroxamine	5_FRAC	spiroketal-amines	amines ("morpholines") (SBI: Class II)	Sterol biosynthesis in membranes. TARGET: Δ14-reductase and Δ8 to Δ7-isomerase in sterol biosynthesis (erg24, erg2)	G2_FRAC								
pyrethrins (ja	pyrethrins (jasm		1172-63-0	O=C(OC)C(=CC2C(C(=O)OC1C(=C(C(=O)C1)C)C=CCC)C)C	-4,459171712	3,4744E-05	0.013	µg/org	PPDB	Insecticide, Acaricide, Veterinar	Plant derived	Contact acting neurotoxins with repellent properties. Sodium channel modulator.	Pyrethroid/Plant derived	**Sodium channel modulators_Na channel(+)**	Classified as "botanical insecticides" even on alanwood. They can be considered as "natural pyrethroids" to distinguish with "syntetic pyrethroids"			pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is	3							

smolinII)	olinII)			2(C)(C))C	52					y substance			/Pyrethrin					responsible for insecticidal effects}									
pyrethrins(pyrethrinII)	pyrethrins (pyrethrin II)		121-29-9	O=C(OC)C(=CC2C(C(=O)OC1C(=C(C(=O)C1)CC=CC=C)C)C2(C)(C))C	-4,45682563 5	3,493E-05	0.013	µg/org	PPDB	Insecticide, Acaricide, Veterinary substance	Plantderived	Contact acting neurotoxins with repellent properties. Sodium channel modulator.	Pyrethroid/Plantderived/Pyrethrin	**Sodium channel modulators_Na channel(+)**	Classified as "botanical insecticides" even on alanwood. They can be considered as "natural pyrethroids" to distinguish with "syntetic pyrethroids"	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3								
pyrethrins(cinerinII)	pyrethrins (cinerin II)		121-20-0	O=C(OC)C(=CC2C(C(=O)OC1C(=C(C(=O)C1)CC=CC)C)C2(C)(C))C	-4,44259272 9	3,609E-05	0.013	µg/org	PPDB	Insecticide, Acaricide, Veterinary substance	Plantderived	Contact acting neurotoxins with repellent properties. Sodium channel modulator.	Pyrethroid/Plantderived/Pyrethrin	**Sodium channel modulators_Na channel(+)**	Classified as "botanical insecticides" even on alanwood. They can be considered as "natural pyrethroids" to distinguish with "syntetic pyrethroids"	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3								
pyreth	pyrethrin		4466-1	O=C(OC1C(=C(C(=O)C1)CC=CCC)C	-4,4044	3,937E	0.013	µg/org	PPDB	Insecticide, Acari	Plantderiv	Contact acting neurotoxins with repellent properties	Pyrethroid/Pl	**Sodium channel modulators_**	Classified as "botanical insecticides" even on alanwood. They can be considered as "natural	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong	3								

