Correlation and Regression Analysis of Cadmium, Lead, Concentrations in Agricultural Soils and Feed Rations of Different Types of Cow Feeding for Prediction and Assessment of Environmental Safety Risk

Serhii Portiannyk[1], Oleksiy Mamenko[2], Alexander Tsereniuk[1], Andriy Onyshchenko[1]

[1]*Institute of Pig Breeding and Agro-Industrial Production of the National Academy of Agrarian Sciences, 36013, 1 Shvedska Mohyla Str., Poltava, Ukraine*

[2]*Institute of Animal Husbandry of the National Academy of Sciences of Ukraine, 61026, 1-A Tvarynnykiv Str., Kharkiv, Ukraine*

Keywords: *cattle, heavy metal, agroecosystem, pollutant.*

Abstract. *Cadmium (Cd) and lead (Pb) as dangerous ecotoxicants pose a significant danger to agroecosystems and all living organisms. They enter the environment due to the action of anthropogenic-technogenic, military factors, global climate change, etc. Data analysis was carried out based on the results of a scientific and economic experiment conducted on dairy cows with different types of feeding in the Forest-Steppe soil-climatic zone of Ukraine. Cows were selected by the method of analogues for live weight and productivity with the third lactation period. The rations included feeds with an excess of Cd and Pb. The ecotoxicity of pollutants affected their transition from soil to plants (feed). The purpose of the research is to analyze the correlational relationship, construct regression equations between the concentration of heavy metals (HM) Cd and Pb in soil samples and feeds of different types of dairy cow diets, which will allow predicting the transfer of pollutants from soil to feed using environmentally safe milk production technologies, carry out environmental monitoring and assess environmental risks on cattle farms in a timely manner. Correlation analysis using the parametric Pearson correlation coefficient and linear regression was performed using the STATISTICA program. In all cases, the analysis established a positive close relationship between the content of heavy metals in soil samples of agroecosystems and feed rations of different types of feeding $r > 0.7$, which indicates a strong correlation dependence and a high predictability of one variable for another. The correlation coefficient for cadmium concentration ranged from 0.799 to 0.971, for lead from 0.790 to 0.869. The highest correlation was established for the content of Cd, where cows were fed a silage-hay-concentrate type ration $r = 0.971$ (very strong correlation). The content of Pb showed a very strong correlation ($r = 0.869$) in a silage-hay ration group. The most acceptable models were for the concentration of Cd in cows with silage-root crops, silage-haylage, and silage-hay-concentrate feeding types, and for the content of Pb in animals with silage-hay and silage-haylage feeding types. A positive close relationship was established between the concentration of Cd and Pb in the soils of agroecosystems and the diets of dairy cows with silage-root crops, silage-hay, silage-haylage, and silage-hay-concentrate feeding type. The best models for predicting Cd in cows on silage-root crops, silage-haylage and silage-hay-concentrate feeding types; Pb ¬ silage-hay and silage-haylage.*

Introduction

Anthropogenic, technogenic, and military impact on the biosphere leads to contamination of agroecosystems with toxic heavy metals (HM) such as cadmium, lead, copper, zinc, chromium, mercury, nickel, etc., which requires constant monitoring, prevention of spread to large territories, solving the problem of accumulation in soil, water, feed, animal bodies, livestock products, organic waste (Sachko et al., 2016; Saurabh et al., 2023).

Some research studies on soil pollution have been published. In a study by Sazal et al. (2021), 54 dangerous toxicants were found. The researchers selected 40 soil samples around five brick kilns. Multivariate statistics were used to process the results of the study. Researchers claim that in addition to burning coal in kilns, soil pollution has a negative impact on agriculture. Monitoring environmental pollution with heavy metals is a complex process. To solve the relevant problems, researchers set the task of developing mathematical models for predicting pollution of ecosystems with heavy metals, tested in scientific experiments. Two- and multivariate correlation analysis between heavy metals in different components of the biosphere were used during the statistical analysis of the obtained research results with the construction of appropriate regression equations to predict the level of pollution (Ashraful et al., 2023). Kuramshina et al. (2014) studied the migration of HM from the soil to plants (feed) and the body of animals in the zone of influence of oil and ore deposits with the aim of biologically indicating the

Corresponding author: portiannyk@gmail.com

state of the ecosystem, assessing the environmental safety of livestock production in various agricultural areas. The concentrations of elements in the soil and feed were determined, and the coefficients of the transition of HM from feed to the body of animals were calculated. The conclusions made by the researchers are not complicated, but extremely important from a practical point of view: the more elements entered the environment, polluting agroecosystems and plants (feed), the more they were consumed by animals. Sachko et al. (2016) also concluded that there is a relationship between the level of HM in the soil and agricultural crops.

Mingtao et al. (2021) studied the correlation between the content of heavy metals in soil and plants grown to assess the risk of contamination of agricultural crops and predict the danger of harmful effects of toxic metals, especially mercury and cadmium on the human body. Researchers mostly paid attention to the contamination of soil and plants with HM, but did not pay attention to the relationship between them. The results showed that the studied toxic metals in the soil had the greatest impact on the level of zinc in plants, while Pb and Cr had a synergistic effect on the absorption of Zn by plants. The experiments prove the importance of correlation analysis for controlling the risk of contamination, which was not taken into account before.

In our previous studies, we analyzed the correlation between the content of HM in feed rations and organic waste with the construction of appropriate regression equations for assessing the environmental risk of contamination of agricultural lands, predicting the transfer of toxicants from feed to organic waste and organic fertilizers applied to fields for agricultural crops, in order to carry out timely and effective monitoring and avoid contamination of agroecosystems where plants are grown and used as feed for dairy cows (Portyannik, 2024a). There are many examples of the use of correlation and regression analysis both in agriculture and other industries related to environmental pollution in different countries of the world, but the study of the correlation between the content of toxic metals in soil samples and feeds of different types of dairy cow diets with the construction of appropriate regression equations was done by us for the first time and is a relevant study that will facilitate the prediction of the transfer of HM from soil to plants. Cow ration feed in the conditions of environmentally safe milk production technology, both today and in the post-war period, on the basis of the correlation relationship and regression equations were verified in a scientific experiment.

The study of the correlation between the content of toxic metals cadmium and lead in soil samples and feed rations of different types of cow feeding is a convenient and reliable indicator of the forecast of environmental safety in agricultural enterprises. The use of proven regression equations complements the existing tools for environmental monitoring of agroecosystems.

The purpose of the research is to analyze the correlational relationship, construct regression equations between the concentration of heavy metals Cd and Pb in soil samples and feed of different types of dairy cow rations, which will allow predicting the transfer of pollutants from soil to feed in the conditions of environmentally safe milk production technology, and carry out environmental monitoring and timely assessment of environmental risks on cattle farms.

Materials and Methods

An experiment on dairy cows of Ukrainian black and red-motley dairy breeds for the production of environmentally safe milk was conducted in farms of the Forest-Steppe soil-climatic zone of Ukraine. For the experiments, 126 cows with silage-hay-concentrate feeding type, 63 with silage-hay, 36 with silage-root and 195 with silage-hay feeding type were selected. Cows of all groups were fed feeds containing heavy metals Cd and Pb. The average live weight of cows was 500–545 kg. The average daily milk yield was 14.0–14.8 kg, which is an average of 4270–4514 kg of milk per lactation. The duration of the comparative period was 42 days. Cows selected by the method of analogues for live weight and productivity in their third lactation period were in the same conditions of feeding and were kept in the production conditions of experimental agricultural enterprises with observance of milk production technology. All manipulations with animals were carried out in accordance with the European Convention for the Protection of Vertebrate Animals Used for Experimental and Scientific Purposes (Strasbourg, 1986). Animals of all experimental groups were fed feeds with an increased content of heavy metals Cd and Pb as part of the diets of four types of feeding: silage-root, silage-hay, silage-hay, silage-hay-concentrate. The experiment lasted 120 days.

