

CORRELATION ANALYSIS IN THE PRE-MODELING STAGE OF HEAVY METALS MIGRATION INTO THE SOIL

UTILIZAREA ANALIZEI DE CORELAȚIE ÎN ETAPA DE PREMODELARE A MIGRĂRII METALELOR GRELE ÎN SOL

George Călin ROGOZAN*, Valer MICLE, Dumitru Dan PORCAR

Technical University of Cluj-Napoca, Faculty of Materials and Environmental Engineering, Department Environmental Engineering and Sustainable Development Entrepreneurship, 103-105 Muncii Ave, Cluj-Napoca, Romania

Abstract: Use of mechanistic mathematical models (consisting of second-degree differential equations with partial derivatives) in the prevention and control actions of heavy metals migration into the soil is an extremely difficult approach. More suitable for this purpose are empirical models obtained by statistical processing of experimental data. Given, however, the high cost of experimentation and time consumption, the number of such experiments should be reduced to the minimum required. An important way to reduce is the simplification of the model equations by retaining only those process parameters with decisive influence on the state variable. Using multiple correlation analysis (in antithesis with the simple correlation analysis), applied to a sample of 150 geological sites, this paper has allowed a classification of soil physicochemical characteristics by the major influence they have on the heavy metals migration phenomenon. This hierarchy can be used, further, to a rigorous selection of variables for a future mathematical model.

Keywords: correlation coefficients, heavy metals, soil characteristics, statistical modeling.

1. Introduction

One of the anthropogenic aggressions exerted constantly by the soil is the heavy metals pollution. But is it a special situation because these metals can also be naturally found in the soil, some of them (such as: copper, lead, cobalt or manganese) having an important role in plant growth [1]. For this reason, specialists were asked in the first instance to find an answer for each of the three fundamental questions related to this topic: what is the threshold value above which soil is considered polluted with a certain heavy metal; what locations are expected to have high natural concentrations for these elements; and

Rezumat: Utilizarea modelelor matematice mecaniciste (formate din ecuații diferențiale de gradul doi cu derivate parțiale) în acțiunile de prevenire și control a migrației metalelor grele în sol este o abordare extrem de dificilă. Mai potrivite pentru acest scop sunt modele empirice obținute prin prelucrarea statistică a datelor experimentale. Având în vedere, însă, costul ridicat și consumul de timp, volumul acestor experimente ar trebui să fie redus la minimum. O modalitate importantă de reducere a acestor costuri este simplificarea modelului, reținând în ecuațiile sale numai acei parametri de proces cu o influență decisivă asupra variabilei de stare. Folosind analiza de corelație multiplă (în antiteză cu analiza de corelație simplă), aplicată pe un eșantion format din 150 de situri geologice, această lucrare a permis o clasificare a caracteristicilor fizico-chimice ale solului cu influență majoră asupra fenomenului de migrație a metalelor grele. Această ierarhizare poate fi utilizată, în continuare, la o selecție riguroasă a variabilelor pentru un viitor model matematic.

Cuvinte cheie: coeficienți de corelație, metale grele, caracteristicile solului, modelare statistică.

1. Introducere

Una dintre agresiunile antropice, exercitate în mod constant asupra solului, o reprezintă poluarea cu metale grele. Dar fenomenul în sine constituie o situație deosebită, deoarece aceste metale se regăsesc și în mod natural în sol, unele dintre ele (cum ar fi: cuprul, plumbul, cobaltul sau manganul) jucând chiar un rol important în creșterea plantelor [1]. Din acest motiv, specialiștii au fost solicitați, într-o primă instanță, de a găsi un răspuns adecvat pentru fiecare dintre cele trei întrebări fundamentale legate de acest subiect: care este valoarea de prag peste care solul va fi considerat poluat cu un anumit metal

what methods must be used to explain the spatial distribution of a certain metal in the soil [2]. A correct response to either of these questions involves a mandatory appeal to mathematical modeling.

The physical approach to the problem involves a model which consists of two second-degree differential equations with partial derivatives [3] that can be solved only by numerical methods [4-6] considering some simplifying hypotheses and a restrictive system of initial and boundary conditions. These substantially reduce the predictability of the model, restricting its applicability. Therefore, for effective actions of prevention and combat of heavy metal pollution, mathematical empiric models are required on a large scale. These models are obtained by statistical processing of experimental data [7-9] or of ground measured values [10,11]. For the latter, specialists can benefit from the support of large databases established to monitor heavy metals, both at a European level [12] and also at national level. In Romania, under the auspices of the Research Institute for Soil Science and Agronomy (I.C.P.A.) and of the Research Institute for Forest Design (I.C.A.S.), was developed and published: Monitoring of soils quality status in Romania [13]. This atlas provides comprehensive data for 942 sites (670 agricultural and 272 forestry) covering a 16 x 16 km network, throughout the country. For each site separately have been determined, inter alia, the content of heavy metals and also several other physical and chemical soil characteristics. Such databases, which associate a measured value of a state variable (concentration in heavy metal) to process parameters (physicochemical characteristics of soil), provides large opportunity of statistical processing in order to obtain empirical performance models.