rins (jasmolin I)	s (jasmolin I)		4-2	)C2C(C=C(C)C)C2(C)(C)	8593556	-05				cide, Veterinary substance	ed	. Sodium channel modulator.	ant derived/Pyrethrin	**Na channel(+)**	pyrethroids" to distinguish with "syntetic pyrethroids"				evidence that action at this protein is responsible for insecticidal effects}												
pyrethrins (pyrethrin I)	pyrethrins (pyrethrin I)		121-21-1	O=C(OC1C(=C(C(=O)C1)CC=CC=C)C)C2C(C=C(C)C)C2(C)(C)	-4,402200312	3,9611E-05	0.013	µg/org	PPDB	Insecticide, Acaricide, Veterinary substance	Plant derived	Contact acting neurotoxins with repellent properties. Sodium channel modulator.	Pyrethroid/Plant derived/Pyrethrin	**Sodium channel modulators_Na channel(+)**	Classified as "botanical insecticides" even on alanwood. They can be considered as "natural pyrethroids" to distinguish with "syntetic pyrethroids"			pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3								
Etofenprox	etofenprox		80844-07-1	O(c1ccccc1)c2ccccc(c2)COCC(c3ccc(OCC)cc3)(C)C	-4,39933197 1	3,9871E-05	0.015	µg/org	OpenFoodTox	Insecticide	Pyrethroid	Broad spectrum with contact and stomach action. Sodium channel modulator.	Pyrethroid/Synthetic	**Sodium channel modulators_Na channel(+)**		etofenprox	3A	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3	Pyrethroids, DDT	Nachannel (+)	Neurotoxic	Pyrethroid	Nachannel (+)	Neurotoxic	0.4	µg/bee
pyre	pyreth		2540	O=C(OC1C(=C(C(=O)C	-4,3	4,11	0.013	µg/org	PPDB	Insecticide,	Plantd	Contact acting neurotoxins with	Pyroi	**Sodium channel**	Classified as "botanical insecticides" even on alanwood. They			pyrethroids	Pyrethroid	Sodium channel modulators Nerve	3								

thrins(cinerinI)	rins(cinerinI)		2-06-6	1)CC=CC)C)C2C(C=C(C)C)C2(C)(C)	860237944	1E-05				Acaricide, Veterinary substance	erived	repellent properties. Sodium channel modulator.	d/Plant derived/Pyrethrin	**modulators_Na channel(+)**	can be considered as "natural pyrethroids" to distinguish with "syntetic pyrethroids"			pyrethrin		action {Strong evidence that action at this protein is responsible for insecticidal effects}									
permethrin	permethrin	3-(2,2-Dichloroethenyl)-2,2-dimethylcyclopropanecarboxylic acid, (3-Phenoxyphenyl)methyl ester	52645-53-1	O=C(OCc2cccc(Oc1ccccc1)c2)C3C(C=C(Cl)Cl)C3(C)(C)	-4,21094127 3	6,153E-05	0.024	µg/org	PPDB|Ecotox	Insecticide, Veterinary substance	Pyrethroid	Broad spectrum with contact and stomach action. Slight repellant effect. Sodium channel modulator.	Pyrethroid/Synthetic	**Sodium channel modulators_Na channel(+)**		permethrin	3A	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3	Pyrethroids, DDT	Nachannel(+)	Neurotoxic	Pyrethroid	Nachannel(+)	Neurotoxic	0.4	µg/bee
prallethrin	prallethrin	2,2-Dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid 2-methyl-4-oxo-3-(2-propynyl)-2-cyclopenten-1-yl ester	23031-36-9	O=C(OC1C(=C(C(=O)C1)CC#C)C)C2C(C=C(C)C)C2(C)(C)	-4,04630527 7	8,9889E-05	0.026|0.028	µg/org	PPDB|Ecotox	Insecticide, Veterinary substance	Synthetic pyrethroid	Contact, with rapid knock down effect, acts on the insects nervous system	Pyrethroid/Synthetic	**Sodium channel modulators_Na channel(+)**	Classified as "Synthetic pyrethroid" in PPDB. I would consider as	prallethrin	3A	pyrethroids pyrethrin	Synthetic pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3	Pyrethroids, DDT	Nachannel(+)	Neurotoxic					
acrinath	acrinath		10100	N#CC(OC(=O)C1C(C=CC(=O)O	-3,800	0,000	0.084	µg/org	PPDB	Insecticide, Ac	Pyrethr	Contact and stomach action.	Pyrethroid/	**Sodium channel modul**	Not in Sanchez-bayo as compound		3A	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action	3	Pyrethroi	Nachannel(+)	Neurotoxic	Pyrethroid	Nachan	Neuroto	0.4	µg/b

