Machine Learning Approach to Identify the Relationship Between Heavy Metals and Soil Parameters in Salt Marshes

Iman Salehi Hikouei[1], S Sonny Kim[2]*, Lori A Sutter[3], Jason Christian[4], Stephan Durham[2] and Jidong J Yang[2]

[1]University of Maryland, Center for Environmental Science, Appalachian Laboratory, USA

[2]University of Georgia, College of Engineering, USA

[3]University of Georgia, Warnell School of Forestry & Natural Resources, USA

[4]MMC Engineering Group LLC, 1711 Meriweather Drive, USA

Submission: March 25, 2021; **Published:** April 14, 2021

*Corresponding author: S Sonny Kim, University of Georgia, College of Engineering, USA

Abstract

Influences from tidal flooding and freshwater inundation from upland watersheds create an environmentally important ecosystem along coastlines, namely salt marshes. Salt marshes have been recognized as effective sinks for organic carbon and heavy metal contaminants. A detailed understanding of the specific binding agents in the soils on the storage of contaminants is investigated herein using two modern machine learning algorithms: extreme gradient boosting (XGboost) and random forest (RF). Results of the current work indicate that Fe is the most important binding agent for As, Cd, Cr and Zn while Mn and organic matter are the most important binding agents for Cu and Pb. Noting the fact that an increase in salinity not only causes heavy metal release into aquatic systems but also leads to a decrease in floral growth and organic matter production, the findings of this study help to formulate proper remediation strategies to contain heavy metals in tidal marshes.

Keywords: Tidal marsh; Heavy metal; Halophyte; Random forest; XGboost

Introduction

Coastal saltmarshes are ecologically sensitive and vital habitats that connect the mainland and the marine environments and provide habitat for a large amount of plants and animals, embracing many substantial biodiversity resource species [1,2]. Saltmarshes enhance the quality of water, maintain the health of estuaries, and act as a buffer through filtering sediments, nutrients, and other dissolved and particulate constituents in runoff [3,4]. With rapid coastal development and population increase for more than two decades, extensive tracts of tidal marshes around the world have been drained and disturbed due to construction activities causing salinity gradient changes [5,6]. Change in salinity gradient alters carbon and nitrogen distribution patterns in tidal marshes [7]; many salt marshes with salinity[1] above 18 (polyhaline) have lower carbon and nitrogen concentrations than freshwater (less than 0.5), oligohaline (0.5-5), and mesohaline (5-18) tidal marshes [8-10].

Tidal marshes are able to absorb heavy metals through physical, chemical, and biological sequestration [11]. Heavy metal concentration in saltmarshes is a function of input sources, soil composition and texture, organic matter content, flooding duration or frequency, riverine circulation, and vegetation community [12,13]. Soils with high clay and organic matter content retain metals due to high cation exchange capacity and surface particle charge [14]. The metals arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), lead (Pb) and zinc (Zn) are the most critical heavy metals causing environmental issues [15], thus their measurement and evaluation are considered to be an important endeavor [16,17]. An excess amount of heavy metals is an ecological threat for the health of vegetation, animals and humans [18]. These heavy metals are potentially capable of reducing biomass production through bioaccumulation and biomagnification processes [19]. Further, they are capable of

---

[1]Salinity is arguably best measured in PSU, which is a unitless metric and will be reported without units throughout this paper.

polluting waterways [20]. The fate and transport of a heavy metal in a hydric soil depends substantially upon the chemical form and speciation of the heavy metal [21]. In soil, heavy metals are adsorbed through initial rapid reactions (occurring over hours) which are followed by gentle adsorption reactions (occurring over days). Then, they are redistributed into varying chemical forms with different bioavailability, mobility, and toxicity [22].

Multiple linear regression method, generalized linear model, partial least squares regression, geographical weighted regression and linear mixed model are popular methods for their simplicity in both application and interpretation but not novel concepts [23,24]. For example, multivariate statistical methods were used to describe the pollution dynamics influencing marine sediments and distinguish between anthropogenic and natural sources of contamination [25]. Despite of this simplicity, these models lead to unsatisfactory performances [26,27]. Lack of important data linked to response variables and self-limitations of these methods as well as complexity of relationships among variables are key reasons for their unsatisfactory estimation [28]. Recently, machine learning has been used solve chemical equilibrium of reactive transport modeling instead of fast-converging numerical methods [29]. According to [29], machine learning had an acceptable performance in solving chemical equilibrium models and significantly reduced the analysis time. Machine learning methods are considered as a reliable technique for soil characterization [30-32], which is able to identify patterns in data that determine a significant unpredictable nonlinearity [31]. Being a data-driven method, a machine learning algorithm only depends upon data quality and model architecture [31]. Machine learning is able to detect the important features of the pattern at different scales explaining both the linear and nonlinear influences [33,34]. Different Machine learning models have been frequently utilized to estimate soil attributes such as particle size fractions [35,36] and soil erosion rates [36]; however, use of machine learning algorithms for estimating soil contamination parameters especially in a saltmarsh environment still remains limited. Machine learning methods assist to predict contamination concentration in complex non-linear contexts [31]. Machine leaning models can be applied to the measured data collected in the monitoring, and then can be utilized to estimate the pollution at the unmonitored areas [37,38]. The ensemble machine learning models including boosted regression trees and random forest were also used for susceptibility mapping of groundwater hardness [39]. According to [39], random forest outperformed boosted regression trees and multivariate discriminant analysis models. Random Forest was utilized to determine the most important explanatory variables influencing nitrate concentration in the fontanili of the Adda and Ticino basins, Italy [40]. Two variables related to source of nitrogen from agriculture (i.e. the percentage of maize cultivated soil and the N from livestock manure) as well as the distance from the rivers were identified as the most influential variables in nitrate concentration [40]. According to [41], random forest had an acceptable performance in modeling heavy metals in

aquatic system and was especially efficient once dataset is small and moving datapoints from the training set to the test dataset is undesirable. They also showed that the tolerance of RF models to noise which is irrelevant descriptors and errors in target property values was acceptable.

In this current study, two popular modern machine learning methods, extreme gradient boosting (XGboost) and random forest (RF), were deployed to investigate the most influential binding agents among organic matter, clay, Fe, and Mn for corresponding heavy metals including As, Cd, Cr, Cu, Pb, Zn.