Before the beginning of the experiment, average soil samples of agricultural lands were taken at a depth of 25 cm, where plants were grown and used to produce feed, included in the diets of different types of feeding for the production of cow's milk. The root system of plants, which absorbs heavy metals from the soil, is located at a depth of 25 cm. Samples were taken before harvesting in the spring on land plots where the corresponding plants from which feed was produced were grown. In total, 25 point samples were taken from one plot with an average area of 5 hectares. The field was divided into equal squares of 2.5 hectares. A soil auger and a shovel were used. All point samples were placed in a clean bucket and mixed thoroughly. After that, 1 kg of soil was taken, as an average sample. It was dried at room temperature without heating. Plant residues and stones were removed. The average sample was packed in a paper bag on which the experimental farm, field number, area, sampling depth, crop, and date were indicated.

At the beginning of the experiment, average samples of the following feeds were selected: cereal-legume hay, wheat straw, corn silage, alfalfa haylage, fodder beet, corn husks in silage-horseradish ration; alfalfa hay, wheat straw, corn silage, alfalfa haylage, fodder beet, barley husks in silage-hay; alfalfa hay, cereal-legume hay, corn silage, alfalfa haylage, pea husks in silage-hay; alfalfa hay, cereal-legume hay, corn silage, alfalfa haylage, corn husks, pea husks in silage-hay-concentrate. Feed samples were selected from the entire batch so that one average sample was representative, reflecting the quality of the entire batch. Clean tools, samplers, and shovels were used. Several spot samples were taken from different places and combined into one average sample weighing 1 kg. The sample was placed in a clean dry bag. An appropriate sampling report was drawn up. The litter was taken from different layers: top, middle, bottom. Hay and straw samples were taken with a special probe from different bales of at least 10% of the batch. Samples from different bales were combined and thoroughly mixed to obtain an average sample. Silage and haylage samples were taken from different depths of the trench, avoiding surface or spoiled layers. They were stored in a sealed bag to avoid moisture loss. The accompanying document indicated the name of the feed, manufacturer, batch number, date of sampling, etc.

Analysis of soil and feed samples for heavy metal content was carried out by atomic absorption spectrophotometry (AAS-30 spectrophotometer). To calculate the correlation dependence, only mobile forms of pollutants in the soil were taken into account.

Data analysis was carried out taking into account the features of the results obtained in the study: sample size and type of data distribution, and the nature of dispersions. Correlation dependence was considered statistically significant at $P < 0.05$. Correlation analysis was carried out using the parametric Pearson correlation coefficient taking into account the Kolmogorov-Smirnov and Lilliefors tests for normality and the Shapiro-Wilk's W test. For this purpose, a sample of six (n = 6) results of laboratory chemical analysis of average soil samples for the content of mobile forms of heavy metals in mg/kg and six (n = 6) average feed samples with the content of heavy metals in mg/kg, which were included in the corresponding basic feeding rations of different types of the first control groups, was taken. For the method of modeling the dependence between the variable U (concentration of heavy metals in feed rations) and the vector variable X (concentration of heavy metals in soils of agroecosystems), the construction of linear regression equations was used. The nature of the relationship between the indicators was checked by constructing scatterplots. The analysis of the residuals was checked for compliance with the law of normal distribution (Gaussian). The assessment of the acceptability of the model as a whole by its level of probability was assessed using the ANOVA method. The quality of the calculated regression was checked using the coefficient of determination R^2. It was considered that if R^2 is less than 0.3 (less than 30% – very low level), the model will most likely not work reliably. The calculation was performed in the STATISTICA software package version 10.0. The interpretation of correlation data was carried out on the scale (Gorkavyi, 2019) where r < 0.3 – no relationship, the variation of one variable almost does not explain the other, often has no practical significance; r = 0.3–0.5 – the relationship is weak, weak correlation; r = 0.5–0.7 – the relationship is average (moderate), a significant proportion of the variation is explained; r > 0.7 – the relationship is close, strong correlation, high predictability of one variable by another.

The choice of linear regression is due to the logical method, a linear relationship is expected between the variables according to physicochemical laws. The more mobile forms of heavy metals are contained in the soil, the greater their amount will be absorbed by the root system of plants, the greater the amount of pollutants will pass into the feed. The theory does not assume curvilinearity, so there is no need to complicate the model. Graphical analysis of the data showed an approximate linearity of the relationship between the variables. Linear regression is easier to interpret and more stable for small samples as in our case. From the point of view of practical interpretation, the coefficient in the linear model has a simple explanation. When X increases by 1 unit, Y changes on average by β units.

Results

Exceeding the maximum permissible concentration of mobile forms of cadmium and lead in the soils of agricultural lands of agricultural enterprises caused their transfer to plants, which were used as feed for dairy cows. The probability according to the Shapiro-Wilk's W test for the concentration of cadmium in the soils of agroecosystems is $P = 0.601$, $P = 0.993$, $P = 0.993$, $P = 0.399$, for the content of the pollutant in the feeds of the main diets $P = 0.103$, $P = 0.654$, $P = 0.782$, $P = 0.871$, respectively, which is significantly greater than 0.05 ($P > 0.05$). The Kolmogorov-Smirnov test in all cases is also greater than 0.05 and is $P > 0.20$. The Lilliefors test for normality is also higher than 0.05, as is the Kolmogorov-Smirnov test ($P > 0.20$), except for the feeds of the silage-root crops diet the probability ($P < 0.20$). For lead concentration in agricultural soils, the probability according to the Shapiro-Wilk's W test is $P = 0.315$, $P = 0.953$, $P = 0.855$, $P = 0.286$, for the pollutant content in the main rations, $P = 0.163$, $P = 0.215$, $P = 0.564$, $P = 0.161$, which is also significantly higher than 0.05 ($P > 0.05$). For cadmium concentration, the Kolmogorov-Smirnov test is also in all cases higher than 0.05 and is $P > 0.20$. The Lilliefors test for normality is higher than 0.05, as is the Kolmogorov-Smirnov test ($P > 0.20$). Only in the soil of agricultural lands where cows were kept on cattle farms with a silage-root crops type of feeding, the probability according to the Lilliefors test for normality is $P < 0.20$, and in the soil and feed of the diet where animals were kept with a silage-haylage-concentrate type of feeding, the probability is also $P < 0.20$ and $P < 0.15$, respectively. For all tests conducted, the error is significantly greater than 0.05 ($P > 0.05$), the hypothesis is correct, and the distribution of data that makes up the corresponding samples does not differ statistically from normal. Such a distribution can be considered normal, one that corresponds to the law of normal distribution (Gaussian). Therefore, for the accuracy of correlation and regression analysis, we choose parametric methods. Correlation analysis is performed by the Pearson method, not Spearman, and regression analysis is performed without logarithmic transformation of data.

Table 1 shows the concentration of cadmium and lead in the selected soil samples and feed, as well as the calculated Pearson correlation coefficients between the content of toxic elements under different types of feeding of dairy cows.

