Journal of Soils and Sediments
https://doi.org/10.1007/s11368-019-02253-2

SOILS, SEC 2 • GLOBAL CHANGE, ENVIRON RISK ASSESS, SUSTAINABLE LAND USE • RESEARCH ARTICLE

Machine learning methods for estimation the indicators of phosphogypsum influence in soil

Maria A. Pukalchik[1] • **Alexandr M. Katrutsa**[1] • **Dmitry Shadrin**[1] • **Vera A. Terekhova**[2,3] • **Ivan V. Oseledets**[1]

Received: 19 September 2018 / Accepted: 15 January 2019

Abstract

Purpose The full understanding of the effect of mineral waste-based fertilizer in soil is still unrelieved, because of the extreme complex chemical composition and plethora of their action pathways. The purposes of this paper is to quantify the input of PG into the soil ecosystem process, considering the direct effects of PG as a whole on soil environment using of a plethora of chemical, toxicological, and biological tests.

Materials and methods Greenhouse experiment includes different PG doses (0, 1%, 3%, 7.5%, 15%, 25%, and 40%) and two-time collection points after treatments—7 and 28 days. For each treatment and each time collection point, we measure (i) soil pH, bioavailable (H_20 and NH_4COOH-extractable) element content (S, P, K, Na, Mg, Ca, Fe, Zn, Sr, Ba, F); (ii) soil enzyme activities—dehydrogenase, urease, acid phosphatase, FDA; (iii) soil CO_2 respiration activity with and without glucose addition; (iv) *Eisenia fetida*, *Sinapis alba*, and *Avena sativa* responses. Finally, we combine the ordinary chemical, toxicology, and biological measuring of soil properties with state-of-the-art mathematical analysis, namely (i) support vector machines (used for prediction), (ii) mutual information test (variable importance tasks), (iii) *t-SNE* and LLE algorithms (used for unsupervised classification).

Results and discussion The results show similarity between the 0%, 1%, and 3% PG treatments in all collection times based on the toxicological and biological properties. Beyond 7.5% PG, some biological test was significantly inhibited in response to trace element stress. Among all tested parameters, soil urease activities, soil respiration activities after glucose addition, *S. alba* root lengths, and *E. fetida* survival rates show sensitivity to PG addition. Furthermore, the machine learning algorithms revealed that only several elements (mobile and water-soluble forms of Ca, Ba, Sr, S, and Na; water-soluble F) could be responsible to elevated soil toxicity for those indicators. SVR models were able to predict soil biological and ecotoxicity properties, and increasing numbers of randomly selected training examples from 50 to 90% of initial experimental data significantly improved model performance.

Conclusions At this study, we demonstrate benefits of unsupervised machine learning methods for investigating toxicity of man-made substances in soil that can be further applied to risk assessments of various toxins, which are of significant interest to environmental protection.

Keywords Bioassay · Biological properties · Feature relevance · Machine learning · Pollution · Regression · Soil · Trace element · Waste

Responsible editor: Xiuping Jia

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11368-019-02253-2) contains supplementary material, which is available to authorized users.

✉ Maria A. Pukalchik
m.pukalchik@skoltech.ru

[1] Skolkovo Institute of Science and Technology, Nobel 1, Moscow, Russia

[2] A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Leninsky pr., 33, Moscow, Russia 119071

[3] M.V. Lomonosov Moscow State University, Leninskiye Gory 1, Moscow, Russia 119992

Published online: 23 January 2019

1 Introduction

Waste production is an increasing global concern that is projected to worsen with the accelerating world's population thus making sustainable waste management a pressing issue. To be adduced just as an example, in the Russian Federation, more than 31.5 billion tons of waste were accumulated and identified by 2016 include 140 million tons of phosphogypsum (PG) with over 100 million tons was landfilled (Russian National Report 2015). Given the large quantities that are produced, and keeping in mind that only 14% of PG is used in the construction industry, it is necessary to dispose the surpluses (Tayibi et al. 2009). For instance, the land application of PG in agricultural fields could be an important recycling alternative aiming to reduce landfilling sites (Saadaoui et al. 2017).

The application of PG as an amendment has generally shown a positive effect on soil chemical properties, including an increase in the available sulfur and phosphorus content (Delgado et al. 2002), improvement of soil structure and crop yield (Vyshpolsky et al. 2010; Carmeis Filho et al. 2017; Kammoun et al. 2017; Ascari and Mendes 2018). Furthermore, PG amendment is recommended in ameliorating salinity in damaged soils, providing a source of Ca to replace the excess Na in cations' exchange (Hurtado et al. 2011). However, there are several difficulties in expanding the use of PG for an agronomy purpose, which is accounting for its complexity structure.

On the one hand, only a fraction of ecotoxicological studies have been performed to evaluate the ecological impact of PG application on soil. PG information is particularly fragmentary especially regarding their inclusion of trace element pollutants and other compounds as its specific composition and characteristics change considerably depending on the geographical origin. This waste typically comprises mainly gypsum and phosphate, but may also include the potentially hazard elements, such as fluoride, strontium, and barium. The presence of the latter at high levels in PG may have hazardous impact on the soil in general and on humans and plants, in particular. Pollutants from PG may adversely affect the soil environment by retarding the plant growth (Al-Hwaiti and Al-Khashman 2015; Ayadi et al. 2015; Elloumi et al. 2015), and enhancing the soil toxicity (Yakovlev et al. 2013; Hentati et al. 2015). The effect of PG on terrestrial ecosystem is largely lacking.

On the other hand, the accurate evaluation of soil conditions in presence of amendments with highly complex chemical composition is one of the most significant research objectives nowadays (Zaman 2014; Liu et al. 2017; Bünemann et al. 2018). This issue has triggered the general focus on more ecologically relevant test designs including responses at different levels of biological organization, and taking into account the chemical, toxicological, and biological parameters describing the structure and functioning of soil ecosystems (International Organization for Standardization ISO:19204 2017). Additionally, as stated by several authors (Alvarenga et al. 2018; Morgado et al. 2018), soil biological and toxicity properties may be suited to measure the impact of PG contamination on the quality of soil. Thus, using the soil enzyme activity as a screening tool to characterize contaminants in a variety of environmental matrices has become a popular, powerful and reliable tool in the environmental toxicology (Burns et al. 2013). Measuring the soil CO_2 emission provides significant data on microbial biomass (Haney and Franzluebbers 2009). Moreover, the earthworms and plants' longevity are generally used for toxicological tests as they are in direct contact with soil and are important in terrestrial food webs, soil productivity, and fertility (Pereira et al. 2018). Consequently, estimating the mentioned above parameters coupled with soil chemistry could provide valuable information of toxicity and help to reveal the potential risk of PG addition in soil. All these facts suggest that identifying the adequate and soil environmental "friendly" dose of PG is in high demand in agrochemistry and should be revealed with a help of mathematical modeling.

The classification of high-dimensional data, such as characteristic of soil environments remains a difficult task (Bouma 2014; Reinwarth et al. 2017). Unsupervised machine learning (ML) methods, in particular dimensionality reduction, is the core research topic in soil science community, is the well-acknowledged solution for this curse of problem. Starting with classical methods, such as principal component analysis, multidimensional scaling offers to reduce the X analysis into k clusters, and those methods have been successfully used in soil and environmental sciences for decades. Modern unsupervised ML methods usually use nonlinear projections of the data into low dimensions to appropriately visualize, rather than preprocess, complex data sets, and these methods including popular algorithms such as locally linear embedding (LLE), t-distributed stochastic neighbor embedding (*t-SNE*) (Bunte et al. 2012). These methods belong to nonlinear dimensionality reduction techniques, enabling the correct visualization of data which lie on curved manifolds or which incorporate clusters of complex shape, as is often the case for real-life examples, thus opening the way towards a visual inspection of nonlinear phenomena in the given data (Gisbrecht and Hammer 2015). The rationale behind the use of advanced nonlinear dimensionality reduction techniques in the analysis of soil data is that they may exploit nonlinear, higher-order relations between the chemical patterns and biological responses that are present in the data.

