



CHALMERS
UNIVERSITY OF TECHNOLOGY

Does the scientific knowledge reflect the chemical diversity of environmental pollution? – A twenty-year perspective

Downloaded from: <https://research.chalmers.se>, 2022-10-11 19:29 UTC

Citation for the original published paper (version of record):

Kristiansson, E., Coria, J., Gunnarsson, L. et al (2021). Does the scientific knowledge reflect the chemical diversity of environmental pollution? – A twenty-year perspective. *Environmental Science and Policy*, 126: 90-98.
<http://dx.doi.org/10.1016/j.envsci.2021.09.007>

N.B. When citing this work, cite the original published paper.



Does the scientific knowledge reflect the chemical diversity of environmental pollution? – A twenty-year perspective

Erik Kristiansson^{a,b,*}, Jessica Coria^{b,c}, Lina Gunnarsson^d, Mikael Gustavsson^{a,b}

^a Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 412 96 Göteborg, Sweden

^b The FRAM Centre for Future Chemical Risk Assessment and Management Strategies, University of Gothenburg, 405 30 Göteborg, Sweden

^c Department of Economics, University of Gothenburg, 405 30 Göteborg, Sweden

^d Biosciences, College of Life & Environmental Sciences, University of Exeter, Exeter, Devon EX4 4QD, UK

ARTICLE INFO

Keywords:

Environmental chemical pollution

Bibliometry

Text mining

Research trends

Environmental legislation

Policymaking

ABSTRACT

Environmental policymaking relies heavily on the knowledge of the toxicological properties of chemical pollutants. The ecotoxicological research community is an important contributor to this knowledge, which together with data from standardized tests supports policy-makers in taking the decisions required to reach an appropriate level of protection of the environment. The chemosphere is, however, massive and contains thousands of chemicals that can constitute a risk if present in the environment at sufficiently high concentrations. The scientific ecotoxicological knowledge is growing but it is not clear to what extent the research community manages to cover the large chemical diversity of environmental pollution. In this study, we aimed to provide an overview of the scientific knowledge generated within the field of ecotoxicology during the last twenty years. By using text mining of over 130,000 scientific papers we established time-trends describing the yearly publication frequency of over 3500 chemicals. Our results show that ecotoxicological research is highly focused and that as few as 65 chemicals corresponded to half of all occurrences in the scientific literature. We, furthermore, demonstrate that the last decades have seen substantial changes in research direction, where the interest in pharmaceuticals has grown while the interest in biocides has declined. Several individual chemicals showed an especially rapid increase (e.g. ciprofloxacin, diclofenac) or decrease (e.g. lindane and atrazine) in occurrence in the literature. We also show that university- and corporate-based research exhibit distinct publication patterns and that for some chemicals the scientific knowledge is dominated by publications associated with the industry. This study paints a unique picture and provides quantitative estimates of the scientific knowledge of environmental chemical pollution generated during the last two decades. We conclude that there is a large number of chemicals with little, or no, scientific knowledge and that a continued expansion of the field of ecotoxicology will be necessary to catch up with the constantly increasing diversity of chemicals used within the society.

1. Introduction

Environmental chemical pollution is caused by emissions from the production, use and disposal of chemicals and chemical-containing materials. In order to reduce the risk from these emissions and protect humans and the environment, both voluntary and mandatory management strategies have been implemented. Many of these management strategies rely heavily on scientific expertise and knowledge about chemical pollutants, including their physical and chemical properties, toxicity, potential adverse effects on wildlife, and fate in the environment is central (Martin et al., 2019; Silbergeld et al., 2015). This knowledge is in part generated during the registration of chemicals,

which, in some jurisdictions, requires a prospective chemical risk assessment based on the results from standardized tests (OECD, 2021). Another important contributor to this knowledge is the ecotoxicological scientific community and its members in academia, government and industry, which perform research, typically by formulation of hypotheses that are tested through experiments and deductive reasoning (Moermond et al., 2017). In contrast to the standardized tests, the scientific community typically addresses ecotoxicological issues in an ecologically broader and more holistic context, e.g. by including non-standard mode-of-actions, endpoints, organisms and chemicals, which enables analysis of risk scenarios that could be overlooked by standardized tests. Regulatory decisions are thus based on both

* Correspondence to: Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 412 96 Göteborg, Sweden.

E-mail address: erik.kristiansson@chalmers.se (E. Kristiansson).

<https://doi.org/10.1016/j.envsci.2021.09.007>

Received 6 April 2021; Received in revised form 9 June 2021; Accepted 14 September 2021

Available online 7 October 2021

1462-9011/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

standardized tests - as required by safety assessment policies - and ecotoxicological scientific knowledge in order to make them as informed, balanced and sustainable as possible (Rudén et al., 2017).

The scientific challenges related to environmental pollution are notorious due to the massive diversity of the tens of thousands of chemicals that individually – or as a mixtures - may constitute a risk to aquatic and terrestrial ecosystems (Escher et al., 2020; Z. Wang et al., 2020). Even though the number of papers published by the research community – and thereby our accumulated scientific knowledge – is continuously growing (Khan and Ho, 2012; M.-H. Wang and Ho, 2011), the information is often deemed insufficient in quantity, quality or interoperability which results in decisions made under (large) uncertainty (McNie, 2007). Indeed, the (eco)toxicological properties and environmental exposure scenarios of only a small fraction of the chemicals on the market have been thoroughly investigated, while limited, or in the worst case, no data is available for the overwhelming majority of chemicals (OECD, 2018). This lack of knowledge negatively affects policy-making, either through decisions that in retrospect turned out to be improper or by the decision to not take any action at all (European Environment Agency, 2013). A well-known example is the biocide tributyltin, which was in use for more than forty years in anti-fouling paint before it, due to an increasing understanding of its adverse effects on shellfish and gastropods, become subject to a series of local and limited restrictions until it was internationally banned in 2001 (Santillo et al., 2001). Similarly, neonicotinoids, a commonly used class of insecticides were banned in 2018 from use outside green-houses within the EU and in some areas in the US, Philippines and Canada, after evidence emerged of adverse effects in bees and other non-target insect groups (Maxim & van der Sluijs 2013; Sgolastra et al., 2020). An even more recent example is per- and polyfluoroalkyl substances (PFASs), which have been produced in large scales for more than seventy years. Studies on the effects of PFASs were initially scarce but research during the last twenty years have shown that several of these substances (e.g. perfluorooctanesulfonic acid (PFOS) and perfluorooctanoic acid (PFOA)) bioaccumulate in humans, animals and plants and there are indications that they are toxic at concentrations observed in individuals and the environment (Blum et al., 2015; Trojanowicz and Koc, 2013; Vestergren and Cousins, 2009). The European Commission's Chemicals Strategy for Sustainability has therefore proposed a comprehensive set of actions to address the environmental contamination of PFAS (European Commission, 2020). Each of these examples shows that the processes that ultimately lead to restrictions and bans were facilitated by our gradually increasing understanding of the adverse effects of the environmental pollution of these chemicals, thus underlining the importance of scientific knowledge for better informed regulatory action (Gold and Wagner, 2020; Holmes and Clark, 2008; Kirchoff et al., 2013).

Even though the scientific body of knowledge is growing, it is not clear to what extent it covers the large diversity of anthropogenic chemicals that is present in the environment today (NORMAN EMPO-DAT Database, 2020). Indeed, the topic of many research projects is heavily influenced by available funding and not directly designed to meet societal needs by covering the chemicals that may negatively affect the ecosystems. A substantial part of the scientific knowledge is, furthermore, generated by corporate-made or corporate-sponsored research projects, which, compared to the research typically done at universities, are fully or partially driven by a commercial agenda (Benbrook, 2019; Boone et al., 2014).