The correlation coefficient between the content of heavy metals cadmium and lead in the soils of agroecosystems and feed rations of different types of feeding of dairy cows in all analyzed cases is positive, the relationship is close $r > 0.7$, i.e., there is a strong

Table 1. Content of cadmium and lead in soils of agroecosystems and feeds of basic diets of different types of cow feeding and Pearson correlation coefficient, mg/kg n = 6

Feeding type	Heavy Metal Concentration				Correlation Coefficient	
	Soil		Feed			
	Cadmium	Plumbum	Cadmium	Plumbum	Cadmium	Plumbum
Silage-root crops	0.58	17.52	0.96	28.59	0.902	0.810
	0.5	12.06	0.71	13.62		
	0.43	14.62	0.69	13.51		
	0.42	11.41	0.75	16.78		
	0.37	11.92	0.64	12.22		
	0.56	16.05	0.96	20.51		
Silage-hay	0.55	16.97	0.51	12.5	0.799	0.869
	0.47	14.92	0.39	9.13		
	0.53	13.44	0.45	9.54		
	0.62	18.27	0.75	17.21		
	0.5	16.13	0.57	11.22		
	0.58	14.75	0.54	11.6		
Silage-haylage	0.34	11.07	0.66	18.5	0.859	0.850
	0.27	9.64	0.48	12.5		
	0.39	14.12	0.54	15.5		
	0.44	12.03	0.63	16.51		
	0.34	11.07	0.66	18.5		
	0.27	9.64	0.48	12.5		
	0.39	14.12	0.54	15.5		
	0.44	12.03	0.63	16.51		
Silage-haylage-concentrate	0.51	16.34	0.75	36.51	0.971	0.790
	0.3	10.64	0.42	10.54		
	0.31	13.54	0.48	14.23		
	0.33	11.21	0.54	18.51		
Silage-haylage-concentrate	0.45	11.08	0.69	16.52	0.971	0.790
	0.39	15.07	0.63	20.53		

Note: the probability of correlation $P < 0.05$.
Source: developed by the authors.

correlation and high predictability of one variable for another. The smallest correlation between the indicators where cows were kept with silage-hay type of feeding was established at the level of r = 0.799 for cadmium concentration, and at the level of r = 0.790 for lead concentration, where cows were kept on a silage-haylage-concentrate ration. The probability of the correlation in all cases is P < 0.05 (Figs. 3–4). The highest correlation coefficient was established for the cadmium content, where the animals were fed a silage-haylage-concentrate type diet r = 0.971 (very high (strong) correlation) (P < 0.05), while for the lead content with silage-hay diet r = 0.869 (very high (strong) correlation) (P < 0.05). Therefore, the correlation coefficients for cadmium ranged from 0.799 to 0.971, and for lead from 0.790 to 0.869 (P < 0.05).

Application of the parametric analysis method in the form of the Pearson correlation coefficient using the STATISTICA program allowed us to calculate and thereby reliably (P < 0.05) in all cases establish a close correlation, a strong correlation between the content of toxic HM Cd and Pb in the soils of the Spodar agricultural lands and the feed of the rations of different types of feeding of productive animals, which facilitates the regression analysis.

We checked the nature of the relationship between the two variables of cadmium and lead concentrations. We constructed scatter diagrams (Figs. 1–2). The location of the points on the diagram had an appropriate dispersion across the canvas, there were no clusters or shifts in any one direction, etc. (Figs. 1–2).

We built a histogram of residuals and checked for compliance with the normal distribution law (Figs. 3–4). The residuals were scattered more or less symmetrically (Figs. 3–4). The normality hypothesis was not rejected.

We looked at the normal-probability plot of residuals. No systematic deviations of the actual data from the theoretical normal line were observed. The residuals were distributed normally. We checked the presence or absence of the dependence of the residuals on the predicted values. No specific systematic direction of movement of the points was observed. Otherwise, we had to state the corresponding dependence. The points did not have a systematic arrangement. They were arranged randomly on the diagram canvas, which proves the independence of the residuals from the predicted values (Figs. 3–4). All conditions are met, and the analysis of residuals shows that our model is quite good for this parameter.

We assessed the acceptability of the model as a whole by the ANOVA parameter and the level

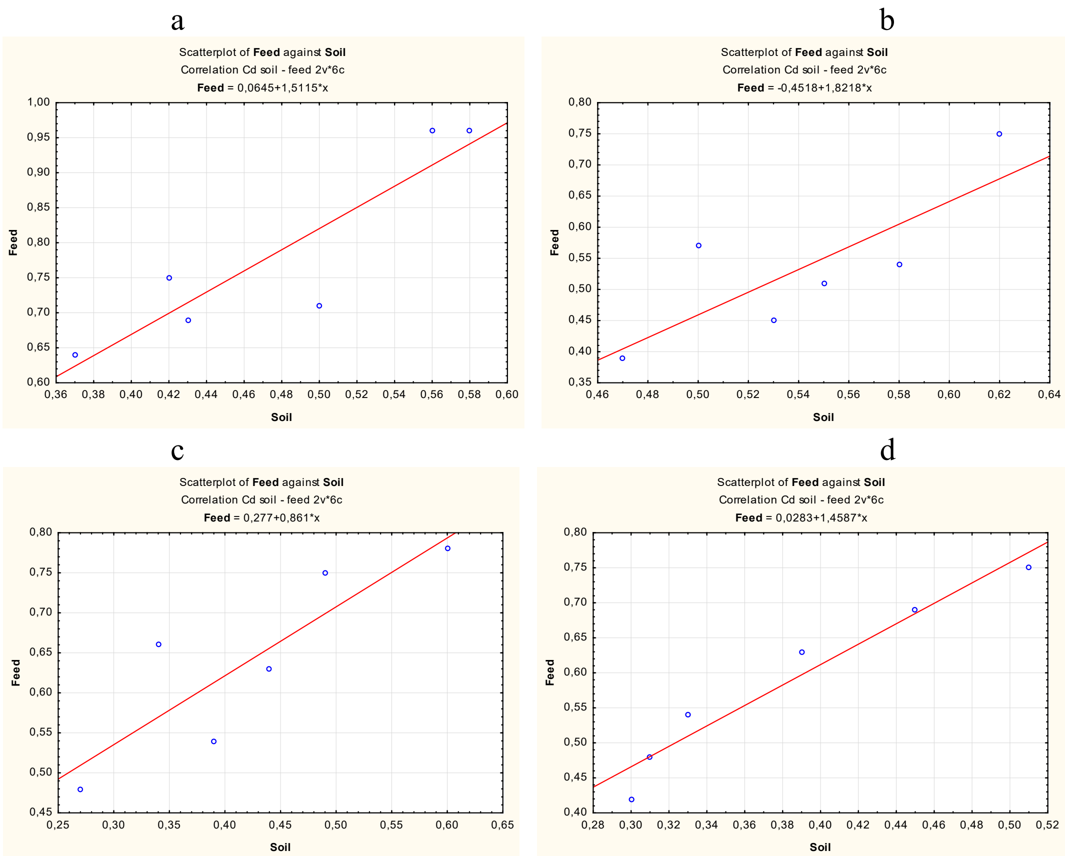

Fig. 1. Scatter diagram for data on Cd content (mg/kg) in agricultural soils and feeds of the main diet with the corresponding regression equation
Types of feeding: a – silage-root crops y = 0.0645 + 1.5115 × x; b – silage-hay y = –0.4518 + 1.8218 × x; c – silage-haylage y = 0.277 + 0.861 × x; d – silage-haylage-concentrate y = 0.0283 + 1.4587 × x.

