Prediction of bioavailability and toxicity of complex chemical mixtures through machine learning models

Cipullo S.[1], Snapir B.[*1], Prpich G.[2], Campo P.[1], Coulon F.[1]

[1]Cranfield University, School of Water, Energy and Environment, Cranfield, MK430AL, UK

[2]University of Virginia, Department of Chemical Engineering, USA

*Corresponding author: b.j.snapir@cranfield.ac.uk, +44(0)1234 750111

Abstract

Empirical data from a 6-month mesocosms experiment were used to assess the ability and performance of two machine learning (ML) models, including artificial neural network (NN) and random forest (RF), to predict temporal bioavailability changes of complex chemical mixtures in contaminated soils amended with compost or biochar. From the predicted bioavailability data, toxicity response for relevant ecological receptors was then forecasted to establish environmental risk implications and determine acceptable end-point remediation. The dataset corresponds to replicate samples collected over 180 days and analysed for total and bioavailable petroleum hydrocarbons and heavy metals/metalloids content. Further to this, a range of biological indicators including bacteria count, soil respiration, microbial community fingerprint, seeds germination, earthworm's lethality, and bioluminescent bacteria were evaluated to inform the environmental risk assessment. Parameters such as soil type, amendment (biochar and compost), initial concentration of individual compounds, and incubation time were used as inputs of the ML models. The relative importance of the input variables was also analysed to better understand the drivers of temporal changes in bioavailability and toxicity. It showed that toxicity changes can be driven by multiple factors (combined effects), which may not be accounted for in classical linear regression analysis

(correlation). The use of ML models could improve our understanding of rate-limiting processes affecting the freely available fraction (bioavailable) of contaminants in soil, therefore contributing to mitigate potential risks and to inform appropriate response and recovery methods.

Keywords: Risk assessment; Machine learning; Bioavailability; Complex chemical mixtures; Compost; Biochar

1. Introduction

Contaminants in the environment are found as a combination of multiple chemicals, often including both organic and inorganic compounds, such as petroleum hydrocarbons and heavy metals/metalloids (EEA, 2012). The contaminants' fraction potentially available to receptors (bioavailable fraction) depends on both chemical properties of pollutants and soil properties (Semple et al., 2003), and is a better proxy than the total concentration of contaminant (Guo et al., 2016) when predicting the actual exposure to organisms and the ecosystem (Gourlay-Francé and Tusseau-Vuillemin, 2013). Bioavailability provides a basis to make robust decision regarding appropriate risk assessment, predicting contaminants' fate, transport, and potential environmental impact (Lehmann and Joseph, 2015). While greater attention has been given to the role of bioavailability of complex contaminants in regard to remediation end points (Brand et al., 2012; Harmsen and Naidu, 2013; Kördel et al., 2013; Ortega-Calvo et al., 2015) as well as its implication for regulatory frameworks (Umeh et al., 2017), its recognition and effective application by European-based environmental regulators is still limited (Harmsen and Naidu, 2013).

The recent shift toward sustainable remediation approaches are spawning a growing interest towards biological amendment such as composting, land farming, bioventing, and biopiling (Bardos et al., 2011). Biochar and compost amendments have been largely applied as effective

2

bioremediation approaches for enhancing recovery of soil contaminated with petroleum hydrocarbons and heavy metals (Lyu et al., 2016; Wu et al., 2014). Even though bioremediation is regarded as a sustainable and economical approach, with minimal disruption on site (often applicable in situ) and great public acceptance, it has some limitations (Boopathy, 2000). For instance, some chemicals may not be degradable, or their degradation may yield by-products, such as oxy-PAHs, which can exhibit greater toxicity than the parent compounds (Hu et al., 2012) and therefore slow down the degradation and reduce remediation efficiency. Because of these challenges, our mechanistic understanding of complex chemical mixtures degradation, the associated toxicity, and the subsequent implication for risk assessment and remediation end-point are still limited.

Accurately monitoring the bioremediation process and supporting risk evaluation and degradation assessment often involve long-term experiments with multiple soil sampling which are labour intensive, time consuming, and expensive. Therefore, the use of Machine Learning (ML) methods trained on empirical data could be advantageous to make predictions on the potential degradation and reduction in toxicity occurring during remediation. ML models are able to learn the relationships between input variables (e.g. soil amendment, soil type) and output variables (e.g. long term changes in contaminants' bioavailability) from a *training dataset*, these relationships can then be generalised to make informed decisions in new cases (Wu et al., 2013). The application of ML to environmental issues, such as waste recovery and degradation studies, have been widely investigated in the literature (Abbasi and El Hanandeh, 2016; Heshmati et al., 2014; Khamforoush et al., 2011; Mason, 2006; Petric and Selimbašić, 2008; Wu et al., 2013), however their implementation is still limited as often these methods are data-specific, or even variable-specific and their performance depends on many factors (Li et al., 2011). In particular, previous research highlighted the potential of ML to determine remediation end-points based on bioavailability predictions (Wu et al., 2013) and showed that

a better understanding of the impacts of bioavailability is necessary to fully comprehend the extent of efficacy of bioremediation and manage the associated risks. Therefore, in this study, we use ML to predict bioavailability but also the actual toxicity of tar-contaminated soils from two former manufactured-gas plant sites before and after lab-based bioremediation with biochar and compost amendments. Specifically, the objectives were to use empirical data from a bioremediation experiment to predict (1) change of complex chemical mixtures bioavailability, (2) change in the associated toxicity, and (3) to assess the input variables which are the most important for the estimation of toxicity.

2. Material and methods

2.1. Mesocosms setup ad data collection

A 6-month mesocosms experiment was carried out using two contaminated soils collected from two former gasworks sites (Soil 1 and 2) located in the UK. The soil physicochemical properties of the two samples used in this study are summarised in Table S1 (supplementary material). Briefly, the soils were amended in duplicate with either 5% w/w biochar (Soil + Biochar) or 15% w/w compost (Soil + Compost). A set of duplicate samples were also left with no physical remedial action (Soil). All samples were stored outdoor for 180 days. Soil samples were collected at day 0, 30, 90, and 180 and subsequently processed for chemical, microbiological, and toxicological analysis. For the chemical analysis, total and bioavailable petroleum hydrocarbons and heavy metals were extracted and analysed for both soil samples at the 4 sampling times. For the estimation of the bioavailable concentration of organic compounds soil subsamples were extracted with mild solvent (methanol) as this has been previously found to be a good representation of the bioavailable fraction, and has been correlated with both the accessibility to earth-worms and bacterial mineralization assays

4

(Kelsey et al., 1997; Reichenberg and Mayer, 2006Yu et al., 2011; Wu and Zhu, 2016;) (details of extraction technique are provided in Supplementary Material).