In this study, we aimed to provide an overview of the scientific knowledge of chemicals associated with environmental pollution generated during the last two decades. Through text mining of bibliometric data and metadata of hundreds of thousands of publications, we generated time series describing the publication trends of chemicals in scientific papers in ecotoxicology since the year 2000. Our results show that research on environmental pollution is highly focused where as few as 65 chemicals corresponded to half of all occurrences in the peer-

reviewed scientific literature. We, furthermore, found that the last two decades have seen substantial changes in research direction, where both chemical classes and individual compounds exhibited significant increasing or decreasing occurrence frequency. We also show that corporate-associated research had a distinctly different publication pattern compared to research generated solely at universities and that there are chemicals where the total accumulated knowledge is dominated by research from the industry. We conclude that there is a large number of chemicals where knowledge is still lacking and that a continued expansion of the field of ecotoxicology will be necessary to meet the constantly increasing diversity of chemicals used within our society.

2. Results and discussion

In total, 18,928 chemicals were selected for this study from four disparate databases containing information on compounds that are currently, or have previously been, used or produced within the society: 1) chemicals registered under the European Union (EU) legislation REACH, 2) currently approved and unapproved active ingredients from plant protection products (PPP) within the EU registered at the European Food Safety Authority (EFSA), 3) currently approved and unapproved active ingredients of biocides within the EU registered at the European Chemicals Agency (ECHA) and 4) pharmaceuticals approved by the United States (US) Federal Drug Administration (FDA). For each chemical, the ten most commonly used names were extracted from PubChem (Sayers et al., 2020), which resulted in 132,518 chemical names in total. Scientific papers from the last 20 years (2000–2019) that were published in 15 renowned international peer-reviewed ecotoxicological journals (Supplementary Table 1) were collected from PubMed. This resulted in 131,227 papers, for which the title, abstract, keywords and chemical lists were searched for the occurrence of any of the chemical names. In total, 93,383 papers were found to contain 153,964 occurrences of a total of 3682 non-redundant chemicals. See Methods for full details of the data analysis.

We first examined the publication frequencies, which differed substantially between chemicals. The distribution of publication frequencies was highly skewed where a limited number of chemicals were very frequently mentioned in the analyzed papers (Fig. 1, Supplementary Table 2). In fact, the 11 chemicals with the highest publication frequency accounted for more than 25% of the occurrences while 65 chemicals accounted for more than 50% of the occurrences. A similar pattern could be seen in the US EPA ECOTOXicology Knowledgebase which collect results from chemical toxicity tests from the scientific literature. From 12,154 reported CAS numbers, 38 and 185 corresponded to 25% and 50% of the reported tests, respectively, thus showing that the information in also this database is highly skewed towards a limited number of chemicals (Supplementary Figure 1).

Furthermore, we investigated if the skewed publication frequencies were also present when the chemicals were stratified according to their use categories (Fig. 2). For REACH-registered chemicals, cadmium was most frequently mentioned (5.4% of the occurrences of chemicals in papers), which was followed by copper (5.0%), mercury (4.0%) and arsenic (3.0%). For PPPs, atrazine (0.98%) was the most frequently mentioned chemical followed by chlorpyrifos (0.82%), lindane (0.72%) and hexachlorobenzene (0.72%). The biocides with the highest publication frequency were triclosan (0.42%), cypermethrin (0.33%) and diuron (0.30%) followed by formaldehyde (0.27%). Finally, the most frequently mentioned pharmaceuticals were estradiol (1.2%), testosterone (0.63%), ethinyl estradiol (0.47%) and triclosan (0.42%). These results show that a substantial part of the scientific knowledge generated by the ecotoxicological research community is focused on a limited group of relatively well-studied chemicals – a pattern that was already observed ten years ago (Grandjean et al., 2011).

Next, we analyzed the yearly publication frequencies, which revealed interesting trends over time (Fig. 3). Among the four chemical

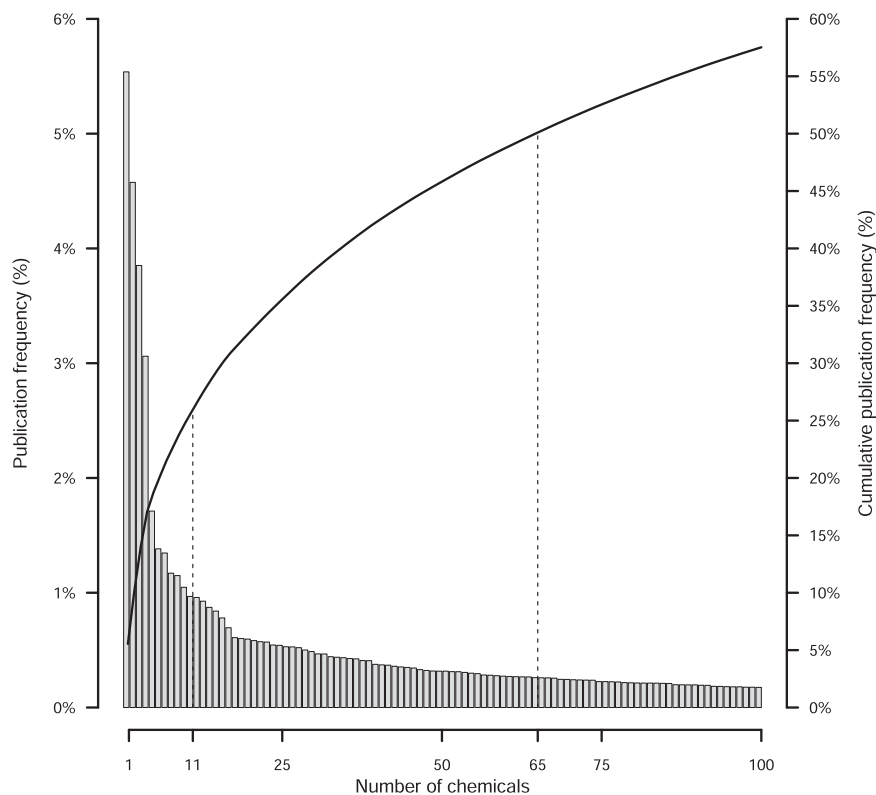


Fig. 1. The histogram shows the publication frequency for the 100 most mentioned chemicals. The distribution is highly skewed with a single chemical corresponding to almost 5% of all occurrences. The black line describes the cumulative number of occurrences measured in percent. The dashed lines correspond to the points where the cumulative number of occurrences reach 25% and 50% respectively. The publication frequency for all chemicals included in this study is available in [Supplementary Table 2](#).