Materials and Methods

Study areas and vegetation communities

Sampling occurred in eight tidal marshes (Figure 1) along the southeast coast of the US in Georgia. These marshes are typical of southeastern salt marshes. At creek banks, low marshes were inundated twice daily by tidal action and vegetated with *Spartina alterniflora* (syn. *Sporobolus alterniflorus*). With a progression landward, the marshes were inundated less frequently, and typical vegetation of the high marsh included short form *S. alterniflora* and *Juncus roemerianus*. At the upland border, there was a zone of species that was less tolerant of flooding, including *Borrichia* frutescens. Three different representative sampling areas (A, B, C) were chosen along these transects based on vegetation dominance, and three replicates of samples at each location were taken to ensure sufficient representation in the dominant vegetation zones. The dominant vegetative species were determined based on which species represented more than 50% total coverage (stem density) in each selected site. For each distinct vegetation community, species richness and number of individuals were estimated utilizing the cover scale of Braun-Blanquet [42]. Transects began several meters from the upland border inside the marsh so that all samples were representative of the marsh itself.

According to a vegetation survey carried out at the study sites in June 2018, *Spartina alterniflora* was the predominant species at sites 1.A, 1.B, 1.C, 2.A, 2.B, 2.C, 6.A, 6.B, 6.C, 8.A, 8.B and 8.C (Table 1). Sites 3.A, 3.B and 3.C had 90%, 60% and 75% cover of *J. roemerianus,* respectively, often interspersed with *S. alterniflora* (Table 1). *B. frutescens* was the predominant species at sites 7.B and 7.C, and *Schoenoplectus tabernaemontani* was observed as a dominant plant at sites 4.A, 4.B, 4.C, 5.A, 5.B, and 5.C (Table 1).

Field and laboratory tests

During the site visit in June 2018, porewater at the study sites was sampled through PushPoint sampler [43], and salinity was measured by HI98194 portable meter (Hanna Instruments-Woonsocket/ Rhode Island/ United States). A total of 24 soil samples were collected from the rooting zone and kept intact in sealed waterproof containers to avoid moisture loss. All samples were transported to a laboratory within four hours and stored at 4°C for measurement of moisture content, organic matter content,

and particle size distribution. Laboratory analyses included particle size distribution, organic matter content, moisture content, metal speciation, x-ray diffraction (XRD), and carbon/ nitrogen determination. Particle size distribution using sieve (American Society for Testing and Materials (ASTM) D1140-17) and hydrometer (ASTM D422) methods were used to determine classification following the USDA soil classification system.

Moisture and organic matter content were also measured based on ASTM D2216–10 and ASTM D2974-87, respectively. Inductively Coupled Plasma (ICP) metal analysis was also performed in accordance with ASTM E1479-99 to measure the elemental constituency of soil samples. Total C & N was determined through dry combustion on a Flash 2000 (CE Elantech, Lakewood, New Jersey (USA)).
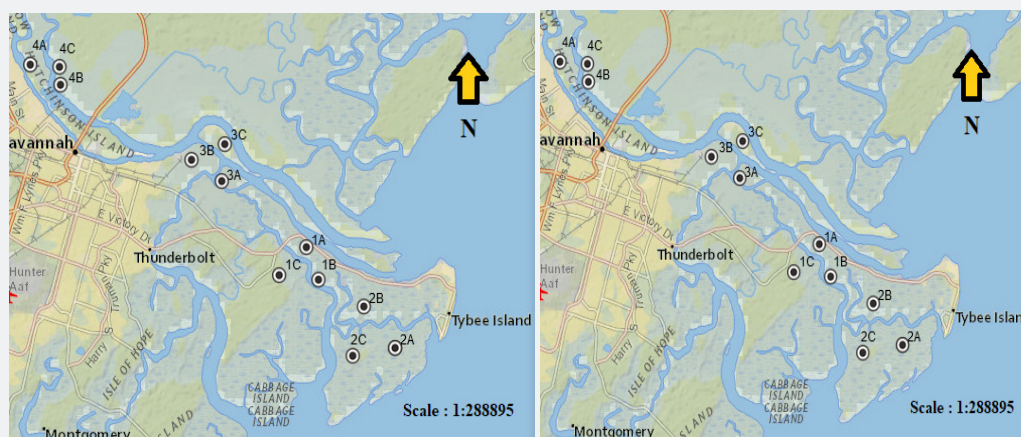

Figure 1: Sampling sites along the southeast coast of the United States.

Table 1: Salinity and vegetative species within eight sampling sites along the Georgia coast.

Site ID	Latitude (Degree)	Longitude (Degree)	Vegetation	Salinity
1.A	32.02745	-80.9253	*S. alterniflora*	27.03
1.B	32.02725	-80.9253	*S. alterniflora*	26.39
1.C	32.02717	-80.9254	*S. alterniflora*	25
2.A	32.01383	-80.8852	*S. alterniflora*	22.5
2.B	32.01387	-80.8851	*S. alterniflora*	31.22
2.C	32.01405	-80.8852	*S. alterniflora*	26.65
3.A	32.05983	-81.0237	*J.roemerianus*	20.29
3.B	32.05982	-81.024	*J.roemerianus*	5.56
3.C	32.0601	-81.0244	*J.roemerianus*	14.88
4.A	32.16555	-81.1574	*S. tabernaemontani*	4.04
4.B	32.16565	-81.1574	*S. tabernaemontani*	4.26
4.C	32.16587	-81.1575	*S. tabernaemontani*	4.84
5.A	31.36427	-81.439	*S. tabernaemontani*	3.38
5.B	31.36443	-81.4391	*S. tabernaemontani*	1.94
5.C	31.36448	-81.4392	*S. tabernaemontani*	2.46
6.A	31.16308	-81.4545	*S. alterniflora*	25.79
6.B	31.1623	-81.4547	*S. alterniflora*	28.76
6.C	31.16222	-81.4546	*S. alterniflora*	28.95
7.A	31.17008	-81.4226	*B. frutescens*	52.19
7.B	31.1701	-81.4226	*B. frutescens*	12.09
7.C	31.16993	-81.4227	*B. frutescens*	25.93

8.A	31.0741	-81.466	*S. alterniflora*	20.57
8.B	31.07413	-81.466	*S. alterniflora*	23.02
8.C	31.07418	-81.466	*S. alterniflora*	24.28