Support vector machine (SVM) is one of the most popular classifiers in the field of pattern recognition. The use of supervised ML methods trained on empirical data could be advantageous to make predictions on the potential toxicity effects of exogenous substances in soil (Deng et al. 2017; Cipullo et al. 2019), properties of drug-like molecules (Palmer et al. 2015), and biomonitoring the pesticide toxicity (Zhu et al. 2018; Niell et al. 2018). ML models are able to learn the relationships between input variables (e.g., soil amendment, soil type) and output variables (e.g., changes in soil toxicity, or bioassay response) from a training dataset, these relationships can then be generalized to make informed decisions in new cases. The interest to ML methods definitely rises, especially when we deal with soil systems, because the traditional statistical extrapolation techniques do not fit well in case of complex environment (Shatar and McBratney 2004; Jager 2011; Fox 2015). Overall, we can conclude that the application of ML to environmental issues, such as waste recovery and degradation studies, is the latest cutting edge research trend. This aroused our strong interest to explore the possibility to use SVM algorithms for predicting soil environment feedback after PG addition.

In our study, we aim to bridge these gaps and set the following objectives:

- to quantify the input of PG into the soil ecosystem process in soil, considering the direct effects of PG as a whole on soil environment by means of a set of chemical, toxicological, and biological tests;
- to provide the effect factors that could be used for subsequent calculating of the characterization factors for assessing terrestrial impacts from PG with a help of ML techniques.

2 Material and methods

2.1 Soil sampling, soil and PG type description

PG was collected at the Voskresensk fertility plant, Moscow Region in spring of 2016. PG was taken directly from the storage from the depth 0–30 cm, ten subsamples of 2 kg were taken within one plot 10 m^2 and joined in a single mixed sample. Mixed sample of PG was dried for 2 days at 60 °C, crashed, sieved through a 1-mm mesh, and stored in airtight polyethylene bags to avoid rehydration. The initial element contents in PG were as follows: CaO 51.3%, SO_3 0.21%, P_2O_5 0.21%, MgO 0.087%, R_2O_3 1.73%, Fe_2O_3 0.7%, Al_2O_3 1.13%, Na_2O 0.46%, K_2O 0.28%, F 3.08%, CO_2 0.1%, and Cl 0.027%.

We collected the top horizon (0–20 cm) of arable sandy loam spodosol, Moscow Region, Russian in spring of 2016. Properties of the tested soil samples were as follows: pH_{KCl} 6.5, TOC, 1.01%; Ntot, 0.11%; Ptot, 0.12%; Ktot, 0.18%; Mgtot, 1.7%; Catot, 10.5%; Na, 0.89%; total Zn, 23.5 ppm; total Cd, 0.1 ppm; total Cr, 16.0 ppm; total Pb, 9.0 ppm, total Ni, 4.3 ppm; and total Sr, 29.8 ppm.

2.2 Experimental design and setup

Greenhouse experiment represented in non-perforated plastic pots about 500 ml volume. The soil was defaunated before placing into pots. It was first kept frozen for at least 24 h at a temperature − 18° and then dried at a room temperature. Then, it was sieved through a 2-mm mesh to remove stones and large organic particles as well as to reach a homogeneous texture.

The rates of PG application were calculated based on our previous research (Yakovlev et al. 2013) and were 0 (NA), 1, 3, 7.5, 15, 25, and 40 w% corresponding to 0, 10, 75,150, 250, and 400 g of PG per kilogram of dry soil. The necessary amount of PG was added into each pot and homogenized with soil. The soil oil in the control pots with “0” treatment was mixed in the similar way as ones with the PG amendment but without adding PG. After mixing, we added the pre-calculated volume of filtered water to achieve the moisture level equal to 70% water holding capacity and homogenized the resulting mixture again.

The pots were left in normal day/night conditions for 28 days at a temperature of 20 ± 2 °C for further stabilization in the climatic chamber; after 7 and 28 days, the series of three samples were collected for further study.

2.3 Soil chemistry

The pH values of the non-amended and amended soils were measured in 1 M KCL at a ratio of 1:4 (*w*/*v*) using WTW pH 340i meter with glass, ion-selective electrode (WTW, Weilheim, Germany). The bioavailable element content (S, P, K, Na, Mg, Ca, Fe, Zn, Sr, Ba) in soil were estimated before and after the experiment through standard methods. We measured the water-soluble (w) and the NH_4COOH-extractable mobile (m) forms of the selected elements according to McBride (1989) and Cheng et al. (2011), respectively. The analyses of the elements were performed using the Inductively Coupled Plasma Mass Spectroscopy at the Agilent 7500a (USA).

The water-soluble fluoride (Fw) content in soil we determined according to Saha and Kundu (2003). The soil extracts were filtered through 0.45-mm membrane filter before analysis through ion chromatograph DIONEX model ICS 2000 (USA).

In routine analyses, three replicates of each sample were analyzed and the trace element concentrations were evaluated as mean of two measurements, with less than 10% repeatability value.

2.4 Soil enzyme activities measuring

Dehydrogenases (DHA) activity was measurement according to Thalmann (1968) and was expressed as μgTPF× g^{-1} soil dwt × 16 h^{-1}. The activities of acid phosphatases (AP) in the soil samples were assayed as outlined in (Eivazi and Tabatabai 1977) and expressed as μgpNP× g^{-1} soil dwt × h^{-1}. Urease activity (URE) was determined according to the method described by to Klose and Tabatabai (2000) and expressed as μgNH_4^+ × g^{-1} soil dwt × 24 h^{-1}. Fluorescein diacetate hydrolysis (FDA) was determined by a modification of the procedure of Inbar et al. (1991) and expressed as μgFDA× g^{-1} soil dwt × h^{-1}. For all enzyme activities, assays were performed in triplicate for each soil sample and were corrected for a blank.

2.5 Soil respiration activity

Carbon mineralization rate was measured in lab conditions using the substrate-induced and microbial basal respiration CO_2 emission parameters. The substrate-induced respiration (SIR) was assessed according to ISO (14240-1:1997 with 40 mg glucose per gram of soil for 4 h. The soil basal respiration rate (SBR), based on ISO (16072:2002) was measured without adding glucose for 24 h. The equipment operates with a constant temperature of 22 ± 2 °C and CO_2 production was determined from vials with a M-3700 gas chromatography with a thermal conductivity detector (Kristall, Granat Co., Russia). The data were calculated as the average rate of CO_2 emitted by each sample and expressed as μg $CO_2 \times g^{-1}$ soil dwt × h^{-1}. All measurements were performed in three replicates from each soil sample.

2.6 Bioassay

The plant-bioassay test was performed in accordance with the Boluda et al. (2011) with mustard (*Sinapis alba*) and oat (*Avena sativa*) seeds. We evaluated the effects of soil samples on the roots' elongation. We chose to grow the mustard and oat in the evaluation of acute phytotoxicity in the soil, because these two species of higher plans were demonstrated to be a valid tool in assessing the effect of different amendment application rate on plant uptake following the recommendation by Nikolaeva and Terekhova (2017). Petri dishes (Ø = 9 cm) were prepared and 10 g (referred to the dry matter) of samples were placed into the Petri dishes, a calculated amount of deionized water was added to obtain a WHC of 70% plus 5 extra milliliters and then a filter paper was placed on top. Ten seeds of mustard or oat were put on the filter paper and then the Petri dishes were closed with parafilm. After the incubation time of 72 h without any light supply at 25 ± 2 °C, every emerged seedling of the four replicates was washed and the root length was measured.

The earthworm (*Eisenia fetida*) toxicity test was conducted according to the procedure described by Organisation for Economic Co-operation and Development, OECD (222:2004) with slight modification. Adult worms were exposed to a range of concentrations of the tested substance mixed into the soil. All the individuals had a weight ranging from 0.4 to 0.7 g. The worms were fed with 5 g of dried cow manure at the onset and weekly thereafter. After 7 days and 4 weeks, the mortality of the adult worms was measured.