Moreover, to derive information on the partitioning and bioavailable concentrations of HM and metalloids in soil, a sequential extraction (Cave et al., 2004) with increasing concentration of nitric acid was used. Details of the technique used are summarised in the Supplementary Materials and relevant descriptive statistics are presented in Table S3. A battery of biological and eco-toxicological indicators representing different trophic levels have been used to assess the ecological health change of the soils undergoing bioremediation treatments (Coulon et al., 2004). The indicators included soil respiration as described previously by Paton et al., (2006), phospholipid fatty acids analysis (PLFA) as per Frostegård et al., (1993) , seeds germination (mustard, rye grass, and pea), earthworm's lethality as described by Dawson et al., (2007) and Coulon et al., (2010), and Microtox® basic soil phase assay as described by Coulon et al., (2004). Information on the determination of the toxicological responses of the model organisms are summarised in Table S4 of the Supplementary Materials.

2.2. Modelling methods and variables input

The ML prediction of soil toxicity for a given time t is done in two stages (Figure 1): (a) first we use ML to predict the bioavailable concentration of some hydrocarbons at time t, then (b) we use these predictions to estimate the toxicity at time t. Our implementation relies on *scikit-learn,* a machine learning library for Python (Pedregosa et al., 2011). We tested Neural Network (NN) and Random Forest (RF) which are two ML techniques often used to model complex and nonlinear environmental problems (Rajaee et al., 2009; Sahoo et al., 2006; Wu et al., 2013).
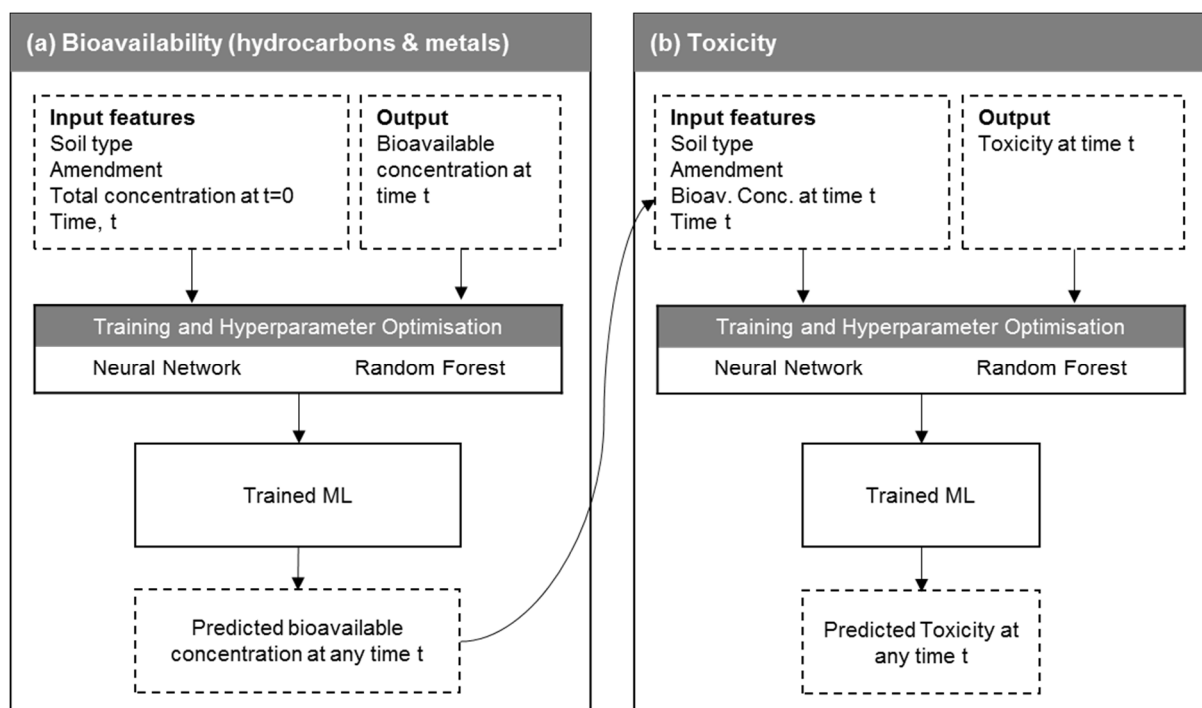
Figure 1: Two-stage approach to predict soil toxicity. (a) A machine learning (ML) algorithm is used to predict the bioavailable concentration of hydrocarbons and heavy metals/metalloids at time t. (b) The predicted concentrations are used as inputs of a second ML algorithm to predict soil toxicity.

For both the NN and RF methods, each input feature was first scaled by removing its average value and dividing by its standard deviation. This was done to avoid features with large values to dominate in the training phase. For the NN, we used a Multi-Layer Perceptron (MLP) which is a category of NN with at least three layers of neurons fully connected (an input layer, an output layer, and one or more hidden layers). NN models require several hyper-parameters to be tuned (Table 1), and are often criticised for being black boxes, as their behaviour cannot easily be explained by inspecting their internal structure (Sturm et al., 2016). That is why RF is also used in this study. RF can be considered as a *grey box* (Prasad et al., 2006), because it allows exploring the relative importance of the different input features . RF also has fewer hyper-parameters than NN (mainly the number of trees and the maximum depth of a tree; see

Table 1). More information on the behaviour of NN and RF can be found in the Supplementary Materials

Table 1: Hyper-parameters of the artificial neural network (NN) and random forest (RF) models.

NN model		RF model	
Hyper-parameters	Values	Hyper-parameters	Values
Activation function	Identity, Relu	Number of trees	20,30,40
Number of hidden layers and neurons	3 layers with 100 neurons,	Maximum depth	20,30,40
	4 layers with 100 neurons,		
	5 layers with 100 neurons		
Regularisation term	0.0001, 0.001, 0.01, 0.1, 1, 10		
Learning rate (constant)	0.001,0.01,0.1		

Relu: rectified linear unit

2.3. Hyper-parameters optimisation of ML models

For both models, the tuning of the hyper-parameters was done automatically using an *exhaustive grid search* where all the combinations of hyper-parameters are tested to find the best combination. The best combination is found through a 4-fold cross-validation. The dataset is split into 4 subsets. Then for each of the 4 folds, the model is trained using 3 of the folds and tested on the 4th fold. The performance of the regression is measured as the mean r^2 value obtained across the 4 folds. After finding the best set of hyper-parameters, the model is retrained on the whole dataset.