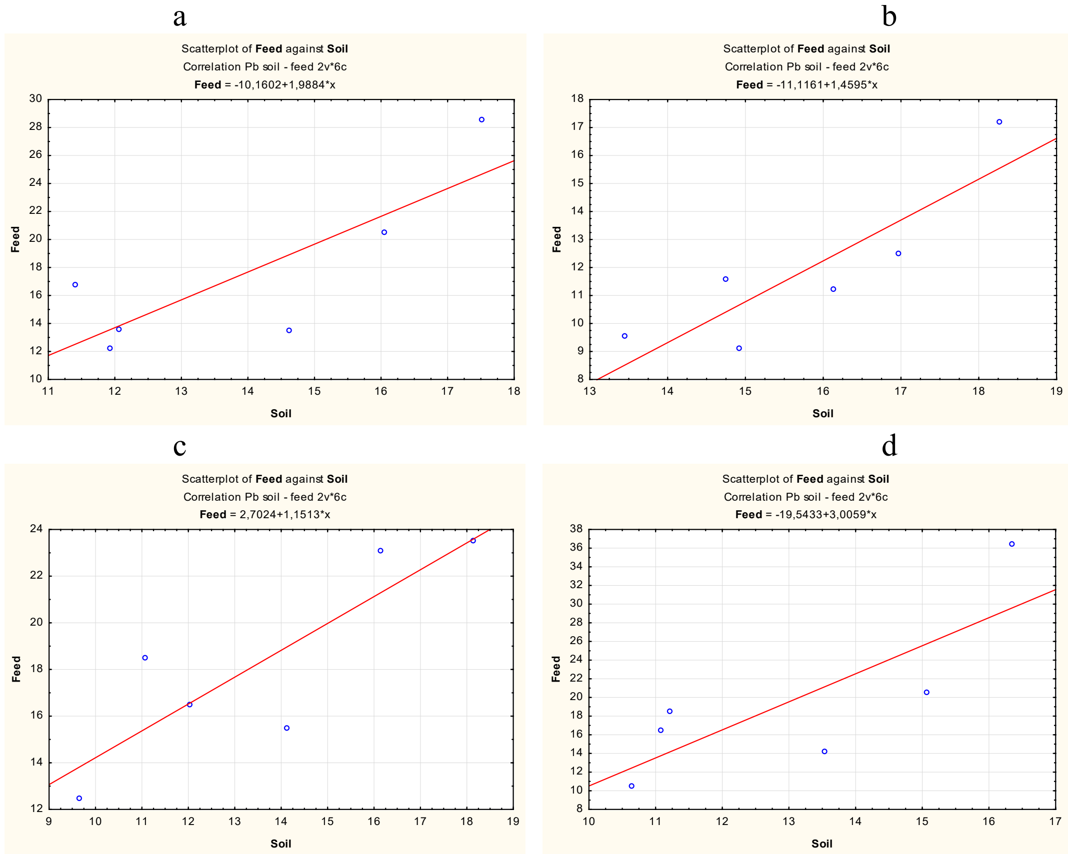

Fig. 2. Scatter diagram for data on Pb content (mg/kg) in agricultural soils and feeds of the main diet with the corresponding regression equation
Types of feeding: a – silage-root crops $y = -10.1602 + 1.9884 \times x$; b – silage-hay $y = -11.1161 + 1.4595 \times x$; c – silage-haylage $y = 2.7024 + 1.1513 \times x$; d – silage-haylage-concentrate $y = -19.5433 + 3.0059 \times x$

of probability. The level of probability for the concentration of cadmium was $P = 0.014$, $P = 0.057$, $P = 0.028$, $P = 0.001$, while for lead it was $P = 0.051$, $P = 0.025$, $P = 0.032$, $P = 0.062$, respectively. The most acceptable model for the concentration of cadmium in animals with silage-root, silage-sage and silage-sage-concentrate feeding type was $P < 0.05$, and for the content of lead, the model in animals with silage-hay and silage-sage feeding type was $P < 0.05$. These models will work much better than the forecast based on average values. All other models do not have a probability level of $P < 0.05$, so the forecast accuracy may be worse.

The coefficient of determination R^2 varies from 0 to 1 and indicates how many factors that affect the response are taken into account in the model. R^2 for Cd concentration was 0.81 (81%) (silage-root crops type of feeding), 0.64 (64%) (silage-hay), 0.74 (74%) (silage-haylage), and 0.94 (94%) (silage-haylage-concentrate). If R^2 is less than 0.3 (less than 30% is a very low level), it will indicate that the model will most likely not work reliably. In our case, there were no such models. Therefore, the probability of predicting the concentration of the pollutant in the feed rations according to the corresponding regression equations is high. R^2 for Pb concentration is 0.66 (66%) for silage-root crops type of feeding, 0.76 (76%) for silage-hay, 0.72 (72%) for silage-haylage, and 0.62 (62%) for silage-haylage-concentrate. For lead content, R^2 less than 0.3 (less than 30% is a very low level) was also not found. The number of factors taken into account when building the model is high. This means that all models will work with sufficient probability. All probability levels that are $P < 0.05$ are acceptable for building the model. In our case, for Cd concentration, $P < 0.014$ (acceptable) for silage-root type of feeding, $P < 0.057$ (unacceptable) for silage-hay type of feeding, $P < 0.028$ (acceptable) for silage-haylage type of feeding, $P < 0.001$ (acceptable) for silage-haylage-concentrate type of feeding. For Pb, the values were $P < 0.051$ (unacceptable), $P < 0.025$ (acceptable), $P < 0.032$ (acceptable), and $P < 0.062$ (unacceptable) for silage-haylage-concentrate type of feeding, respectively.

Discussion

The consequences of heavy metal contamination of the trophic chain are critical. Pollutants, getting into the soil, are actively absorbed by plants, accumulate in root crops, cereals and leafy vegetables. Contaminated feed becomes a source of heavy metal accumulation in animal products, such as meat, milk and eggs, which poses a threat to food security and public health, as noted by Gulich et al. (2025), with

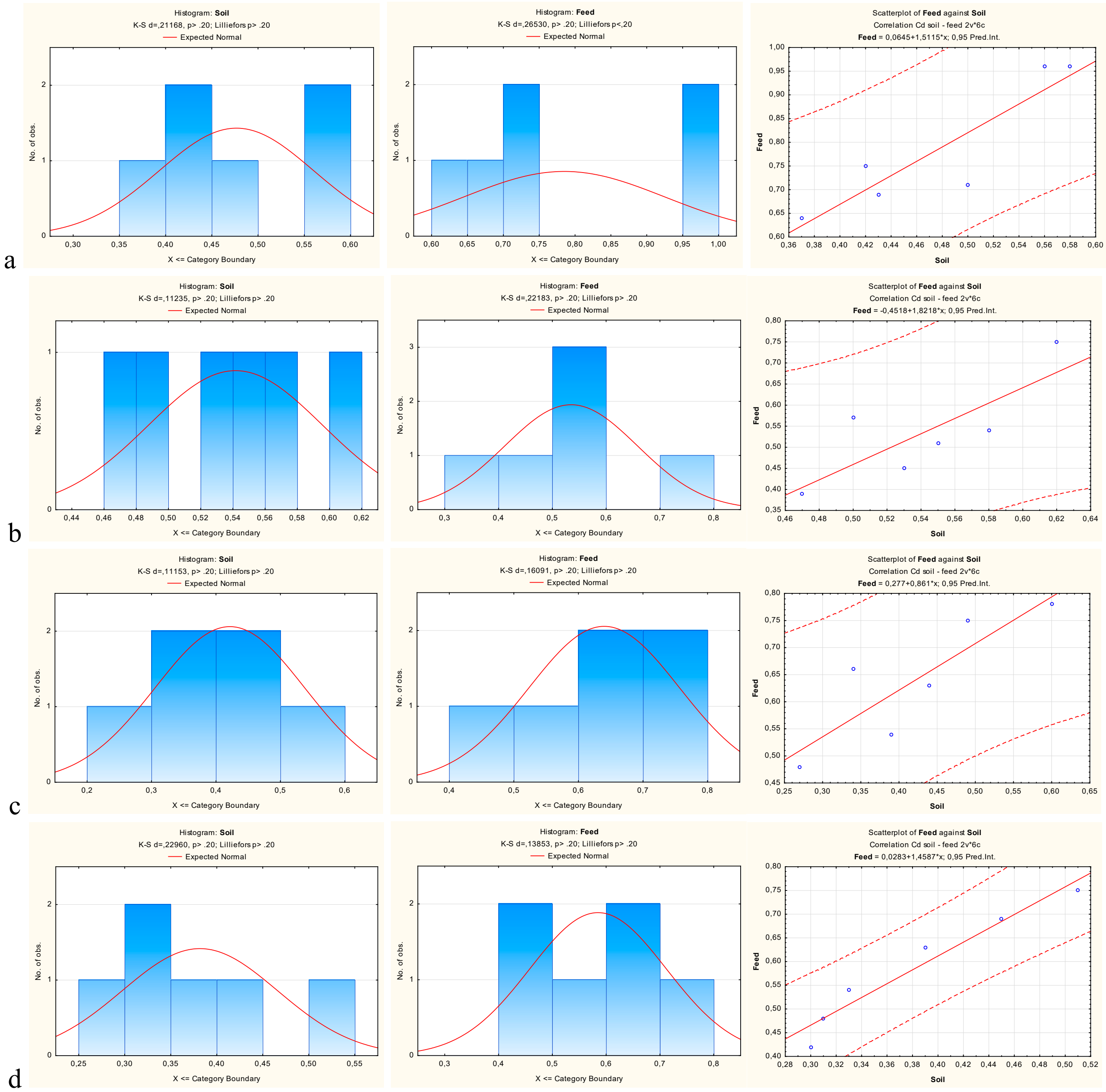

Fig. 3. Analysis of residues by Cd concentration (mg/kg) in soils of agroecosystems and feeds of rations of different types of cow feeding
types of feeding: a – silage-root crops; b – silage-hay; c – silage-haylage; d – silage-haylage-concentrate.