2.7 Statistical analysis and machine learning

The collected data for gas emission, enzyme activity, and bioassay responses were pre-checked for the outliers prior to the analysis. The normality of residuals and adherence to the final model assumptions were checked using the Statistica 8.0 software residuals tool. No drastic deviations from the normal distribution of the model were revealed.

There is a huge variety of machine learning supervised and unsupervised algorithms existing nowadays. In this study, the following algorithms were used: (i) support vector machines (used for regression and prediction), (ii) mutual information test (variable importance tasks), and (iii) *t-SNE* and LLE algorithms (used for unsupervised classification). All the calculations were done using Python, which is a free object-oriented data analysis language and software environment for statistical computing. The packages used while working on this research were from Scikit-learn Python library (Pedregosa et al. 2011).

2.7.1 Support vector machine

Support vector machine is a supervised machine learning algorithm based on statistical learning theory (Wu et al. 2004; Li et al. 2014), developed for data classification in (Vapnik 1995) and later extended to solve regression problems (Wu et al. 2004). Support vector regression **(**SVR) is a learning regression algorithm extended from the SVM (Vapnik et al. 1997). Mathematical formulation of SVR is explained in detail by Twarakavi et al. (2009). The strength of SVR is to model the complex nonlinear relationships in the multidimensional or hyperdimensional feature space and estimate the linear dependency of the variables to be predicted on the predictive covariates by fitting an optimal approximating hyperplane to the training data.

In this research, the dataset was randomly split into a training and validation datasets by using of a *train_test_split* method from Scikit-learn Python library with different ratios (90% or 116 observations, 70% or 88 observations, and 50% or 63 observations were used as a training datasets). The training datasets were used for SVR model calibration and testing datasets (10%, 70%, and 50% observations, respectively) were used after for models validation. The input data (all measured chemical, biological, and ecotoxicity data) were log-transformed prior to model development, biological, and ecotoxicity data were scaling from 0 to 100 scale in compare with NA control samples.

We used the training dataset to initially fitting the SVR model with the linear kernel function through epsilon regression in the Python and the optimal model's hyper parameters were obtained by solving of an optimization (minimization of the RMSE) problem on a grid of hyper parameters: C and Gamma. The parameter search space was a priori set to $0.001 \leq C \leq 1000$ at an incremental ratio of 10 and $0 \leq$ gamma ≤ 0.3 at steps of 0.001.

After training, the derived SVR models were applied to the validation datasets to produce the apparent soil biological and ecotoxicity properties. We evaluated the performance of SVR modeling by the root mean square error (RMSE) which can shed light on the goodness of fit between the prediction and measurement. Mathematically, the latter can be expressed as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

where n is the number of training compounds, $\hat{y}_i$ and y_i are the estimated and observed responses, respectively.

2.7.2 Mutual information test

The relevance factors investigation was carried out mutual information test (Kraskov et al. 2004). This method studies probabilistic dependencies between the target vectors and considered factors. These measures complement to each other and can be useful to analyze the data from different angles of view. In contrast to correlation analysis, this test allows identifying nonlinear relations between given factors and target vectors. Moreover, it gives a degree of dependencies for every pair of considered factors or between factors and target vector. These degrees help eliminate the most redundant and the least relevant factors. The elimination can be based on some threshold number of required factors or the threshold value of mutual information score. In this study, with a help of mutual information test, we practically assessed the load of individual chemical variables in biological and ecotoxicity responses. The calculated values are presented in a heatmap, where the greatest loads are colored in red, and the lowest are in blue.

2.7.3 Unsupervised classification

To visualize the obtained data, we used two nonparametric approaches—*t-SNE* algorithm (Maaten and Hinton 2008) and LLE (Roweis and Saul 2000). These methods are the useful analytical tools in unsupervised clusterization, where the data is classified by the algorithm into specified amount of classes based on internal patterns. Practically, it can be used to search for the subtypes and subclasses for researched process, value, or compound.

In more details, *t-SNE* algorithm chooses two similarity measures between pairs of points - one for the high-dimensional data and one for the two-dimensional embedding. It then attempts to construct a two-dimensional embedding that minimizes the of Kullback-Leib divergence between the vector of similarities between pairs of points in the original dataset and the similarities between pairs of points in the embedding. This is a non-convex optimization problem and *t-SNE* employs gradient descent with random initialization to compute a reasonable solution to it. The perplexity in *t-SNE* was equals to 7, and this value of perplexity provided the best looking plot.

LLE is an eigenvector methods designed for the problem of nonlinear dimensionality reduction, and is carried out in three main steps: select neighbors, compute weight matrix, and compute the low-dimensional coordinates by using the reconstruction weights.

3 Results

3.1 PG influence on soil chemical, biological, and ecotoxicity properties

The water-soluble and mobile-concentrations of S, P, Ca, Fe, Zn, Sr, Ba, K, Na, Mg, and water-soluble F ratios in the tested soil with different PG treatments are reported in Table 1 and Table A in the Electronic Supplementary Material. According to our results, all soil chemical properties evaluated were significantly influenced by increasing phosphogypsum rates in the range of 7.5 to 40%. Significant decrease in the soil pH and increase in Ca, S, P, Sr, Ba, and F content with PG was expected, while changes in Zn and Fe contents were surprising. The most marked changes in soil chemical properties with PG treatments were observed for Ca, S, P, and Sr. For example, the concentrations of mobile species of Ba and Sr in the studied soils with PG varied from 9.4 ± 2.6 to 13.3 ± 9.8 g/kg and from 52.6 ± 6.7 to $784.1 \pm 23.11.8$ g/kg, respectively. The higher contents of Ca(m) and S(m) were registered in 40% PG treatments. At the same time, no statistically significant differences were observed between the TE's contents in 0 (NA), 1 and 3% PG.

PG differentiated the biological activities of soil URE, SIR, and SBR, while the other parameters (AP, FDA, and DHA) showed negligible difference or light stimulation tendency with the increased dose of PG (Fig. 1). In particular, URE activities in soil were markedly different, yet the other enzymes parameters comprising AP, FDA, and DHA showed a negligible difference or light stimulation tendency with the elevated dose of PG The highest URE activity were recorded for the NA soil (387.5 ± 64.9 μg$NH_4^+ \times g^{-1}$ soil dwt × 24 h^{-1} and 394.2 ± 96.7 μg$NH_4^+ \times g^{-1}$ soil dwt × 24 h^{-1} in 7-day and 28-day samples, respectively) and the lowest in the soil with 40% PG (150.7 ± 19.9 μg$NH_4^+ \times g^{-1}$ soil dwt × 24 h^{-1} and 126.1 ± 24.6 μg$NH_4^+ \times g^{-1}$ soil dwt × 24 h^{-1}in 7-day and 28-day samples, respectively). The difference between FDA, AP, and DHA in treatments was less visible compared to URE. Moreover, AP and FDA showed a hormesis trend, clearly indicating at the initial beneficial situation and a subsequent intoxication after the peak. Carbon dioxide production in the samples without the glucose additive (SBR) had a trend to stimulate with low PG content and markedly decreased with high PG content. Generally, phosphogypsum stimulate CO_2 production up to 30% with 1–25% treatments, and no significant changes in compare with NA were observed for 40% PG in soil. Opposite this, the glucose-induced respiration data

Table 1 Influence of phosphogypsum on soil chemical properties (mean ± SD, $n = 6$ and summarized the 7 days and 28 days of samples collection data)