2.4. Estimation of bioavailable concentrations and toxicity

ML algorithms were trained and tested for each metal (As, Cd, Cr, Cu, Hg, Ni, Pb, Se, and Zn) and each of the following hydrocarbon: (1) aliphatic fraction including all individual compounds in the range of EC_{10} to EC_{40}, (2) aromatic fraction including: Acenapthene (AE), Fluorene (F), Anthracene (A), Phenantrene (P), Pyrene (PY), Chrysene (C),

Benzo(a)anthracene (BA), Benzo(a)pyrene (BaP), Benzo(b)fluoranthrene (BB), Benzo(k)fluoranthrene (BK), Dibenzo(a,h)anthracene (DA), Benzo[g,h,i]perylene (BP), Indeno[1,2,3-c,d]pyrene (IP). To estimate the concentration of a hydrocarbon/metal at time t, the inputs of the ML models were the soil type (Soil 1 or Soil 2), the amendment (un-amended, biochar, or compost), and the initial (t=0) total concentration of the hydrocarbon/metal. After training/testing via the 4-fold cross-validation, the NN models with $r^2 > 0.7$ (good model fitting) were selected to generate the inputs for the estimation of the toxicity (Figure 1 b). Using models with a lower r^2 may generate inaccurate inputs which would confuse the estimation of the toxicity.

NN was used instead of RF, because RF provides a discrete output dictated by the finite number of split points in each tree as explained in section 2.2. Since measurements are only available at time t=0, 30, 60, and 180 days, the RF output does not vary continuously in between these data points which may lead to unrealistic variations in bioavailable concentrations. Here, RF is mainly used (i) as a benchmark against NN for which finding an appropriate combination of hyper-parameters is less straightforward, and (ii) because it provides access to the relative importance of the input variables.

2.5. Estimation of toxicity

NN and RF algorithms were then trained and tested for each biological and eco-toxicological indicators listed in Table S4 of the Supplementary Materials. To estimate a given indicator at time t, the ML models take again as inputs the soil type (Soil 1 or Soil 2), the amendment (un-amended, biochar, or compost), but also the bioavailable concentration of the hydrocarbons/metals predicted with NN having $r^2>0.7$. The latter is anticipated to be a valuable input for the toxicity estimation, as bioavailable concentration is reported to be a good proxy for toxicity estimation (Gourlay-Francé and Tusseau-Vuillemin, 2013).

2.6. Drivers of bioavailability and toxicity

To investigate which of the input features are driving the estimation of the bioavailability/toxicity variables, we rely on the RF model which can output the relative importance of each input feature. To do so, the Skicit-learn library implements the method based on the Mean Decrease in Impurity (MDI) described by Breiman (1984). In a decision tree, every node corresponds to a logical if-then condition on a single input feature, which splits the input dataset into 2 subsets. The choice of the feature and its splitting condition is based on a measure called impurity. For regression trees, the impurity in the data is quantified by the variance, and the training phase corresponds to finding the splits which lead to the greatest reduction in variance. To reflect the feature importance, the MDI accounts for the number of times a feature is selected to split the data and how many data samples it splits (i.e. the importance of the splits). In practice, an input feature is important if it is often used by the trees to make decisions and if these decisions are concerning many data samples.

3. Results and discussion

3.1. Prediction of the bioavailable concentration of hydrocarbons and metals

The accuracy of the ML models (r^2 value) for the prediction of bioavailability is shown in Figure 2. A thorough inspection of the data revealed that noisy measurements, i.e. when two replicates are different from each other, are one of the main reasons for the lower r^2 values obtained for some hydrocarbons and elements. The corresponding combinations of hyper-parameters returned by the grid search are shown in Table S5 of the Supplementary Materials. The models with high r^2 can be used to estimate bioavailability at any time step. For example, Figure 3 shows the predicted bioavailable concentration of Fluorene (F) (Figure 3a and b), benzo[a]anthracene (BA) (Figure 3c and d), benzo[a]pyrene (BaP) (Figure 3e and f), and Copper (Cu) (Figure 3g and h) with RF and NN, respectively. The corresponding r^2 values

were $r^2 = 0.93$ (RF) and $r^2 = 0.89$ (NN) for F, $r^2 = 0.97$ (RF) and $r^2 = 0.91$ (NN) for BA, $r^2 = 0.96$ (RF) and $r^2 = 0.90$ (NN) for BaP, and values of $r^2 = 0.96$ (both RF and NN) for Cu.
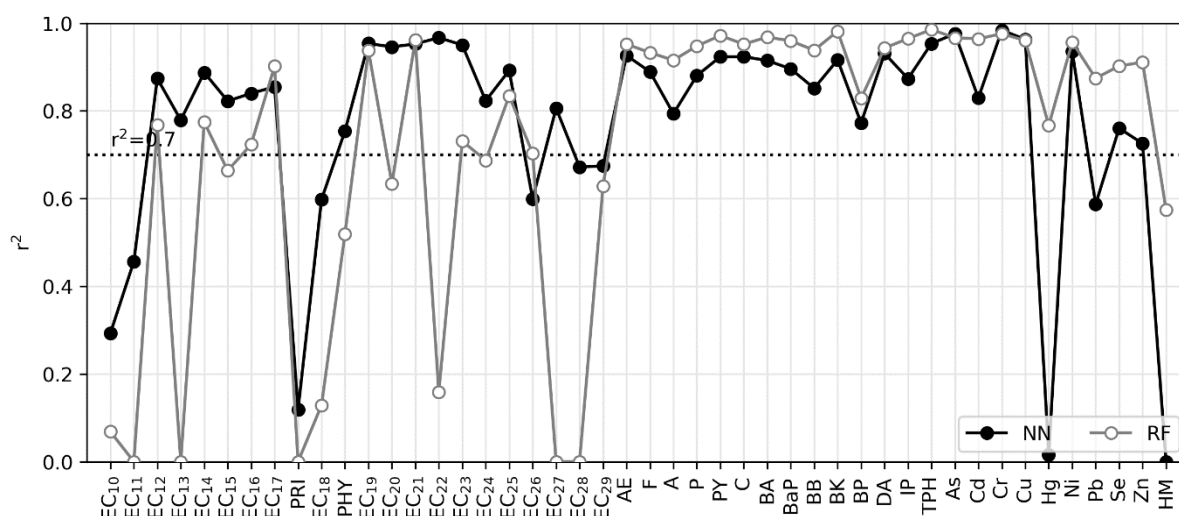

Figure 2: Prediction accuracy (r^2) for the bioavailable concentration of hydrocarbons and heavy metals/metalloids using Neural Network (black) and Random Forest (grey).

