PROCEEDINGS OF THE YEREVAN STATE UNIVERSITY
УЧЕНЫЕ ЗАПИСКИ ЕРЕВАНСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
Geology and Geography **59**(2), 2025 ***Геология и география***

Geography

УДК 504.064.2.001.18; 504.054; 004.942

PREDICTING CONCENTRATION CHANGE OF SOME TMS IN SOIL–WATER ECOSYSTEM USING MACHINE LEARNING

A. R. SUKIASYAN [*], P. A. YESAYAN [**], Z. G. KHANAMIRYAN [***], A. A. KIRAKOSYAN [****]

National Polytechnic University of Armenia (NPUA), Armenia

This article provides a discussion of the applicability of machine learning methods, with a particular focus on linear regression, to predict the total concentration of TMs in the soil–water ecosystem. Despite the requirement of only minute quantities, TMs have the potential to exert a detrimental effect on the environment. For prediction, seasonal and geographical parameters along with metal concentrations in the soil and their irrigation water were used. A key focus of the study was the normalization of data, a process that has been shown to improve the identification of linear relationships between variables. The developed linear regression model demonstrated a high degree of precision as evidenced by the coefficient of determination 0.9945, the average absolute error of 0.1, and the average percentage error of 5.5%. These findings substantiate the feasibility of employing the proposed methodology to monitor water quality, evaluate pollution risks, and identify potential threats at an early stage in ecosystems that anthropogenic factors have been impacted.

https://doi.org/10.46991/PYSUC.2025.59.2.574

***Keywords*:** transition metal, soil–water chain, correlation coefficients, machine learning, linear regression, Python.

Introduction. Transition metals (TMs), also known as d-elements, are distinguished by partially filled d-orbitals, a property that enables them to exhibit varying degrees of oxidation. Despite the lack of synonymy between the terms "heavy metals" and "transition metals", it is evident that numerous d-elements exhibit properties characteristic of heavy metals [1]. TMs represent a critical category of elements that have been the subject of scientific concern due to their environmental impact, particularly within soil–water systems used for irrigation [2]. Although these elements are essential in minimal quantities for living organisms, their anthropogenic accumulation in soil and water leads to ecological imbalance, which poses a serious threat to ecosystem sustainability [3]. The presence of TMs in irrigation water has been shown to alter soil chemistry and fertility, as well as affect plant physiology [4]. There is also a potential for human health to be affected by bioaccumulation along the food chain [5]. The mobility and bioavailability of TMs

[*] E-mail: sukiasyan.astghik@gmail.com
[**] E-mail: patyesayan@gmail.com
[***] E-mail: zavenkhanamiryan99@gmail.com
[****] E-mail: arm.kirakosyan71@gmail.com

in the soil–water ecosystem are influenced by various factors, including pH, organic matter content, soil structure, oxidation-reduction potential and microbiological activity [6].

Soil irrigation is particularly vulnerable to changes in the concentration of transition metals [7]. A comprehensive understanding and reliable prediction of the dynamics of these concentrations is imperative for the implementation of effective prevention and recovery strategies for contaminated soil [8]. However, conventional monitoring methodologies entail costly and time-consuming sampling and laboratory analysis [9]. To address these limitations, recent research has increasingly focused on the use of machine learning (ML) algorithms. These algorithms can analyze large multidimensional datasets and identify hidden patterns, offering a promising alternative to model the behaviour of TMs in complex ecological systems [10]. In particular, regression models, artificial neural networks (ANN), and collective learning methods have demonstrated great potential in predicting metal concentrations based on the physico-chemical parameters of soil and water [11–13].

The aim of the presented work is to investigate the dynamics of concentration changes as well as the prediction of the total concentration of some TMs in the soil–water ecosystem using advanced machine learning methods.