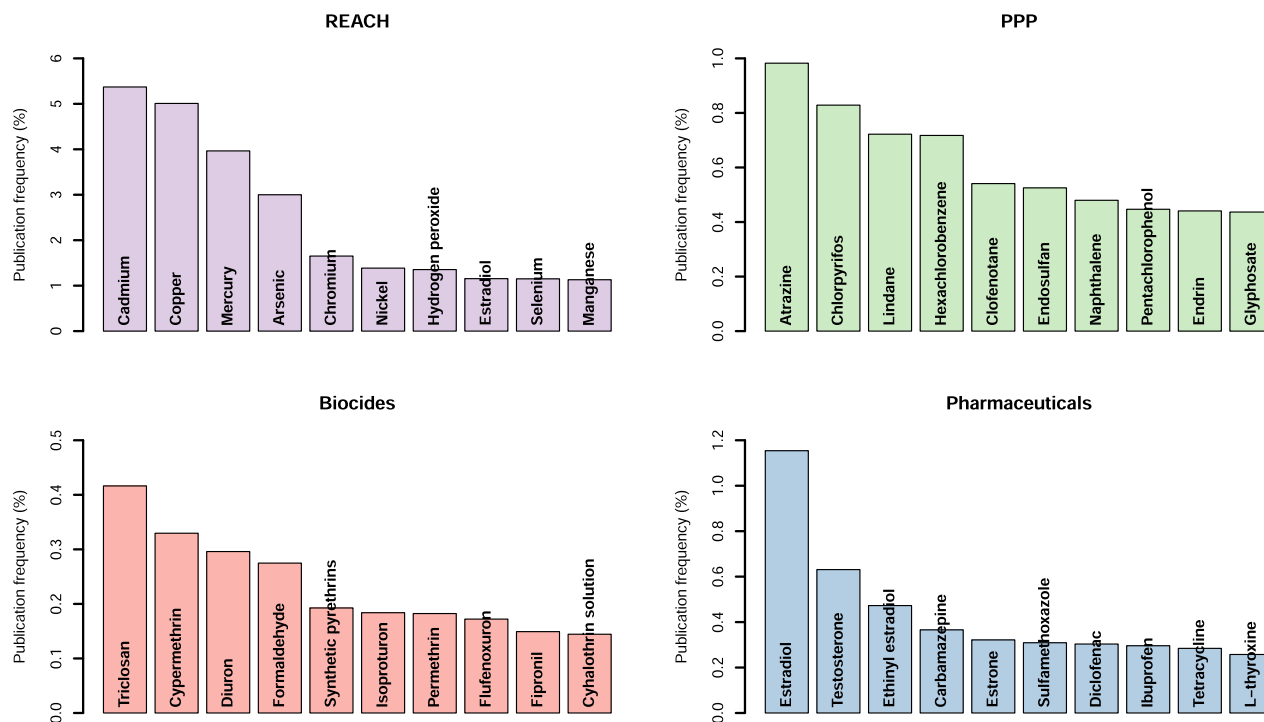


Fig. 2. Bar plots describing the 15 most commonly occurring chemicals in the four categories a) REACH, b) plant production products (PPPs), c) biocides and d) pharmaceuticals. The publication frequency (y-axis) is calculated from the yearly number of publications between 2000 and 2019. For panels b, c and d, only chemicals with primary use within the respective category were included.

categories, pharmaceuticals showed the largest and most consistent relative increase with a publication frequency that was 65% higher in 2019 compared to 2000 (Fig. 3a). The occurrences of biocides also increased (19%), which was in contrast to PPPs that showed an overall

decrease (23%). The publication frequency of chemicals registered under REACH was approximately constant until 2010, from where it showed a small but consistent increase, resulting in an overall twenty-year growth of 2.8%. Large temporal changes could also be seen

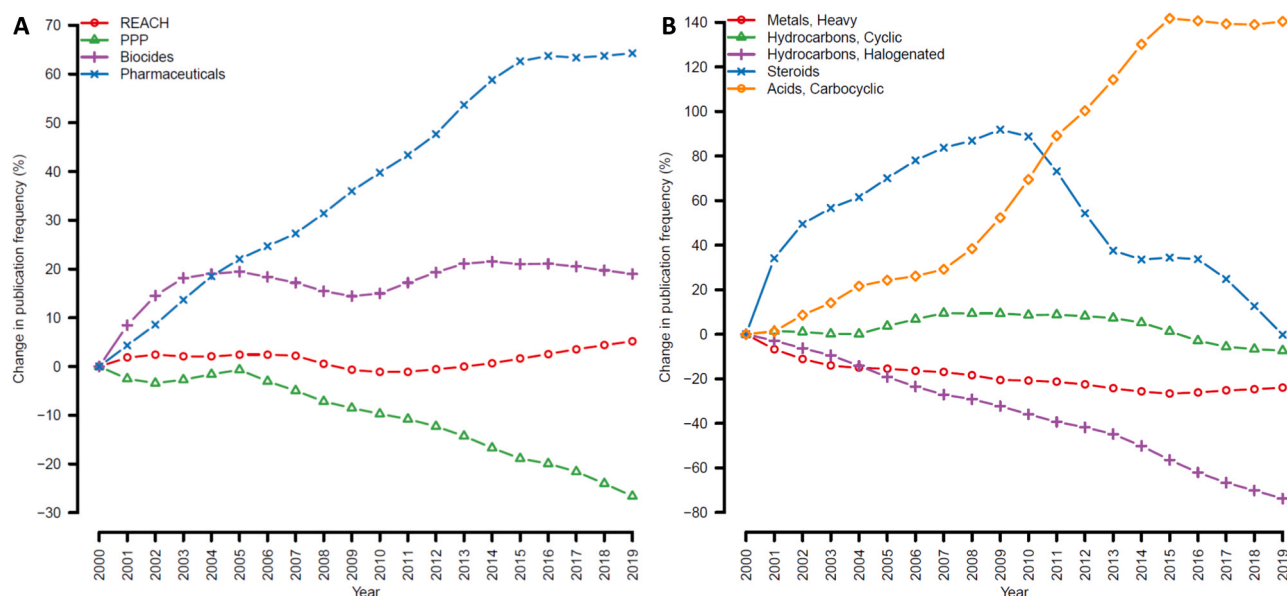


Fig. 3. Time trends describing changes in publication frequencies for a) categories and b) MeSH classes of chemicals. The curves correspond to the yearly change in publication frequency in relation to a reference value calculated as the median publication frequency between 2000 and 2004. The curves were smoothed to reduce yearly fluctuations. In (a), red (circle), green (triangle) and purple (plus) curves correspond to chemicals registered under REACH, plant protection products (PPPs) and biocides registered by EU, respectively. The blue curve (cross) corresponds to FDA-approved pharmaceuticals. In (b), the curves correspond to the five chemical classes (MeSH) with the highest publication frequencies.

between the different chemical classes (Fig. 3b). Among the most commonly mentioned MeSH classes, carbocyclic acids - which contains both frequently mentioned drugs like diclofenac and ibuprofen as well as the phthalates - showed a twenty-year increase of 152%. In contrast, the class of halogenated hydrocarbons, which is dominated by PPPs such as hexachlorobenzene and lindane, showed a sharp decline with a reduction in publication frequency with as much as 72%. The class of steroids, containing e.g. estradiol, testosterone and ethinyl estradiol, showed an increasing trend between 2000 and 2010, while the trend was decreasing between 2010 and 2019, resulting in a small overall decrease of 3.4%. The most published class of chemicals, heavy metals, showed a small but consistent decrease with a twenty-year reduction in publication frequency of 24%.

The last twenty years have thus resulted in substantial changes in the research focus of the ecotoxicological scientific community, significantly affecting the knowledge generated for specific use categories and classes of chemicals. To further pinpoint these changes, we analyzed the time trends for the 253 chemicals that were mentioned in at least 100

publications since 2000. Among these, 123 (48.6%) showed a significant change in publication frequency, where 70 of those (57%) were increasing while the remaining 53 (43%) were decreasing (Table 1, Supplementary Table 3, Supplementary Results 1). Four examples of chemicals with highly significant trends are shown in Fig. 4. Copper was, after cadmium, one of the most commonly mentioned chemicals and showed a decreasing trend, which was more significant than other metals (Fig. 4a). A strongly decreasing trend could also be seen for the PPP (insecticide) lindane (Fig. 4b) while the publication frequencies for the biocide (fungicide) tebuconazole (Fig. 4c) and the anti-inflammatory drug diclofenac increased (Fig. 4d). Interestingly, neither tebuconazole nor diclofenac was mentioned a single time in 2000 but twenty years later they were two of the most frequently mentioned chemicals. The fungicide tebuconazole has received attention for its possible endocrine-disrupting effects (Li et al., 2020; Lv et al., 2017) and it was recently added to the watch list under the Water Framework Directive (WFD) (Cortes et al., 2020). The nonsteroidal anti-inflammatory drug diclofenac was found to be the major cause of the collapse of the population of

Table 1

The 15 chemicals with most significant increase or decrease in publication frequency over time.