	Phosphogypsum dose (% dry weight soil)						
	0	1	3	7.5	15	25	40
pH_{KCl}	6.5 ± 0.1 a	4.9 ± 0.2 a	4.4 ± 0.3 ab	4.13 ± 0.32 b	3.7 ± 0.2 bc	3.3 ± 0.4 bc	3.1 ± 0.3 c
F_w, mg/kg	0.1 ± 0.0 a	0.4 ± 0.1 a	0.6 ± 0.1 a	1.6 ± 0.3 ab	2.8 ± 0.2 bc	3.8 ± 0.7 c	6.8 ± 1.3 d
S_w, mg/kg	30.2 ± 29.1 a	2254.0 ± 302.4 b	5485.3 ± 311.9 c	8910.5 ± 1329.6 d	8593.1 ± 277.2 d	9341.4 ± 1638.2 e	9016.1 ± 434.6 de
P_w, mg/kg	2.9 ± 0.16 a	7.9 ± 1.8 b	23.6 ± 5.9 c	39.3 ± 11.2 d	69.0 ± 6.1 e	115.6 ± 12.5 f	189.3 ± 43.2 g
K_w, mg/kg	33.9 ± 6.0 a	34.0 ± 4.5 ab	40.1 ± 3.6 bc	46.5 ± 2.2 c	53.2 ± 2.4 d	55.3 ± 4.5 d	55.0 ± 5.1 d
Na_w, mg/kg	0.5 ± 0.5 a	8.8 ± 7.9 b	6.4 ± 1.4 b	17.6 ± 6.8 c	53.7 ± 13.2 d	70.4 ± 23.3 e	88.4 ± 15.1 d
Mg_w, mg/kg	24.8 ± 2.5 a	113.8 ± 31.7 b	180.0 ± 33.4 c	151.7 ± 30.3 bc	178.6 ± 32.7 c	169.5 ± 12.0 c	163.7 ± 7.6 c
Ca_w, mg/kg	69.5 ± 10.4 a	1923.4 ± 346.2 b	4811.5 ± 408.3 c	7497.2 ± 976.7 d	7427.3 ± 327.9 d	7919.9 ± 1346.8 d	7755.6 ± 460.3 d
Fe_w, mg/kg	6.9 ± 2.4 a	16.7 ± 7.2 b	45.5 ± 5.8 c	54.8 ± 11.7 cd	57.2 ± 3.2 d	60.8 ± 6.0 d	60.9 ± 5.2 d
Zn_w, mg/kg	0.2 ± 0.2 a	0.5 ± 0.3 ab	0.7 ± 0.6 b	2.1 ± 0.9 c	2.7 ± 1.2 cd	3.6 ± 1.5 d	4.1 ± 0.7 d
Sr_w, mg/kg	0.8 ± 0.3 a	81.1 ± 14.6 b	205.4 ± 19.4 c	315.3 ± 50.4 d	295.5 ± 6.7 d	345.5 ± 130.3 d	333.2 ± 44.8 d
Ba_w, mg/kg	0.1 ± 0.1 a	3.5 ± 0.5 b	3.8 ± 0.9 bc	2.5 ± 0.3 c	2.0 ± 0.5 cd	1.6 ± 0.1 d	1.4 ± 0.2 e
S_m, mg/kg	0.0 ± 0.0 a	255.9 ± 99.8 b	1107.9 ± 88.8 c	4827.2 ± 1302.9 d	7827.5 ± 989.4 f	12,469.7 ± 1760.0 g	1516.2 ± 834.2 m
P_m, mg/kg	8.2 ± 2.8 a	9.4 ± 6.3 a	16.7 ± 5.6 bc	17.7 ± 6.11 abc	13.1 ± 1.2 c	14.8 ± 7.5 acb	13.3 ± 9.8 acb
K_m, mg/kg	33.9 ± 0.6 a	33.0 ± 4.5 a	40.1 ± 3.6 b	46.5 ± 3.2 c	53.2 ± 2.4 d	55.3 ± 4.5 d	55.6 ± 5.1 d
Na_m, mg/kg	20.1 ± 2.5 a	47.9 ± 13.9 b	99.5 ± 6.9 c	141.2 ± 18.6 d	228.0 ± 8.5 e	324.8 ± 26.3 f	414.6 ± 41.4 g
Mg_m, mg/kg	66.8 ± 20.0 ab	56.5 ± 4.2 b	59.3 ± 4.4 b c	61.5 ± 4.3 c	71.0 ± 25.1 abc	74.7 ± 17.4 abc	62.9 ± 6.1 abc
Ca_m, mg/kg	776.7 ± 120.5 a	1117.2 ± 129.3 b	2203.4 ± 171.6 c	5459.6 ± 1276.4d	12,120.9 ± 1669.2e	15,006.7 ± 1121.4 ef	18,166.7 ± 663.0 f
Fe_m, mg/kg	0.7 ± 1.2 a	2.3 ± 1.0 a	12.1 ± 1.3 b	49.1 ± 10.2 c	121.7 ± 15.8 d	138.8 ± 18.3 de	160.4 ± 10.0 e
Zn_m, mg/kg	3.7 ± 1.4 a	4.0 ± 1.9 a	4.1 ± 1.1 a	5.0 ± 1.7 a	4.5 ± 1.0 a	4.8 ± 1.4 a	4.4 ± 2.2 a
Sr_m, mg/kg	4.8 ± 1.0 a	52.6 ± 6.7 b	155.9 ± 12.6 c	422.8 ± 73.5 d	668.0 ± 62.0 e	704.1 ± 28.9 f	784.0 ± 23.1 g
Ba_m, mg/kg	8.2 ± 2.8 ac	9.4 ± 2.6 ac	16.7 ± 5.6 b	17.7 ± 6.1 b	13.1 ± 1.2 c	14.8 ± 7.5 cb	13.3 ± 9.8 cb

w water soluble species of elements, *m* mobile NH4COOH species of elements. Different letters in the row denote significant differences at the 0.05 confidence level between the samples according to the Tukey test ($a = 0.05$)

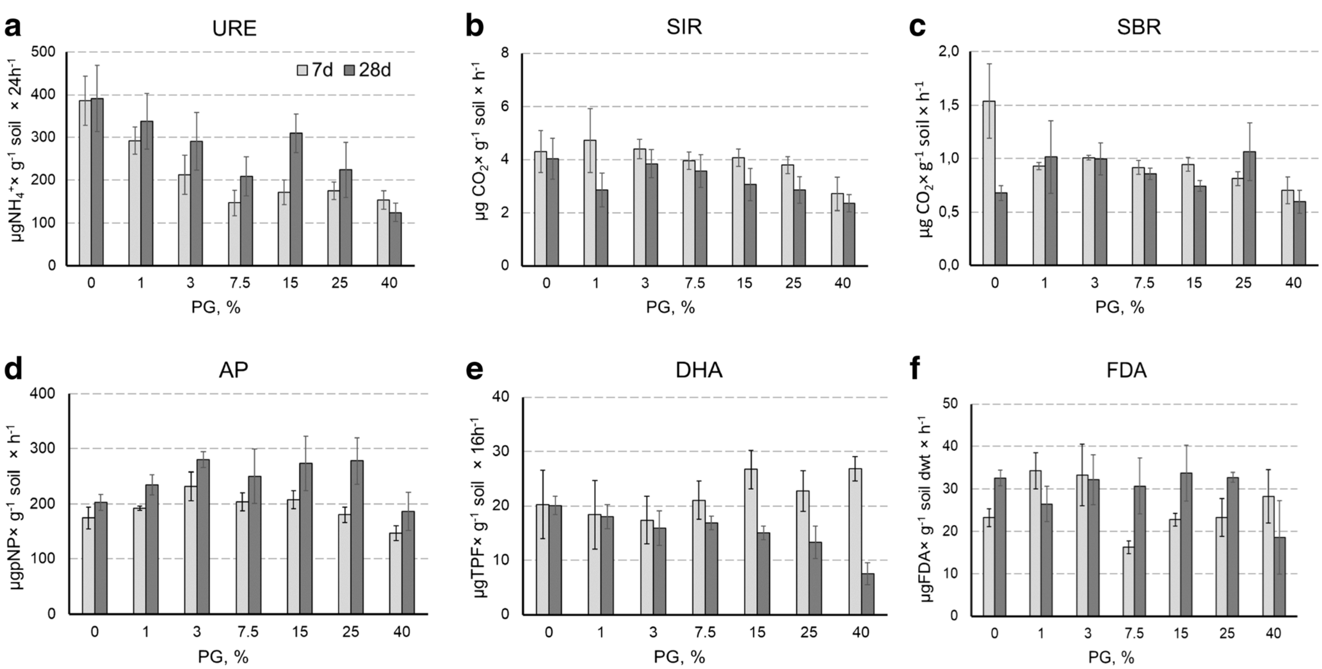

Fig. 1 Biological parameters in soil with varying doses of PG after 7 and 28 days of exposure: **a** soil urease activity, **b** soil substrate-induced respiration, **c** soil basal respiration, **d** acid phosphatase activity, **e** dehydrogenase activity, and **f** FDA-hydrolysis activity

(SIR) revealed more sensitivity to phosphogypsum in soil, the inhibitory effect was observed with more than 7.5% PG.