which we also agree (Portiannyk et al., 2024b). The study of heavy metals as a factor of environmental pollution during war is extremely important, since they constitute one of the most dangerous groups of pollutants. Military operations significantly increase the anthropogenic load on ecosystems, leading to the accumulation of toxic elements in soil, water and biological objects. In the post-war period, the accumulation of heavy metals in the soils of various regions of Ukraine may significantly exceed safety standards; global warming and climate change also lead to an increase in the concentration of heavy metals in the soils of ecosystems (Xiao et al., 2024; Wrzecińska et al., 2021). Agricultural lands in the pre-war period were subjected to significant technogenic pollution, while today's hostilities only increase the pollution with heavy metals. Mining, air defense work, when unmanned aerial vehicles and missiles crash mainly somewhere in the field, rather than over residential buildings, avoiding a threat to human life, ultimately lead to the dispersion and accumulation of mercury, cadmium, lead, copper, zinc, chromium and other HMs in the soil of agricultural lands.

Measuring environmental parameters in the broadest sense – monitoring, sampling, laboratory analysis, etc. – often entails significant challenges. Significant costs, time constraints, inherent uncontrolled conditions of on-site measurements can lead to insufficient and/or low-quality data with a small number of repetitions, which needs to be taken

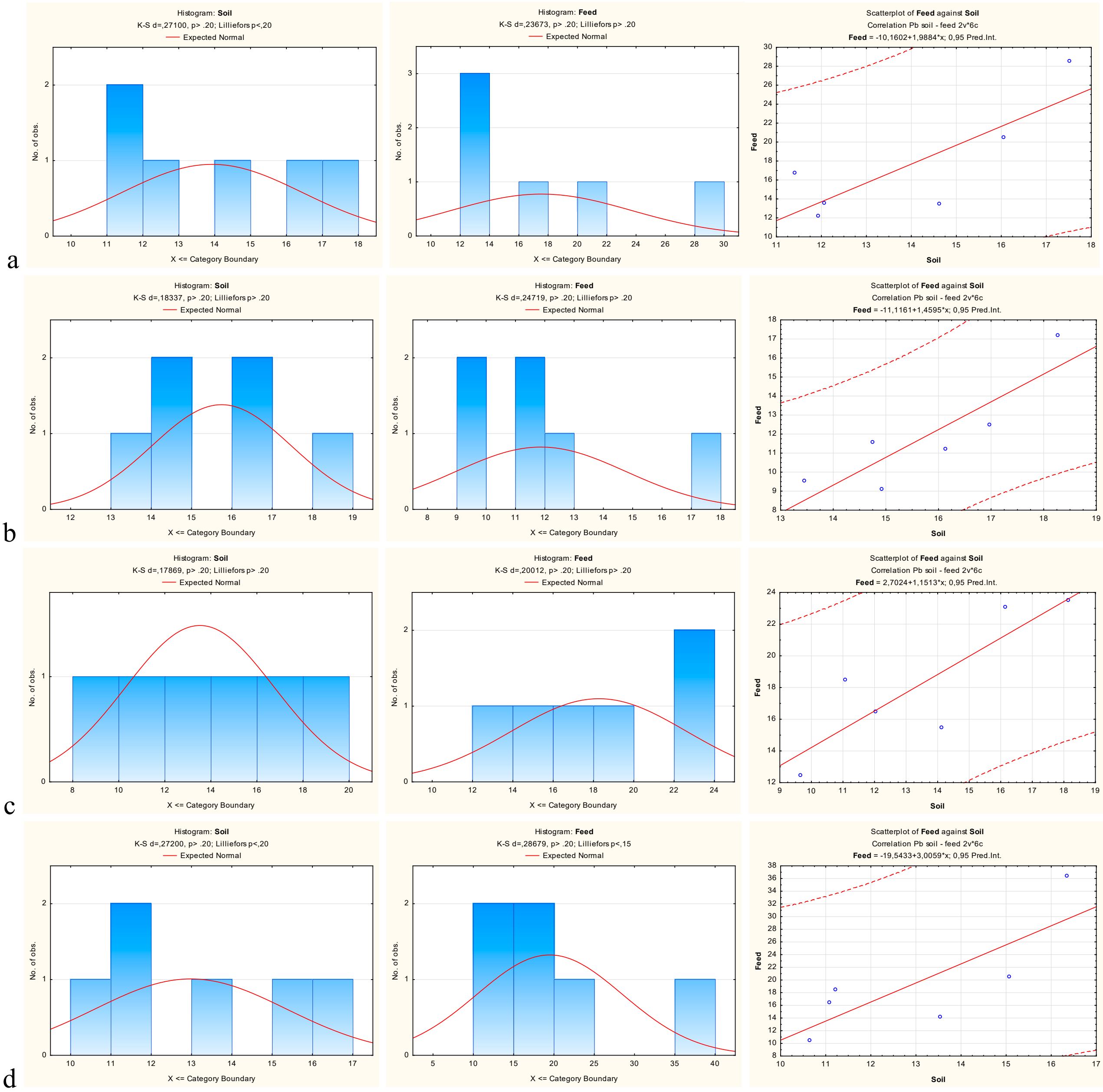

Fig. 4. Analysis of residues by Pb concentration (mg/kg) in soils of agroecosystems and feed rations of different types of cow feeding
Types of feeding: a – silage-root crops; b – silage-hay; c – silage-haylage; d – silage-haylage-concentrate.

into account by researchers. In general, methods based on correlation and regression aim to go beyond (usually graphical) evidence of a certain relationship between variables and, possibly, to consolidate them, with different goals: from quantifying the degree of the relationship (for example, using the correlation coefficient) to developing a predictive model. A number of methods have been developed for the above-mentioned goals, which, like any statistical model, are based on their own fundamental assumptions, and are now widely used and implemented in most existing data analysis tools. For example, Damien et al. (2025) in their work consider a number of important methodological approaches that researchers need to take into account in order to develop qualitative models during correlation and regression data analysis. Analyzing a large number of publications related to correlation and regression analysis, researchers come to the conclusion that most environmental scientists face the inherent problem of establishing certain evidence about complex processes, which are often estimated using limited, partial and/or uncertain data. To this end, methods inherited from statistical modeling open up a wide range of possibilities, overcoming the mechanistic description of phenomena when this is practically impossible to implement. Therefore, the introduction of correlation and regression approaches in environmental research is understandable and in many cases fully justified. It is also likely that these approaches have been facilitated by the democratization of statistical tools and software, which facilitate data processing and analysis. The STATISTICA program we used allows not only performing all statistical and analytical calculations, but also using it for convenience in practice. Of course, it is possible to use other similar programs, which are many. It is also recommended not to violate the assumption of normality for data testing. When using and distributing the model, in order to avoid unnecessary confusion, researchers (Damien et al., 2025) recommend employing the coefficient of determination when the goal is prediction, and correlation when describing the relationships between variables. The conclusion in regression is based on the theoretical properties of the residuals, which should be checked, which we took into account in the process of this analysis.