Chemical name	CAS number	Category ^a	MeSH class	Publications	Trend (pp/yr) ^b	P-value
Ciprofloxacin	85721-33-1	Pharmaceutical	Heterocyclic compounds	435	0.028	8.70×10^{-11}
Norfloxacin	70458-96-7	Pharmaceutical	Heterocyclic compounds	161	0.011	2.02×10^{-10}
Lindane	319-84-6	PPP, pharmaceutical	Halogenated hydrocarbons	886	-0.062	2.91×10^{-10}
Pentachlorophenol	87-86-5	PPP	Cyclic hydrocarbons	625	-0.039	3.56×10^{-10}
Sulfamethoxazole	129378-89-8	Pharmaceutical	Aniline compounds, sulfonamides, sulfones	555	0.033	5.03×10^{-10}
Chlordane	5103-74-2	PPP	-	476	-0.027	2.09×10^{-9}
Acetaminophen	103-90-2	REACH, pharmaceutical	Anilides, aniline compounds	333	0.020	3.33×10^{-9}
Diclofenac	15307-86-5	REACH, pharmaceutical	Carbocyclic acids	515	0.029	1.28×10^{-8}
Atrazine	102029-43-6	REACH, PPP	Triazines	1239	-0.043	5.04×10^{-8}
Endrin	72-20-8	PPP	Halogenated hydrocarbons	781	-0.035	7.27×10^{-8}
Ceric oxide	1306-38-3	REACH	-	143	0.010	1.00×10^{-7}
Copper	7440-50-8	REACH, biocide	Heavy metals, transition elements	6747	-0.157	1.04×10^{-7}
Carbamazepine	298-46-4	Pharmaceutical	Heterocyclic compounds	603	0.036	1.95×10^{-7}
Tetracycline	60-54-8	REACH, pharmaceutical	Cyclic hydrocarbons, naphthacenes	548	0.031	2.04×10^{-7}
Tebuconazole	107534-96-3	REACH, PPP, biocide	-	161	0.011	3.01×10^{-7}

^a PPP = plant protection products,

^b percent points per year

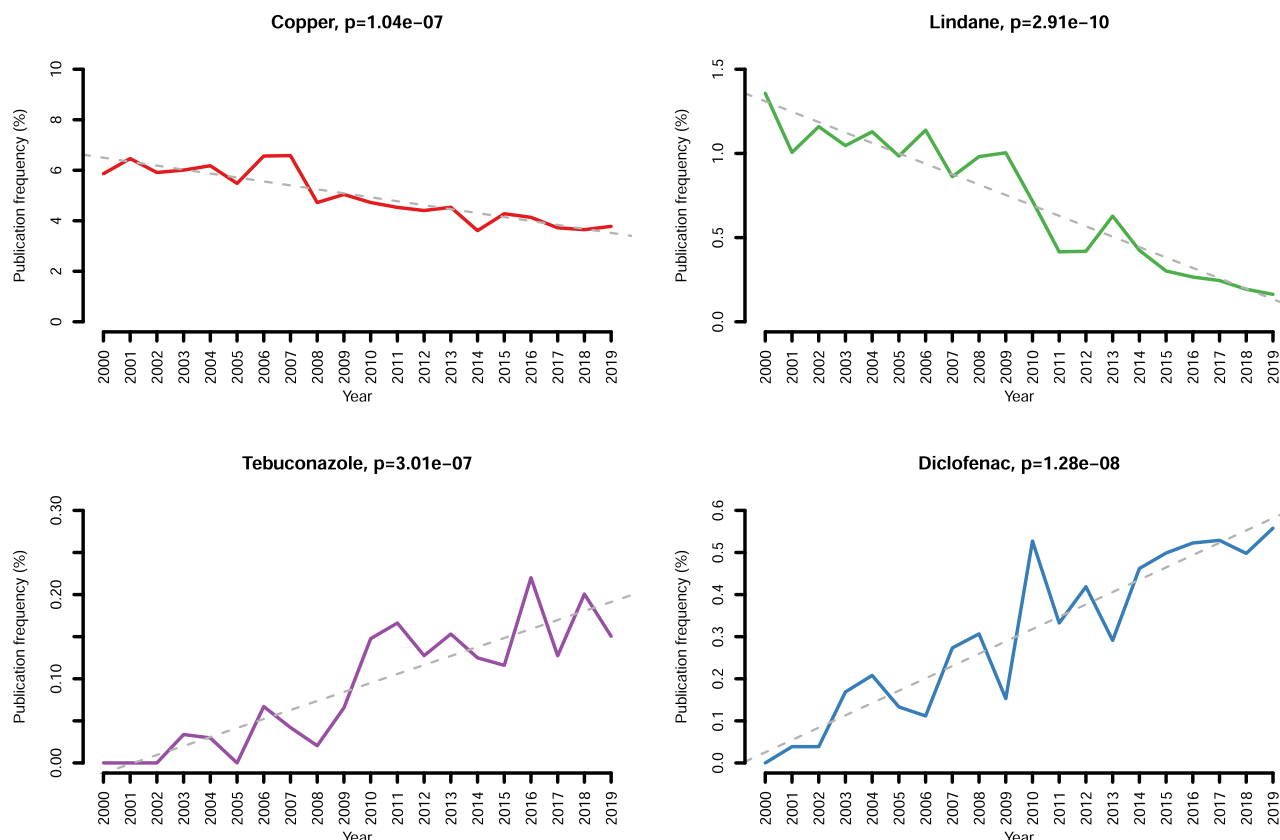


Fig. 4. Time trends describing the publication frequency between 2000 and 2019 for four selected chemicals, a) copper, b) glyphosate, c) tebuconazole and d) diclofenac. The publication frequencies (y-axis) were calculated for each year. The dashed line describes the overall trend and was estimated by linear regression. Trends for all chemicals mentioned in at least 100 publications can be found in [Supplementary Results 1](#).

three vulture species in Pakistan (Oaks et al., 2004; Taggart et al., 2007) and has since then been shown to cause adverse effects in other aquatic and terrestrial organisms (Lonappan et al., 2016). Moreover, we noted that chemicals such as lindane, pentachlorophenol, chlordane and atrazine, all with strong decreasing trends (Table 1), have been either banned or heavily regulated in large parts of the world. Lindane, which was internationally banned for use in 2009 as a persistent organic pollutant under the Stockholm Convention had a publication frequency as high as 1.5% in 2000 which decreased to 0.05% in 2019. A substantial decrease was also seen for atrazine, which was banned for use in EU in 2004 and subsequently considered as a priority pollutant under the WFD watch list in 2008. For atrazine, the publication frequency in 2000 was 0.56% which had, twenty years later, been reduced to 0.17%. All these examples suggest that the scientific community adapts by changing its focus away from chemicals that due to legislation or other policy changes reduce their potential environmental impact. It should, however, be emphasized that chemicals that have been banned at a global level may still have important ecotoxicological knowledge gaps, e.g. due to their persistence in the environment or due to their presence in man-made structures or materials (Cousins et al., 2019).