If, for establishing the model equations, regression analysis method is used [14], it is particularly useful, as a preliminary step, to carry out an analysis of the correlation [7]. This allows a simplification of the mathematical model equation by retaining as process factors only those physicochemical characteristics that have a decisive influence on the state variable. Quantifying this influence will be achieved unequivocally by determining the simple correlation coefficients and especially those of multiple correlations between the physicochemical characteristics of soils and heavy metals contents.

greu?; ce locații sunt de așteptat să aibă concentrații naturale ridicate pentru aceste elemente?; și: ce metode trebuie să fie utilizate pentru a explica distribuția spațială a unui anumit metal greu în sol? [2]. Un răspuns corect la oricare dintre aceste întrebări implică un apel obligatoriu la modelarea matematică.

Abordarea fizică a acestor probleme implică un model matematic ce constă din două ecuații diferențiale de gradul doi cu derivate parțiale [3], care pot fi rezolvate numai prin metode numerice [4-6], introducând numeroase ipoteze simplificatoare și un sistem restrictiv de condiții inițiale și la limită. Toate acestea reduc, însă substanțial predictibilitatea modelului, limitând drastic aplicabilitatea sa. Prin urmare, pentru acțiuni eficiente de prevenire și combatere a poluării solului cu metale grele, modelele matematice empirice sunt mult mai adecvate, fiind utilizate pe scară largă. Aceste modele sunt obținute prin prelucrarea statistică a datelor experimentale [7-9] sau a valorilor măsurate în sol [10,11]. Pentru acestea din urmă, specialiștii pot beneficia de sprijinul unor mari baze de date create pentru monitorizarea metalelor grele, atât la nivel european [12] cât și la nivel național. În România, sub auspiciile Institutului de Cercetări pentru Pedologie și Agronomie (ICPA) și al Institutului de Cercetări și Amenajări Silvice (ICAS), a fost elaborat și publicat: Monitoringul Stării de Calitate a Solurilor din România [13]. Acest atlas oferă date complete pentru 942 de sit-uri (670 agricole și 272 forestiere) ce acoperă cu o rețea de 16 x 16 km, întreaga suprafață a țării. Pentru fiecare sit în parte au fost determinate, printre altele, conținutul de metale grele, precum și o serie de alte caracteristici fizice și chimice ale solului. Aceste baze de date, care asociază o valoare măsurată a unei variabile de stare (concentrația în metale grele) unor parametri de proces (caracteristici fizico-chimice ale solului), oferă o mare posibilitate de prelucrare statistică în scopul obținerii unor modele empirice performante.

În cazul în care, pentru stabilirea ecuațiilor modelului, este utilizată metoda analizei de regresie [14], este deosebit de util, ca într-o etapă preliminară, să se efectueze și o analiză de corelație [7]. Acest lucru permite o simplificare a ecuației modelului matematic, reținând ca factori de proces doar acele caracteristici fizico-chimice ale solului care au o influență decisivă asupra variabilei de stare. Cuantificarea acestei influențe se va realiza, fără echivoc, prin determinarea coeficienților de corelație simplă dar mai ales a celor de corelație multiplă dintre caracteristicile fizico-chimice ale solurilor și concentrațiile de metale grele.

2. Materials and methods

When studying a particular natural phenomenon governed by two variables y and x , (y is called the state variable and x is a process parameter), are relatively rare situations where there is a functional dependence of the type: $y = f(x)$ that binds between these two variables. Even more rare are situations where this function is known. In all cases, however, if we have a sufficiently large number of x, y pairs we can proceed to determine a *simple correlation coefficient* (noted: $r_{y,x}$) whose modulus value indicates the extent to which there is a covariance between the two values [15]. The formula for calculating this coefficient is:

$$r_{y,x} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y} \quad (1)$$

where: n is the number of pairs of values $x-y$ that we have, \bar{x}, \bar{y} are the arithmetic means and s_x, s_y are the standard deviations of these values.

If a null value resulting from the calculation of the simple correlation coefficient will conclude that the two variables are perfectly independent. If, however, we find a modulus value equal to the unit, may conclude that there is a functional dependence of the type $y=f(x)$ between the two variables. A value in between these extremes is a lesser or greater degree of correlation between the two variables.

Due to facility of this method, the simple correlation coefficient is frequently used also when the state variable (y) is connected not only to a parameter, but a series of "k" parameters ($x_1 \dots x_k$). The simple correlation coefficients: $r_{yx_1} \dots r_{yx_k}$, will be calculated with the same equation which emphasizes the relationship between the state variable and every individual process parameter. Unfortunately, in this case the accuracy of the simple correlation coefficients is diminished by the fact that Eq. (1) ignores two important aspects in the state variable dependence of more parameters with simultaneous action. These aspects are: the mutual interactions between these parameters (absolutely independent only in few cases) and that the influence of some parameters on the state variable may be different in the presence of other parameters, than if they would act individual.

2. Materiale și metode

Atunci când se studiază un anumit fenomen natural guvernat de două variabile y și x , (y fiind variabila de stare iar x un parametru de proces), sunt relativ rare situațiile în care există o dependență funcțională de tip: $y = f(x)$, care să lege între ele aceste două variabile. Și mai rare sunt situațiile în care această funcție este cunoscută. În toate cazurile, însă, dacă avem un număr suficient de mare de perechi $x-y$ putem determina un coeficient de corelație simplă (notat: $r_{y,x}$) a cărui valoare indică măsura în care există o covarianță între cele două valori [15]. Formula de calcul a acestui coeficient este:

în care: n este numărul de perechi de valori $x-y$, având mediile aritmetice: \bar{x}, \bar{y} și abaterile medii pătratice: s_x și s_y .