hrin	rin		7-06-1	C(C(F)(F)F)C(F)(F)F)C1(C)(C))c3cccc(Oc2ccccc2)c3	902424245	1552				aricide	oid	Neurotoxin.	Synthetic	**ators_Na channel(+)**					{Strong evidence that action at this protein is responsible for insecticidal effects}		ds, DDT				nel(+)	xic		e e
resmethrin	resmethrin		10453-86-8	O=C(OCc1coc(c1)Cc2ccccc2)C3C(C=C(C)C)C3(C)(C)	-3,72981789 4	0,0001863	0.063	µg/org	PPDB	Insecticide	Pyrethroid	Non-systemic with contact action. Sodium channel modulator.	Pyrethroid/ Synthetic	**Sodium channel modulators_Na channel(+)**		resmethrin	3A	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3	Pyrethroids, DDT	Nachannel (+)	Neurotoxic	Pyrethroid	Nachannel(+)	Neurotoxic	0.4	µg/bee

Chemical Name	PPDB_CommonName	Othernames	CASNumber	SMILES (VEGA)	Exp_log	Exp_Value	Original Value_Unique conc at enate	Original Unit_Unique conc at enate	Sourced dataset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical group	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No.Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees	Sanchez_value
2,2-Dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid (3-phenoxyphenyl)methyl ester	phenothrin		26002-80-2	O=C(OCc2cccc(Oc1ccccc1)c2)C3C(C=C(C)C)C3(C)(C)	-3,718226699	0,000191313	0.067	µg/org	Ecotox	Insecticide, Veterinary substance	Pyrethroid	Non-systemic with rapid contact and stomach action	Pyrethroid/Synthetic	**Sodium channel modulators_Nachannel(+)**		phenothrin	3A	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3	Pyrethroids, DDT	Nachannel(+)	Neurotoxic	Pyrethroid	Nachannel(+)	Neurotoxic	0.4	µg/bee
2,2-Dimethyl-3-(1,1,2,2-tetrabromoethyl)cyclopropanecarboxylic acid cyano(3-phenoxyphenyl)methyl ester	tralomethrin	tralomethrin	66841-25-6	N#CC(OC(=O)C1C(C(C(Br)(Br)Br)Br)C1(C)(C))c3ccccc(Oc2ccccc2)c3	-3,707809971	0,000196	0.129|0.13	µg/org	Ecotox|PPDB	Insecticide; Veterinary substance	Pyrethroid	Non-systemic with contact and stomach action	Pyrethroid/Synthetic	**Sodium channel modulators_Nachannel(+)**		tralomethrin	3A	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3	Pyrethroids, DDT	Nachannel(+)	Neurotoxic	Pyrethroid	Nachannel(+)	Neurotoxic	0.4	µg/bee

(1,3,4,5,6,7-Hexahydro-1,3-dioxo-2H-isoindol-2-yl)methyl ester 2,2-dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid	tetramethrin		7696-12-0	O=C(OCN1C(=O)C2=C(C1(=O))CCCC2)C3C(C=C(C)C)C3(C)	-3,329730251	0,000468	0.155	µg/org	Ecotox	Insecticide	Pyrethroid	Non-systemic with rapid contact action. Sodium channel modulator.	Pyrethroid/Synthetic	**Sodium channel modulators_Na channel(+)**		tetramethrin	3A	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3	Pyrethroids, DDT	Na channel(+)	Neurotoxic	Pyrethroid	Na channel(+)	Neurotoxic	0.4	µg/bee
momfluorothrin	momfluorothrin		609346-29-4	N#CC(=CC2C(C(=O)OCc1c(F)c(F)c(c(F)c1(F))CO C)C2(C)(C))C	-3,284577474	0,000519 3	0.2	µg/org	PPDB	Insecticide	Pyrethroid	Axonic poison; keeps sodium channels in the neuronal membranes open, leading to hyperexcitation and paralysis.	Pyrethroid/Synthetic	**Sodium channel modulators_Na channel(+)**	Not in Sanchez-bayo as compound NOR in IRAC.												
tefluthrin	tefluthrin		79538-32-2	O=C(OCc1c(F)c(c(F)c(c(F)c1(F))C)C2C(C=C(C(F)(F)F)Cl)C2(C)(C)	-3,174077525	0,000669 8	0.28	µg/org	PPDB	Insecticide	Pyrethroid	Contact and respiratory action with some repellant effects. Sodium channel modulator.	Pyrethroid/Synthetic	**Sodium channel modulators_Na channel(+)**		tefluthrin	3A	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3	Pyrethroids, DDT	Na channel(+)	Neurotoxic	Pyrethroid	Na channel(+)	Neurotoxic	0.4	µg/bee
imiprothrin	imiproth		72963-	O=C(OCN1C(=O)N(CC#C)CC1(	-2,9000	0,0012	0.4	µg/org	PPDB	Insecticide	Pyrethroid	Similar to other synthetic pyrethoids, acts by over	Pyrethroid/Sy	**Sodium channel mo**		imiproth	3A	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong	3	Pyrethroids, DDT	Na channel(+)	Neurotoxic	Pyrethroid	Na chan	Neurot	0.4	µg/bee