Next, we analyzed the total number of chemicals mentioned in the literature and found an overall increase over time. In 2000, 515 different chemicals were mentioned in at least one publication, which increased to 941 in 2009 and 1716 in 2019, corresponding to an overall increase of 471% over two decades – or, in average, 24 percentage points per year. As expected, these numbers correlated well with the total number of publications within ecotoxicology ($r = 0.80$, $p = 1.8 \times 10^{-5}$) and, thus, the increasing number of chemicals mentioned each year could largely be explained by the overall growth of the field. However, when the yearly publication rate was normalized to the same level as in 2000 (see Methods), the number of chemicals mentioned each year in scientific

literature still increased. In 2019 there were, in average, 8.3 chemicals mentioned per 100 publications, which was 11% higher than twenty years earlier (Supplementary Figure 2). Thus, taken together, the field of ecotoxicology today generates knowledge that is significantly more diverse than in 2000 and this increase is explained both by the growth of the field itself but also by the tendency to address more chemicals per publication. It should be emphasized that of the 18,928 chemicals included in the study, only 3682 (19%) were mentioned in any paper from the 15 included major ecotoxicological journals and as few as 1118 (5.9%) were mentioned ten or more times (Supplementary Table 2), suggesting that we still far from covering the chemosphere. However, the number of chemicals has been estimated to have an approximate annual growth rate of ~4% (Llanos et al., 2019), which, if sustained over a 20 year period, becomes an accumulated increase of approximately 200%. This suggests that the ecotoxicological research community may actually be reducing the gap – or at least expanding the knowledge at a rate comparable to the number of new chemicals introduced on the market. Maintaining this trend would, however, require a sustained growth of the scientific output in the coming decades and beyond.

Corporations are known to contribute significantly to the scientific literature and thus also to the knowledge about environmental pollution. We therefore investigated if there were any differences in which chemicals that were mentioned in studies made solely on the universities compared to those made in association with corporations. First, each publication was classified into the distinct groups ‘university’ or ‘corporate’ based on the affiliation of the authors (see Methods). This resulted in 47,895 publications classified as ‘university’ and 3112 publications classified as ‘corporate’ while 42,376 publications could not be assigned to any of the groups. Statistical analysis showed that 35 chemicals showed distinct differences in publication patterns between the university- and corporate-associated publications ($p < 0.01$, Fisher’s

exact test) (Supplementary Table 4). Of these, 27 chemicals were overrepresented in corporate-associated publications and was thus mentioned at a significantly higher frequency than in publications from the university. In contrast, only eight were underrepresented and thus more frequent in publications made solely in universities. The overrepresentation was especially high for decamethylcyclotetrasiloxane (D₅), octamethylcyclotetrasiloxane (D₄) and dodecamethylcyclohexasiloxane (D₆), three chemicals commonly used as additives in cosmetics, where as much as 58%, 54% and 50% of the publications were classified as ‘corporate’ (Fig. 5) ($p = 3.2 \times 10^{-17}$, $p = 6.2 \times 10^{-10}$ and $p = 5.7 \times 10^{-6}$ respectively). Interestingly, these chemicals have recently been scrutinized under REACH and were put on the list of substances of very high concern in 2018. Our results show, furthermore, that the publication frequency of siloxanes was the highest between 2013 and 2015, i.e. one year after the Committee for Risk Assessment at ECHA concluded that these chemicals are persistent, bioaccumulative and toxic and thus may be restricted under REACH (Committee for Risk Assessment ECHA, 2016). A high overrepresentation of corporate-associated publications was also found for the pesticide oxasulfuron (21.6%, $p = 2.4 \times 10^{-8}$), the herbicides thifensulfuron-methyl (26%, $p = 3.5 \times 10^{-8}$) and 2,4,5-trichlorophenoxyacetic acid (50%, $p = 5.2 \times 10^{-8}$), the insecticide clothianidin (23%, $p = 8.7 \times 10^{-6}$) and the explosive hexogen (24%, $p = 8.6 \times 10^{-6}$). The most significant chemicals overrepresented in papers from the university were cadmium (3.6%, $p = 9.3 \times 10^{-11}$) and clofenotane, a major component of DDT (1.1%, 2.9×10^{-6}). These results show that compared to studies made at universities, corporate-associated publications target a smaller but more specific group of chemicals that are likely of direct or indirect commercial interest. The results also show that for several of these chemicals, publications produced in association with corporations constitute a large part, and in some cases even the majority, of the total scientific knowledge.

In this study we have shown that the publication patterns related to environmental chemical pollution have changed significantly during the last twenty years. These trends can be seen at a broad level, where pharmaceuticals and PPPs showed a large relative increase and decrease in publication frequency, respectively. Also, several classes of chemicals showed strong increasing (e.g. carbocyclic acids) or decreasing (e.g.

halogenated hydrocarbons) publication trends. Changes could also be pinpointed to individual compounds, where some chemicals, which were not mentioned at all twenty years ago, today are commonly investigated in ecotoxicological research studies. These changes in publication patterns are likely caused by several factors. Ecotoxicology is, similarly to other research fields, influenced by governmental and/or industry-specific funding which often is targeted towards emerging pollutants that are considered especially problematic. This has, for example, been the case for pharmaceuticals in the environment, which have seen several directed funding efforts during the last two decades (European Commission, 2009). There are also technical aspects that influence which chemicals that can be studied. For example, methods for analytical chemistry have seen a constant improvement and can today be used to find chemicals at a range of ng/L in aquatic environments (Aceña et al., 2015; Hernández et al., 2019). Such low detection limits are essential in the study of highly potent pollutants, such as synthetic steroids and pyrethroids (Adeel et al., 2017; Tang et al., 2018). Our results also demonstrate that the publication patterns are affected by legislation and policy decisions. Indeed, several of the chemicals with the most rapidly decreasing publication frequency had been banned or introduced on regulatory watchlists during the last twenty years – indicating that there exists sufficient information for taking significant policy decisions. This also demonstrates the interaction between the societal stakeholders and the scientific community, where researchers change their focus towards chemicals where knowledge is more warranted (Kirchhoff et al., 2013). However, based on the observed decrease in publication frequency, this process is relatively slow and, based on our results, seems to span over at least a decade.

Due to the many factors that influence ecotoxicological research, there is no guarantee that the generated scientific knowledge meets the needs of the society. In particular, our results showed a large number of chemicals were, during the last twenty years, only mentioned a few times or not at all. There is thus an ‘unexplored’ part of the chemosphere containing chemicals for which the ecotoxicological aspects have not been addressed by the scientific research community. Indeed, several studies have suggested that a potential environmental impact cannot be excluded for many of these chemicals (Gustavsson et al., 2017; Scherlinger, 2017; Stempel et al., 2012; Wang et al., 2020). In addition, many chemicals demonstrated a large proportion of corporate-associated publications while university-based studies were underrepresented, suggesting that, due to different incentives, there is a risk that the generated knowledge may be partially biased which may lead to improper or inconsistent policies (Boone et al., 2014). Finally, it should be emphasized that it is not sufficient to only have knowledge on the toxicological properties of individual chemicals to ensure that the environment is protected. Indeed, as chemicals co-occur in the environment at concentrations that vary both spatially and temporally, knowledge of the combined mixture effects and their synergies with other stressors are also essential (Lemm et al., 2021; van Gils et al., 2020). Taken together, our findings underscore the need to generate scientific information for more chemicals and align the generated knowledge more closely to the needs of society.