Materials and Methods. The concentrations of scandium (Sc), titanium (Ti), vanadium (V), chromium (Cr), manganese (Mn), iron (Fe), cobalt (Co), nickel (Ni), copper (Cu) and zinc (Zn) were measured in soil and irrigation water samples collected during each season over a year. Samples were collected from several regions of Armenia, including Hrazdan (cement plant territory), Gavar (Noratus community), and Martuni (Yeranos community) from a depth of 20 *cm* using non-metallic instruments in dry weather conditions. Instrumental measurements were conducted within the laboratory setting in accordance with [14]. The standard ISO 5667-1:2006 methodology was used for the collection of irrigation water samples. This method describes the analysis of water samples during four different seasons. The elemental analysis of all irrigation water samples was conducted using a portable Termo Scientific™ Niton™ X-ray analyzer, which utilized direct X-ray radiation [15]. In modeling the relationship between transition metal concentrations in soil and water samples, linear and multivariate regression were used according to [16].

Results and Discussion. Tabs. 1 and 2 present the concentrations of TMs in soil and water samples from the designated locations in the four seasons. The features are therefore the input variables that were used to predict the target variable (“Water” in this case), which is defined as the total concentration of certain metals in the water samples. Similarly, the “Soil” feature indicates the summed concentration of the same or related metals in soil samples. The input features can be categorized into two main groups: seasonal features (Autumn, Winter, Spring, and Summer) that capture temporal variations in the presence of metal and geographical features (Gavar, Martuni, and Hrazdan) that account for spatial differences across various regions. The “Water” characteristic, which being the output variable, is the primary focus of the analysis in terms of how it is influenced by these input variables.

In our case, each element concentration in the soil had a different range and magnitude. As a case in point, the range of iron concentrations can vary from 20 to

40 *g*/*kg*, while cobalt concentrations can only range from 0 to 300 *mg*/*kg*. Absent scaling, features with more expansive numeric ranges have the potential to exert an undue influence over regression coefficients, a circumstance that may give rise to the promulgation of a biased or misleading model. To address this challenge, Min-Max normalization was implemented to scale all input features to a uniform [0,1] range [17]:

$$x_i^{\text{norm}} = \frac{x_i - x_{i,\min}}{x_{i,\max} - x_{i,\min}}.$$

The normalized data are presented in Tabs. 3 and 4.

Table 1

Concentrations of some transition metals in samples of mountain chernozem, mg/kg

Transition metall	The territory of the Hrazdan Cement Plant				Gavar, Noratus community				Martuni, Yeranos community			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Sc	98.17	trace	137.90	171.8	162.0	trace	173.5	166.5	177.65	131.1	213.25	171.6
Ti	3550.80	3928.7	4115.33	4050.15	4272.80	4065.35	3967.55	4650.7	3999.8	3239.80	3631.10	4076.45
V	124.87	125.93	153.10	131.2	130.55	128.25	121.85	122.8	116.55	104.55	114.65	115.25
Cr	97.73	102.33	138.50	118.9	113.30	146.25	127.20	109.9	135.30	111.95	129.00	131.95
Mn	849.07	858.6	799.10	401.8	819.65	810.90	792.50	780.8	737.40	776.05	725.30	741.5
Fe	29633.10	29432.4	29639.13	20903.4	38515.0	36379.35	36027.95	37116.6	31548.45	29325.95	31199.50	30847.7
Co	84.0	trace	trace	162.2	137.0	280.05	93.70	216.55	78.80	139.80	138.10	182.7
Ni	62.0	65.77	65.80	64.2	80.70	84.25	86.65	81.65	57.0	84.25	65.20	76.85
Cu	85.10	58.47	67.30	71.58	84.40	69.95	65.45	69.8	71.80	58.65	53.60	54.45
Zn	99.27	109.97	118.13	116.9	87.50	126.50	89.50	90.1	85.45	98.60	85.40	83.2

Table 2

Concentrations of some transition metals in water samples near mountain chernozem, mg/L