Dacă din calculul coeficientului de corelație simplă rezultă o valoare nulă, concluzia care se va trage este aceea că: cele două variabile sunt perfect independente. Dacă, însă, rezultă o valoare egală în modul cu unitatea, concluzia va fi aceea că există o dependență funcțională de tip $y = f(x)$ între cele două variabile. O valoare între aceste două extreme exprimă un nivel mai mic sau mai mare de corelare între respectivele variabile.

Datorită facilității acestei metode, coeficientul de corelație simplă este frecvent utilizat chiar și în situația în care variabila stare (y) este legată, nu numai de un singur parametru, ci de o serie de "k" parametri ($x_1 \dots x_k$). Coeficienții de corelație simplă: $r_{yx_1} \dots r_{yx_k}$, vor fi calculați cu aceeași relație de mai sus care pune accentul pe legătura dintre variabila de stare și fiecare parametru de proces în parte. Din păcate, în acest caz, acuratețea coeficienților de corelație simplă este mult diminuată de faptul că relația (1) ignoră două aspecte importante legate de dependența variabilei de stare de mai mulți parametri cu acțiune simultană. Aceste aspecte sunt: interacțiunile frecvente dintre acești parametri (cazurile în care ei sunt absolut independenți fiind relativ rare) și faptul că influența unui parametru asupra variabilei de stare poate fi diferită în prezența altor parametri, decât în situația în care acesta ar acționa individual.

To eliminate these deficiencies, it is recommended to determine the *multiple correlation coefficients*. Although the algorithm is much more laborious, this approach removes the shortcomings noted above and, in addition, it offers the possibility of a thorough analysis of the multiple interdependencies that can appear between the state variable and the set of process parameters.

Relationship for calculating the multiple correlation coefficient between state variable and the set parameters $x_1... x_k$ is the following [15]:

$$r_{y \cdot x_1 x_2 \dots x_k} = \sqrt{1 - P / P_{11}} \quad (2)$$

where P is the correlation determinant of the form:

$$p = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_k} \\ r_{x_1y} & 1 & r_{x_1x_2} & \dots & r_{x_1x_k} \\ r_{x_2y} & r_{x_2x_1} & 1 & \dots & r_{x_2x_k} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ r_{x_ky} & r_{x_kx_1} & r_{x_kx_2} & \dots & 1 \end{vmatrix} \quad (3)$$

and P_{11} represents the algebraic complement of determinant P to the simple correlation coefficient $r_{yy}=1$. Each of the simple correlation coefficients that appear in the determinant will be calculated, in turn, using Eq. (1).

Next will show how these relationships will be used to determine the correlations between heavy metals contents and the physicochemical characteristics of soils, starting from the results of the measurements published in the *Monitoring of the Quality Status of Soils in Romania*.

Since the use of data from all 942 sites, published in the atlas above would have required an excessive amount of calculation, a representative sample was chosen, containing a total of only 150 sites, enough to reveal the correlations followed. For establishing this sample a program was used to generate random numbers, which picked the 150 sites. From these, 104 were arable sites and 46 were forestry sites. For each of these, the concentrations of heavy metals (Cd, Co, Cr, Cu, Mn, Ni, Pb and Zn) were retained and also the values of six physicochemical characteristics: clay content [%], compacting degree [%], hydraulic conductivity at saturation [mm/h], structural hydro stability [%], pH and organic matter content [%].

Pentru eliminarea acestor deficiențe, se recomandă recurgerea la calculul coeficienților de corelație multiplă. Deși, în acest caz, algoritmul este mult mai laborios, această abordare elimină neajunsurile menționate mai sus și în plus oferă posibilitatea unei analize aprofundate a multiplelor interdependențe care pot apărea între variabila de stare și ansamblul parametrilor de proces.

Relația de calcul a coeficientului de corelație multiplă între variabila de stare și setul parametrilor de proces $x_1... x_k$ este următoarea [15]:

unde P este determinantul de corelație și are forma:

iar P_{11} reprezintă complementul algebric al determinantului P în raport cu coeficientul de corelație simplă $r_{yy}=1$. Fiecare dintre coeficienții de corelație simplă care apar în acest determinant va fi calculat, la rândul său, cu ajutorul relației (1).

În continuare va fi prezentat modul în care aceste relații au fost utilizate pentru determinarea corelațiilor dintre conținutul de metale grele și caracteristicile fizico-chimice ale solurilor, pornind de la rezultatele măsurătorilor publicate în *Monitoringul Stării de Calitate a Solurilor din România*.

Deoarece utilizarea datelor din toate cele 942 de situri, publicate în atlasul de mai sus, ar fi necesitat o cantitate excesivă de calcule, a fost ales un eșantion reprezentativ, care conține un total de doar 150 de sit-uri, suficiente însă pentru a pune în evidență corelațiile urmărite. Pentru stabilirea acestui eșantion a fost utilizat un program de generare a numerelor aleatoare, prin intermediul căruia au fost extrase cele 150 de sit-uri. Dintre acestea, 104 au fost sit-uri agricole iar 46 au fost sit-uri forestiere. Pentru fiecare sit în parte au fost reținute concentrațiile în metale grele (Cd, Co, Cr, Cu, Mn, Ni, Pb și Zn) precum și valorile a șase caracteristici fizico-chimice: conținut de argilă[%], grad de tasare [%], conductivitate hidraulică la saturație [mm/h], hidrostabilitate

In order to verify the law of distribution of the sample values for each of the six features above, was applied *Kolmogorov* test [16]. These consist of comparing the cumulative absolute frequencies of the sample values with the normal theoretical repartition function (*Gaussian* type). Following this test it was found that, besides the organic matter content which shows a minor deviation, the values of all the other physicochemical characteristics comply the *Gaussian* normal distribution. These positive results confirm the representativeness of the sample and provide that any conclusions on the sample can be extrapolated to the entire statistic population.