The current study is based on the analysis of bibliometric data and was made possible by using combined information gathered from multiple data sources. It should, however, be pointed out that a large part of this data has been assembled by multiple actors over many years and it is therefore likely to contain errors and inconsistencies that may affect our results. One example is the non-standard notation of chemicals used by the scientific community. Even though the majority (78.4%) of the publications published in recent years declare a ‘list of chemicals’, only a minority (5.0%) of the listed chemicals use a systematic nomenclature, e.g. CAS or EC number notation. We addressed this issue by also considering the ten most popular synonyms, which enabled us to find significantly more chemicals in the text analysis. It should, however, be noted that due to the many inconsistencies in reporting chemicals in the scientific literature, we cannot guarantee that every single publication is

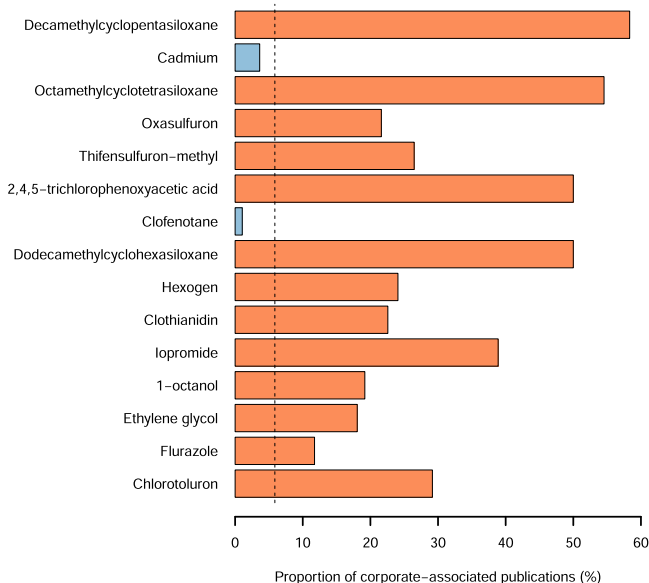


Fig. 5. The 15 most over- and under-represented chemicals in corporate-associated publications. For each chemical, the bar corresponds to the proportion of corporate-associated publications. The chemicals are ordered based on significance according to Fisher's exact test. Results for all chemicals are available in Supplementary Table 4.

properly identified using our approach. Another uncertainty is the actual quality of the included scientific publications. Particularly, the increasing number of predatory journals often contain publications where the peer-review process is flawed or, in the worst case, completely ignored. In order to ensure that our results are based on high-quality research, we limited our analysis to a limited set of highly esteemed international peer-reviewed ecotoxicological journals. This means that our results should not be considered as complete in the sense that it encompasses every single ecotoxicological study published the last twenty years. However, since the included journals all have high yearly volumes of published papers and since our results are based on relative increase and decrease, we argue that our results correspond to a substantial part of the papers published by the ecotoxicological research community and are thus, to a large extent, representative for the field as a whole.

3. Conclusion

In this study, we investigated the scientific knowledge on environmental chemical pollution generated by the research community over the last two decades. Our results show significant changes in the research agenda with decreasing publication frequency of chemicals used as plant protection products while the publication frequency of pharmaceuticals increased. We could, furthermore, conclude that the ecotoxicological research community is highly focused on a few well-studied chemicals, especially heavy metals, and this raises concerns about our ability to sufficiently cover the large chemical diversity of environmental pollutants. There is, indeed, a large number of chemicals for which no, or very little, knowledge is available or where the knowledge is, to a large extent, generated through corporate-associated research. We conclude that a continued expansion and/or a reprioritization of the ecotoxicology research is necessary to meet the challenges associated with the increasing chemical diversity of the expanding chemosphere and to ensure that the need for independent and objective scientific knowledge – as requested by the society – are properly met.

4. Methods

Chemicals and their corresponding CAS numbers were collected from four sources: 1) chemicals registered within the EU under REACH (<https://echa.europa.eu/information-on-chemicals/registered-substances>, retrieved November 2019), 2) approved and unapproved active ingredients of plant protection products (PPP) from within the EU as registered by EFSA (<https://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/active-substances/?event=search.as>, retrieved July 2020), 3) approved and unapproved active ingredients of biocides from within the EU as registered by ECHA (<https://echa.europa.eu/information-on-chemicals/biocidal-active-substances>, retrieved July 2020) and 4) pharmaceuticals approved by the Federal Drug Administration (FDA) registered by DrugBank (<http://drugbank.ca>, retrieved March 2020) (Wishart et al., 2018). Chemicals without a valid CAS number were excluded. The list of chemicals was then manually inspected and those not directly related to anthropogenic exposure and/or not of ecotoxicological relevance were removed (Supplementary Table 5). All elements were excluded from the pharmaceuticals.

A list of names was established for each chemical by matching each individual CAS number to the PubChem filtered synonym list (<ftp://ftp.ncbi.nlm.nih.gov/pubchem>, retrieved November 2019) (Sayers et al., 2020) and extracting up to ten of the most popular synonyms. If a CAS number appeared for multiple PubChem entries, the union of the extract synonyms was used. The synonyms that were mentioned in at least 100 publications were manually inspected and words that may have multiple and/or ambiguous meanings were removed (see Supplementary Table 6 for a list of removed synonyms). Each CAS number was furthermore associated with a classification based on the MeSH database (<https://www.nlm.nih.gov/databases/download/mesh.html>, retrieved

March 2020). For chemicals where the MeSH database did not contain any CAS number, the matching was done using either Unique Ingredient Identifiers (UNII, <https://fdasis.nlm.nih.gov/srs/jsp/srs/uniiListDownload.jsp>, retrieved April 2020) or the chemical name.

The full PubMed database (release 20) was downloaded from the NCBI FTP site (<https://ftp.ncbi.nlm.nih.gov/pubmed/>, retrieved April 2020). A selection of journals was created based on the following criteria: 1) representation in the field of ecotoxicology, 2) publish mainly original research, 3) international focus and a peer-review principle and 4) overall high scientific standard (thus excluding any potentially predatory journals) (Supplementary Table 1). The criteria were manually assessed based on information from the journals' homepages and bibliometric information from the Scopus and Web of Science databases.

All papers published in the selected journals between 2000 and 2019 were extracted and the following metadata was stored in a database: 1) title, 2) abstract, 3) author list, 4) author affiliation, 5) publication year, issue and volume, 6) list of keywords and 7) list of chemicals. After that, the abstract, title, keywords and chemical list were searched for the CAS number and/or its synonyms. For each chemical, all matching publications were stored. If two or more overlapping matches were found, i.e. matches from two or more chemicals containing partially the same words, the chemical with the name containing the highest number of words (and thus the more specific chemical name) was kept while the other matches were discarded. If the overlapping matches had the same number of words, one match was selected randomly. A yearly publication frequency for each chemical was then calculated based on the total number of publications in all selected journals. To remove redundancy where a chemical can be represented by more than one CAS number, all chemicals were clustered based on their set of synonyms using the Jaccard index distance metric together with complete linkage using a similarity cut-off of 0.2. For each cluster, the chemical with the CAS number that was mentioned the most time was kept while all other CAS numbers were discarded from the analysis. In order to reduce yearly fluctuations, time trends for categories and MeSH classes of chemicals were smoothed using the lowess algorithm (Cleveland, 1981) (using parameter $f=0.30$). The increase and decrease of individual chemicals were analyzed using linear regression and ranked based on the significance of the slope parameter, which was assessed using a *t*-test.

Scientific papers were classified as university- or corporate-associated as follows. First each author affiliation was classified as 'university', 'corporate', 'both' or 'other' based on two sets of keywords (Supplementary Table 7). If none of the keywords were found, the affiliation was classified as 'other'. The keywords were collected by manually examining all affiliations that appeared at least ten times during the last twenty years. Papers where all authors' affiliations were classified as 'university', were classified as 'university'. Papers where any of the first two authors or any of the last two authors, or, if the number of authors was less than four, any author, were classified as 'corporate', were classified as 'corporate associated'. Papers that did not match any of these two criteria were classified as 'other' and were not included in this part of the analysis. Over- and under-representation of chemicals in corporate-associated publications that occurred at least 10 times over the last twenty years were statistically assessed using Fisher's exact test. A test with a *p*-value less than 0.05 was considered significant.