Transition metall	The territory of the Hrazdan Cement Plant				Gavar, Noratus community				Martuni, Yeranos community			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Sc	0.09	0.11	0.18	0.078	0.09	0.04	0.09	0.043	0.04	–	0.06	0.048
Ti	0.09	0.51	1.46	0.106	0.03	0.04	Trace	0.055	trace	–	trace	0.021
V	0.003	0.02	0.04	0.007	0.03	0.02	0.03	0.020	0.001	–	0.004	trace
Cr	trace	trace	0.05	0.003	trace	0.01	0.02	0.007	trace	–	0.03	trace
Mn	0.09	0.07	0.18	–	0.02	0.01	trace	trace	trace	–	trace	trace
Fe	1.06	3.63	13.91	1.318	0.65	0.91	0.39	1.280	0.14	–	0.24	0.284
Co	trace	trace	trace	trace	trace	trace	trace	trace	trace	–	trace	trace
Ni	trace	trace	trace	trace	trace	trace	trace	trace	trace	–	trace	trace
Cu	0.07	0.04	0.07	0.019	0.02	0.07	0.02	trace	0.012	–	0.02	0.013
Zn	0.07	0.05	0.09	0.023	0.02	0.12	0.02	0.012	0.007	–	0.01	0.011

Table 3

Normalized values of some transition metals in samples of mountain chernozems

Transition metall	The territory of the Hrazdan Cement Plant				Gavar, Noratus community				Martuni, Yeranos community			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Sc	0.46	0.00	0.65	0.81	0.76	0.00	0.81	0.78	0.83	0.61	1.00	0.80
Ti	0.22	0.49	0.62	0.57	0.73	0.59	0.52	1.00	0.54	0.00	0.28	0.59
V	0.42	0.44	1.00	0.55	0.54	0.49	0.36	0.38	0.25	0.00	0.21	0.22
Cr	0.00	0.09	0.84	0.44	0.32	1.00	0.61	0.25	0.77	0.29	0.64	0.71
Mn	0.98	1.00	0.87	0.00	0.91	0.90	0.86	0.83	0.73	0.82	0.71	0.74
Fe	0.50	0.48	0.50	0.00	1.00	0.88	0.86	0.92	0.60	0.48	0.58	0.56
Co	0.30	0.00	0.00	0.58	0.49	1.00	0.33	0.77	0.28	0.50	0.49	0.65
Ni	0.17	0.30	0.30	0.24	0.80	0.92	1.00	0.83	0.00	0.92	0.28	0.67
Cu	1.00	0.15	0.43	0.57	0.98	0.52	0.38	0.51	0.58	0.16	0.00	0.12
Zn	0.37	0.62	0.81	0.78	0.10	1.00	0.15	0.16	0.05	0.36	0.05	0.00

Table 4

Normalized values of some transition metals in water samples near mountain chernozems

Transition metall	The territory of the Hrazdan Cement Plant				Gavar, Noratus community				Martuni, Yeranos community			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Sc	0.50	0.61	1.00	0.44	0.50	0.22	0.50	0.22	0.22	0.00	0.33	0.28
Ti	0.06	0.35	1.00	0.07	0.02	0.03	0.00	0.04	0.00	0.00	0.00	0.01
V	0.00	0.50	1.00	0.25	0.75	0.50	0.75	0.50	0.00	0.00	0.00	0.00
Cr	0.00	0.00	1.00	0.00	0.00	0.20	0.40	0.20	0.00	0.00	0.60	0.00
Mn	0.50	0.39	1.00	0.00	0.11	0.06	0.00	0.00	0.00	0.00	0.00	0.00
Fe	0.08	0.26	1.00	0.09	0.05	0.07	0.03	0.09	0.01	0.00	0.02	0.02
Co	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ni	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cu	1.00	0.57	1.00	0.29	0.29	1.00	0.29	0.00	0.14	0.00	0.29	0.14
Zn	0.58	0.42	0.75	0.17	0.17	1.00	0.17	0.08	0.08	0.00	0.08	0.08

To evaluate the impact of normalization, a comparison was made between the Pearson correlation coefficients between the concentrations of soil metals and the total concentrations of metal present in water before and after the normalization process [18]. The results of this study are demonstrated in Figs. 1 and 2 below.