Differences in the sensitivity of the test species to PG in ecotoxicological studies have been revealed (Fig. 2). The acute toxicity to earthworm (more than 50% of earthworm's mortality) was detected in 40% PG for samples collected after 7 days and in 15, 25, and 40% PG treatments after 28 days of exposure on compare with NA soil. Lower doses of PG (1 and 3%) had positive influence to *A. sativa* and *S. alba* root lengths; 40% PG decreased root lengths of both plant species (Fig. 2). Overall, *S. alba* roots were more affected by PG influence than *A. sativa* roots.

3.2 Visualization with *t-SNE* and *LLE* algorithms

Figure 3 shows the result of visualization with *t-SNE* and *LLE* algorithms for given data. As can be seen from *LLE*, samples were clustered very well according to PG treatments, forming a total of four well-defined clusters. Furthermore, the samples from different collections point but with the same PG dose qualitatively clustered together (Fig. 3a). We indicated that chemical characterization of the soil after mixing 1 and 3% PG had specific patterns and could be recognized automatically with similarity to 0 (NA) based on the soil biological response. In contrast, the points corresponding to the higher percentages of PG cannot be grouped similar to the previous case and make up the single cloud of points (Fig. 3b). It means that chemical description of the soil after mixing with higher than 3% of PG are not different according to chemistry with ML algorithms and pointed at a similar chemical pattern.

3.3 Mutual information test results

The most important chemical features for the estimation of the PG biological and ecotoxicological influence in soil are represented in Fig. 4 and they were identified by mutual information algorithm. The heatmap shows that different features from PG were dominated for each biological and ecotoxicity variables.

For example, the earthworm acute toxicity was mainly driven by F(w), P(m), and Sr(m). According to the bioassay data (Fig. 2), we detected an acute toxicity from treated soils to *E. fetida* with a 15%, 25%, and 40% of PG for 28-day treated soil samples. And this applications of PG increased the F(w), P(m), and Sr(m) contents in soil more than 28, 1.6, and 139 times, respectively, in comparison with NA treatment.

The heatmap also suggests that *S. alba* root length toxicity effects from PG were mainly related to S(m), Ca(m), and Na(m) features. The acute toxicity effects were also observed for soils with 15% PG and more, and the excess of mentioned above elements in soil reached 7000 times for S(m), 15 times for Ca(m), and 11.4 times for Na(m) in comparison with NA (Table 1).

3.4 SVR modeling

Based on the methodology described above, we trained the SVR algorithm to predict the soil biological and ecotoxicity properties in a present of different PG doses. We made two series of ML performance. The first one included the SVR-1 models, which were studied on 90% observation training set and the second performance SVR-2 models included only 50% of initial observation for training (64 observations). Table 2 shows the performance indicators for SVR-1, SVR-2, and SVR-3 models with varying input size on training datasets. The RMSE values that ranged from 4.25 to 12.15 for SVR-1 model provide the best accuracy for modeling parameters than SVR-2 and SVR-3 models. Figure 5 shows predicted and experimental values for the selected biological and ecotoxicological parameters (as an example, we choose the URE, *S. alba*, and *E. fetida*) based on the SVR-1. As can be seen, visual correlation between measured and predicted values for the random selected samples was satisfactory.

4 Discussion

4.1 PG influence on soil chemical properties

Adding PG may drastically change the soil chemical properties including acidification process, changes in soil chemical compositions. Soil solution pH is one of the major factors

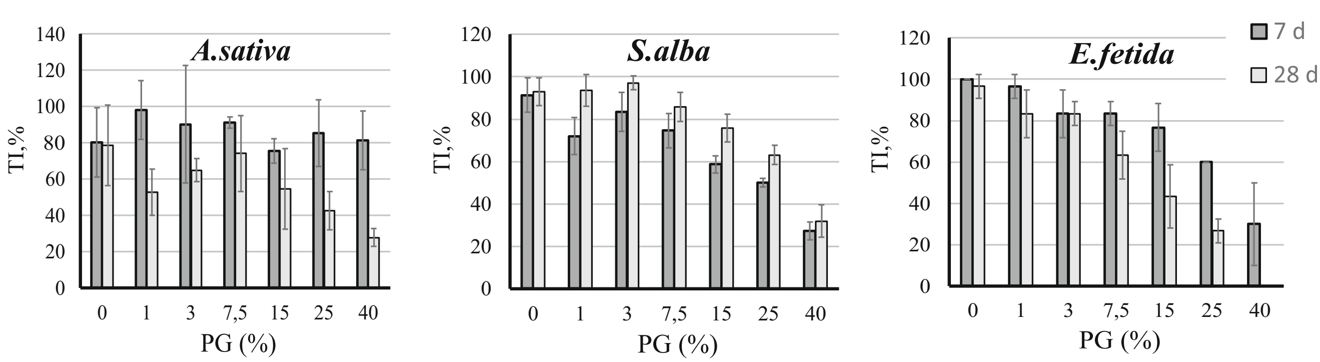

Fig. 2 Effects of the PG on *Sinapis alba*, *Avena sativa* roots length and *Eisenia fetida* mortality after the 7 and 28 days of soil-PG exposure

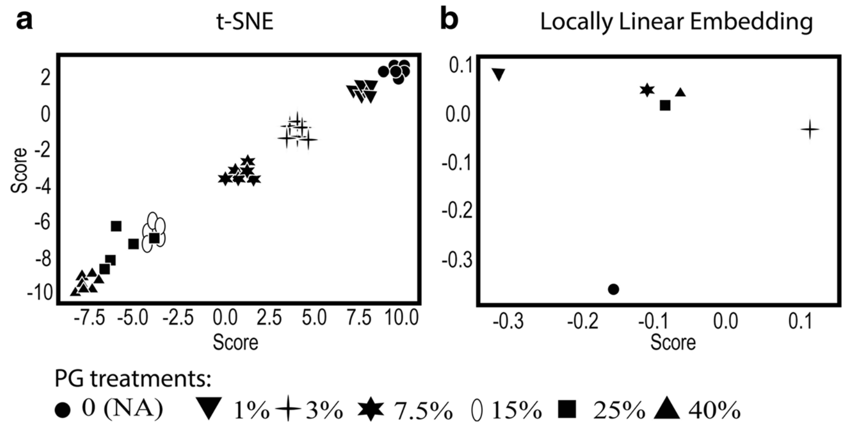

Fig. 3 Visualization of tested PG treatments using the *t-SNE* (**a**) and *LLE* algorithms (**b**)

controlling soil properties of variable charge components (Pan et al. 2014). The pH affects the surface charge through the supply of H+ for adsorption onto the metal oxides and the dissociation of the functional groups in the soil organic matter. A decrease in pH can elevate the concentration of trace elements, taking up a greater proportion of the cation exchange sites, reducing base saturation, and promoting soil toxicity (Liu et al. 2018). It is well-known that mineral waste like phosphogypsum contain residues of sulfuric and phosphoric acid, which are easy hydrolysable in soil solution and induced acidification (International Atomic Energy Agency Report 2013). Our data are in line with it; we observed marked changes in soil pH, and the ΔpH dropped from 0.50 with 1% PG to 2.27 with the 40% in compare with control (Table 1).