	rin		72-5	=O))C2C(C=C(C)C)C2(C)(C)	582798	572						stimulation of the nervous system. Sodium channel modulator.	nthetic	**dulators_Nachannel(+)**		rin				evidence that action at this protein is responsible for insecticidal effects}						el(+)	oxic		
DDT	DDT		50-29-3	c1cc(ccc1C(c2ccc(cc2)Cl)C(Cl)(Cl)Cl)Cl	-2,8140436 35	0,001534 5	0.54	µg/org	PPDB	Insecticide	Organochlorine	Non-systemic stomach and contact action. Sodium channel modulator.	Organochlorine	**Sodiumchannelmodulators_Nachannel(+)**		DDT	3B	ddtmethoxychlor	Organochlorine	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3	Pyrethroids, DDT	Nachannel(+)	Neurotoxic					

Chemical Name	PPDB_Common Name	Other names	CAS Number	SMILES (VEGA)	Exp_log	Exp_Value	Original Value	Original Unit_Unique concatenate	Source dataset	PPDB_Pesticide type	PPDB_Substance group	PPDB_Mode of action	Harmonised Chemical gorup	Harmonised MoA	NOTE	IRAC_Active Ingredient	IRAC_No.Chemical Sub-group or exemplifying Active Ingredient	IRAC_Chemical Sub-group or exemplifying Active Ingredient	IRAC_Substance group	IRAC_Main Group and Primary Site of Action	IRAC_No. Main Group and Primary Site of Action	Insecticide class_Sanchez_bayo2018	Mode of action_Sanchez_bayo2018	Group_Sanchez_bayo2018	Sanchez_Insecticide class	Sanchez_Mode of action	Sanchez_Group	Sanchez_Bees	Sanchez_value
bioallethrin	allethrin	2,2-Dimethyl-3-(2-methyl-1-propenyl)cyclopropanecarboxylic acid 2-methyl-4-oxo-3-(2-propenyl)-2-cyclopenten-1-yl ester	584-79-2	O=C(OC1C(=C(C(=O)C1)C=C)C)C2C(C=C(C)C)C2(C)(C)	-1,9487985777	0,01125133	3.4	µg/org	PPDB|Ecotox	Insecticide, Veterinary substance	Pyrethroid	Stomach and respiratory action, non-systemic, paralyses insects before killing them.	Pyrethroid/Synthetic	**Sodium channel modulators_Na channel(+)**			3A	pyrethroids pyrethrin	Pyrethroid	Sodium channel modulators Nerve action {Strong evidence that action at this protein is responsible for insecticidal effects}	3	Pyrethroids, DDT	Na channel(+)	Neurotoxic	Pyrethroid	Na channel(+)	Neurotoxic	0.4	µg/bee
1,1'-(2,2,2-Trichloroethylidene)bis[4-metho	methoxychl		72-43-5	O(c1ccc(cc1)C(c2ccc(OC)cc2)C(Cl)(Cl)Cl)C	-1,16421627	0,06851447	23.57	µg/org	Ecotox	Insecticide, Veterin	Organochlori	Contact and stomach action. Central nervous stimulant, producing hyperactivity,	Organochlorine	**Sodium channel modulators_Na**		methoxychlor	3B	ddt methoxychlor	Organochlorine	Sodium channel modulators Nerve action {Strong evidence that action at this protein is	3	Pyrethroids, DDT	Na channel(+)	Neurotoxic					