Soil	1.00								
Water	–0.16	1.00							
Fall	0.18	–0.21	1.00						
er	–0.03	–0.07	–0.33	1.00					
Spring	0.06	0.44	–0.33	–0.33	1.00				
Summer	–0.22	–0.17	–0.33	–0.33	–0.33	1.00			
Hrazdan	–0.65	0.58	0.00	0.00	0.00	0.00	1.00		
Gavar	0.83	–0.22	0.00	0.00	0.00	0.00	–0.50	1.00	
Martuni	–0.18	–0.36	0.00	0.00	0.00	0.00	–.50	–0.50	1.00
	Soil	**Water**	**Fall**	**Winter**	**Spring**	**Summer**	**Hrazdan**	**Gavar**	**Martuni**

Fig. 1. Correlation matrix before normalization.

Soil	1.00								
Water	0.29	1.00							
Fall	0.00	–0.13	1.00						
Winter	-0.12	–0.02	–0.33	1.00					
Spring	0.07	0.48	–0.33	–0.33	1.00				
Summer	0.06	–0.32	–0.33	–0.33	–0.33	1.00			
Hrazdan	–0.38	0.58	0.00	0.00	0.00	0.00	1.00		
Gavar	0.83	–0.02	0.00	0.00	0.00	0.00	–0.50	1.00	
Martuni	–0.45	–0.56	0.00	0.00	0.00	0.00	–0.50	–0.50	1.00
	Soil	**Water**	**Fall**	**Winter**	**Spring**	**Summer**	**Hrazdan**	**Gavar**	**Martuni**

Fig. 2. Correlation matrix after normalization.

The matrices illustrate the linear relationship between each input feature and “Water”. In this context, normalization refers to the process of scaling the input data, so that disparate variables can be analyzed comparatively on an equal footing, particularly when their raw values are expressed in different units or ranges.

A significant insight derived from these matrices is that specific relationships experience substantial enhancement after normalization. For example, Spring consistently shows a positive correlation with water TMs concentrations, suggesting that seasonal processes – such as snowmelt or increased surface runoff – are major contributors to metal transport into water bodies during this period. In contrast, Summer has been observed to exhibit a negative correlation, attributable to reduced runoff and elevated evaporation rates. Among the geographical features, one location reveals a notably strong positive correlation with “Water” in both matrices, indicating persistent regional factors influencing metal presence, such as geological composition or local industrial activity.

Further quantifies how these relationships change due to normalization (Fig. 3) as the difference matrix, highlighting the increase or decrease in correlation values between features and the target variable after normalization. A particularly significant change is observed in the correlation between “Soil” and “Water”, which increases by +0.45, increasing from –0.16 to 0.29. This shift underscores the impact of normalization in revealing a more accurate linear relationship between soil and water metal concentrations – two variables expected to be intrinsically connected in natural systems. Other features, such as “Martuni” and “Gavar”, also show changes in their relationships with “Water”, but the magnitude is less pronounced. Interestingly, while the strength of some correlations increases, others show a slight decrease, reflecting the nuanced effects of normalization on different feature types (such as Summer).

Soil	0.00								
Water	0.45	0.00							
Fall	–0.19	0.08	0.00						
Winter	–0.09	0.05	0.00	0.00					
Spring	0.01	0.03	0.00	0.00	0.00				
Summer	0.27	–0.16	0.00	0.00	0.00	0.00			
Hrazdan	0.27	0.00	0.00	0.00	0.00	0.00	0.00		
Gavar	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	
Martuni	–0.27	–0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Soil	**Water**	**Fall**	**Winter**	**Spring**	**Summer**	**Hrazdan**	**Gavar**	**Martuni**

Fig. 3. Normalized and original data correlation difference matrix.