Researchers widely use methods of correlation and regression analysis of research results. Various computer programs and analysis algorithms are used. Not only the analysis of the strength of the correlation between indicators is important, but also the correct choice of indicators themselves. Correlation analysis ha been used by researchers to identify the relationship between HM in marine sediments of Hong Kong (Lili et al., 2015). The Spearman rank correlation was used in the first stage, followed by the Pearson correlation in the second. The bay and harbor (Victoria) were polluted with Pb and Cu, which correlated with Cr, Ni and Zn pollution (Lili et al., 2015). The researchers believe that they were the first to apply such analysis to the study of environmental pollution. In addition, such methods have already been widely used in biology, sociology, computer science, etc. We believe that correlation and regression analysis can be applied not only to marine sediments or any other environmental studies. In our case, the importance of such analysis increases due to the complex, multifaceted processes occurring in agroecosystems, where living organisms, especially dairy cows, play a major role in producing environmentally safe dairy products; meanwhile, the main criteria remain quality and accuracy, which we ensured by incorporating the methodological developments of other scientists as much as possible.

We carried out the entire analysis in stages. First, we established the correspondence of the research results to the law of normal distribution (Gaussian), and then applied the powerful and reliable Kolmogorov-Smirnov and Lilliefors test for normality and Shapiro-Wilk's W test. These tests are effective for small samples, as in our case ($n = 6$). According to all the tests performed, the error $P > 0.05$. The hypothesis is correct, the data distribution does not statistically differ from normal, i.e., it obeys the law of normal distribution (Gaussian). Next, we applied the Pearson parametric method and calculated the correlation coefficients between the content of cadmium and lead in agricultural soils and feed rations of different types of feeding of productive animals. In all cases, the coefficient is positive $r > 0.7$, which indicates a strong (high) correlation, and the probability of correlation in all cases is $P < 0.05$. The correlation coefficients ranged from 0.799 to 0.971 for Cd, and from 0.790 to 0.869 for Pb ($P < 0.05$). The results of our correlation analysis are consistent with other studies (Sachko et al., 2016), which also established a high correlation at the level of 0.80–0.93 between the content of HM in soil and forages, in particular meadow hay and wheat straw.

Before conducting the regression analysis, we checked the nature of the relationship between the two variables for both cadmium and lead. The constructed scatter plots do not show any clusters in any one direction. Residual analysis is important in regression analysis. Therefore, we constructed a histogram of the residuals and checked for compliance with the normal distribution law. The normality hypothesis was not rejected. The residuals were normally distributed. It is important that there is no dependence of the residuals on the predicted values, so the model is quite good and we can further assess its acceptability as a whole. Taking into account the methodology (Damien et al., 2025), checking the significance of the model is of great importance. We applied the ANOVA parameter. It turned out that the most acceptable model for predicting the transfer of the element from soil to feed in animals with silage-root crops, silage-haylage and silage-haylage-concentrate diets for cadmium concentration was with silage-hay and silage-haylage diets for lead ($P < 0.05$). Other models do not have such a level of probability. Therefore, their application may have a greater error in use. We performed an analysis of R^2 determination, which is recommended (Damien et al., 2025; Boldizsar et al., 2017). The biosorption characteristics of Cd and Zn ions from monocomponent aqueous solutions by the macromushroom Agaricus bisporus were studied. Initial concentration of metal ions, contact time, initial pH and temperature were the parameters that affected biosorption. Experimental data were analyzed using kinetic models, and various models were calculated in linear and nonlinear (CMA-ES optimization algorithm) regression. Langmuir and Freundlich coefficients were calculated using linear regression for Cd and Zn biosorption on Agaricus bisporus macrofungus biomass. Boldizsar et al. (2017) provide the corresponding graphs that indicate the R^2 determination index and the corresponding regression equation. In their experiment, R^2 is very high, 97–99%. In our case, R^2 for Cd concentration ranged from 0.64 to 0.94 (64–94%) and Pb from 0.62 to 0.76 (62–76%). However, we do not have models with a very low coefficient of determination R^2 below 30% in our analysis. This means that all models will work with sufficient probability. In terms of probability, the models for cadmium concentration for silage-root crops feeding type ($P < 0.014$), silage-haylage

($P < 0.028$), silage-haylage-concentrate ($P < 0.001$), lead in silage-hay ($P < 0.025$) and silage-haylage ($P < 0.032$) are acceptable.

After constructing the regression equations, we tested them on examples to verify the accuracy and effectiveness of the developed models. Laboratory analysis of selected soil samples from agricultural lands where plants are grown for fodder and bark is used in the rations of different types of feeding of dairy cows established the concentration of cadmium at 0.48 mg/kg; however, the concentration in the feed is unknown. Animals are fed the silage-root crops type of feeding. We substituted this indicator into the program, or the corresponding regression equation $y = 0.0645 + 1.5115 \times x$ and obtained the result $0.790 = 0.0645 + 1.5115 \times 0.48$. The predicted concentration of Cd in the feed of the corresponding type of diet is 0.790 mg/kg. This result is consistent with the actual results of laboratory analysis of the feed given in the original tabular data. We also checked all other forecast models, including Pb. We checked the forecast of the concentration of heavy metal Cd in the feed of the silage-haylage diet. The concentration of the pollutant in soil samples was 0.44 mg/kg ($0.350 = -0.4518 + 1.8218 \times 0.44$). The predicted concentration of Cd in the feed of the ration that will be fed to dairy cows according to the corresponding type of feeding is 0.350 mg/kg. This result is consistent with the actual results of laboratory analysis of feeds given in the original tabular data. Next, we checked the forecast of the concentration of Cd in the feed of the silage-haylage-concentrate type of diet. The concentration of the pollutant in soil samples was 0.52 mg/kg ($0.725 = 0.277 + 0.861 \times 0.52$). The predicted concentration of Cd in the feed of this cow ration was 0.725 mg/kg. This result is consistent with the actual results of laboratory analysis of the feed of the ration given in the original tabular data. We checked the forecast of the concentration of Cd in the feed of the ration of dairy cows with the silage-hay-concentrate type of feeding. The concentration of the pollutant in the soil samples was 0.37 mg/kg ($0.568 = 0.0283 + 1.4587 \times 0.37$). The predicted concentration of Cd in the cow feed was 0.568 mg/kg. This result is consistent with the actual concentration obtained after laboratory analysis of feed, which is given in the original tabular data. Similarly, we checked the values for lead. Chemical analysis of soil samples from agroecosystems where plants were grown for feed included in the diets of various types of feeding of dairy cows, established the concentration of lead at 13.45 mg/kg; however, the concentration in the feed rations is unknown. The animals were fed according to the silage-root crops type of feeding. We substituted this indicator into the program, or the corresponding regression equation $y = -10.1602 + 1.9884 \times x$ and obtained the result $16.584 = -10.1602 + 1.9884 \times 13.45$. The predicted concentration of Pb in the feed of the diet of the corresponding type was 16.584 mg/kg (the result agrees). We checked the prediction of the concentration of Pb in the feed of the diet of cows with the silage-hay type of feeding. The concentration of the pollutant in the soil samples was 12.32 mg/kg ($6.865 = -11.1161 + 1.4595 \times 12.32$). The predicted concentration of Pb in the feed of the diet of this type was 6.865 mg/kg (the result is consistent). Next, we checked the prediction of the concentration of Pb in the feed of the diet of cows with the silage-haylage feed type. The concentration of the pollutant in the soil samples was 13.67 mg/kg ($18.440 = 2.7024 + 1.1513 \times 13.67$). The predicted concentration of Pb in the feed of the diet of the corresponding type of feed was 18.440 mg/kg (the result is consistent). We checked the prediction of Pb concentration in the feed of the diet of cows with silage-haylage-concentrate type of feeding. The concentration of the pollutant in soil samples was 12.24 mg/kg ($17.249 = -19.5433 + 3.0059 \times 12.24$). The predicted concentration of Pb in the feed of the diet of this type was 17.249 mg/kg and the result is consistent with the actual concentration obtained after laboratory analysis of the feed, which is given in the original tabular data.