Machine learning algorithms

Two tree-based ensemble methods, i.e., Random Forest (RF) and Extreme Gradient Boosting (XGboost), were utilized to determine the most important binding agents for the aforementioned heavy metals (Figure 2). RF was considered as an ensemble since many trees were trained from bootstrapped samples in parallel and the results were aggregated (e.g., averaged for regression or majority vote for classification) for final prediction. By amalgamating individual tree models, the ensemble model was generally less biased with lower variance [44]. Besides the bootstrapping technique, RF adopted a random selection of a limited number of feature variables at each node split in order to decouple the trees. As such, these generated trees did not have collinearity issue with each other [45]. As a result, the prediction variance is greatly reduced. On the other hand, XGboost derived individual tree models in a sequential fashion and each individual tree model learned from the output obtained by the previous model [46]. Previous studies suggested that XGboost and RF models outperformed other machine learning techniques [47]. In this study, RF and XGboost were employed to identify the key drivers of heavy metals in hydric soils.
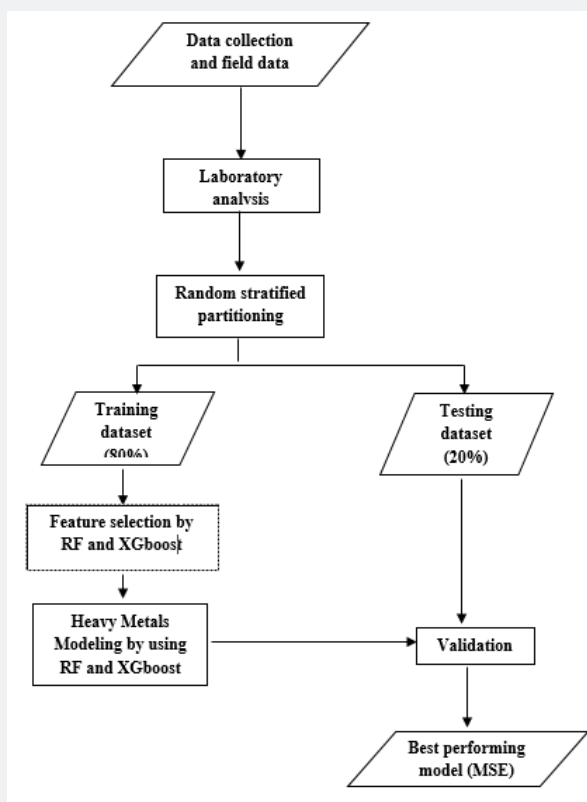

Figure 2: Flowchart illustrating steps for heavy metals modeling by RF and XGboost.

The RF model adopted in this paper included 200 trees on the training dataset, and the model assessment was carried out by using a separate test dataset that was not considered for the model training. The commonly used metric, mean squared error (MSE), was considered for model evaluation.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left[y_{predicted_i} - y_{Actual_i}\right]^2 \qquad (1)$$

Where, $y_{predicted_i}$ was the predicted value by a machine learning algorithm, y_{Actual_i} was the observed value, and N was the size of test dataset.

For illustration purposes, an exemplar decision tree is plotted in Figure 3, where an internal node represents feature (binding agents), the branch represents a decision rule, and each leaf

node represents the outcome (heavy metals concentration). Fe was the most important element tested for estimating arsenic concentration (Figure 3). Each individual tree in RF model was fitted to a random sample of the data and the results were then averaged for the final prediction. Tuning a machine learning model by setting the related hyperparameters was important to control the complexity of the model and combat overfitting.

The hyperparameters for our RF model, including the minimum number of samples required for each leaf, the minimum number of samples required to split each node, the maximum number of levels in each decision tree, and the number of trees in the forest, are chosen to be 4, 6, 3, and 200, respectively. The XGboost model was tuned with the hyperparameters of 200 trees in the ensemble, a maximum tree depth of 3 and a learning rate of 0.5.
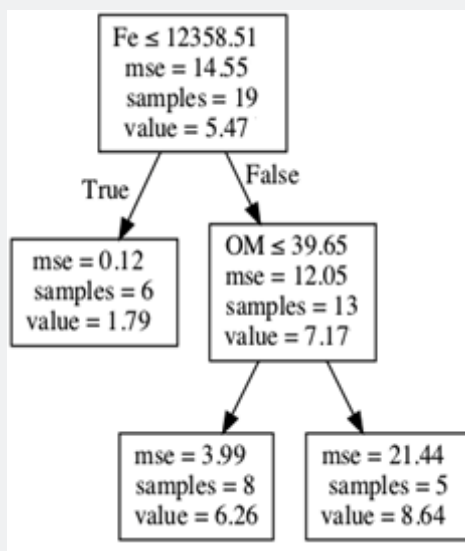

Figure 3: Decision tree for Arsenic concentration estimation.

Threshold Effects Levels (TELs) and Probable Effects Levels (PELs) were used to assess the ecological risks of heavy metals of the study samples [17]. These two levels categorize the negative ecological impacts of metal contamination into rarely (< TELs), occasionally (between TELs and PELs), and frequently (> PELs)

occurring [48]. Arsenic concentration in these marshes was lower than TEL and PEL values (Table 2) and there was no significant difference (Table 2) in mean arsenic concentration in oligohaline-mesohaline and polyhaline marshes.

Table 2: TELs and TPLs for heavy metals.

	As	Cd	Cr	Cu	Pb	Zn
TELs (mg/kg)	7.24	0.68	52.3	18.7	30.2	12.4
TPLs (mg/kg)	41.6	4.21	160	108	112	271

Results

Organic matter as a binding agent for heavy metals

A strong positive linear association (r (Spearman's correlation coefficient) =0.90, 0.85 and 0.91) was found between the concentration of the metals Cr, Cu, and Pb and soil organic matter content (Table 3). Soil organic matter retained a high amount of heavy metals due to forming metal-organic complexes [49]. Organic matter content affects the mobility and bioavailability of Cu in soil and water [50]. Because Cu had a high tendency to bind to soil organic matter, hydric soils containing organic matter likely retained Cu through forming metal-organic complexes, and as such, no free Cu was present in the porewater. This suggests

that organic matter plays a key role in Cu accumulation in soils. Although Cu is considered as an important nutrient for flora and fauna [51], it is toxic to living organisms at high concentrations [52]. Typical concentrations for copper in soil solution range from 0.025 to 0.140mg/kg [13], and Cu concentrations in the current research sampling sites were considerably higher than the typical range (Table 3).