3. Results and Interpretations

In a first step, using Eq. (1), the simple correlation coefficients between the heavy metals contents and each of the six physicochemical characteristics of soils were determined. These coefficients were calculated both globally (G) for all the 150 sites of the sample and separately for arable sites (A), respectively for forestry sites (F) (Table 1).

Table 1.
Simple correlation coefficients (module values).

r		Cd	Co	Cr	Cu	Mn	Ni	Pb	Zn
Hydraulic conductivity at saturation	G	0.00137	0.04338	0.12525	0.09953	0.02468	0.13029	0.16054	0.13598
	A	0.05949	0.01182	0.01916	0.17221	0.05046	0.09176	0.11388	0.08130
	F	0.11540	0.10168	0.00236	0.01852	0.13312	0.11173	0.15350	0.38392
Compacting degree	G	0.18473	0.03542	0.14525	0.27648	0.08066	0.04458	0.04022	0.02765
	A	0.09991	0.05690	0.04579	0.34363	0.07520	0.08707	0.07502	0.04001
	F	0.28770	0.57571	0.12184	0.10768	0.21982	0.78083	0.21156	0.09631
Structural hydro stability	G	0.56343	0.02038	0.19256	0.04847	0.11607	0.10950	0.32096	0.00669
	A	0.55252	0.09644	0.17332	0.00497	0.08188	0.26719	0.47183	0.03394
	F	0.59298	0.33273	0.08108	0.36281	0.12729	0.49269	0.30948	0.35086
Clay content	G	0.01367	0.15708	0.00498	0.16508	0.00737	0.48248	0.10986	0.18272
	A	0.24908	0.24348	0.16903	0.23394	0.21222	0.55191	0.09105	0.16197
	F	0.04789	0.00353	0.31124	0.01431	0.04492	0.28582	0.12322	0.08071
pH	G	0.24115	0.13087	0.17066	0.01726	0.10525	0.24219	0.38532	0.06723
	A	0.06578	0.13196	0.24075	0.04252	0.14370	0.22011	0.26029	0.07783
	F	0.07414	0.13457	0.01807	0.10432	0.01814	0.19207	0.31107	0.34258
Organic matter content	G	0.25526	0.02183	0.19787	0.01924	0.02936	0.02051	0.38576	0.10663
	A	0.17662	0.06759	0.08317	0.06029	0.14735	0.15882	0.42637	0.04499
	F	0.08795	0.04955	0.05167	0.16330	0.09327	0.02828	0.20572	0.12866

structurală [%], pH și conținut de materie organică [%].

În scopul verificării legii de distribuție a valorilor eșantionului, pentru fiecare dintre cele șase caracteristici de mai sus, a fost aplicat testul Kolmogorov [16]. Acesta constă în compararea frecvențelor absolute cumulate ale valorilor eșantionului cu funcția normală de repartiție teoretică (de tip Gaussian). În urma acestui test s-a constatat că, în afară de conținutul de materie organică, ce prezintă o abatere minoră, valorile tuturor celorlalte caracteristici respectă legea de distribuție normală. Aceste rezultate pozitive confirmă reprezentativitatea eșantionului prelevat și garantează că orice concluzie referitoare la acesta poate fi extrapolată la nivelul întregii populații statistice.

3. Rezultate și interpretări

Într-o primă etapă, folosind relația (1), s-au determinat coeficienții de corelație simplă între conținuturile de metale grele și fiecare dintre cele șase caracteristicilor fizico-chimice ale solurilor. Acești coeficienți au fost calculați atât la nivel global (G), pentru toate cele 150 de sit-uri ale probei, cât și separat pentru sit-urile agricole (A), respectiv pentru sit-urile forestiere (F), (vezi Tab. 1).

For an objective verification of the significance of simple correlation coefficient, its value can be compared with size criteria: $r'_{\alpha,v}$, available in tables [16]. This criterion is chosen according to the level of significance α , which has a commonly adopted value of 0.05 and the number of liberty degrees: $v=n-2$, where n is the sample volume. In the studied case, for the global level ($\alpha=0.05$ and $v=148$) the criterion has the following value: $r'_{0.05,148}=0.1825$. For agricultural sites ($\alpha=0.05$ and $v=102$) the criterion has the value: $r'_{0.05,102}=0.1946$, and for the forestry sites ($\alpha=0.05$ and $v=44$) the following value is adopted: $r'_{0.05,44}=0.2875$.

For all simple correlation coefficients from Table 1, whose values exceed these limits criteria (note the characters "bold"), we can conclude, with a probability of 95%, that between the respective content of heavy metals and the physicochemical characteristic for which they were calculated there is a significant correlation and, therefore, these parameters must be retained in the equation of an mathematical model. Such significant correlations can be distinguished, in the case of forestry sites for example, between the majority of contents in heavy metals and the structural hydro stability or in the case of agricultural sites between the same majority of state variables and clay content. For values under these criterion limits (such as the vast majority of correlations between heavy metal contents and hydraulic conductivity at saturation), the conclusion, with the same probability, will be that these correlation levels are insignificant.