As explained in the methodology section, RF generates an output with discrete values which may lead to unrealistic variations due to the measurements being available only at a few time steps. In comparison, NN generates a continuous output which can be used to estimate the concentration at intermediate time steps, but also to forecast the potential evolution of the concentration for time steps beyond 180 days. Therefore, in the next section, we use NN, and not RF, to generate the estimated bioavailable concentrations used as inputs of the toxicity models. The selected hydrocarbons are EC_{12}, EC_{14}, EC_{16}, EC_{17}, EC_{19}, EC_{21}, EC_{23}, EC_{25}, EC_{26}, AE, F, A, P, PY, C, BA, BaP, BB, BK, BP, DA, IP, P, and the selected elements are As, Cd, Cr, Cu, Hg, Ni, Se, Pb, and Zn. (Figure 2; NN $r^2 > 0.7$).

In this mesocosms experimental setup the decrease in hydrocarbons bioavailable concentration was 2 times higher in Soil 1 (high contamination) than in Soil 2 (low contamination). The most significant changes were observed for the aromatics EC_{16-21} and aliphatic EC_{16-35} fractions; where readily available /bioavailable (methanol extracted) concentration decreased by 71, 58%

(Soil 1 + Compost) and by 11, 91% (Soil 2 + Compost). Fewer changes in bioavailable concentration were observed in Soil 2, which overall had lower concentrations of contaminants. This is clearly highlighted in Figure 3, where Fluorene (F) was plotted as representative aromatic compound for the EC_{16}-EC_{21} fraction. The concentration of F significantly decreased after 180 days incubation for Soil 1 + Compost. It was also clear that the soils amended with compost had F concentration 10 times lower than the non-treated (Soil 1) soil at the end of the simulation period.

(a)   (b)

(c)   (d)

(e)   (f)

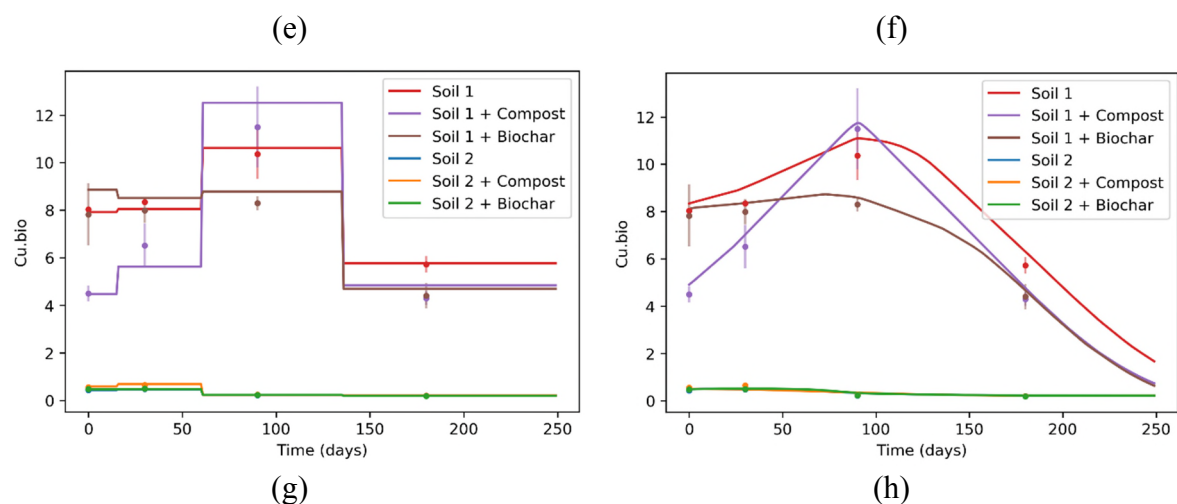(g)   (h)
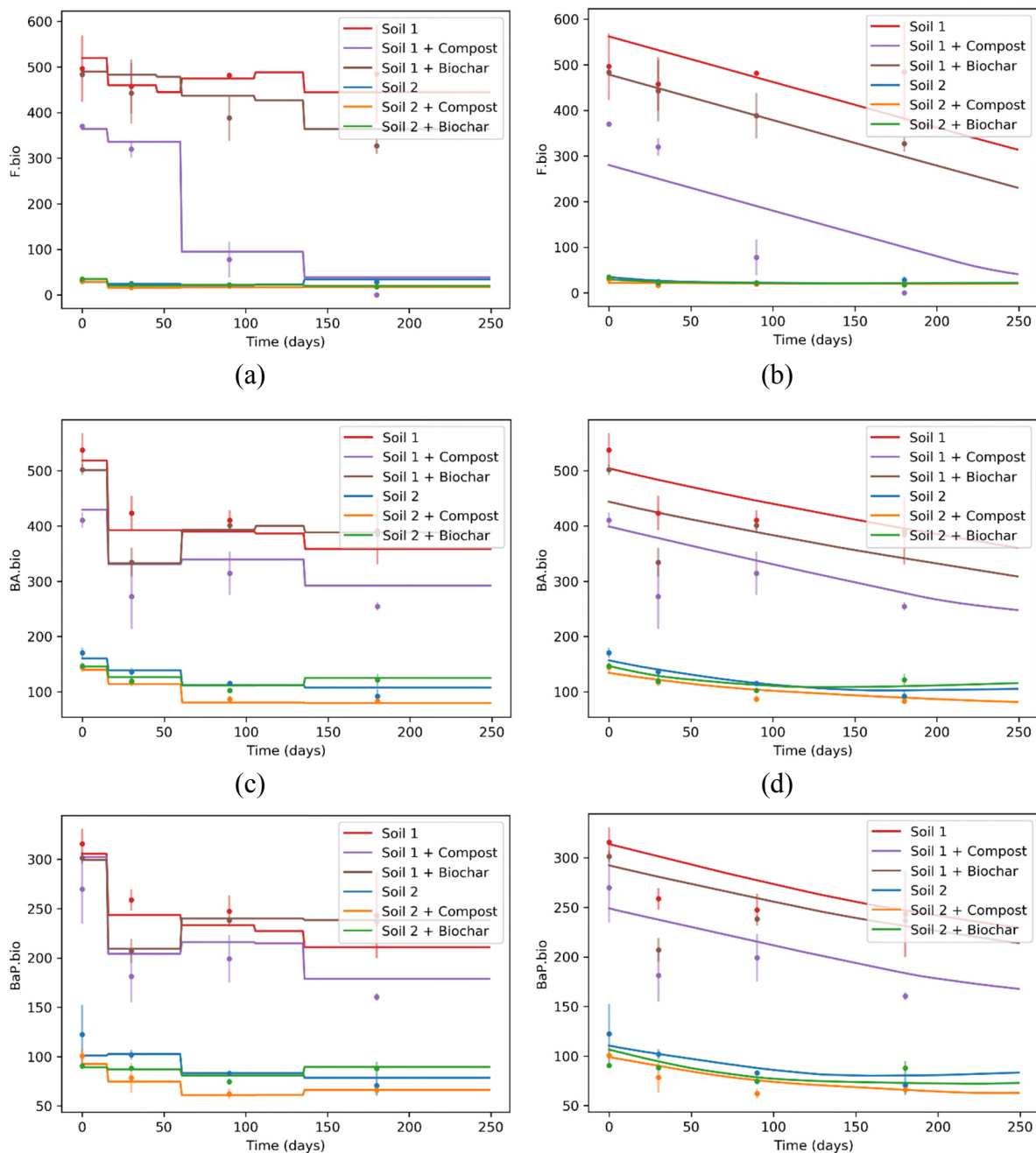
Figure 3: Prediction of the bioavailable concentration (mg/kg) of Fluorene (F), benzo[a]anthracene (BA), benzo[a]pyrene (BaP), and Copper (Cu) using random forest (left) and neural network (right). The continuous lines are the predicted concentrations for the different soils (Soil 1 and Soil 2), amendments (biochar, compost) and un-amended. The dots are the available measurements after averaging the replicates (the error bars are the standard deviation across the replicates).