Heavy metal analysis

Simple t-tests showed that there is a significant difference in mean concentration of organic matter, Cd, Cu, Pb and Zn in oligohaline-mesohaline and polyhaline marsh soils (Table 4). The mean concentrations of other metals in study marshes did not

pass the thresholds (both TELs and TPLs), except Cd which had a concentration of 0.94 (mg/kg) and 0.65 (mg/kg) in oligohaline-mesohaline and polyhaline marshes, respectively. The mean concentration of Cd in oligohaline-mesohaline marshes passed TELs for Cd (0.94mg/kg > 0.68mg/kg); which could be a threat and ecological risk for the aquatic system. The present study found a negative and linear association (r=-0.529) between Cd and salinity, and as salinity increases, Cd concentration in soil decreases. Mobilization and concentration of Cd varied along the salinity gradient in flooded marshes because Cd had a great tendency for chloride which is abundant in seawater, and high concentration of chloride in salt caused Cd mobilization from soil to interstitial water [53].

Besides, according to Table 2 & 5, mean concentration of Zn and Cd pass the TELs but mean concentration of other heavy metals is lower than TELs and PELs in all sampling soils. The coefficients of skewness (Table 5) of As, Cd, Cr, Cu, Pb and Zn are much higher than zero, which indicates high values in the samples. In other words, the positive skewness shows that the tail is long on the right side of distribution and heavy metals concentration distribution is not perfectly symmetric (Figure 4). Heavy metals concentrations distribution can be vividly demonstrated with the help of histograms, normal quantile (QQ) plot and boxplots (Figure 4). Normal quantile plot of heavy metals (Figure 4) show that heavy metals concentration distribution does not perfectly follow a normal distribution. The outliers, datapoints located outside the whiskers of the box plot, are observed on the right side of As, Cr, Cu and Pb concentration distribution (Figure 4), representing soils samples with very high heavy metals concentration which passes the TELs values.
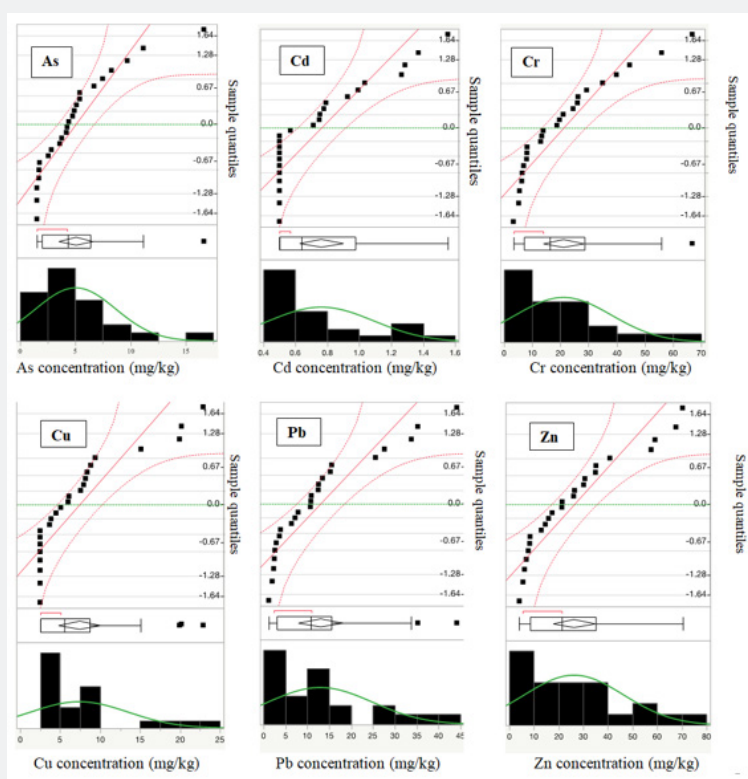

Figure 4: Histograms, boxplots and normal quantile (QQ) plots for heavy metals.

Table 3: Spearman's correlation coefficients between metals and binding agents in soil system according to the laboratory results.

Spearman's Correlation Coefficient				
	OM %	Fe (mg/kg)	Mn (mg/kg)	Clay %
As (mg/kg)	0.22	0.77	0.33	0.56
Cd (mg/kg)	0.78	0.88	0.67	0.61
Cr (mg/kg)	0.9	0.86	0.79	0.57
Cu (mg/kg)	0.85	0.82	0.74	0.56
Pb (mg/kg)	0.91	0.84	0.83	0.45
Zn (mg/kg)	0.87	0.9	0.83	0.59

Table 4: Difference in means of soil properties and metal concentration (in mg/kg) in bulk saltmarsh soil samples from oligohaline-mesohaline and polyhaline marshes.

	Mean in the Oligohaline-Mesohaline Marshes	Mean in the Polyhaline Marshes	T-test	P-value
Bulk density(g/cm^3)	0.70±0.60	0.83±0.46	0.58	0.56
Clay (%)	33.73±19.35	23.77±12.61	1.53	0.13
Fe (mg/kg)	18352±12093.11	14791±8889.21	0.83	0.41
Mn (mg/kg)	134±122.32	122.9±71.92	0.28	0.78
Organic matter (%)	10.99±10.58	4.18±2.93	2.37	0.02*
C:N	13.32±4.48	14.07±3.09	0.48	0.63
P (mg/kg)	508±378.68	423±235.15	0.67	0.5
As (mg/kg)	4.42±2.96	5.40±4.00	0.63	0.53
Cd (mg/kg)	0.94±0.42	0.65±.21	2.3	0.03*
Cr (mg/kg)	29.39±23.03	16.11±10.06	1.96	0.06
Cu (mg/kg)	11.30±8.33	5.06±2.51	2.73	0.01*
Pb (mg/kg)	18.87±16.17	9.40±6.87	2	0.05*
Zn (mg/kg)	36.75±27.07	19.91±11.52	2.13	0.04*

Note: * significant at the 0.05 level.

Table 5: Statistics of heavy metals concentration.