The extent to which these simple correlation coefficients describe the covariance between a state variable and a certain process parameter can be even more strongly emphasized by using its graphical representation. Figure 1 illustrates, for comparison, two of the correlations presented in Table 1. A certain point from the graphical field represents a pair of values: state variable – process parameter and characterizes one of the 150 sites of the sample. Increasing the correlation coefficient (from 0.004983 to 0.482477) leads to an alignment of the 150 points in a certain preferential direction.

Pentru o verificare obiectivă a semnificației coeficientului de corelație simplă, valoarea acestuia poate fi comparată cu o mărime criterială: $r'_{\alpha,v}$, disponibilă sub formă tabelară în literatura de specialitate [16]. Acest criteriu se alege în funcție de nivelul de semnificație: α , pentru care se adoptă uzual valoarea de 0,05 și numărul de grade de libertate: $v=n-2$, unde n reprezintă volumul eșantionului. În cazul studiat, pentru nivelul global ($\alpha=0,05$ și $v=148$) criteriul are valoarea: $r'_{0.05,148}=0,1825$. Pentru siturile agricole ($\alpha=0,05$ și $v=102$) criteriul are valoarea: $r'_{0.05,102}=0,1946$, iar pentru cele forestiere ($\alpha=0,05$ și $v=44$) se adoptă: $r'_{0.05,44}=0,2875$.

În cazul tuturor coeficienților de corelație simplă din tabelul 1, ale căror valori depășesc aceste limite criteriale, (cei notați cu caractere "bold") se poate concluziona, cu o probabilitate de 95%, că între respectivul conținut în metal greu și caracteristica fizico-chimică pentru care au fost calculați există o corelație semnificativă și prin urmare, acești parametri vor trebui să fie reținuți, în mod obligatoriu, în ecuația unui eventual model matematic. Asemenea corelații semnificative se pot evidenția – în cazul siturilor forestiere spre exemplu – între majoritatea conținuturilor în metale grele și hidrostabilitatea structurală sau în cazul siturilor agricole – între aceeași majoritate a variabilelor de stare și conținutul de argilă. Pentru valorile situate sub aceste limite criteriale (cum ar fi cazul mării majorități a corelațiilor dintre conținuturile în metale grele și conductivitatea hidraulică la saturație) concluzia – cu aceeași probabilitate – va fi că aceste nivele de corelație sunt nesemnificative.

Măsura în care acești coeficienți de corelație simplă descriu covarianța dintre o variabilă de stare și un anumit parametru de proces poate fi mai bine evidențiată prin folosirea reprezentărilor grafice. Fig. 1 ilustrează, pentru comparație, două dintre corelațiile prezentate în tabelul 1. Un anumit punct din câmpul grafic reprezintă o pereche de valori: variabilă de stare - parametru de proces și caracterizează unul dintre cele 150 de situri ale probei. Creșterea valorii coeficientului de corelație (de la 0,004983 la 0,482477) conduce la alinierea celor 150 de puncte pe o anumită direcție preferențială.

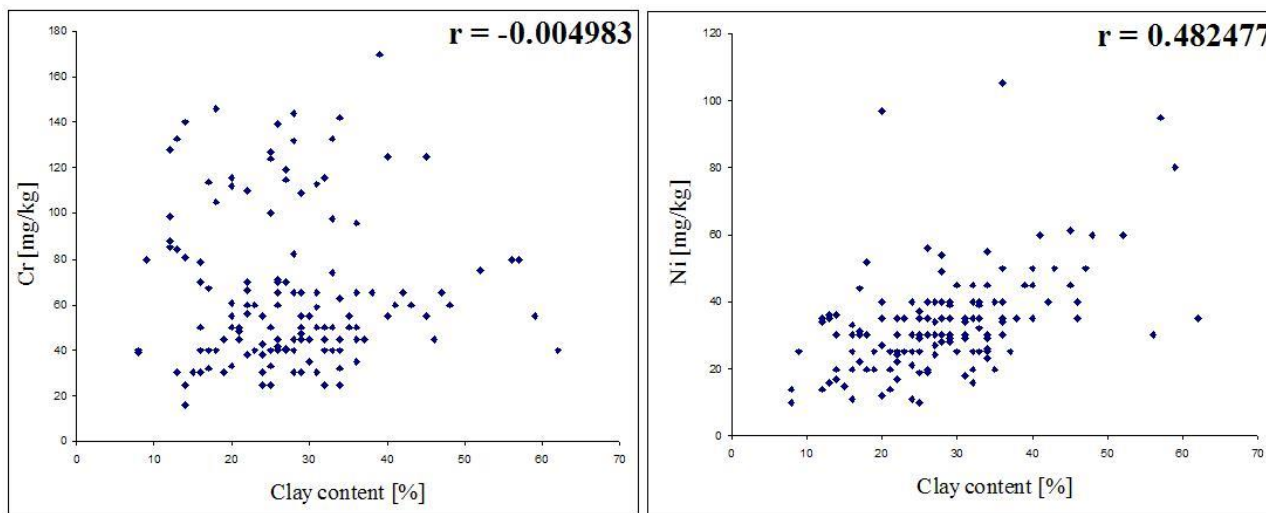


Figure 1. Covariance between the state value and the process parameter in the case of Ni and Cr.