Overall, the improved correlations after normalization suggest that this preprocessing step improves the clarity of the linear relationships in the data set. It makes the features more comparable, reduces the bias caused by scale differences, and ultimately contributes to more effective regression modelling. This is particularly important in environmental studies, where multivariate influences, such as seasonal cycles and regional soil properties, interact to affect outcomes such as water quality. By improving the internal consistency of the data, normalization increases the reliability of predictive models and supports more accurate interpretations of environmental interactions.

As shown in the correlation analysis, many of the input features, in particular “Soil” and several seasonal indicators, have approximately linear relationships with the target variable. Linear regression is a well-established method for quantifying such relationships, offering interpretability, computational efficiency, and robustness in situations, where linearity is present [19]. In addition, the normalization process further improves the suitability of the data for linear modelling by aligning feature scales and enhancing linear patterns. These factors make linear regression a logical and effective choice for modelling metal concentrations in water based on environmental and spatial indicators. The relationship between metal concentrations in soil and total metal concentrations in water can be modelled using a linear regression approach:

$$C_{water} = \beta_0 + \sum_{i=1}^{n} \beta_i \, x_i^{\text{norm}},$$

where n is the number of metal types (in our case $n = 10$); β_0 is the intercept of the model; β_i is the weight (coefficient) of each input feature x_i^{norm}; x_i^{norm} is the normalized concentration of the i-th metal in the soil. Normalization is a crucial preprocessing step in regression modeling, particularly when input features (predictors) vary in scale or unit.

The final regression model is formulated follows:

$$C_{water} = \beta_0 + \sum_{i=1}^{n} \beta_i \frac{x_i - x_{i,\min}}{x_{i,\max} - x_{i,\min}},$$

where x_i is the original concentration of the i-th metal in the soil; $x_{i,\max}$ and $x_{i,\min}$ are the maximum and minimum observed values for that metal in the soil data set.

The learned coefficients and intercepts of the model, which define the impact of soil metal concentration, are summarized in Tab. 5.

Table 5

Model coefficients

N	Coefficients, β	Value
1.	Intercept	–4.3892
2.	Sc–Soil	1.7642
3.	Ti–Soil	–1.1038
4.	V–Soil	7.2705
5.	Cr–Soil	0.1539
6.	Mn–Soil	4.1672
7.	Fe–Soil	–0.6444
8.	Co–Soil	0.0174
9.	Ni–Soil	–0.2106
10.	Cu–Soil	–1.1616
11.	Zn–Soil	1.9906

The effectiveness of the linear regression model is further evaluated using four common error metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). As shown in Tab. 6, the model achieves an MAE of 0.1, meaning that, on average, the predicted total metal concentration in water deviates from the actual values by only 0.1 units. The MSE and RMSE values are 0.02 and 0.14, respectively – both relatively low, indicating that large prediction errors are rare and that the model maintains a good overall fit.

Table 6

Error metrics

MAE	MSE	RMSE	R^2
0.1	0.02	0.14	0.9945

In particular, the R^2 score is 0.9945, which implies that the model explains more than 99% of the variance in the target variable. This exceptionally high R^2 value confirms that the model captures the underlying relationships between soil metal concentrations and water metal concentration with great accuracy.

Together, these metrics demonstrate that the linear regression model is not only interpretable and efficient but also highly reliable in predicting metal concentration levels in water based on the given set of input features. This level of performance makes the model suitable for practical applications, such as environmental monitoring and early-warning systems for water contamination.

To further assess the accuracy of the model, trained regression models are used to predict water metal concentrations using the same data set. The actual values, predicted values, and absolute errors are displayed in Tab. 7 for each metal.