Heavy Metals	Mean	Standard Deviation	Minimum	Maximum	Skewness
As	5.03	3.61	1.52	16.63	1.68
Cd	0.76	0.32	0.51	1.55	1.11
Cr	21.09	17.01	3.32	66.76	1.25
Cu	7.39	6.12	2.5	22.88	1.47
Pb	12.94	11.91	1.21	44.12	1.29
Zn	26.22	20.13	4.07	70.22	0.97

Machine learning algorithms for heavy metals characterization.
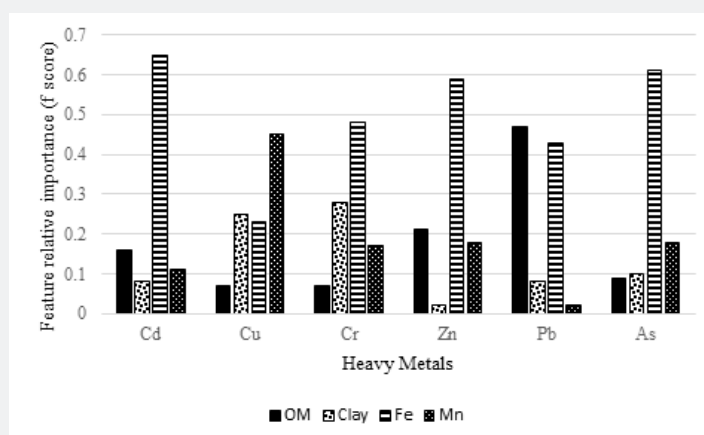

Figure 5: Heavy metals binding agents prioritized by RF.

The feature importance of the top four binding agents (i.e., clay, organic matter, Mn and Fe) for heavy metals are shown in Figures 5 & 6. The feature importance analysis was carried out based on comparable measures, feature relative importance. In this analysis, RF and XGboost were used for determining the most important binding agents (i.e., organic matter, clay, Fe and Mn)

for heavy metals, which infers the relative contribution of each binding agent to heavy metal model predictions. As a result, the feature relative importance from the two models shows similar patterns.

Both methods selected Fe as the most important binding agent for heavy metals such As, Cd, Cr and Zn (Figures 5 & 6). Organic matter had the highest sorption capacity for Pb. Mn was the most

important binding agent for Cu concentration modeling. Fe was identified as the first top binding agent for As, Cd, Cr and Zn while organic matter had the highest sorption capacity for Pb (Figures 5 & 6). Overall, the XGboost model performed more accurately (lower MSE) than RF for modeling the concentration of As, Cr, Cu, Cd, Pb and Zn (Figure 7).
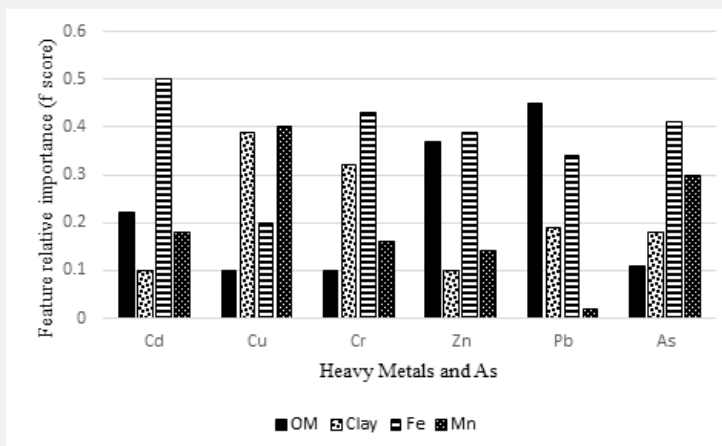

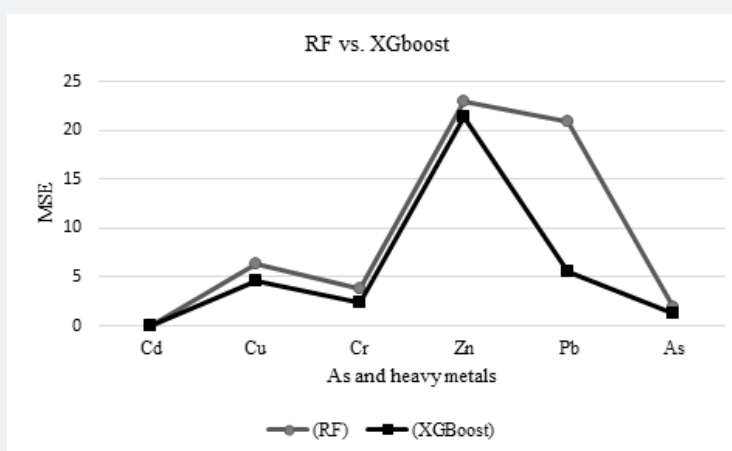**Figure 6:** Heavy metals binding agents prioritized by XGboost.

Figure 7: Comparison of RF and XGboost models for heavy metals prediction.

Discussion

Machine learning algorithms, such as RF and XGboost, can identify the key drivers of heavy metals in hydric soils. Iron (Fe) was the most important feature for estimating As concentration (Figures 5 & 6). A high concentration of Fe in soil and water can precipitate As and reduce its bioavailability. However, high concentrations of As and Fe can also reduce a plant production [54]. The relationship between soil geochemistry and As concentrations is not yet fully understood [54], but Fe reduces the

lability of As and effectively attenuates As in arsenic-polluted soils [55]. According to [25], As is not correlated with other elements available in marine sediments, but its spatial concentration pattern is a function of As concentration of groundwater and subaerial/submarine geothermal spring. Fe reduces nearly fifty percent extractable As in soils [56]. Goethite (consisting of Fe(III) oxide-hydroxide) is effective to reduce arsenic toxicity in contaminated soil [57]. Further, water-soluble iron-hydrous oxides regulate the arsenic adsorption–desorption reaction in sludge [58]. Ferrous sulfate ($FeSO_4$) [59] and amorphous Fe hydroxide (am-$Fe(OH)_3$)

[60] also have a high adsorptive capacity for As. If Fe concentration is more than 12358.51 (mg/kg), organic matter is used as the second most important feature for predicting As concentration.

The top four binding agents for heavy metals were determined to be clay, organic matter, Mn and Fe. Feature relative importance from the two models (Figures 5 & 6) showed similar patterns. For instance, Fe was the most important binding agent for Cd according to both models (Figures 5 & 6). The mean concentration of Cd in oligohaline-mesohaline marshes passed TELs, and both XGboost and RF suggested that the cycle of Fe should be controlled to enhance the health of the aquatic system in such areas. One approach to control Fe cycle is to inhibit saltwater intrusion into low salinity areas because sulfur abundance in seawater plays the key role in pyrite (FeS_2) formation which is a vital part of Fe cycle in marsh soil system [61]. Pyrite is considered as a common compound in saltmarsh hydric soils because of saltwater intrusion into saltmarsh environments saltwater [62]. In other words, saltwater contains a substantial amount of sulfur (reduced to sulfate and then sulfide) which reacts with iron to form pyrite [63]. In such environments, this reaction decreases the concentration of soil phosphorus (bound with Fe) and yields more available phosphorus in porewater and as such, negatively impacts the ecosystem [64]. Pyrite burial in marine environments is considered as a long-term preservation and plays an important role in sustaining the alkalinity of the soil-water system [65].