For advancing to the second stage: determining multiple correlation coefficients, the interaction coefficients (correlation coefficients which describe the existing interdependencies only between the process parameters) were calculated in advance (Table 2.).

In this case we can also notice very high levels of correlations, some predictable, such as the correlation between the hydraulic conductivity and the compacting degree or between pH and clay content, respectively organic matter and others more subtle such as the correlation between pH level and hydraulic conductivity at saturation. An overview of the data from Table 2 is enough to support a previous assertion: situations where process parameters could be perfectly independent are rare.

For this reason, we went further to determine the multiple correlation coefficients. They were calculated using Eqs. (2-3) and represent the covariance between each of the eight heavy metals content and the block of six process parameters, taken as a whole.

Pentru trecerea la cea de-a doua etapă: determinarea coeficienților de corelație multiplă, au fost calculați în prealabil (vezi Tab. 2) coeficienții interacțiunilor (adică acei coeficienți de corelație care descriu interdependențele existente numai între parametrii de proces).

Și în acest caz, pot fi observate, deasemenea, niveluri foarte ridicate de corelație, unele previzibile, cum ar fi: corelația dintre conductivitatea hidraulică și gradul de tasare sau cea dintre pH și conținutul de argilă, respectiv materie organică, iar altele mai subtile, cum ar fi: corelația între nivelul pH-ului și conductivitatea hidraulică la saturație. O privire de ansamblu asupra datelor din tabelul 2 este suficientă, însă, pentru a susține afirmația anterioară cu privire la faptul că situațiile în care parametrii de proces sunt perfect independenți sunt foarte rare.

Din acest motiv, s-a trecut mai departe la determinarea coeficienților de corelație multiplă. Aceștia au fost calculați cu ajutorul relațiilor (2-3) și reprezintă covarianța dintre fiecare conținut de metal greu în parte și blocul celor șase parametri de proces, luați în ansamblu.

 Table 2.
Interaction coefficients.

r	Hydraulic conductivity	Compacting degree	Hydro stability	Clay content	pH	Organic matter
Hydraulic conductivity	1	0.45887	0.15991	0.21406	0.37051	0.32982
Compacting degree	0.45887	1	0.09647	0.12508	0.01893	0.18790
Hydro stability	0.15991	0.09647	1	0.18630	0.39774	0.33657
Clay content	0.21406	0.12508	0.18630	1	0.34693	0.14763
pH	0.37051	0.01893	0.39774	0.34693	1	0.41430
Organic matter	0.32982	0.18790	0.33657	0.14763	0.41430	1

Their values are summarized in Table 3, along with the calculated sizes of the Fischer criterion, which serves to objectively testing the significance of these coefficients. The value of the Fischer criterion is calculated, for each multiple correlation coefficient using the following relation [15]:

Valorile lor sunt sintetizate în tabelul 3, alături de mărimile calculate ale criteriului Fischer (care servește la testarea obiectivă a semnificației acestor coeficienți). Valoarea criteriului Fischer a fost calculată, pentru fiecare coeficient de corelație multiplă în parte folosind următoarea relație [15]:

$$F = \frac{n - k - 1}{k} \cdot \frac{r^2_{y \cdot x_1 x_2 \dots x_k}}{1 - r^2_{y \cdot x_1 x_2 \dots x_k}} \quad (4)$$

where n = 150 is the volume of the sample and k = 6 is the number of process parameters.

unde n = 150 reprezintă volumul eșantionului iar k = 6 este numărul total al parametrilor de proces.

Table 3. Multiple correlation coefficients.

	Cd	Co	Cr	Cu	Mn	Ni	Pb	Zn
r	0.61813	0.18718	0.27253	0.33483	0.16843	0.50982	0.47998	0.25555
F	14.737	0.86534	1.9122	3.0094	0.69586	8.3703	7.1343	1.6652
F _{α,v1,v2}	2.09							
Conclusion	correlation	-	-	correlation	-	correlation	correlation	-

For Cd, Cu, Ni and Pb contents, the Fischer criterion values, calculated with Eq. (4), are superior to the tabular size F_{α,v1,v2} [16] (for α=0.05, v1=k=6 and v2=n-k-1=143). Therefore, with the same probability of 95%, we can say that between the contents of these heavy metals and the physicochemical characteristics of soils there is a very high level of correlation which justifies the development of further mathematical models.

Pentru conținuturile de Cd, Cu, Ni și Pb valorile criteriului Fischer, calculate cu relația (4), sunt superioare mărimii F_{α,v1,v2} [16] (pentru α=0,05, v1=k=6 și v2=n-k-1=143). Prin urmare, se poate afirma, cu aceeași probabilitate de 95%, că între conținuturile în aceste metale grele și caracteristicile fizico-chimice ale solurilor există un nivel de corelație foarte ridicat ce justifică elaborarea unor eventuale modele matematice ulterioare.

In the case of elements: Cr, Zn, but mostly Co and Mn, the low values of the Fischer criterion indicate much lower significance levels of existing correlations between the contents of these heavy metals and all the physicochemical characteristics of soils.