Addition of 7.5% PG brings the potentially hazard content of several elements into the soil like Ba, Sr, Zn, and Ca and also increased the S, P, Ca, and F contents. Our data is confirmed by several studies (Konarbaeva 1997; Blum et al. 2013), in which a significant increase in trace elements and fluoride contents were pointed in soil after PG treatments. For example, the average F content in natural soil is normally less than 1 mg × g^{-1} soil dwt (Pickering 1985), while in our experiment more than 1.6 mg F × g^{-1} soil dwt with 7.5% PG was detected. Overall, these levels of soil contamination with

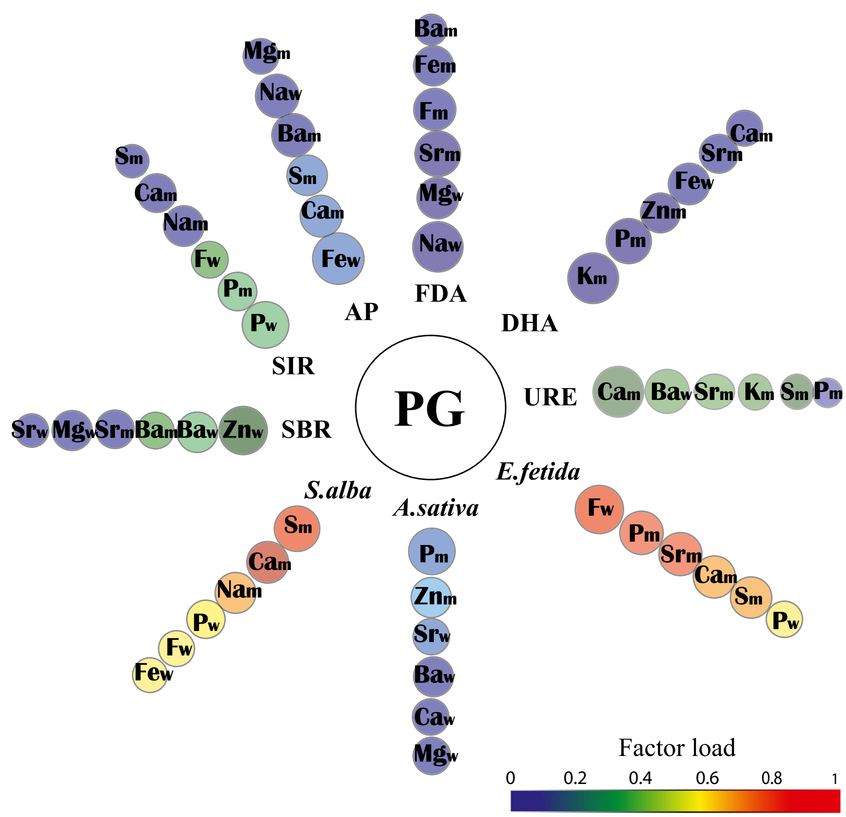

Fig. 4 Influence of measured chemical elements in soil after PG addition to soil biological and toxicological responses from the mutual information test. Balls are colored according to calculated load (from 0 to 1): the higher values are colored in red, and the lowest values are in blue

Table 2 Influence of training set size on SVR models performance to predict soil biological and ecotoxicity properties after PG addition

	DHA	URE	AP	FDA	SBR	SIR	*E. fetida*	*S. alba*	*A. sativa*
SVR-1. Training dataset 116 observations and validation datasets 12 observations									
RMSE	8.64	7.04	12.15	8.13	4.25	6.18	9.19	6.01	8.57
SVR-2. Training dataset 88 observations and validation datasets 38 observations									
RMSE	9.35	10.37	11.64	11.91	6.13	11.48	12.33	8.10	8.12
SVR-3. Training dataset 63 observations and validation datasets 63 observations									
RMSE	13.55	11.94	12.97	11.93	11.16	12.16	12.50	9.84	12.64

fluorine induce phytotoxicity effect and could influence on crop yield (Cui et al. 2011; Telesiński et al. 2012).

4.2 Drivers for changes in soil biological and ecotoxicity properties

Among the nine measured soil biological and ecotoxicity variables, only URE and SIR soil activities, *S. alba* root lengths, and *E. fetida* survival rate were negatively affected by PG treatments. Our results further suggest that only a few elements may be dependent on the exacerbating effects of PG on mentioned above variables, in particular, F(w), P(m), and Sr(m) have the greatest load on earthworm's toxicity; S(m), Ca(m), and Na(m) influenced on *S. alba* root lengths toxicity; Ca(m), Ba(w), and Sr(m) mostly affected on soil URE activities; finally, F(w), P(w), and P(m) affected in SIR values. These results were supported by the mutual information scores (Fig. 4). Previous studies highlighted the key role of the exacerbated F, Sr, and P soil content in toxicity to earthworms. In particular, fluorine and strontium may have led cytotoxicity effects (Morgan and Morgan 1988; Chae et al. 2018), and phosphorus addition with fertilizer may also induced earthworm's mortality (Chaudhari 2016). The sulfur, calcium and sodium phytotoxicity effects may be connected with their possible accumulation in roots and ion relations effects (Negrão et al. 2017). Inhibition activities of barium and strontium to soil URE activities were earlier observed by Tabatabai (1977).

The lack of effect of PG on soil enzyme activities like AP, FDA, and DHA, looks controversial, yet it provides the evidence in favor of the sensitivity of these enzymes to soil contamination that could be overvalued for soil monitoring purposes. In general, the enzyme activities are considered to be the first to respond to soil contamination; due to their high sensitivity to react to environmental changes. Moreover, they play a fundamental role in the dynamics of C, N, P, and S (Caldwell 2005). However, our results in general make it possible to assume that a high amount of fertilizer elements could interfere the effect of trace elements on hydrolysis enzymes. As could be seen from the mutual information scores shown in Fig. 4, the P, K, Na, Mg, and S addition had the highest load in AP, DHA, and FDA responses among all the other elements. Thus, we conclude that the chosen machine learning techniques are useful to further studies in the issues in questions and potentially help elucidate quite "in-obvious" relations.

4.3 Limitations of SVR modeling for prediction soil biological and toxicity properties after PG addition

Models based on biological indicators could become a powerful tool in soil ecotoxicology and could help to reduce the amount of analysis needed to the adequate monitoring of soil

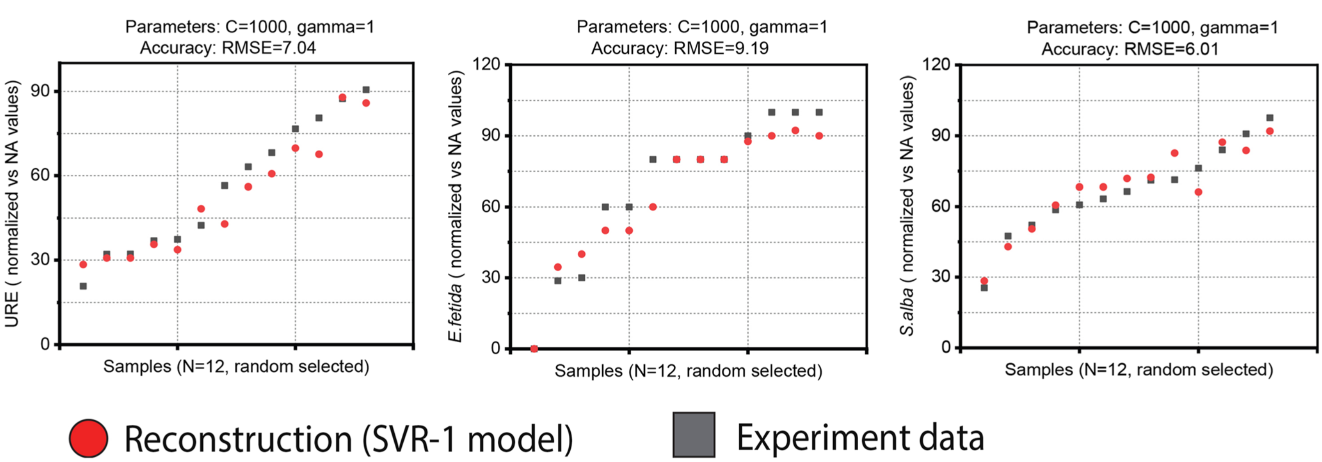

Fig. 5 Prediction accuracy for the selected soil toxicity data using SVR-1 model

systems quality (Cipullo et al. 2019). The results of our SVR performances revealed that model prediction ability consistently improved with increasing size of training sets. The SVR model was able to predict the toxicity and biological properties with adequate accuracy only in case when 90% of received data was used as a training dataset (Fig. 5), and when we reduced the training dataset to 70 or 50% of experimental data, the accuracy of modeling dramatically decreased (Table 2). Similar influence of varying training set size on SVM-based prediction previously was published by Rodríguez-Pérez et al. (2017). Meanwhile, our model is a priory valid only 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 different from tested spodosol.