Table 7

Predicted values of total metal concentrations in water

Actual concentration	Predicted concentration	Absolute error	Error percent from actual value
2.721181	2.530304	0.190877	7.0%
3.098374	3.132100	0.033726	1.1%
7.750000	7.809350	0.059350	0.8%
1.317064	1.301301	0.015762	1.2%
1.880769	2.155323	0.274554	14.6%
3.070596	3.045271	0.025325	0.8%
2.130418	1.819638	0.310781	14.6%
1.138672	1.078421	0.060251	5.3%
0.458477	0.511329	0.052851	11.5%
0.000000	0.183271	0.183271	–
1.319635	1.306441	0.013194	1.0%
0.537796	0.550233	0.012437	2.3%

To further evaluate the predictive performance of the model at individual data points, Tab. 7 compares the actual metal concentrations in water with the predicted values produced by the linear regression model. In addition to these, the table lists the absolute error and the percentage error relative to the actual value for each instance. The absolute error represents the direct difference between the actual and predicted values. Lower values indicate that the model predictions are closer to the true concentrations. Most absolute errors in the table are small, often below 0.1, which reflects strong prediction accuracy. The percentage of error from the actual value provides a normalized view of these errors, making it easier to compare across samples of different magnitudes. In this data set, the majority of percentage errors are impressively low. Several instances show errors below 2%, with many below 1%, indicating that the model can accurately replicate observed water metal concentrations even at different concentration scales. There are a few samples with relatively higher errors (e.g., 14.6%), but these are likely influenced by outliers or edge cases where input features may not fully capture variability in water metal concentration. Despite these, the average error across the dataset is just 5.5%, which is a very strong result in environmental modeling, particularly for systems as complex and variable as soil-to-water metal transfer. This detailed comparison confirms that the model not only performs well on average metrics like MAE and R^2, but also maintains high accuracy at the individual prediction level – making it suitable for both overall trend analysis and practical environmental decision-making.

The absolute error provides insight into the deviation between the actual and predicted values. A lower value indicates a better fit to the model. The average error is 5.5%, which is a real good result for the metal–water system.

The analysis, including data calculations, preprocessing, normalization, and training and evaluation of the linear regression model, was conducted in the Python environment using scikit-learn for machine learning workflows and stats models for statistical analysis and inference.

Conclusion. The results suggest that linear regression can serve as a reliable tool for environmental monitoring, allowing early detection of pollution risks and aiding in the formulation of mitigation strategies. Given the robustness and precision of the model, it has significant potential for practical applications, including real-time assessment of water quality and ecosystem management. However, future studies should explore the incorporation of more advanced machine learning techniques, such as deep learning and ensemble methods, to further refine predictions and account for nonlinear relationships in complex environmental systems.

In general, this research contributes to the growing body of knowledge on the application of data-driven approaches to environmental science, demonstrating that machine learning can be an efficient and cost-effective alternative to traditional monitoring techniques.

This work was supported by the Science Committee of the MESCS RA, in the frames of the research project No. 21T-2H216.

Received 04.04.2025
Reviewed 04.06.2025
Accepted 15.08.2025

REFERENCES

1. Duffus J.H. "Heavy Metals" a Meaningless Term? (IUPAC Technical Report). *Pure and Applied Chemistry* **74** (2002), 793–807. https://doi.org/10.1351/pac200274050793
2. Bilčíková J., Fialková V., et al. Influence of TMs on Animal and Human Health: A Review. *The Serbian Journal of Agricultural Sciences* **67** (2018), 187–195. https://doi.org/10.1515/contagri-2018-0027
3. Ali H., Khan E., Ilahi I. Environmental Chemistry and Ecotoxicology of Hazardous Heavy Metals: Environmental Persistence, Toxicity, and Bioaccumulation. *Journal of Chemistry* **2019** (2019). https://doi.org/10.1155/2019/6730305
4. Gupta N., Yadav K.K., et al. Trace Elements in Soil-Vegetables Interface: Translocation, Bioaccumulation, Toxicity and Amelioration. A Review. *Science of The Total Environment* **651** (2019), 2927–2942. https://doi.org/10.1016/j.scitotenv.2018.10.047
5. Singh J., Kalamdhad A.S. Effects of Heavy Metals on Soil, Plants, Human Health and Aquatic Life. *International Journal of Research in Chemistry and Environment (IJRCE)* **1** (2011), 15–21. https://www.ijrce.org/index.php/ijrce/article/view/78
6. Alloway B.J. *Heavy Metals in Soils: Trace Metals and Metalloids in Soils and Their Bioavailability* (3rd ed.). Environmental Pollution Series, Springer (2013). https://doi.org/10.1007/978-94-007-4470-7