În cazul elementelor: Cr, Zn, dar mai ales Co și Mn valorile scăzute ale criteriilor Fischer indică nivele de semnificație mult mai reduse ale corelațiilor existente între conținuturile acestor metale grele și ansamblul caracteristicilor fizico-chimice ale solurilor.

The histogram in figure 2 brings together the simple correlation coefficients previously calculated, for each of the six physicochemical characteristics, with multiple correlation coefficients, for all the eight heavy metals that have been studied.

Histograma din figura 2 reunește, într-o manieră sintetică, coeficienții de corelație simplă, calculați anterior, pentru fiecare din cele șase caracteristici fizico-chimice, cu coeficienții de corelație multiplă, în cazul tuturor celor opt metale grele studiate.

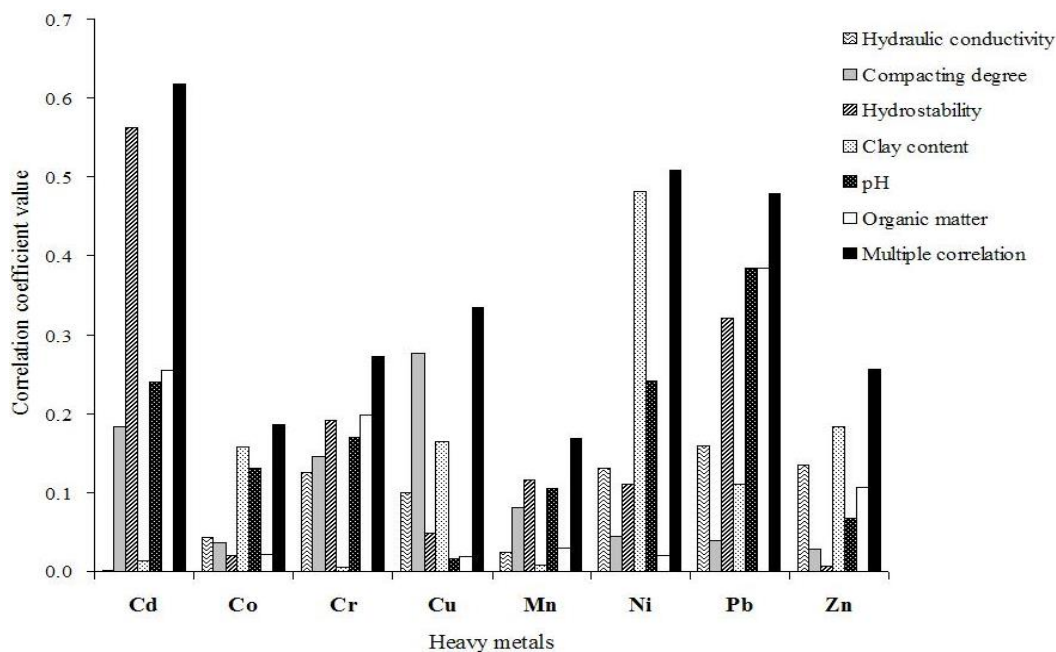


Figure 2. Synthesis of simple and multiple correlation coefficients.

4. Conclusions

Analyzing and connecting all data provided by this preliminary study, performed using correlation analysis, can be summarized some general conclusions. They can play an important role in subsequent stage of mathematical modeling and can be used as general information on prevention and control activities of heavy metal pollution in soils.

A first important observation, that can be made by analyzing Table 1, is that there are notable differences between the values of the same calculated correlation coefficients, for each of the two categories of sites separately (agricultural and forestry). Most significant is the case of structural hydro stability for which, in forestry sites, we can observe a much higher level of correlation than in the agricultural sites. This conclusion prefigures the necessity to develop one particular mathematical model for each of the two site categories.

Further, if these results are interpreted in terms of physicochemical properties of soils, it can make a ranking of these features on the role they play in the heavy metal retention process in the soil. According to this hierarchy, the structural hydro stability and pH of the soil are more closely correlated to the heavy metal quantities retained by the soil. A moderate relation was highlighted in the case of clay and organic matter content, while the compacting degree and the hydraulic conductivity at saturation have an insignificant influence. This latter observation supports the

4. Concluzii

Analizând și cuplând toate datele oferite de acest studiu preliminar, efectuat cu ajutorul analizei de corelație, pot fi sintetizate câteva concluzii cu caracter general. Acestea pot juca un rol important într-o etapă ulterioară de prefigurare a unor modele matematice sau pot fi utilizate ca informații cu caracter general în activitățile de prevenție sau combatere a poluării solurilor cu metale grele.

O primă observație importantă – ce se poate formula analizând tabelul 1 – este aceea că există diferențieri notabile între valorile aceluiași coeficienți de corelație calculați, în mod separat, pentru fiecare din cele două categorii de situri (agricole și forestiere). Cel mai pregnant se observă acest lucru în cazul hidrostabilității structurale unde, în cazul siturilor forestiere, se remarcă un nivel de corelație mult mai ridicat decât în cazul celor agricole. Această concluzie prefigurează necesitatea ca în etapa de modelare să se elaboreze câte un model matematic aparte pentru fiecare din cele două categorii de situri.