xybenzene]	or				35					arysubstance	ne	convulsions and death.		**channel(+)**						responsible for insecticidal effects}									
4-(2,4-Dichlorophenoxy)butanoic acid	2,4-DB		94-82-6	O=C(O)CCCOc1ccc(cc1Cl)Cl	-1,2330849044	0,058467676	14.5	µg/org	Ecotox	Herbicide	Aryloxyalkanoic acid	Selective, systemic. Synthetic auxin causing stem and leaf malformations leading to death.	Phenoxy-carboxylic acid, benzoic acid, pyridine carboxylic acid, quinolone carboxylic acid, and other	**Synthetic auxins (action like indole acetic acid)**			Group_O_(HRAC)	Group_4_(WSSA)	Synthetic auxins (action like indole acetic acid)										
4-Bromo-2-(4-chlorophenyl)-1-(ethoxymethyl)-5-(trifluoromethyl)-1H-pyrrole-3-carbonitrile	chlorfenapyr		122453-73-0	N#Cc2c(c1ccc(cc1)Cl)n(c(c2Br)C(F)(F)F)COCC	-3,529312201	0,000029556	0.12	µg/org	Ecotox	Insecticide, Acaricide, Miticide	Pyrrole	Limited systemic activity, mainly stomach but some contact action. Uncoupler of oxidative phosphorylation.	Pyrrole (azole)	**Uncoupler of oxidative phosphorylation**			13	Pyrroles Dinitrophenols Sulfluramid	Pyrroles Dinitrophenols Sulfluramid	Uncouplers of oxidative phosphorylation via disruption of the proton gradient Energy metabolism	13								
2-(1-Methylpropyl)-4,6-dinitro	dinos		88-85	O=[N+]([O-])c1cc(c(O	-0,87	0,134	32.26	µg/org	Ecot	Herbici	Dinitro	Contact. Inhibits respiratory electron transport	Dinitrophenol	**Uncoupling (membr**			Group_M_(HRAC)	Group_24_(WSSA)	Uncoupling (membrane disruption)									

phenol	eb		-7	)c(c1)C(C)CC)[N+](=O)[O-]	1681899	3749			ox	de	phenol	and is an uncoupler of oxidative phosphorylation via disruption of proton gradient - membrane disruption.		**anedisruption)**															
indoxacarb	indoxacarb	(4aS)-7-Chloro-2,5-dihydro-2-[[(methoxycarbonyl)[4-(trifluoromethoxy)phenyl]amino]carbonyl]indeno[1,2-e][1,3,4]oxadiazine-4a(3H)-carboxylic acid methyl ester	173584-44-6	O=C(OC)N(C(=O)N2N=C3c1ccc(cc1CC3(OC2)(C(=O)OC))Cl)c4ccc(OC(F)(F)F)cc4	-3,607668705	0,0002468	0.094|0.18	µg/org	PDB|Ecotox	Insecticide	Oxadiazine	Contact and stomach action. Voltage-dependent sodium channel blocker.	Oxadiazine	**Voltage-dependent sodium channel blocker_Nachannel(-)**		indoxacarb	22A	oxadiazine	Oxadiazine	Voltage-dependent sodium channel blockers Nerve action {Good evidence that action at this protein complex is responsible for insecticidal effects}	22	Indoxacarb	Nachannel(-)	Neurotoxic	Indoxacarb	Nachannel(-)	Neurotoxic	