7. Chaoua S., Boussaa S., et al. Impact of Irrigation with Wastewater on Accumulation of Heavy Metals in Soil and Crops in the Region of Marrakech in Morocco. *Journal of the Saudi Society of Agricultural Sciences* **18** (2019), 429–436. https://doi.org/10.1016/j.jssas.2018.02.003
8. Kumar M., Seth A., et al. Remediation Strategies for Heavy Metals Contaminated Ecosystem: A Review. *Environmental and Sustainability Indicators* **12** (2021), 100155. https://doi.org/10.1016/j.indic.2021.100155
9. Monjardin C.E.F., Power C., et al. Application of Machine Learning for Prediction and Monitoring of Manganese Concentration in Soil and Surface Water. *Water* **15** (2023). https://doi.org/10.3390/w15132318
10. Padarian J., Minasny B., Mc Bratney A.B. Machine Learning and Soil Sciences: A Review Aided by Machine Learning Tools. *SOIL* **6** (2020), 35–52. https://doi.org/10.5194/soil-6-35-2020
11. Azizi K., Ayoubi S., et al. Predicting Heavy Metal Contents by Applying Machine Learning Approaches and Environmental Covariates in West of Iran. *Journal of Geochemical Exploration* **233** (2022), 106921. https://doi.org/10.1016/j.gexplo.2021.106921
12. Luo N. Methods for Controlling Heavy Metals in Environmental Soils Based on Artificial Neural Networks. *Scientific Reports* **14** (2024), 1–13. https://doi.org/10.1038/s41598-024-52869-9
13. Yang H., Huang K., et al. Predicting Heavy Metal Adsorption on Soil with Machine Learning and Mapping Global Distribution of Soil Adsorption Capacities. *Environmental Science & Technology* **55** (2021), 14316–14328. https://doi.org/10.1021/acs.est.1c02479
14. Sukiasyan A.R., Kirakosyan A.A. Seasonal Aspects of Macro, Trace, and Ultra Trace Element Changes in Soils with Different Anthropogenic Loads. *Sustainable Development of Mountain Territories* **16** (2024), 789–802. https://doi.org/10.21177/1998-4502-2024-16-2-789-802
15. Thermo Scientific Sample Collection and Preparation Tools for Exploration and Mining. https://www.malvernpanalytical.com/en/assets/pn12871_le_sample_prep_of_geological_or_mining_and_other_inorganic_samples_tcm50-95064.pdf
16. James G., Witten D., et al. An Introduction to Statistical Learning: with Applications in Python. *Springer Nature* **1** (2023), 596. https://doi.org/10.1007/978-3-031-38747-0
17. De Amorim L.B.V., Cavalcanti G.D.C., Rafael M.O. Cruz the Choice of Scaling Technique Matters for Classification Performance. *Applied Soft Computing* **133** (2023), 109924. https://doi.org/10.1016/j.asoc.2022.109924
18. *Field Andy Discovering Statistics Using IBM SPSS Statistics* (6th ed.). Sage Publications (2024), 1144.
19. Montgomery D.C., Peck E.A., Vining G.G. *Introduction to Linear Regression Analysis*. John Wiley & Sons (2011), 645.