We checked all the regression equations we calculated by introducing, as an example, instead of X, the concentration of cadmium or lead in soil samples, conditionally known from laboratory analysis, and obtained the corresponding predicted concentration of the pollutant in the feed of rations of different types of cow feeding. The calculation was carried out in two ways: simple mathematical calculation and using the STATISTICA program, which includes a built-in function for this purpose. All regression equations showed an acceptable result, but it is more necessary to apply those models where high probability is established. In the conditions of our model, only two parameters are set: the first is the concentration of cadmium and lead in soil samples of agroecosystems and the second is the concentration of cadmium and lead in the feed of different animal rations. Model testing using the Kolmogorov-Smirnov and Lilliefors test for normality and Shapiro-Wilk's W test, Pearson parametric correlation analysis, construction of scatter plots, residual analysis, assessment of model acceptability using the ANOVA parameter and the R^2 coefficient of determination allows us to select the most effective and most accurate regression equations for forecasting. After the comprehensive analysis and at the final stage of the corresponding verification, we recommend using the constructed regression equations for the concentration of Cd in the silage-root crops diet $y = 0.0645 + 1.5115 \times x$, silage-hay $y = -0.4518 + 1.8218 \times x$, silage-haylage $y = 0.277 + 0.861 \times x$, silage-haylage-concentrate $y = 0.0283 + 1.4587 \times x$, and the following for the concentration of Pb $y = -10.1602 + 1.9884 \times x$, $y = -11.1161 + 1.4595 \times x$, $y = 2.7024 + 1.1513 \times x$,

$y = -19.5433 + 3.0059 \times x$, respectively, as a tool for environmental monitoring, forecasting (modeling) the transition of toxic HM cadmium and lead in feed rations of various types of feeding of dairy cows on cattle farms.

Multiple linear regression was used (Weibin et al., 2022) in the study of HM pollution in an industrial park. Researchers have

created a method that combines principal component analysis, geodetector, and multiple linear regression of the distance to the sources of pollutant emissions, which affects the level of soil pollution. The main sources of pollution with specific elements were identified by researchers using regression equations. Identification of potential sources of pollution is the basis of environmental risk control. Analysis of sources of HM soil pollution is one of the most important problems in recent years (Yanxue et al., 2017; Xufeng et al., 2020; Geng et al., 2020; Weibin et al., 2022) as well as the problem of producing environmentally safe milk. Milk and dairy products are important in human nutrition. For example, milk from animals fed on feed grown on agricultural land irrigated with wastewater poses a risk for Cd, Cr, Cu, Mn, Ni and Pb (Sena et al., 2023). A study conducted in the Algerian region analyzed 88 raw milk samples obtained from local animal breeds. According to the results of the study, there is a potential risk of HM contamination, especially Pb, for infants due to the consumption of raw milk (Boudebbouz et al., 2023). Another study conducted in Egypt analyzed a total of 75 dairy products, including raw cow milk, raw buffalo milk, condensed milk, infant formula and powdered milk. The results showed that the levels of toxic metals Pb and Cd were higher than the recommended daily intake (RDI) (Sena et al., 2023). In another study comparing heavy metal levels in cow and buffalo milk, the mean concentrations of Pb, Cd, Cu, and Ni in cow milk samples were 0.62 ± 0.25, 0.25 ± 0.22, 0.31 ± 0.20, and 21 ± 2 mg/kg, respectively, while in buffalo milk, they were 0.60 ± 0.3, 0.33 ± 0.15, 0.27 ± 0.11, and 18 ± 2.5 mg/kg, respectively. The researchers attributed the situation to excessive environmental contamination with heavy metals (Al-Rudha et al., 2021). Ribeiro et al. (2021) noted that dairy products from the market in San Luis were a source of heavy metal (Hg, Pb, Se, Cu, Ni) contamination (Ribeiro et al., 2021; Sena et al., 2023). Therefore, we believe that one of the main tasks of cow milk production specialists is the timely and high-quality prediction of the level of contamination of HM feeds and the environmental safety of the produced dairy raw materials according to different types of animal feeding. The regression models we developed with the appropriate correlation analysis will help specialists in production conditions.

The problem of ecosystem pollution is associated with the toxicity of HMs, which have a harmful effect on living organisms. The determining indicator of the negative impact of chemicals is concentration. It is the concentration of pollutants that we included in our regression model. The highest priority for chemical-toxicological analysis is Pb and Cd, which exhibit high toxicity and migration ability. This has been proven by our studies, including previous research on predicting the transition of HMs from feed rations to organic waste (Portyannik, 2024a) and studies by other scientists. During the experiment, we determined the concentration of copper and zinc in the soil and feed rations, but in this publication we chose the most ecocidal and dangerous Pb and Cd for analysis. In subsequent works, analysis of copper and zinc will be performed. The behavior of ecotoxicants in various natural environments is determined by their specificity and corresponding properties: toxicity, accumulation, mobility, complexation, etc.

Thus, having established a correlation and obtained the corresponding regression equations using a mathematical model, it is possible to predict the corresponding ecological changes and assess the ecological risks in this case of contamination of animal rations with BM in the process of environmentally safe milk production technology. Our research took into account the experience and methodology of both domestic and foreign researchers. The regression equations calculated and tested by us will help to predict the concentration of BM in the rations of different types of feeding of dairy cows and, therefore, to predict the risk of contamination of produced milk with pollutants depending on the amount of toxicants that will enter the animal body with the feed. In the post-war period, there will be a restoration of agricultural production in Ukraine, an increase in the number of cattle, including dairy cows; milk production will increase significantly. Therefore, modeling the body load of dairy cows that will arrive with feed rations in agroecosystems that have been subjected to technogenical or military pollution will be of great importance, especially in the conditions of organic-biological farming and production of environmentally safe livestock and crop products. This study has some limitations, including a small sample size and the lack of independent validation of the model. Future studies may address these limitations.

Conclusions

A positive close relationship and a high correlation between the concentration of cadmium and lead in the soils of agroecosystems and the feed of dairy cows with silage-root crops, silage-hay, silage-haylage, silage-haylage-concentrate type of feeding was established. A very high correlation was found for the cadmium content when feeding animals with silage-haylage-concentrate rations and lead in silage-hay. To predict the concentration of toxic heavy metals in the feed of diets of different types of dairy cows' feeding and to assess the environmental safety of the feed factor, the corresponding linear regression equations were constructed. The most acceptable models that will give the most accurate prediction result for cadmium in cows on silage-root crops, silage-haylage and silage-haylage-concentrate type of feeding, and for lead in silage-hay and silage-haylage.

Further research is aimed at conducting correlation and regression analysis between the content of heavy metals in feeds and cow milk with an assessment of environmental safety and risks of dairy farming in the Forest-Steppe soil-climatic zone of Ukraine.