Yakovlev et al. (2013) and Hentati et al. (2015) have already investigated phosphogypsum toxicity using the simple linear probit model. In particular, Hentati et al. (2015) proposed that no observed effect concentration of PG in soil were determined from 1.24 (*F. candida*) to 24.61% (*E. crypticus*), and no toxic effect was detected for *Zea mays* and *Lactuca sativa* up to 25% PG. According to Yakovlev et al. (2013), the most sensitive indicator of an ecosystem stress for PG application was a microbial respiration activity, and the calculated no observed effect concentration were 10.8% in artificial soil.

We believe that our approach to identify the PG influence in soil with advanced ML models looks beneficial in compare with previous studies, which can be explained by better applicability of received knowledge because we used both qualitative data (biological and ecotoxicological properties) and quantitative data (chemical properties of soils with different doses of PG) for models training. Keeping in mind that the relationship among pollutants and even the chemical composition of waste is highly nonlinear and very complex, it was mandatory to use more accurate analysis tools based on statistical learning such as the above-mentioned support vector regression.

5 Conclusions

Empirical data from a 2-month greenhouse experiment were used to assess the ability and performance of unsupervised and supervised ML methods to differentiate PG treatments and to predict temporal biological and ecotoxicological changes in soils after PG addition. We clearly revealed the multiplicity influence of PG in soil chemical, toxicological, and biological properties. And these results confirm that the PG addition capable to induce additive toxicity and might contribute to severe negative process when it is applied in wrong dose. Results obtained from the mutual information test illustrated that among all tested biological and toxicological parameters, only several, namely, urease activities, soil respiration activities after glucose addition, *S. alba* root lengths, and *E. fetida* survival rates, showed as good indicators for early-stage PG risk assessment. Thus, ML methods could be very helpful to understand complex mixtures fate, and identify the key variables affecting their behavior and the environmental risks posed by the various pools of contaminants. Finally, we note that size of training datasets significantly influenced on the SVR model performance and even a "small" data could be enough to training SVM models. At the time when the ecological monitoring programs are declining in a cost-effective manner and we are not always having a possibility to receive a "Big Data" in soil environment, the usage of ML methods may be a promising candidate tools to prevent soil degradation and contamination.

Acknowledgments The research was supported by the Skoltech Next Generation Program and the Russian Found of Basic Research (Project No. 16-34-00063 mol_a). We are grateful to Anastasia Sharapkova (Rosetta Stone MSU) for proofreading the final version of the manuscript.

References

Al-Hwaiti M, Al-Khashman O (2015) Health risk assessment of heavy metals contamination in tomato and green pepper plants grown in soils amended with phosphogypsum waste materials. Environ Geochem Health 37:287–304

Alvarenga P, Clemente R, Garbisu C, Becerril JM (2018) Indicators for monitoring mine site rehabilitation. In: Bio-Geotechnologies for Mine Site Rehabilitation. Elsevier, pp 49–66

Ascari JP, Mendes IRN (2018) Desenvolvimento agronômico e produtivo da soja sob diferentes doses de gesso agrícola. Revista Agrogeoambiental 9

Ayadi A, Chorriba A, Fourati A, Gargouri-Bouzid R (2015) Investigation of the effect of phosphogypsum amendment on two *Arabidopsis thaliana* ecotype growth and development. Environ Technol 36:1547–1555

Blum SC, Caires EF, Alleoni LRF (2013) Lime and phosphogypsum application and sulfate retention in subtropical soils under no-till system. J Soil Sci Plant Nutr 13(2):279–300

Boluda R, Roca-Pérez L, Marimón L (2011) Soil plate bioassay: an effective method to determine ecotoxicological risks. Chemosphere 84:1–8

Bouma J (2014) Soil science contributions towards sustainable development goals and their implementation: linking soil functions with ecosystem services. J Plant Nutri Soil Sci 177(2):111–120

Bünemann EK, Bongiorno G, Bai Z, Creamer RE, De Deyn G, de Goede R, Fleskens L, Geissen V, Kuyper TW, Mäder P, Pulleman M, Sukkel W, van Groenigen JW, Brussaard L (2018) Soil quality – a critical review. Soil Biol Biochem 120:105–125

Bunte K, Biehl M, Hammer B (2012) A general framework for dimensionality reducing data visualization mapping. Neural Comp 24(3):771–804

Burns RG, DeForest JL, Marxsen J, Sinsabaugh RL, Stromberger ME, Wallenstein MD, Weintraub MN, Zoppini A (2013) Soil enzymes in a changing environment: Current knowledge and future directions. Soil Biol Biochem 58:216–234

Caldwell BA (2005) Enzyme activities as a component of soil biodiversity: a review. Pedobiologia 49(6):637–644

Carmeis Filho ACA, Penn CJ, Crusciol CAC, Calonego JC (2017) Lime and phosphogypsum impacts on soil organic matter pools in a tropical Oxisol under long-term no-till conditions. Agr Ecosyst Environ 241:11–23

Chae Y, Kim D, An YJ (2018) Effects of fluorine on crops, soil exoenzyme activities, and earthworms in terrestrial ecosystems. Ecotoxicol Environ Saf 151:21–27

Chaudhari MS (2016) Acute toxicity of diammonium phosphate to earthworm (Eudrilus eugeniae). J Entomol Zool Stud 4(6):501–503

Cheng Z, Lee L, Dayan S, Grinshtein M, Shaw R (2011) Speciation of heavy metals in garden soils: evidences from selective and sequential chemical leaching. J Soils Sediments 11:628–638

Cipullo S, Snapira D, Prpich G, Campo P, Coulona F (2019) Prediction of bioavailability and toxicity of complex chemical mixtures through machine learning models. Chemosphere 215:388–395

Cui X, Wang XD, Fan WH, Wang JM, Cui KY (2011) Effects of fluoride on soil properties and yield and quality of maize. Chin J Eco-Agric 19(4):897–901

Delgado A, Madrid A, Kassem S, Andre L, del Carmen del Campillo M (2002) Phosphorus fertilizer recovery from calcareous soils amended with humic and fulvic acids. Plant Soil 2:277–286

Deng J, Chen X, Wang R et al (2017) LS-SVM data mining analysis: how does biochar influence soil net nitrogen mineralization in the field? J Soils Sediments 17(3):827–840

Eivazi F, Tabatabai MA (1977) Phosphatases in soils. Soil Biol Biochem 9:167–172

Elloumi N, Zouari M, Chaari L, Abdallah FB, Woodward S, Kallel M (2015) Effect of phosphogypsum on growth, physiology, and the antioxidative defense system in sunflower seedlings. Environ Sci Pollut Res 22:14829–14840

Fox DR (2015) Selection bias correction for species sensitivity distribution modeling and hazardous concentration estimation: correction for SSD modeling. Environ Toxicol Chem 34:2555–2563

Gisbrecht A, Hammer B (2015) Data visualization by nonlinear dimensionality reduction. Wiley Interdiscip Rev Data Min Knowl Discov 5(2):51–73

Haney RL, Franzluebbers AJ (2009) Soil CO_2 evolution: response from arginine additions. Appl Soil Ecol 42:324–327

Hentati O, Abrantes N, Caetano AL, Bouguerra S, Gonçalves F, Römbke J, Pereira R (2015) Phosphogypsum as a soil fertilizer: ecotoxicity of amended soil and elutriates to bacteria, invertebrates, algae and plants. J Hazard Mater 294:80–89