Ա. Ռ. ՍՈՒՔԻԱՍՅԱՆ, Պ. Ա. ԵՍԱՅԱՆ, Զ. Գ. ԽԱՆԱՄԻՐՅԱՆ, Ա. Ա. ԿԻՐԱԿՈՍՅԱՆ

ՄԵՔԵՆԱՅԱԿԱՆ ՈՒՍՈՒՑՄԱՆ ԿԻՐԱՌՄԱՄԲ ՀՈՂ–ՋՈՒՐ ԷԿՈՀԱՄԱԿԱՐԳՈՒՄ ՈՐՈՇ ԱՆՑՈՒՄԱՅԻՆ ՄԵՏԱՂՆԵՐԻ ԿՈՆՑԵՆՏՐԱՑԻՈՆ ՓՈՓՈԽՈՒԹՅԱՆ ԿԱՆԽԱՏԵՍՈՒՄԸ

Ամփոփում

Հոդվածում քննարկվում է մեքենայական ուսուցման մեթոդների կիրառումը՝ հատուկ ուշադրություն դարձնելով գծային ռեգրեսիային՝ հող–ջուր էկոհամակարգում անցումային մետաղների գումարային կոնցենտրացիան կանխատեսելու համար: Չնայած նրան, որ անցումային մետաղները անհրաժեշտ են չնչին քանակությամբ, սակայն դրանք կարող են վնասակար ազդեցություն ունենալ շրջակա միջավայրի վրա: Կանխատեսման համար օգտագործվել են սեզոնային և աշխարհագրական պարամետրերը, ինչպես նաև հողում և ոռոգման ջրում մետաղների կոնցենտրացիաները: Տվյալների նորմալացումն այս հետազոտության առանցքային փուլերից էր, որը ցույց է տվել իր արդյունավետությունը՝ փոփոխականների միջև գծային կախվածության բացահայտման համար: Մշակված գծային ռեգրեսիայի մոդելը ցույց է տվել բարձր ճշգրտություն. դետերմինացիայի գործակիցը կազմել է 0,9945, միջին բացարձակ սխալը՝ 0,1, իսկ միջին տոկոսային սխալը՝ 5,5%: Այս արդյունքները հաստատում են առաջարկված մեթոդի կիրառելիությունը ջրի որակի մոնիտորինգի, աղտոտման ռիսկերը գնահատելու և էկոհամակարգերում մարդածին ազդեցությամբ պայմանավորված հնարավոր սպառնալիքները վաղ փուլում բացահայտելու համար:

А. Р. СУКИАСЯН, П. А. ЕСАЯН, З. Г. ХАНАМИРЯН, А. А. КИРАКОСЯН

ПРОГНОЗИРОВАНИЕ КОНЦЕНТРАЦИОННОГО ИЗМЕНЕНИЯ НЕКОТОРЫХ ПЕРЕХОДНЫХ МЕТАЛЛОВ В ЭКОСИСТЕМЕ ПОЧВА–ВОДА С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

Резюме

В статье обсуждается применимость методов машинного обучения с особым упором на линейную регрессию для прогнозирования общей концентрации переходных металлов в экосистеме почва–вода. Несмотря на то, что переходные металлы требуются в незначительных количествах, они потенциально могут оказывать пагубное воздействие на окружающую среду. Для прогнозирования были использованы сезонные и географические параметры, а также концентрации металлов в почве и оросительной воде. Ключевым направлением исследования была нормализация данных – процесс, который, как было показано, улучшает выявление линейных зависимостей между переменными. Разработанная модель линейной регрессии продемонстрировала высокую степень точности, о чем свидетельствуют коэффициент детерминации 0,9945, средняя абсолютная погрешность 0,1 и средняя процентная погрешность 5,5%. Эти результаты подтверждают целесообразность использования предложенной методологии для мониторинга качества воды, оценки рисков загрязнения и выявления потенциальных угроз на ранней стадии в экосистемах, подвергшихся воздействию антропогенных факторов.