The ECOTOXicology Knowledgebase was downloaded from the United States Environmental Protection Agency website (July 2020). The number tests for each CAS-number were counted using a custom script.

All statistical analyses were done in the statistical programming language R v 3.5.1 (R Core Team, 2019).

Funding

This research is funded by the Centre for Future Chemical Risk Assessment and Management Strategies (FRAM) at University of

Gothenburg, the Swedish Research Council FORMAS (2020-01895) and the Swedish Environmental Protection Agency (2020-00073).

CRedit authorship contribution statement

Erik Kristiansson: Conceptualization, Methodology, Formal analysis, Validation, Writing – original draft, Visualization, Funding acquisition, **Coria Jessica:** Conceptualization, Validation, Writing – review & editing, Funding acquisition, **Lina Gunnarsson:** Conceptualization, Validation, Writing – review & editing, **Mikael Gustavsson:** Conceptualization, Methodology, Data Curation, Validation, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We are grateful to Professor Sverker Molander, Chalmers University of Technology, Sweden, for fruitful discussion about key aspects of the manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.envsci.2021.09.007](https://doi.org/10.1016/j.envsci.2021.09.007).

References

- Aceña, J., Stampachiachiere, S., Pérez, S., Barceló, D., 2015. Advances in liquid chromatography - high-resolution mass spectrometry for quantitative and qualitative environmental analysis. *Anal. Bioanal. Chem.* 407, 6289–6299. <https://doi.org/10.1007/s00216-015-8852-6>.
- Adeel, M., Song, X., Wang, Y., Francis, D., Yang, Y., 2017. Environmental impact of estrogens on human, animal and plant life: a critical review. *Environ. Int.* 99, 107–119. <https://doi.org/10.1016/j.envint.2016.12.010>.
- Benbrook, C.M., 2019. How did the US EPA and IARC reach diametrically opposed conclusions on the genotoxicity of glyphosate-based herbicides? *Environ. Sci. Eur.* 31 (1), 2. <https://doi.org/10.1186/s12302-018-0184-7>.
- Blum, A., Balan, S.A., Scheringer, M., Trier, X., Goldenman, G., Cousins, I.T., Diamond, M., Fletcher, T., Higgins, C., Lindeman, A.E., Peaslee, G., de Voogt, P., Wang, Z., Weber, R., 2015. The Madrid statement on poly- and perfluoroalkyl substances (PFASs). *Environ. Health Perspect.* 123 (5), A107–A111.
- Boone, M.D., Bishop, C.A., Boswell, L.A., Brodman, R.D., Burger, J., Davidson, C., Gochfeld, M., Hoverman, J.T., Neuman-Lee, L.A., Relyea, R.A., Rohr, J.R., Salice, C., Semlitsch, R.D., Sparling, D., Weir, S., 2014. Pesticide regulation amid the influence of industry. *BioScience* 64 (10), 917–922. <https://doi.org/10.1093/biosci/biu138>.
- Cleveland, W.S., 1981. LOWESS: a program for smoothing scatterplots by Robust locally weighted regression. *Am. Stat.* 35, 54. <https://doi.org/10.2307/2683591>.
- Committee for Risk Assessment (ECHA). (2016). Background Document to the Opinion on the Annex XV dossier proposing restrictions on Octamethylcyclotetrasiloxane (D4) and Decamethylcyclopentasiloxane (D5).
- Cortes, L.G., Marinov, D., Sanseverino, I., Cuenca, A.N., Niegowska, M., Porcel, E., & Lettieri, T. (2020). Selection of substances for the 3rd Watch List under the Water Framework Directive. <https://doi.org/10.2760/194067>.
- Cousins, I.T., Ng, C.A., Wang, Z., Scheringer, M., 2019. Why is high persistence alone a major cause of concern? *Environ. Sci. Process. Impacts* 21, 781–792. <https://doi.org/10.1039/c8em00515j>.
- Escher, B.I., Stapleton, H.M., Schymanski, E.L., 2020. Tracking complex mixtures of chemicals in our changing environment. *Science (N.Y.)* 367 (January), 388–392.
- European Commission. (2009). Communication from the Commission to the European Parliament and the Council on the Establishment of a Joint EU Resettlement Programme (COM(2009) 447 final).
- European Commission. (2020). Chemicals Strategy for Sustainability Towards a Toxic-Free Environment.
- European Environment Agency. (2013). Late lessons from early warnings: science, precaution, innovation. In *Report No 1/2013* (Issue 1). (<https://doi.org/10.2800/73322>).
- Gold, S.C., Wagner, W.E., 2020. Filling gaps in science exposes gaps in chemical regulation. *Science* 368 (6495), 1066–1068.
- Grandjean, P., Eriksen, M.L., Ellegaard, O., Wallin, J.A., 2011. The Matthew effect in environmental science publication: A bibliometric analysis of chemical substances in journal articles. *Environmental Health: A Global Access Science Source* 10 (1), 1–8. <https://doi.org/10.1186/1476-069X-10-96>.
- Gustavsson, M.B., Helleloh, A., Backhaus, T., 2017. Evaluating the environmental hazard of industrial chemicals from data collected during the REACH registration process. *Sci. Total Environ.* 586, 658–665. <https://doi.org/10.1016/j.scitotenv.2017.02.039>.
- Hernández, F., Bakker, J., Bijlsma, L., de Boer, J., Botero-Coy, A.M., Bruinen de Bruin, Y., Fischer, S., Hollender, J., Kasprzyk-Hordern, B., Lamoree, M., López, F.J., Laak, T.L., te, van Leerdam, J.A., Sancho, J.V., Schymanski, E.L., de Voogt, P., Hogendoorn, E. A., 2019. The role of analytical chemistry in exposure science: focus on the aquatic environment. *Chemosphere*. <https://doi.org/10.1016/j.chemosphere.2019.01.118>.
- Holmes, J., Clark, R., 2008. Enhancing the use of science in environmental policy-making and regulation. *Environ. Sci. Policy* 11, 702–711. <https://doi.org/10.1016/j.envsci.2008.08.004>.
- Khan, M.A., Ho, Y.S., 2012. Top-cited articles in environmental sciences: Merits and demerits of citation analysis. *Sci. Total Environ.* 431, 122–127. <https://doi.org/10.1016/j.scitotenv.2012.05.035>.
- Kirchhoff, C.J., Carmen Lemos, M., Dessai, S., 2013. Actionable knowledge for environmental decision making: broadening the usability of climate science. *Annu. Rev. Environ. Resour.* 38 (1), 393–414. <https://doi.org/10.1146/annurev-environ-022112-112828>.
- Lemm, J.U., Venohr, M., Globevnik, L., Stefanidis, K., Panagopoulos, Y., van Gils, J., Posthuma, L., Kristensen, P., Feld, C.K., Mahnkopf, J., Hering, D., Birk, S., 2021. Multiple stressors determine river ecological status at the European scale: towards an integrated understanding of river status deterioration. *Glob. Change Biol.* 27 (9), 1962–1975. <https://doi.org/10.1111/gcb.15504>.
- Li, S., Jiang, Y., Sun, Q., Cof, S., Chen, L., Qiao, K., Gui, W., Zhu, G., 2020. Tebuconazole induced oxidative stress related hepatotoxicity in adult and larval zebrafish (*Danio rerio*). *Chemosphere* 241, 125129. <https://doi.org/10.1016/j.chemosphere.2019.125129>.
- Llanos, E.J., Leal, W., Luu, D.H., Jost, J., Stadler, P.F., Restrepo, G., 2019. Exploration of the chemical space and its three historical regimes. *Proc. Natl. Acad. Sci. USA* 116 (26), 12660–12665. <https://doi.org/10.1073/pnas.1816039116>.
- Lonappan, L., Kaur, S., Kumar, R., Verma, M., Surampalli, R.Y., 2016. Diclofenac and its transformation products: environmental occurrence and toxicity - A review. *Environ. Int.* 96, 127–138. <https://doi.org/10.1016/j.envint.2016.09.014>.
- Lv, X., Pan, L., Wang, J., Lu, L., Yan, W., Zhu, Y., Xu, Y., Guo, M., Zhuang, S., 2017. Effects of triazole fungicides on androgenic disruption and CYP3A4 enzyme activity. *Environ. Pollut.* 222, 504–512. <https://doi.org/10.1016/j.envpol.2016.11.051>.
- Martin, O.V., Adams, J., Beasley, A., Belanger, S., Breton, R.L., Brock, T.C.M., Buonsante, V.A., Galay Burgos, M., Green, J., Guiney, P.D., Hall, T., Hanson, M., Harris, M.J., Henry, T.R., Huggett, D., Junghans, M., Laskowski, R., Maack, G., Moermond, C.T.A., Ågerstrand, M., 2019. Improving environmental risk assessments of chemicals: steps towards evidence-based ecotoxicology. In: *Environment International*, 128. Elsevier Ltd, pp. 210–217. <https://doi.org/10.1016/j.envint.2019.04.053>.
- Maxim, L., van der Sluijs, J., ‘Seed-dressing systemic insecticides and honeybees’, in: EEA, 2013 Late lessons from early warnings: science, precaution, innovation, Environmental issue report No. 1/2013, pp. 369–406, European Environmental Agency.
- McNie, E.C., 2007. Reconciling the supply of scientific information with user demands: an analysis of the problem and review of the literature. *Environ. Sci. Policy* 10 (1), 17–38. <https://doi.org/10.1016/j.envsci.2006.10.004>.
- Moermond, C., Beasley, A., Breton, R., Junghans, M., Laskowski, R., Solomon, K., Zahner, H., 2017. Assessing the reliability of ecotoxicological studies: an overview of current needs and approaches. *Integr. Environ. Assess. Manag.* 13, 640–651. <https://doi.org/10.1002/ieam.1870>.
- , 2020NORMAN EMPODAT Database. (2020). <https://www.norman-network.com/norm/empodat/chemicalSearch.php>.
- Oaks, J.L., Gilbert, M., Virani, M.Z., Watson, R.T., Meteyer, C.U., Rideout, B.A., Shivaprasad, H.L., Ahmed, S., Jamshed, M., Chaudhry, I., Arshad, M., Mahmood, S., Ali, A., Khan, A.A., 2004. Diclofenac residues as the cause of vulture population decline in Pakistan. *Nature* 427 (February), 2002–2005.
- OECD, 2021. OECD Guidelines for the Testing of Chemicals. (https://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals_72d77764-en).
- OECD. (2018). OECD workshop on Managing Contaminants of Emerging Concern in Surface Waters: Scientific developments and cost-effective policy responses, 5 February 2018 Summary Note. February.
- R Core Team, 2019. R: A Language and Environment for Statistical Computing (3.5.3). R Foundation for Statistical Computing. (<https://www.r-project.org/>).
- Rudén, C., Adams, J., Ågerstrand, M., Brock, T.C., Poulsen, V., Schleka, C.E., Wheeler, J. R., Henry, T.R., 2017. Assessing the relevance of ecotoxicological studies for regulatory decision making. *Integr. Environ. Assess. Manag.* 13 (4), 652–663. <https://doi.org/10.1002/ieam.1846>.
- Santillo, D., Johnston, P., and Langston, W.J., 2001, ‘Tributyltin (TBT) antifoulants: a tale of ships, snails and imposex’, in: EEA, 2001 Late lessons from early warnings: the precautionary principle 1896–2000, Environmental issue report No. 22, pp. 135–148, European Environmental Agency.
- Sayers, E.W., Beck, J., Brister, J.R., Bolton, E.E., Canese, K., Comeau, D.C., Funk, K., Ketter, A., Kim, S., Kimchi, A., Kitts, P.A., Kuznetsov, A., Lathrop, S., Lu, Z., McGarvey, K., Madden, T.L., Murphy, T.D., O’Leary, N., Phan, L., Ostell, J., 2020. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 48 (D1), D9–D16. <https://doi.org/10.1093/nar/gkz899>.
- Scheringer, M., 2017. Environmental chemistry and ecotoxicology: in greater demand than ever. *Environ. Sci. Eur.* 29, 3. <https://doi.org/10.1186/s12302-016-0101-x>.
- Sgolastra, F., Medrzycki, P., Bortolotti, L., Mainia, S., Porrinia, C., Simon-Delsoc, N., Bosch, J., 2020. Bees and pesticide regulation: Lessons from the neonicotinoid