The relative feature importance when estimating bioavailability with RF for compounds modelled with $r^2 > 0.7$ is presented in Figure S1 of the Supplementary Materials. The bar chart shows that for each compound different features (time, soil, or amendment) dominate. For example, the prediction of the AE, P, PY, BA, BaP, and DA were mainly driven by soil properties and the total concentration of the compounds at onset. The importance of the soil feature reflects the differences observed in bioavailable concentrations between Soil 1 (heavily contaminated) and Soil 2 (low contamination). For most of the HMs/metalloids (As, Cd, Cr, Cu, Ni, Zn) time (ageing) and amendment (compost or biochar) were not important variables influencing bioavailability prediction (Figure S1 of the Supplementary Materials). Indeed HMs/metalloids bioavailable concentration did not change significantly over incubation time.

3.2. Prediction of the toxicity

Overall, RF performed slightly better than NN to predict the toxicity (Figure 4), and was particularly good at predicting seed germination, condition index, and EC_{50}. Lower r^2 values were obtained for the PLFA, which can be explained by different factors including: (1) the use of quantitative data (concentrations) over qualitative data (relative abundance of individual PLFA) could have provided a better input for r^2 estimation; (2) it was not possible to conclusively attribute these PLFA changes based on the input parameters of this model as there are multiple factors (or additional factors) affecting community composition. Further to this, we did not evaluate the effects of other chemicals (e.g metabolites) which may have impacted the microbial community at early experimental stages (e.g 2, 5, 10, 20 days), thus sampling regime established (30, 90, 180 days) was not significant for assessing smaller but significant toxic effects on the microbial community. The corresponding combinations of hyper-parameters returned by the grid search are shown in Table S6 of the Supplementary Materials.
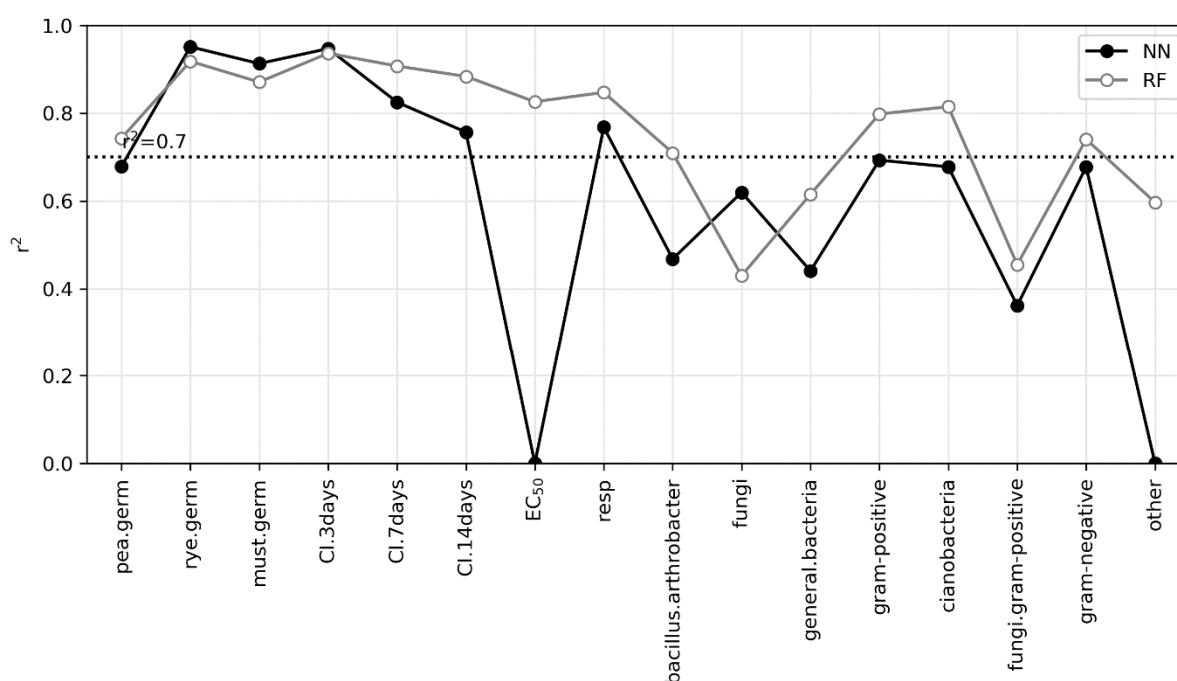
Figure 4: Prediction accuracy for the toxicity using Neural Network (NN) (black) and Random Forest (RF) (grey). Germ: germination, CI: condition index, EC$_{50}$: Half maximal effective concentration, resp: soil respiration (mg CO$_2$/g soil).