Acknowledgements

This work was funded by the European Food Safety Authority (EFSA) under the contract number NP/EFSA/AFSCO/2016/02 (Edoardo Carnesecchi), and OC/EFSA/SCER/2018/01 (Mario Negri Institute for Pharmacological Research, S-IN Soluzioni Informatiche, Molecular Networks GmbH and Altamira); and the European Commision under the contract number LIFE16 ENV/IT/000167 "LIFE-VERMEER". The views expressed in this manuscript do not reflect the views of the European Food Safety Authority and are the authors only.

I gratefully acknowledge my promoter Prof. Juliette Legler and co-promote dr. Nynke Kramer for the constructive and helpful discussions throughout the past three years. I would like to express my sincere gratitude to my advisors dr. Emilio Benfenati and dr. Jean Lou Dorne for the continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge – "merci beaucoup my guru"!

Finally, to my best friends Cesa, Fede, Ilaria, Meo, Mirko and Popi: you should know that your support and encouragement was worth more than I can express on paper.

To my parents, Claudio and Tiziana, my lovely family: "un grazie speciale", I could not have made it without you: tutto questo è anche vostro!

About the author

Edoardo Carnesecchi was born on February 27th, 1991, in Massa Marittima (Italy). In 2010, he graduated from secondary school at Istituto B. Lotti (Massa Marittima) and started his study Environmental Health and Safety at the University of Pisa (Pisa, Italy). In 2014, Edoardo obtained his bachelor degree and in the same year he joined the Department of Hygiene and Epidemiology of Pisa University as a research fellow. In May 2015, he joined the European Food Safety Authority (EFSA) as a trainee in the Application Desk unit, while enrolling in the master Food Science & Technology at the Univerisità Cattolica del Sacro Cuore (Milan, Italy). In 2016, he worked as Junior Scientific Officer in EFSA's Animal and Plant unit. In 2017, Edoardo obtained his master degree while working as consultant in the Scientific Committee and Emerging Risk Unit (SCER) of EFSA. From January 2018 until March 2020, he worked as a researcher at the Istituto di Ricerche Farmacologiche Mario Negri, Laboratory of Chemistry and Environmental Toxicology under the supervision of Dr. Emilio Benfenati (Milan, Italy). In this context, Edoardo carried out his PhD thesis in joint collaboration between the Mario Negri Institute, the Institute for Risk Assessment Sciences of Utrecht University, supervised by Dr. Nynke I. Kramer and Prof. Dr. ir. Juliette Legler, and EFSA, supervised by Dr. Jean Lou Dorne. The research presented in this thesis was part of an EFSA funded project entitled "Openfoodtox integrated with non-testing methods for toxicity evaluation", which was coordinated by Dr. Benfenati. During his PhD Edoardo has been awarded international prizes such as the best oral presentation at the European Commission Joint Research Centre' Summer School on "Non-animal approaches in sciences" (Ispra, 2019) and the Lush Prize for young researchers (London, 2020). In April 2020, Edoardo took up duties at the Organisation for Economic Co-operation and Development (OECD, Paris) as a Junior Policy Analyst – Chemicals working within the Environment, Health and Safety (EHS) division. Since March 2021, Edoardo has joined the Evidence Management Unit of the European Food Safety Auhtority (EFSA) as a Data Officer.