Both machine learning algorithms (i.e., XGboost and RF) selected Fe as the most important binding agent for heavy metals such As, Cd, Cr and Zn (Figures 5 & 6). Fe compounds considerably influence the behavior of some heavy metals in a soil substrate [66]. The level to which soil Fe is responsible for heavy metal solubility and availability is greatly determined by some soil factors. On the other hand, heavy metals are also known to affect the bioavailability of Fe [67]. Fe has a high sorption capacity, especially for heavy metals [67]. The mechanisms of sorption involve the isomorphic substitution of divalent or trivalent cations for Fe ions, the cation exchange reactions, and the oxidation effects at the surface of the oxide precipitates [67].

Organic matter was selected as the most important binding agent for Pb by XGboost and RF because organic matter forms complexes with Pb and plays a key role in Pb cycling [68]. Mn was selected as the most important binding agent for Cu concentration modeling by XGboost and RF because Cu has tendency to form strong ionic bond with Mn [69]. Based on the MSE values of both methods (Figure 7), XGboost lead to a better accuracy than RF to predict Pb concentration in saltmarsh soils because XGboost repetitively leveraged the patterns in residuals and strengthened the model with predictions made through sequential analysis.

Soils physical remediation mainly includes soil replacement procedure which is based on the use of clean soil to replace the polluted soil with the aim of diluting the pollutant concentration and increasing the soil environmental capacity for the remediation [70]. Soil replacement procedure including soil replacement, soil spading and new soil importing is recommended for treatment of small-scale contamination. However, this technology is costly and time-consuming for soil in large area, especially a saltmarsh ecosystem. Among various chemical remediation procedures used for polluted soils, in situ immobilization of heavy metals using a chemical amendment can be considered as a cost-effective and environmentally sustainable remediation approach by reducing the mobility and availability of metals. Therefore, the results from this study lead to the most effective procedure for remediating polluted soils. According to our results, identifying heavy metals available in a soil substrate as well as having knowledge about their binding agents could improve remediation effort of polluted soils.

Heavy metals concentration in hydric soil system is influenced by wide range of factors, and it is very difficult to include all potential factors into a model. Limited by economic and labor costs, factors such as meteorological and climate considerations as well as soil and water properties were not included in our models. This might lead to the poor performance of our model for certain elements like Pb and As. Although the machine-learning models yielded reasonable predictions for a wide range of soil heavy metals, their accuracies are expected to be enhanced though introducing environmental variables that are closely related to dynamics of soil heavy metals. Therefore, further research is needed to take more related factors into consideration and improve model performance in order to determine the most possible important parameters in heavy metals cycling. Moreover, potential effects of interactions between different heavy metals on their bioaccumulation in soil-plant ecosystems were not taken into consideration in this study. This may also lead to bias in our results. Finally, the applicability of our model to other regions should also be investigated.

By comparing RF to XGboost, we found that both methods produced comparable results for our variables of interest, heavy metals concentration. XGboost outperformed RF for variables of interest with relatively low MSE. The model's error was typically observed as the mismatch between the observed and the estimated heavy metal concentrations. In the context of heavy metals modeling, these two machines learning models are unavoidable owing to the inherent uncertainties in the process. There are various sources of uncertainty; it can be related to predictors, model parameters, and model structure, etc. [71]. Since the contribution of different sources of errors is not completely known and separating their roles is difficult especially in studying heavy metals in aquatic environments, an overall assessment of uncertainty is more preferable than determining an exact source of uncertainty. Understanding the total model uncertainty rather than the uncertainty resulting from individual sources is more important for decision-makers, particularly those in natural

resources management [71]. These two machines learning models lead to more accurate (less uncertain) results which can be an acceptable representation of reality if such complex sophisticated algorithms with many hyperparameters are tuned and validated properly.

Conclusion

Alteration in salinity gradients exerts an influence on vegetation and soil. An increase in salinity results in a decrease in soil organic matter budget, total C, total N, total P and aboveground biomass production as well as an increase in soil bulk density. Salinity increases also alter the concentration of the metal-binding agents like organic matter in tidal marsh soils. It is observed that oligohaline-mesohaline marsh soils have a greater mean in Cd, Cu, Pb and Zn concentrations than in polyhaline marshes. According to both RF and XGboost, Fe is the most important binding agent for As, Cd, Cr and Zn. An alteration in Fe cycling in hydric soil systems causes As, Cd, Cr and Zn release into the aquatic environments. The knowledge of sorption and desorption of the heavy metals by individual soil components enhances the understanding of soils behavior once polluted by heavy metals. This knowledge guides ecologists to select the best amendments for removing or fixing heavy metals in hydric soil substrates. Besides the consistent results from two different ensemble methods for modeling heavy metals, XGboost outperforms RF in terms of MSE. Further, Mn and organic matter are determined as the most important binding agent for Cu and Pb, respectively, through the feature selection analysis.

References

1. Belluco E, Camuffo M, Ferrari S, Modenese L, Silvestri S, et al. (2006) Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing. Remote sensing of environment 105(1): 54-67.

2. Christian J, Kim S, Stephan AD, Lori S, Salehi HI, et al. (2020) Best Management Practices for Post-construction Restoration of Rights-of-way in Saltwater Marshes, Estuaries, and Other Tidally Influenced Areas.

3. Corcoran, JM, Knight JF, Gallant AL (2013) Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in Northern Minnesota. Remote Sensing 5(7): 3212-3238.

4. Deegan LA, Johnson DS, Warren RS, Peterson BJ, Fleeger JW, et al. (2012) Coastal eutrophication as a driver of salt marsh loss. Nature 490(7420): 388-392.

5. Streever B (1999) An international perspective on wetland rehabilitation. 1999, Dordrecht: Springer Science & Business Media.

6. Adam P (2002) Saltmarshes in a time of change. Environmental conservation 29(1): 39-61.

7. Casselman ME, Patrick W, DeLaune R (1981) Nitrogen Fixation in a Gulf Coast Salt Marsh. Soil Science Society of America Journal 45(1): 51-56.