- experience. *Biological Conservation*. <https://doi.org/10.1016/j.biocon.2019.108356>.
- Silbergeld, E.K., Mandrioli, D., Cranor, C.F., 2015. Regulating chemicals: law, science, and the unbearable burdens of regulation. *Annu. Rev. Public Health* 36, 175–191. <https://doi.org/10.1146/annurev-publhealth-031914-122654>.
- Stempel, S., Scheringer, M., Ng, C.A., Hungerbühler, K., 2012. Screening for PBT chemicals among the “existing” and “new” chemicals of the EU. *Environ. Sci. Technol.* 46, 5680–5687. <https://doi.org/10.1021/es3002713>.
- Taggart, M.A., Cuthbert, R., Das, D., Sashikumar, C., Pain, D.J., 2007. Diclofenac disposition in Indian cow and goat with reference to Gyps vulture population declines. *Environ. Pollut. (Barking Essex 1987)* 147, 60–65. <https://doi.org/10.1016/j.envpol.2006.08.017>.
- Tang, W., Wang, D., Wang, J., Wu, Z., Li, L., Huang, M., Xu, S., Yan, D., 2018. Pyrethroid pesticide residues in the global environment: an overview. *Chemosphere* 191 (308), 990–1007. <https://doi.org/10.1016/j.chemosphere.2017.10.115>.
- Trojanowicz, M., Koc, M., 2013. Recent developments in methods for analysis of perfluorinated persistent pollutants. *Mikrochim. Acta* 180, 957–971. <https://doi.org/10.1007/s00604-013-1046-z>.
- van Gils, J., Posthuma, L., Cousins, I.T., Brack, W., Altenburger, R., Baveco, H., Focks, A., Greskowiak, J., Kühne, R., Kutsarova, S., Lindim, C., Markus, A., van de Meent, D., Munthe, J., Schueder, R., Schüürmann, G., Slobodnik, J., de Zwart, D., van Wezel, A., 2020. Computational material flow analysis for thousands of chemicals of emerging concern in European waters. *J. Hazard. Mater.* 397 (February), 122655 <https://doi.org/10.1016/j.jhazmat.2020.122655>.
- Vestergren, R., Cousins, I.A.N.T., 2009. Tracking the pathways of human exposure to perfluorocarboxylates. *Environ. Sci. Technol.* 43, 5565–5575.
- Wang, M.-H., Ho, Y.-S., 2011. Different durations of corticosteroid therapy for exacerbations of chronic obstructive pulmonary disease. *Cochrane Database Syst. Rev.*, 006897.
- Wang, Z., Walker, G.W., Muir, D.C.G., Nagatani-Yoshida, K., 2020. Toward a global understanding of chemical pollution: a first comprehensive analysis of national and regional chemical inventories. *Environ. Sci. Technol.* <https://doi.org/10.1021/acs.est.9b06379>.
- Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maclejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Wilson, M., 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>.