As done for the hydrocarbons, metals and metalloids, ML models with $r^2 >0.7$ were successfully used to predict the toxicity at intermediate time step (e.g. Figure 5 shows the NN prediction obtained for mustard germination and the earthworm toxicity assay (14-day condition index)). Good predictions were obtained for CI 3 days ($r^2 = 0.95$), CI 7 days ($r^2 = 0.83$), CI 14 days ($r^2 = 0.76$), mustard germination ($r^2 = 0.91$), ryegrass germination ($r^2 = 0.95$), and soil respiration ($r^2 = 0.77$).
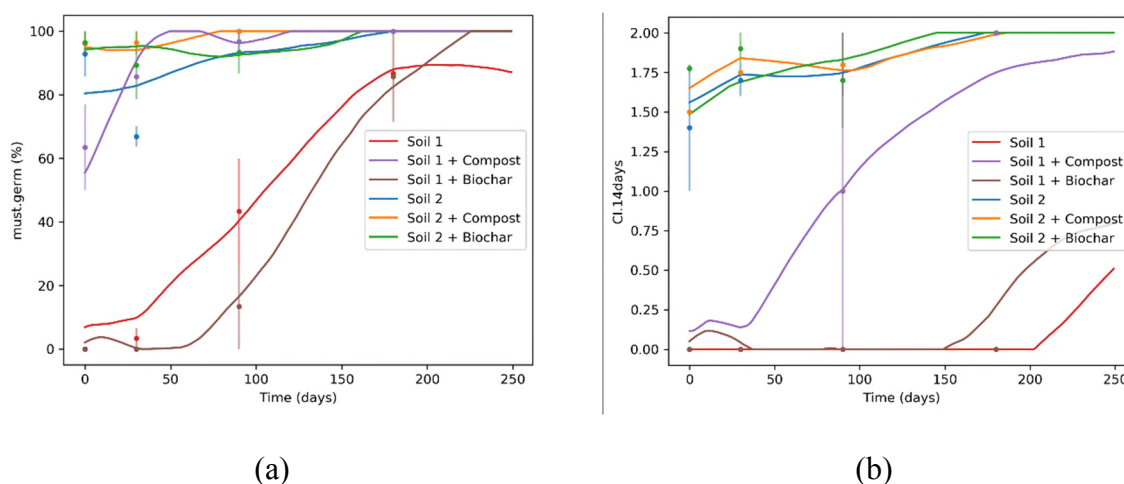

(a)   (b)

Figure 5: Prediction of (a) mustard seeds germination (r^2=0.91) and (b) earthworms 14-day condition index (r^2=0.75) using Neural Network. The continuous lines are the predicted values for the different soils and amendments, the dots are the available measurements after averaging the replicates (the error bars are the standard deviation across the replicates). The minimum value allowed is 0, and the maximum values allowed are 100 and 2 for mustard germination and condition index, respectively.

Empirical data, obtained from the mesocosms experimental setup showed that Soil 1 had a greater toxic effect on mustard seed germination than Soil 2, likely because concentration of contaminants was 4 times lower in Soil 2 (TPH > 2000 mg/kg) compared to Soil 1 (TPH >

8000 mg/kg). Interestingly, seed germination showed a good ecological recovery for Soil 1 + Compost and even for Soil 1 which reached similar recovery after 180 days. The model predictions highlighted the potential for all soil samples to achieve 100% germination after 250 days (Figure 5a).

Similar ecological recovery for Soil 1 were observed from empirical data obtained for earthworm lethality assay (*E fetida*); while the CI was similar at the onset of the experiment (CI = 0, mortality), and only Soil 1 + Compost showed greater ecological recovery (CI was 2 times higher than biochar and un-amended conditions). The predicted ecotoxicity of Soil 1+Biochar and Soil 1 significantly decreased after 250 days (CI = 0.4 - 1) (Figure 5b). The progressive positive effect of Soil 1 + Compost was also reflected by the prediction where at 250 days the sample ceased to be lethal (CI = 2) to earthworms assay. For the less contaminated samples, a higher CI was recorded during the experimental setup for both Soil 2 + Biochar and Soil 2 + Compost compared with Soil 2 (Figure 5b). The earthworms in Soil 2 + Compost and Soil 2+ Biochar remained in the healthiest condition throughout the 14 days assay as no significant differences were observed in the empirical data which was also reflected by the model.

3.3. Drivers for toxicity

The most important input features for the estimation of the toxicity are shown in Figure 6. They were identified by looking at the relative feature importance from the RF modelwith r^2 >0.7. This allows assessing the influence of each feature on the prediction.
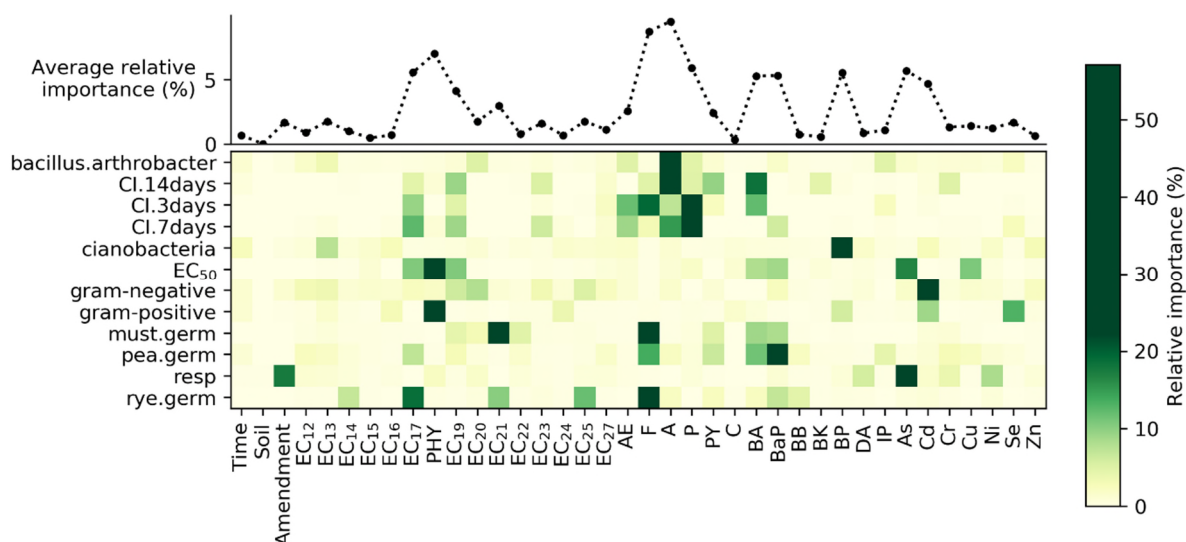
Figure 6: Relative feature importance when estimating toxicity with Random Forest. The heatmap shows the relative feature importance for all the toxicity variables which were modelled with an r^2 value above 0.7. The top graph shows the relative feature importance averaged across the toxicity variables.

The heatmap shows that, for each toxicity variable, different features dominate. For example, the prediction of the CI (earthworm acute toxicity index) was mainly driven by EC_{17}-EC_{19}, AE, F, A, P, and PY. A previous study highlighted the key role of the mid-chain length aliphatic compounds in toxicity to earthworms (Dorn and Salanitro, 2000). In particular, these compounds are the most volatile, soluble, and biodegradable constituents of the tar fraction remaining after the initial removal of the more soluble/volatile components (e.g low molecular weight compounds). Thus, mid-chain length aliphatic compounds can be a valuable indicator of acute toxicity to soil dwelling organisms (e.g., earthworms). Furthermore, this fraction, along with small aromatic compounds such as phenanthrene and acenaphtene, represented the least aged fraction (more bioavailable) with the highest toxicity potential, and should therefore be of special concern with respect to remediation/natural attenuation monitoring (Yang et al., 2016).