8. Craft C (2007) Freshwater input structures soil properties, vertical accretion, and nutrient accumulation of Georgia and US tidal marshes. Limnology and oceanography 52(3): 1220-1230.

9. Loomis MJ, Craft C (2010) Carbon sequestration and nutrient (nitrogen, phosphorus) accumulation in river-dominated tidal marshes, Georgia, USA. Soil Science Society of America Journal 74(3): 1028-1036.

10. Sutter LA (2014) Effects of Saltwater Intrusion on Vegetation Dynamics and Nutrient Pools in Low-Salinity Tidal Marshes, Pamunkey River (Virginia, USA), in Virginia Institute of Marine Science. College of William and Mary.

11. Bai J, Zhoa Q, Wang W, Wang X, Jia J, et al. (2019) Arsenic and heavy metals pollution along a salinity gradient in drained coastal wetland soils: Depth distributions, sources and toxic risks. Ecological Indicators 96: 91-98.

12. Roychoudhury AN (2007) Spatial and seasonal variations in depth profile of trace metals in saltmarsh sediments from Sapelo Island, Georgia, USA. Estuarine, Coastal and Shelf Science 72(4): 675-689.

13. Williams T, Bubb J, Lester J (1994) Metal accumulation within salt marsh environments: a review. Marine pollution bulletin 28(5): 277-290.

14. Horowitz AJ (1985) A primer on trace metal-sediment chemistry. 1985: US Government Printing Office Washington, DC.

15. Vu CT, Lin C, Shern CC, Yeh G, Le Van G, et al. (2017) Contamination, ecological risk and source apportionment of heavy metals in sediments and water of a contaminated river in Taiwan. Ecological indicators 82: 32-42.

16. Tessier A, Campbell P (1987) Partitioning of trace metals in sediments: relationships with bioavailability, in Ecological Effects of *In Situ* Sediment Contaminants. 1987, Springer. pp. 43-52.

17. Ustaoğlu F, Islam MS (2020) Potential toxic elements in sediment of some rivers at Giresun, Northeast Turkey: A preliminary assessment for ecotoxicological status and health risk. Ecological Indicators 113: 106237.

18. Zhuang W, Liu Y, Chen Q, Wang Q, Zhou F (2016) A new index for assessing heavy metal contamination in sediments of the Beijing-Hangzhou Grand Canal (Zaozhuang Segment): a case study. Ecological Indicators 69: 252-260.

19. Wuana RA, Okieimen FE (2011) Heavy metals in contaminated soils: a review of sources, chemistry, risks and best available strategies for remediation. Isrn Ecology 2011.

20. USEPA (1996) Report: recent developments for in situ treatment of metals contaminated soils. 1996, US Environmental Protection Agency, Office of Solid Waste and Emergency Response.

21. Vane CH, Kim AW, Hayes VM, Turner G, Mills G, et al. (2020) Organic pollutants, heavy metals and toxicity in oil spill impacted salt marsh sediment cores, Staten Island, New York City, USA. Marine Pollution Bulletin 151: 110721.

22. Shiowatana J, McLaren RG, Chanmekha N, Samphao A (2001) Fractionation of arsenic in soil by a continuous-flow sequential extraction method. Journal of environmental quality 30(6): 1940-1949.

23. Kumar S, Lal R, Liu D (2012) A geographically weighted regression kriging approach for mapping soil organic carbon stock. Geoderma 189: 627-634.

24. Thompson JA, Yewtukhiw EMP, Grove JH (2006) Soil–landscape modeling across a physiographic region: Topographic patterns and model transportability. Geoderma 133(1-2): 57-70.

25. Giglioli S, Colombo L, Contentabile P, Musco L, Armiento G, et al. (2020) Source apportionment assessment of marine sediment contamination in a post-industrial area (Bagnoli, Naples). Water 12(8): 2181.

26. Guo PT, Li MF, Luo W, Tang QF, Liu ZW, et al. (2015) Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. Geoderma 237-238: 49-59.

27. Tajik S, Ayoubi S, Nourbakhsh F (2012) Prediction of soil enzymes activity by digital terrain analysis: comparing artificial neural network and multiple linear regression models. Environmental Engineering Science 29(8): 798-806.

28. Lark R (1999) Soil–landform relationships at within-field scales: an investigation using continuous classification. Geoderma 92(3-4): 141-165.

29. Leal AM, Kulik DA, Saar MO (2017) Ultra-fast reactive transport simulations when chemical reactions meet machine learning: chemical equilibrium. arXiv :1708.04825.

30. Zeissler K, Hertwig T (2011) Artificial Neural Network instead of Kriging? A Case Study with Soil Contamination of Complex Sources. Landwirtschaft und Geologie.

31. Sergeev A, Buevich AG, Baglaeva EM, Shichkin AV (2019) Combining spatial autocorrelation with machine learning increases prediction accuracy of soil heavy metals. Catena 174: 425-435.

32. Zhong B, Liang T, Wang L, Li K (2014) Applications of stochastic models and geostatistical analyses to study sources and spatial patterns of soil heavy metals in a metalliferous industrial district of China. Science of the total environment 490: 422-434.

33. Heung B, Ho HC, Zhang J, Knudby A, Bulmer CE, et al. (2016) An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265: 62-77.

34. Mosavi A, Hosseini FS, Choubin B, Taromideh F, Rahi G, et al. (2020) Susceptibility mapping of soil water erosion using machine learning models. Water 12(7): 1995.

35. Priori S, Bianconi N, Costantini EA (2014) Can γ-radiometrics predict soil textural data and stoniness in different parent materials? A comparison of two machine-learning methods. Geoderma 226-227: 354-364.

36. Barman U, Choudhury RD (2020) Soil texture classification using multi class support vector machine. Information Processing in Agriculture 7(2): 318-332.

37. Hu B, Xue J, Zhou Y, Shao S, Fu Z, Li Y, et al. (2020) Modelling bioaccumulation of heavy metals in soil-crop ecosystems and identifying its controlling factors using machine learning. Environmental Pollution 262: 114308.

38. Jia X, Hu B, Marchant BP, Zhou L, Shi Z, et al. (2019) A methodological framework for identifying potential sources of soil heavy metal pollution based on machine learning: A case study in the Yangtze Delta, China. Environmental Pollution 250: 601-609.