În continuare, dacă aceste rezultate sunt interpretate prin prisma caracteristicilor fizico-chimice ale solurilor, se poate opera o ierarhizare a acestor caracteristici după rolul pe care ele îl joacă în procesul de retenție a metalelor grele în sol. Potrivit acestei ierarhii: hidrostabilitatea structurală și pH-ul solului prezintă cea mai strânsă corelație cu cantitățile de metale grele reținute în sol. O legătură moderată a fost pusă în evidență în cazul conținuturilor de argilă și materie organică, pe când gradul de tasare și conductivitatea hidraulică la I

theories [3,9] that the pollutant sorption is instantaneous and independent of the hydraulic conductivity of the analyzed soil.

If we follow the influence that soil characteristics have on the retention of each type of heavy metal, we can easily notice that a higher dependency is recorded for cadmium. The contents of nickel and lead are also highly influenced. Copper, chrome and zinc are moderately influenced, while in the case of the manganese content this influence is absolutely minor.

For situations in which the correlation coefficients values are below the criterion limits, for final removal of uncertainty, a dispersion analysis (single or multi-factorial) can be organized. This analysis could highlight an eventual complex dependence (highly non-linear) between contents of heavy metals and the soil characteristics, dependence that may go unnoticed in the case of a correlation analysis.

saturație au o influență nesemnificativă. Această ultimă constatare vine în sprijinul teoriilor [3,9] conform cărora sorbția poluantului este una instantanee, ce nu depinde de conductivitatea hidraulică a solului analizat.

Dacă se urmărește influența pe care aceste caracteristici ale solului o au asupra retenției fiecărui tip de metal greu în parte, atunci se poate ușor observa că cea mai mare dependență se înregistrează în cazul cadmiului. Conținuturile de nichel și plumb sunt, de asemenea, puternic influențate, cuprul, cromul și zincul beneficiază de influențe moderate, în timp ce în cazul conținutului de mangan această influență este una absolut minoră.

Pentru situațiile în care s-au obținut valori ale coeficienților de corelație sub limitele critice, în vederea înlăturării definitive a incertitudinilor, se pot organiza analize dispersionale mono sau multi-factoriale. Acestea ar putea evidenția eventuale dependențe complexe (puternic non-lineare) între conținuturile în metale grele și respectivele caracteristici ale solurilor, dependențe ce pot trece neobservate în cazul unei analize de corelație.

References

- [1] Palumbo B., Angelone M., Bellanca A., Dazzi C., Hauser S., Neri R., Wilson J., (2000), Influence of inheritance and pedogenesis on heavy metal distribution in soils of Sicily, Italy, *Geoderma*, **95**:247-266.
- [2] Lado L., Hengl T., Reuter I., (2008), Heavy metals in European soils: A geostatistical analysis of the FOREGS Geochemical database, *Geoderma*, **148**:189-199.
- [3] Mermoud A., (1998), *Elements of soil physics (Elements de physique du sol)*, H.G.A., Bucharest, Romania.
- [4] Araya R., Behrens E., Rodríguez R., (2005), An adaptive stabilized finite element scheme for the advection-reaction-diffusion equation, *Applied Numerical Mathematics*, **54**:491-503.
- [5] Javadi A.A., AL-Najjar M., (2007), Finite element modeling of contaminant transport in soils including the effect of chemical reactions, *Journal of Hazardous Materials*, **143**:690-701.
- [6] Sun T., Yuan Y., (2009), An approximation of incompressible miscible displacement in porous media by mixed finite element method and characteristics-mixed finite element method, *Journal of Computational and Applied Mathematics*, **228**:391-411.
- [7] Borůvka L., Vacek, O., Jehlička J., (2005), Principal component analysis as a tool to indicate the origin of potentially toxic elements in soils, *Geoderma*, **128**:289-300.
- [8] Merdun H., Cinar Ö., Meral R., Apan M., (2006), Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity, *Soil & Tillage Research*, **90**:108-116.
- [9] Vidal M., Santos M.J., Abrão T., Rodríguez J., Rigol A., (2009), Modeling competitive metal sorption in a mineral soil, *Geoderma*, **149**:189-198.
- [10] Hengl T., Heuvelink G., Rossiter D., (2007), About regression-kriging: From equations to case studies, *Computers & Geoscience*, **33**:1301-1315.
- [11] Imrie C.E., Korre, A., Munoz-Melendez G., Thornton I., Durucan S., (2008), Application of factorial kriging to the FOREGS European topsoil geochemistry database, *Science of the Total Environment*, **393**:96-110.
- [12] FOREGS, (1998), *Geochemical Atlas of Europe*, On line at: <http://weppi.gtk.fi/publ/foregsatlas/index.php>
- [13] Dumitru M., Ciobanu C., Motelică D.M., Mashali A.M., Dumitru E., Cojocaru G., Enache R., Gament E., Plaxienco D., Radnea C., Cârstea St., Manea A., Vrînceanu N., Calciu I., (2000), *Monitoring of soils quality status in Romania (Monitoringul Stării de Calitate a Solurilor din România)*, GNP F.A.O., Bucharest, Romania.
- [14] Rajkai K., Kabos S., van Genuchten M.T., (2004), Estimating the water retention curve from soil properties: comparison of linear, nonlinear and concomitant variable methods, *Soil & Tillage Research*, **79**:145-152
- [15] Taloi D., (1987), *Optimization of technological processes (Optimizarea proceselor tehnologice)*, Academy Press, Bucharest, Romania.
- [16] Taloi D., Berceanu E., Bratu C., Florian E., (1983), *Optimization of metallurgical processes (Optimizarea proceselor metalurgice)*, E.D.P., Bucharest, Romania.