List of publications

Carnesecchi E., Toma C., Roncaglioni, A., Kramer N., Benfenati E., Dorne J.L.C.M., (2020). Integrating QSAR models predicting acute contact toxicity and mode of action profiling in honey bees (A. mellifera): Data curation using open source databases, performance testing and validation. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.139243

Carnesecchi E., Raitano G., Gamba A., Benfenati E., Roncaglioni A., (2020). Evaluation of non-commercial models for genotoxicity and carcinogenicity in the assessment of EFSA's databases. SAR and QSAR in Environmental Research. 31(1), 33-48. https://doi.org/10.1080/1062936X.2019.1690045

Carnesecchi E., Toropova A. P., Toropov A. A., Benfenati E., Dorne J.L.C.M., (2020). Predicting acute contact toxicity of organic binary mixtures in honey bees (A. mellifera) through innovative QSAR models. Science of The Total Environment, 135302. https://doi.org/10.1016/j.scitotenv.2019.135302

Carnesecchi E., Svendsen C., Lasagni S., Grech A., Quignot N., Amzal B., Toma C., Tosi S., Rortais A., Cortinas-Abrahantes J., Capri E., Kramer N., Benfenati E., Spurgeon D., Guillot G., Dorne J.L.C.M., (2019). Investigating combined toxicity of binary mixtures in bees: Meta-analysis of laboratory tests, modelling, mechanistic basis and implications for risk assessment. Environment International, 133 (2019). https://doi.org/10.1016/j.envint.2019.105256

Benfenati E., **Carnesecchi E.**, Roncaglioni A., Baldin R., Ceriani L., Ciacci A., Kovarich S., Sartori L., Mostrag A., Magdziarz T., Yang, C., (2020). Maintenance, update and further development of EFSA's Chemical Hazards: OpenFoodTox 2.0. EFSA supporting publication 2020:EN-1822. 36 pp. doi:10.2903/sp.efsa.2020.EN-1822

Como F., **Carnesecchi E.**, Volani S., Dorne J.L., Richardson J., Bassan A., Benfenati E. Predicting acute contact toxicity of pesticides in honeybees (Apis mellifera) through a k-nearest neighbor model.. (2017). Chemosphere, 166, 438-444. https://doi.org/10.1016/j.chemosphere.2016.09.092

Benfenati E., Roncaglioni A., **Carnesecchi E.**, Mazzucotelli M., Marzo M., Toropov A., Toropova A., Baldin R., Ciacci A., Hobocienski B., Kovarich S., Sartori L., Mostrag A., Magdziarz T., Yang, C., (2021). Maintenance, update and further development of EFSA's Chemical Hazards: OpenFoodTox 2.0. EFSA supporting publication 2020:EN-1822. 36 pp. doi:1 https://doi.org/10.2903/sp.efsa.2021.EN-6476

Ghosh S., Ojha P.K., **Carnesecchi E.**, Lombardo A., Roy K., Benfenati E., (2020). Exploring QSAR modeling of toxicity of chemicals on earthworm. Ecotoxicol. Environ. Saf. 190. https://doi.org/10.1016/j.ecoenv.2019.110067

Toropov A.A., Toropova A.P., **Carnesecchi E.**, Benfenati E., Dorne J.L.C.M., 2020. The index of ideality of correlation and the variety of molecular rings as a base to improve model of HIV-1 protease inhibitors activity. Struct. Chem. https://doi.org/10.1007/s11224-020-01525-9

Toropova A. P., Toropov A. A., **Carnesecchi E.**, Benfenati E., Dorne J.L.C.M., (2020). The index of ideality of correlation: models for flammability of binary liquid mixtures. Chem. Pap. https://doi.org/10.1007/s11696-019-00903-w

Toropova A.P., Toropov A.A., **Carnesecchi E.**, Benfenati E., Dorne J.L.C.M., (2020). The using of the Index of Ideality of Correlation (IIC) to improve predictive potential of models of water solubility for pesticides. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-020-07820-6

Roy J., Ojha P. K., **Carnesecchi E.**, Lombardo A., Roy K., Benfenati, E., (2019). First report on a classification-based QSAR model for chemical toxicity to earthworm. Journal of Hazardous Materials, 121660. https://doi.org/10.1016/j.jhazmat.2019.121660