39. Mosavi A, Hosseini FS, Choubin B, Adolshahnejad M, Gharechaee H, et al. (2020) Susceptibility Prediction of Groundwater Hardness Using Ensemble Machine Learning Models. Water 12(10): 2770.

40. Balestrini R, Delconte CA, Sacchi E, Buffagni A (2020) Groundwater-dependent ecosystems as transfer vectors of nitrogen from the aquifer to surface waters in agricultural basins: The fontanili of the Po Plain (Italy). Science of the Total Environment 753: 141995.

41. Polishchuk PG, Murotov EN, Artemenko AG, Kolumbin OG, Murotov NN, et al. (2009) Application of random forest approach to QSAR prediction of aquatic toxicity. Journal of chemical information and modeling 49(11): 2481-2488.

42. Blanquet JB (1932) Plant sociology. The study of plant communities. Plant sociology. The study of plant communities. (1st edn), 1932.

43. Cleveland D, Brumbaugh WG, MacDonald DD (2017) A comparison of four pore water sampling methods for metal mixtures and dissolved organic carbon and the implications for sediment toxicity evaluations. Environmental toxicology and chemistry 36(11): 2906-2915.

44. Zhou ZH (2009) Ensemble Learning. Encyclopedia of biometrics 1: 270-273.

45. James G, et al. (2013) An introduction to statistical learning. Vol. 112, Springer.

46. Chen T, et al. (2015) Xgboost: extreme gradient boosting. R package version 0.4-2, pp. 1-4.

47. Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial. Frontiers in Neurorobotics 7: 21.

48. Macdonald DD, Carr RS, Calder FD, Long ER, Ingersoll CG (1996) Development and evaluation of sediment quality guidelines for Florida coastal waters. Ecotoxicology 5(4): 253-278.

49. He Y, Men B, Yang X, Xu H, Wang De (2019) Relationship between heavy metals and dissolved organic matter released from sediment by bioturbation/bioirrigation. Journal of Environmental Sciences 75: 216-223.

50. Inaba S, Takenaka C (2005) Effects of dissolved organic matter on toxicity and bioavailability of copper for lettuce sprouts. Environment International 31(4): 603-608.

51. Soetan KO, Olaiya CO, Oyewole OE (2010) The importance of mineral elements for humans, domestic animals and plants: A review. African journal of food science 4(5): 200-222.

52. Martinez MCC, Smith BD, Luoma SN, Rainbow PS (2010) Metal toxicity in a sediment-dwelling polychaete: threshold body concentrations or overwhelming accumulation rates? Environmental Pollution 158(10): 3071-3076.

53. Du Laing G, Rinklebe J, Vandecasteele B, Meers E, Tack FMG (2009) Trace metal behaviour in estuarine and riverine floodplain soils and sediments: a review. Science of the total environment 407(13): 3972-3985.

54. Hartley W, Lepp NW (2008) Remediation of arsenic contaminated soils by iron-oxide application, evaluated in terms of plant productivity, arsenic and phytotoxic metal uptake. Science of the Total Environment 390(1): 35-44.

55. Wang N, Xue XM, Juhasz AL, Chang ZZ, Li HB, et al. (2017) Biochar increases arsenic release from an anaerobic paddy soil due to enhanced microbial reduction of iron and arsenic. Environmental Pollution 220(Pt A): 514-522.

56. Pillai P, Nilaksh K, Zeel T, Swapnil D, Mika S. (2020) Removal of arsenic using iron oxide amended with rice husk nanoparticles from aqueous solution. Materials Today: Proceedings.

57. Sun X, Doner HE (1998) Adsorption and oxidation of arsenite on goethite. Soil Science 163(4): 278-287.

58. Barrachina AC, Jugsujinda A, Burlo F, Delaune RD, Patrick Jr WH (2000) Arsenic chemistry in municipal sewage sludge as affected by redox potential and pH. Water Research 34(1): 216-224.

59. Artiola JF, Zabcik D, Johnson SH (1990) *In situ* treatment of arsenic contaminated soil from a hazardous industrial site: laboratory studies. Waste management 10(1): 73-78.

60. Shaibur MR, Kitajima N, Huq SMI, Kawai S (2009) Arsenic–iron interaction: Effect of additional iron on arsenic-induced chlorosis in barley grown in water culture. Soil Science and Plant Nutrition 55(6): 739-746.

61. Reddy KR, DeLaune RD (2008) Biogeochemistry of wetlands: science and applications. CRC press.

62. Fanning D, Rabenhorst M, Bigham J (1993) Colors of acid sulfate soils. Soil color 31: 91-108.

63. Antler G, Mills JV, Hutchings AM, Redeker KR, Turchyn AV (2019) The Sedimentary Carbon-Sulfur-Iron Interplay–A Lesson From East Anglian Salt Marsh Sediments. Frontiers in Earth Science.

64. Bai J, Yu L, Ye X, Yu Z, Guan Y, et al. (2020) Organic phosphorus mineralization characteristics in sediments from the coastal salt marshes of a Chinese delta under simulated tidal cycles. Journal of Soils and Sediments 20(1): 513-523.

65. Peiffer S, Stubert I (1999) The oxidation of pyrite at pH 7 in the presence of reducing and nonreducing Fe (III)-chelators. Geochimica et cosmochimica acta 63(19-20): 3171-3182.

66. Bartlett RJ, James BR (1993) Redox chemistry of soils. Adv Agron 50(151208): 7.

67. Sipos P, Choi C, Nemeth T, Szalai Z, Poka T (2014) Relationship between iron and trace metal fractionation in soils. Chemical Speciation & Bioavailability 26(1): 21-30.

68. Chen B, Zhu YG (2006) Humic acids increase the phytoavailability of Cd and Pb to wheat plants cultivated in freshly spiked, contaminated soil (7 pp). Journal of Soils and Sediments 6(4): 236-242.

69. Arulanandan K, Sargunam A, Loganathan P, Krone RB (1973) Application of chemical and electrical parameters to prediction of erodibility. Soil Erosion: Causes and Mechanisms, Prevention and Control, pp. 42-51.

70. Nejad ZD, Jung MC, Kim KH (2018) Remediation of soils contaminated with heavy metals with an emphasis on immobilization technology. Environmental geochemistry and health 40(3): 927-953.

71. Solomatine DP, Shrestha DL (2009) A novel method to estimate model uncertainty using machine learning techniques. Water Resources Research 45(12).

Your next submission with Juniper Publishers
will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php