In comparison, the heatmap suggests that respiration was mainly related to arsenic (As) and soil amendment (biochar or compost). Toxicity to As is caused by its ability to inhibit basic cellular functions, and disrupt microbial metabolism (Walker et al., 2000). A number of studies previously highlighted the adverse effects of As contamination on microbial biomass C and respiration rates (Edvantoro et al., 2003; Van Zwieten et al., 2003) and in particular bioavailable arsenic exerted greater inhibitory effect rather than to total arsenic concentration (Ghosh et al., 2004). On the opposite, the model showed that EC_{15}-EC_{20}, Cd, Cr, Cu, Ni, Se, and Zn ranked among the least important features for respiration.

Overall, Time, soil type, amendment (biochar and compost) were not identified as important features. This is probably because these features were already used as inputs of the ML models to estimate the bioavailable compounds, thus their effects may be embedded in the variations of bioavailable hydrocarbons and metals. In summary, these findings reinforce the idea that the bioavailability of multiple hydrocarbons and metals drives the soil toxicity.

3.4. Implications and limitations of ML for risk assessment and remediation end-point evaluation

The two ML models used in this study were trained using a limited dataset. The models are a priory only valid for the values of the input variables which are captured by the training dataset. For example, the models may not accurately predict the toxicity for soils which are too different from Soil 1 and Soil 2. Similarly, the accuracy of the models may decrease when predicting toxicity at time steps larger than 180 days. This remark also applies to the analysis of feature importance. For example, our results suggest that the bioavailable concentration of arsenic is important to predict respiration. Although, this is true for our dataset, applying our method to a different soil type and contamination may suggest that other variables are also driving changes in respiration. Nonetheless, our study indicates that ML models can help us understand

complex mixtures fate, and identify the key variables affecting their behaviour and the environmental risks posed by the various pools of contaminants. Future studies should indeed investigate the importance of other input variables such as pH, soil organic matter and dissolved organic carbon on contaminant bioavailability and toxicity changes. Recovery of contaminated soils in urban areas is nowadays a priority because of the increased urbanization which requires suitable and safe land for construction. At several stages, relative costs and benefits must be considered when evaluating potential remedial actions. Field-scale studies are often implemented in real-case scenario to provide relevant empirical evidence to support the design of a full-scale remediation strategy.

4. Conclusions

Empirical data from a 6-month mesocosms experiment were used to assess the ability and performance of two ML models to predict temporal bioavailability changes of polycyclic aromatic hydrocarbons, aliphatic hydrocarbons, and heavy metals/metalloids in contaminated soils amended with compost or biochar. In addition, ML was used to predict the toxicity changes, mainly based on the knowledge of some of the bioavailable concentrations. Results obtained showed that both NN and RF were able to model the bioavailability of various contaminants and should be used in combination as: NN model provides a realistic continuous output, while RF can explain which input measurements are actually important to predict the toxicity. This study suggests that ML models are good candidate tools to support remediation monitoring of multi-contaminated sites, in a cost-effective manner. ML capabilities should be further investigated with larger datasets encompassing a representative range of soil types and contaminations.

Acknowledgement

5. References

Abbasi, M., El Hanandeh, A., 2016. Forecasting municipal solid waste generation using artificial intelligence modelling approaches. Waste Manag. 56, 13–22.

Bardos, P., Bone, B., Boyle, R., Ellis, D., Evans, F., Harries, D.N., Smith, W.. J., 2011. Applying sustainable development principles to contaminated land management using the SuRF-UK framework. Remediat. J. 21, 77–100.

Boopathy, R., 2000. Factors limiting bioremediation technologies. Bioresour. Technol. 74, 63–67.

Brand, E., Smedes, F., Jonker, M.T.O., Harmsen, J., Peijnenburg, W.J.M., Lijzen, J.P.A., 2012. Advice on implementing bioavailability in the Dutch soil policy framework - User protocols for organic contaminants. RIVM Report 711701102/2012 73.

Breiman, L., 1984. Classification and regression trees.

Cave, M.R., Milodowski, A.E., Friel, E.N., 2004. Evaluation of a method for identification of host physico-chemical phases for trace metals and measurement of their solid-phase partitioning in soil samples by nitric acid extraction and chemometric mixture resolution. Geochemistry Explor. Environ. Anal. 4, 71–86.

Coulon, F., Pelletier, E., St Louis, R., Gourhant, L., Delille, D., 2004. Degradation of petroleum hydrocarbons in two sub-antarctic soils: influence of an oleophilic fertilizer. Environ. Toxicol. Chem. 23, 1893–1901.

Coulon, F., Whelan, M.J., Paton, G.I., Semple, K.T., Villa, R., Pollard, S.J.T., 2010. Multimedia fate of petroleum hydrocarbons in the soil: Oil matrix of constructed biopiles. Chemosphere 81, 1454–1462.

Coulon, F.F., Al Awadi, M., Cowie, W., Mardlin, D., Pollard, S., Cunningham, C., Risdon, G., Arthur, P., Semple, K.T., Paton, G.I., 2010. When is a soil remediated? Comparison of biopiled and windrowed soils contaminated with bunker-fuel in a full-scale trial. Environ. Pollut. 158, 3032–3040.

Dawson, J.J.C., Godsiffe, E.J., Thompson, I.P., Ralebitso-Senior, T.K., Killham, K.S., Paton, G.I., 2007. Application of biological indicators to assess recovery of hydrocarbon impacted soils. Soil Biol. Biochem. 39, 164–177.

Dorn, P. B. and Salanitro, J.P., 2000. Temporal ecological assessment of oil contaminated soils before and after bioremediation. Chemosphere 40, 419–426.

Edvantoro, B.B., Naidu, R., Megharaj, M., Singleton, I., 2003. Changes in microbial properties associated with long-term arsenic and DDT contaminated soils at disused cattle dip sites. Ecotoxicol. Environ. Saf. 55, 344–351.

Eurpoean Environment Agency, 2012. Overview of contaminants affecting soil and groundwater in Europe [WWW Document]. Publ. 12 Nov 2009. URL http://www.eea.europa.eu/data-and-maps/figures/overview-of-contaminants-affecting-soil-and-groundwater-in-europe (accessed 12.8.15).

Frostegård, Å., Tunlid, A., Bååth, E., Frostegård, A., Tunlid, A., Bååth, E., 1993. Phospholipid Fatty Acid composition, biomass, and activity of microbial communities from two soil types experimentally exposed to different heavy metals. Appl. Environ. Microbiol. 59, 3605–17.

Ghosh, A.K., Bhattacharyya, P., Pal, R., 2004. Effect of arsenic contamination on microbial biomass and its activities in arsenic contaminated soils of Gangetic West Bengal, India. Environ. Int. 30, 491–499.