Hurtado MD, Enamorado SM, Andreu L, Delgado A, Abril JM (2011) Drain flow and related salt losses as affected by phosphogypsum amendment in reclaimed marsh soils from SW Spain. Geoderma 161:43–49

Inbar Y, Boehm MJ, Hoitink HJ (1991) Hydrolysis of fluorescein diacetate in sphagnum peat container media for predicting suppressiveness to damping-off caused by *Pythium ultimum*. Soil Biol Biochem 23:479–483

International Atomic Energy Agency Report (2013) Radiation protection and management of norm residues in the phosphate industry. 308 P. https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1582_web.pdf. checked 19.11.2018

ISO 14240-1:1 (1997) Soil quality - determination of soil microbial biomass - Part 1: substrate-induced respiration method

ISO 16072 (2002) Soil quality - laboratory methods for determination of microbial soil respiration

ISO 19204 (2017) Soil quality - procedure for site-specific ecological risk assessment of soil contamination (soil quality TRIAD approach)

Jager T (2011) Some Good Reasons to ban EC x and related concepts in ecotoxicology. Environ Sci Technol 45:8180–8181

Kammoun M, Ghorbel I, Charfeddine S, Kamoun L, Gargouri-Bouzid R, Nouri-Ellouz O (2017) The positive effect of phosphogypsum-supplemented composts on potato plant growth in the field and tuber yield. J Environ Manag 200:475–483

Klose S, Tabatabai M (2000) Urease activity of microbial biomass in soils as affected by cropping systems. Biol Fertil Soils 31:191–199

Konarbaeva G (1997) Fluorine in the crusty solonetzes of Western Siberia and the impact of phosphogypsum on its content. Eurasian Soil Sci 30:977–981

Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. Phys Rev E 69. https://doi.org/10.1103/PhysRevE.69.066138

Li H, Leng W, Zhou Y, Chen F, Xiu Z, Yang D (2014) Evaluation models for soil nutrient based on support vector machine and artificial neural networks. Sci World J 478569:7. https://doi.org/10.1155/2014/478569

Liu Z, Rong Q, Zhou W, Liang G (2017) Effects of inorganic and organic amendment on soil chemical properties, enzyme activities, microbial community and soil quality in yellow clayey soil. PLOS One 12. https://doi.org/10.1371/journal.pone.0172767

Liu J, Liu M, Wu M et al (2018) Soil pH rather than nutrients drive changes in microbial community following long-term fertilization in acidic Ultisols of southern China. J Soils Sediments 18:1853–1864

Maaten LVD, Hinton G (2008) Visualizing data using *t-SNE*. J Mach Learn Res 9:2579–2605

McBride M (1989) Reactions controlling heavy metal solubility in soils. Adv Soil Sci 26:1–56

Morgado RG, Loureiro S, González-Alcaraz MN (2018) Changes in Soil Ecosystem Structure and Functions Due to Soil Contamination. Soil Pollution. Elsevier, pp 59–87

Morgan JE, Morgan AJ (1988) Calcium-lead interactions involving earthworms. Part 2: The effect of accumulated lead on endogenous calcium in *Lumbricus rubellus*. Environ Pollut 55(1):41–54

Negrão S, Schmöckel SM, Tester M (2017) Evaluating physiological responses of plants to salinity stress. Ann Bot 119(1):1–11

Niell S, Jesús F, Díaz R, Mendoza Y, Notte G, Santos E, Gérez N, Cesio V, Cancela H, Heinzen H (2018) Beehives biomonitor pesticides in agroecosystems: simple chemical and biological indicators evaluation using Support Vector Machines (SVM). Ecol Ind 91:149–154

Nikolaeva OV, Terekhova VA (2017) Improvement of laboratory phytotest for the ecological evaluation of soils. Eurasian Soil Sci 50:1105–1114

OECD Guideline for testing chemicals 222 (2004). Earthworm reproduction test (*Eisenia fetida/Eisenia andrei*)

Palmer DS, Mišin M, Fedorov MV, Llinas A (2015) Fast and general method to predict the physicochemical properties of druglike molecules using the integral equation theory of molecular liquids. Mol Pharm 12:3420–3432

Pan Y, Koopmans GF, Bonten LTC et al (2014) Influence of pH on the redox chemistry of metal (hydr)oxides and organic matter in paddy soils. J Soils Sediments 14:1713–1726

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blindel M, Prettenhofer P, Wiess R, Dubourg V et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Pereira R, Cachada A, Sousa JP, Niemeyer J, Markwiese J, Andersen CP (2018) Ecotoxicological effects and risk assessment of pollutants. Soil Pollution. Elsevier, pp 91–216

Pickering WF (1985) The mobility of soluble fluoride in soils. Environ Pollut Series B 9:281–308

Reinwarth B, Miller JK, Glotzbach C et al (2017) Applying regularized logistic regression (RLR) for the discrimination of sediment facies in reservoirs based on composite fingerprints. J Soils Sediments 17(6): 1777–1795

Rodríguez-Pérez R, Vogt M, Bajorath J (2017) Influence of varying training set composition and size on support vector machine-based prediction of active compounds. J Chem Inf Model 57(4):710–716. https://doi.org/10.1021/acs.jcim.7b00088

Roweis ST, Saul L (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326. https://doi.org/10.1126/science.290.5500.2323

Russian National Report (2015) On the state and protection of the environment issued annually by the Ministry of Natural Resources and Environment (in Russian)

Saadaoui E, Ghazel N, Ben Romdhane C, Massoudi N (2017) Phosphogypsum: potential uses and problems – a review. Int J Environ Studies 74:558–567

Saha JK, Kundu S (2003) Determination of fluoride in soil water extract through ion chromatography. Commun Soil Sci Plant Anal 34:181–188

Shatar TM, Mcbratney AB (2004) Boundary-line analysis of field-scale yield response to soil properties. J Agric Sci 142:553–560

Tabatabai MA (1977) Effects of trace elements on urease activity in soils. Soil Biol Biochem 9(1):9–13

Tayibi H, Choura M, López FA, Alguacil FJ, López-Delgado A (2009) Environmental impact and management of phosphogypsum. J Environ Manag 90:2377–2386

Telesiński A, Siwczyk F, Zakrzewska H (2012) An attempt to determination of the 50% phytotoxicity threshold for different fluoride concentrations affecting the spring wheat (*Triticum aestivum* L.) and white mustard (*Sinapis alba* L.) seedlings. Fluoride 45(3/1):213–214

Thalmann A (1968) Zur methodic derestimung der. Dehydrogenaseaktivität i. Boden mittels. Triphenyltetrazoliumchlorid (TTC). Landwirdschaft. Forschung 21:249–258

Twarakavi NCK, Šimůnek J, Schaap MG (2009) Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. Soil Sci Soc Am J 73(5):1443–1452

Vapnik VN (1995) The nature of statistical learning theory. Springer, New York

Vapnik VN, Golowich S, Smola A (1997) Support vector method for function approximation, regression estimation, and signal processing. In: Mozer MC, Jordan MI, Petsche T (eds) Advances in neural information processing systems, vol 9. MIT Press, Cambridge, pp 281–287

Vyshpolsky F, Mukhamedjanov K, Bekbaev U, Ibatullin S, Yuldashev T, Noble AD, Mirzabaev A, Aw-Hassan A, Qadir M (2010) Optimizing the rate and timing of phosphogypsum application to magnesium-affected soils for crop yield and water productivity enhancement. Agric Water Manag 97:1277–1286. https://doi.org/10.1016/j.agwat.2010.02.020

Wu CH, Ho JM, Lee DT (2004) Travel time prediction with support vector regression. IEEE Trans Intell Transp Syst 5(4):276–281

Yakovlev AS, Kaniskin MA, Terekhova VA (2013) Ecological evaluation of artificial soils treated with phosphogypsum. Eurasian Soil Sci 46:697–703

Zaman AU (2014) Identification of key assessment indicators of the zero waste management systems. Ecol Ind 36:682–693

Zhu J, Wang J, Ding Y, Liu B, Xiao W (2018) A systems-level approach for investigating organophosphorus pesticide toxicity. Ecotox Environ Saf 149:26–35