References

1. Al-Rudha, A. M., Khalil, N. K., & Altaai, N. A. (2021). Evaluation of bacterial contaminants and heavy metals in cow and buffalo raw milk sold in Baghdad governorate. Iraqi Journal of Veterinary Sciences, 35 (Supplement I-III), 101-105. https://doi.org/10.33899/ijvs.2021.131744.1999
2. Boudebbouz A, Boudalia S, Bousbia A, Gueroui Y, Boussadia MI, Chelaghmia ML, Zebsa R, Affoune AM, Symeon GK. (2023). Determination of Heavy Metal Levels and Health Risk Assessment of Raw Cow Milk in Guelma Region, Algeria. Biol Trace Elem Res. Apr; 201(4):1704-1716. Epub 2022 Jun 6. PMID: 35666385. https://doi.org/10.1007/s12011-022-03308-1
3. Boldizsar Nagy, Carmen Mânzatu, Andrada Măicăneanu, Cerasella Indolean, Lucian Barbu-Tudoran, Cornelia Majdik. (2017). Linear and nonlinear regression analysis for heavy metals removal using Agaricus bisporus macrofungus, Arabian Journal of Chemistry, Volume 10, Supplement 2, Pages S3569-S3579, ISSN 1878-5352, https://doi.org/10.1016/j.arabjc.2014.03.004
4. Damien Tedoldi, Boram Kim, Santiago Sandoval, Nicolas Forquet, Bruno Tassin. (2025). Position paper: Common mistakes and solutions for a better use of correlation- and regression-based approaches in environmental sciences, Environmental
5. Modelling & Software, Volume 192, 106526, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2025.106526
6. Geng Liu, Ying Shi, Guanlin Guo, Long Zhao, Junjie Niu, Chao Zhang. (2020). Soil pollution characteristics and systemic environmental risk assessment of a large-scale arsenic slag contaminated site, Journal of Cleaner Production, Volume 251, 119721, ISSN 0959-6526, https://doi.org/10.1016/j.jclepro.2019.119721
7. Gulich M. P., Petrenko O. D., Kharchenko O. O. (2025). Determination of heavy metal content in agricultural products
8. from regions affected by hostilities: first results of the study. Environment & Health, №1, 24-30. https://doi.org/10.32402/dovkil2025.01.024
9. Gorkavyi V. K. (2019). Statistics: Textbook. Third edition, revised and supplemented / V. K. Gorkavyi. K.: Alerta, 644 p.
10. Kuramshina, N.G., Kuramshin, E.M., Nikolaeva, S.V., Imashev, Y.B. (2014). The biogeochemical characteristics of the content of heavy metals in soil, plants and animals in different natural areas of Bashkortostan, Journal of Geochemical Exploration, Volume 144, Part B, Pages 237-240. DOI: https://doi.org/10.1016/j.gexplo.2014.01.027.
11. Lili Liu, Zhiping Wang, Feng Ju, Tong Zhang. (2015). Co-occurrence correlations of heavy metals in sediments revealed using network analysis, Chemosphere, Volume 119, Pages 1305-1313. https://doi.org/10.1016/j.chemosphere.2014.01.068
12. Mingtao Xiang, Yan Li, Jiayu Yang, Kaige Lei, Yi Li, Feng Li, Daofu Zheng, Xiaoqian Fang, Yu Cao. (2021). Heavy metal contamination risk assessment and correlation analysis of heavy metal contents in soil and crops, Environmental Pollution, Volume 278, 116911, ISSN 0269-7491, https://doi.org/10.1016/j.envpol.2021.116911
13. Portyannik S.V. (2024a). Using regression analysis in modeling the ecological situation by predicting the transfer of toxic heavy metals from the feed of dairy cows to organic waste and the accumulation of pollutants in the soil of agricultural livestock enterprises. Ukrainian Journal of Natural Sciences, Ivan Franko Zhytomyr State University. Zhytomyr, No. 7. P. 210-228. https://doi.org/10.32782/naturaljournal.7.2024.23
14. Ribeiro Sant'Ana, M. A., de Carvalho, T. C., & da Silva, I. F. (2021). Concentration of heavy metals in UHT dairy milk available in the markets of São Luís, Brazil, and potential health risk to children. Food chemistry, 346, 128961. https://doi.org/10.1016/j.foodchem.2020.128961
15. Sachko, R., Lesyk, J., Luchka, I., & Nevostruyeva, I. (2016). Contents of heavy metals in food, organism and animal products in the zacarpathian biogeochemical province. Scientific Messenger of LNU of Veterinary Medicine and Biotechnologies. Series: Veterinary Sciences, 18(3(71), 87-90. https://doi.org/10.15421/nvlvet7120
16. Saurabh Shukla, Grace Mbingwa, Sakshum Khanna, Jyoti Dalal, Deeksha Sankhyan, Anindita Malik, Neha Badhwar. (2023). Environment and health hazards due to military metal pollution: A review, Environmental Nanotechnology, Monitoring & Management, Volume 20, 100857, ISSN 2215-1532, https://doi.org/10.1016/j.enmm.2023.100857
17. Sazal Kumar, Meixia Zhao, Haiyang Zhang, Md Aminur Rahman, Chuanxiu Luo, Mohammad Mahmudur Rahman. (2021). Distribution, contamination status and source of trace elements in the soil around brick kilns, Chemosphere, Volume 263, 2021, 127882, ISSN 0045-6535, https://doi.org/10.1016/j.chemosphere.2020.127882
18. S. M. Ashraful Abedin Asha, Md. Moshfequr Rahman, Md. Tahmid Tayef, Md. Reazul Islam, Anti Sorker, Sazal Kumar. (2023). Comprehensive pollution and ecological risk of heavy metals in an industrial region of south-west Bangladesh, Environmental Nanotechnology, Monitoring & Management, Volume 20, 100899, ISSN 2215-1532, https://doi.org/10.1016/j.enmm.2023.100899
19. Serhii Portiannyk, Oleksiy Mamenko, Valentyn Rybalko, Andriy Onyshchenko. (2024b). Application of Effective Technological Methods for the Production of Environmentally Safe Cow's Milk, Vet Med Zoot 82, no. 2, 70-80, ISSN 1392-2130 (Print) ISSN 2669-2511 (Online) https://doi.org/10.5281/zenodo.15718492
20. Sena Özbay, Emrah Dikici, Caner Soylukan. (2023). Evaluation of biological (feed, water), seasonal, and geological factors affecting the heavy metal content of raw milk, Journal of Food Composition and Analysis, Volume 121, 105401, ISSN 0889-1575, https://doi.org/10.1016/j.jfca.2023.105401
21. Weibin Zeng, Xiaoming Wan, Lingqing Wang, Mei Lei, Tongbin Chen, Gaoquan Gu. (2022). Apportionment and location of heavy metal(loid)s pollution sources for soil and dust using the combination of principal component analysis, Geodetector, and multiple linear regression of distance, Journal of Hazardous Materials, Volume 438, 129468, ISSN 0304-3894, https://doi.org/10.1016/j.jhazmat.2022.129468
22. Wrzecińska, M., Kowalczyk, A., Cwynar, P., & Czerniawska-Piątkowska, E. (2021). Disorders of the Reproductive Health of Cattle as a Response to Exposure to Toxic Metals. Biology, 10(9), 882. https://doi.org/10.3390/biology10090882
23. Xufeng Fei, Zhaohan Lou, Rui Xiao, Zhouqiao Ren, Xiaonan Lv. (2020). Contamination assessment and source apportionment of heavy metals in agricultural soil through the synthesis of PMF and GeogDetector models, Science of The Total Environment, Volume 747, 141293, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2020.141293
24. Xiao, W., Zhang, Y., Chen, X., Sha, A., Xiong, Z., Luo, Y., Peng, L., Zou, L., Zhao, C., & Li, Q. (2024). The Easily Overlooked Effect of Global Warming: Diffusion of Heavy Metals. Toxics, 12(6), 400. https://doi.org/10.3390/toxics12060400
25. Yanxue Jiang, Sihong Chao, Jianwei Liu, Yue Yang, Yanjiao Chen, Aichen Zhang, Hongbin Cao. (2017). Source apportionment and health risk assessment of heavy metals in soil for a township in Jiangsu Province, China, Chemosphere, Volume 168, Pages 1658-1668, ISSN 0045-6535, https://doi.org/10.1016/j.chemosphere.2016.11.088