Gourlay-Francé, C., Tusseau-Vuillemin, M.-H., 2013. Bioavailability of Contaminants BT - Encyclopedia of Aquatic Ecotoxicology. In: Férard, J.-F., Blaise, C. (Eds.), . Springer Netherlands, Dordrecht, pp. 181–190.

Guo, M., Gong, Z., Allinson, G., Tai, P., Miao, R., Li, X., Jia, C., Zhuang, J., 2016. Variations in the bioavailability of polycyclic aromatic hydrocarbons in industrial and agricultural soils after bioremediation. Chemosphere 144, 1513–1520.

Harmsen, J., Naidu, R., 2013. Bioavailability as a tool in site management. J. Hazard. Mater. 261, 840–846.

Heshmati R, A.A., Mokhtari, M., Shakiba Rad, S., 2014. Prediction of the compression ratio for municipal solid waste using decision tree. Waste Manag. Res. 32, 64–69.

Hu, J., Nakamura, J., Richardson, S.D., Aitken, M.D., 2012. Evaluating the Effects of Bioremediation on Genotoxicity of Polycyclic Aromatic Hydrocarbon-Contaminated Soil Using Genetically Engineered, Higher Eukaryotic Cell Lines. Environ. Sci. Technol. 46, 4607–4613.

Kelsey, J.W., Kottler, B.D., Alexander, M., 1997. Selective chemical extractants to predict bioavailability of soil-aged organic chemicals. Environ. Sci. Technol. 31, 214–217.

Khamforoush, M., Rahi, M., Hatami, T., Rahimzadeh, K., 2011. The Use of Artificial Neural Network (ANN) for Modeling of Diesel Contaminated Soil Remediation by Composting Process. Ind. Eng. Eng. Manag. (IEEM), 2011 IEEE Int. Conf. 585–589.

Kördel, W., Bernhardt, C., Derz, K., Hund-Rinke, K., Harmsen, J., Peijnenburg, W., Comans,

R., Terytze, K., 2013. Incorporating availability/bioavailability in risk assessment and decision making of polluted sites, using Germany as an example. J. Hazard. Mater. 261, 854–862.

Lehmann, J., Joseph, S., 2015. Biochar for Environmental Management: Science, Technology and Implementation. Taylor & Francis.

Li, J., Heap, A.D., Potter, A., Daniell, J.J., 2011. Application of machine learning methods to spatial interpolation of environmental variables. Environ. Model. Softw. 26, 1647–1659.

Lyu, H., Gong, Y., Gurav, R., Tang, J., 2016. Chapter 9 - Potential Application of Biochar for Bioremediation of Contaminated Systems BT - Biochar Application. Elsevier, pp. 221–246.

Mason, I.G., 2006. Mathematical modelling of the composting process: A review. Waste Manag. 26, 3–21.

Ortega-Calvo, J.-J.J., Harmsen, J., Parsons, J.R., Semple, K.T., Aitken, M.D., Ajao, C., Eadsforth, C., Galay-Burgos, M., Naidu, R., Oliver, R., Peijnenburg, W.J.G.M.G.M., Römbke, J., Streck, G., Versonnen, B., 2015. From Bioavailability Science to Regulation of Organic Chemicals. Environ. Sci. Technol. 49, 10255–10264.

Paton, G.I., Viventsova, E., Kumpene, J., Wilson, M.J., Weitz, H.J., Dawson, J.J.C., 2006. An ecotoxicity assessment of contaminated forest soils from the Kola Peninsula. Sci. Total Environ. 355, 106–117.

Pedregosa, F., Weiss, R., Brucher, M., 2011. Scikit-learn : Machine Learning in Python 12, 2825–2830.

Petric, I., Selimbašić, V., 2008. Development and validation of mathematical model for aerobic composting process. Chem. Eng. J. 139, 304–317.

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. Ecosystems 9, 181–199.

Rajaee, T., Mirbagheri, S.A., Zounemat-Kermani, M., Nourani, V., 2009. Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. Sci. Total Environ. 407, 4916–4927.

Reichenberg, F., Mayer, P., 2006. Two complementary sides of bioavailability: accessibility and chemical activity of organic contaminants in sediments and soils. Environ. Toxicol. Chem. 25, 1239.

Sahoo, G.B., Ray, C., Mehnert, E., Keefer, D.A., 2006. Application of artificial neural networks to assess pesticide contamination in shallow groundwater. Sci. Total Environ. 367, 234–251.

Semple, K.T., Morriss, a. W.J., Paton, G.I., 2003. Bioavailability of hydrophobic organic contaminants in soils: fundamental concepts and techniques for analysis. Eur. J. Soil Sci. 54, 809–818.

Sturm, I., Lapuschkin, S., Samek, W., Müller, K., 2016. Interpretable deep neural networks for single-trial EEG classification. J. Neurosci. Methods 274, 141–145.

Umeh, A.C., Duan, L., Naidu, R., Semple, K.T., 2017. Residual hydrophobic organic contaminants in soil: Are they a barrier to risk-based approaches for managing contaminated land? Environ. Int. 98, 18–34.

Van Zwieten, L., Ayres, M.R., Morris, S.G., 2003. Influence of arsenic co-contamination on DDT breakdown and microbial activity. Environ. Pollut. 124, 331–339.

Walker, C., Goodyear, C., Anderson, D., Titball, R.W., 2000. Identification of arsenic-

resistant bacteria in the soil of a former munitions factory at Löcknitz, Germany. L. Contam. & Reclam. 8, 13–18.

Wu, G., Kechavarzi, C., Li, X., Wu, S., Pollard, S.J.T.T., Sui, H., Coulon, F., 2013. Machine learning models for predicting PAHs bioavailability in compost amended soils. Chem. Eng. J. 223, 747–754.

Wu, G., Li, X., Kechavarzi, C., Sakrabani, R., Sui, H., Coulon, F., 2014. Influence and interactions of multi-factors on the bioavailability of PAHs in compost amended contaminated soils. Chemosphere 107, 43–50.

Wu, X., Zhu, L., 2016. Evaluating bioavailability of organic pollutants in soils by sequential ultrasonic extraction procedure. Chemosphere 156, 21–29.

Yang, X., Yu, L., Chen, Z., Xu, M., 2016. Bioavailability of Polycyclic Aromatic Hydrocarbons and their Potential Application in Eco-risk Assessment and Source Apportionment in Urban River Sediment. Sci. Rep. 6, 23134.

Yu, X.Z., Wu, S.C., Wu, F.Y., Wong, M.H., 2011. Enhanced dissipation of PAHs from soil using mycorrhizal ryegrass and PAH-degrading bacteria. J. Hazard. Mater. 186, 1206–1217.