Sperandio G., Simonetto A., **Carnesecchi E.**, Costa C., Hatjina F., Tosi S., Gilioli G., (2019). Beekeeping and honey bee colony health: A review and conceptualization of beekeeping management practices implemented in Europe. Science of the Total Environment, 696(13), 133795. https://doi.org/10.1016/j.scitotenv.2019.133795

Toropov A.A., Toropova A.P., Marzo M., **Carnesecchi E.**, Selvestrel G., Benfenati, E., (2020). Pesticides, cosmetics, drugs: identical and opposite influences of various molecular features as measures of endpoints similarity and dissimilarity. Mol. Divers. https://doi.org/10.1007/s11030-020-10085-3

Dorne J.L.C.M., Richardson J., Livaniou A., **Carnesecchi E.**, Ceriani L., Baldin R., Kovarich S., Pavan M., Saouter E., Biganzoli F., Pasinato L., Zare Jeddi M., Robinson TP., Kass, GEN., Liem AKD., Toropov A.A., Toropova A.P., Yang C., Tarkhov A., Georgiadis N., Di Nicola M., Mostrag A., Verhagen H., Roncaglioni A., Benfenati E., Bassan A., (2021). EFSA's OpenFoodTox: An open source toxicological database on chemicals in food and feed and its future developments. Environ. Int. https://doi.org/10.1016/j.envint.2020.106293

EFSA Panel on Animal Health and Welfare, (2017). EFSA Staff: Beltran-Beck B., **Carnesecchi E.**, Casier P., Czwienczek E., Dhollander S., Georgiadis M., Gogin A., Pasinato L., Richardson J., Riolo F., Rossi G., Watts M., Lima E. and Stegeman JA. Scientific opinion on Vector-borne diseases. EFSA Journal 2017; 15(5):4793, 91 pp. https://doi.org/10.2903/j.efsa.2017.4793

Broglia A., Beltran B., **Carnesecchi E.**, Garcia A., Gogin A., Czwienczek E., Baldinelli F., Cortinas Abrahantes J., Kohnle L., Dhollander S., (2017). Bluetongue: control, surveillance and safe movement of Animals. EFSA Journal 2017; 15(3):4698, 126 pp. doi:10.2903/j.efsa.2017.4698

Balaz V., Gortazar Schmidt C., Murray K., **Carnesecchi E.**, Garcia A., Gervelmeyer A., Martino L., Munoz Guajardo I., Verdonck F., Zancanaro G., and Fabris C., (2017). Scientific and technical assistance concerning the survival, establishment and spread of Batrachochytrium salamandrivorans (Bsal) in the EU. EFSA Journal 2017;15(2):4739, 73 pp. doi:10.2903/j.efsa.2017.4739

EFSA Panel on Animal Health and Welfare, (2016). EFSA staff: Auteri D., **Carnesecchi E.**, Guillot G., Lima E., Rortais A., Sperandio A., Streissl F., Verdonck F., Volani S., and Vos S. Scientific Opinion on assessing the health status of managed honeybee colonies (HEALTHY-B): a toolbox to facilitate harmonised data collection. EFSA Journal 2016; 14(10):4578, 241 pp. doi:10.2903/j.efsa.2016.4578

Casini B., **Carnesecchi E.**, Totaro M., Privitera G., Baggiani A., (2015). Diffusion model applied to assessment of children health impact. European Journal of Public Health, 25 (suppl_3)

Totaro M., **Carnesecchi E.**, Valentini P., Porretta A., Bruni B., Privitera G., Baggiani A., (2015). Presence of Legionella in water networks of Italian residential buildings. European Journal of Public Health, 25 (suppl_3)

D'ascenzi, C., Formato, G., **Carnesecchi E.**, Mannaioni G., Caracciolo I., (2015). Valutazione del rischio alimentare applicata alla contaminazione del polline da residui di prodotti fitosanitari. Libro Bianco della Società italiana di Medicina Veterinaria Preventiva, Roma, 2015. https://arpi.unipi.it/handle/11568/844595#.Xcves1f0mUk