Information-analytical modeling of soil composition based on spectral analysis

E M Basarygina, V S Zybalov, N A Pakhomova and O E Akulich*
Federal State Budgetary Educational Institution of Higher Education «South Ural State Agrarian University», Chelyabinsk, Russia

* E-mail: akulich-olga@yandex.ru

Abstract. Over the past decade, public interest in precision farming technologies, monitoring of production and technological processes and environmental protection issues has noticeably increased. Nowadays, agricultural producers are increasingly using technologies based on digital data, which are used to manage and optimize production. Digital maps and information models are one of the ways to present a field passport, that is, a visual database obtained on the basis of an analysis of land resources and reflecting reliable information about the state of a particular land plot. As a result of the research carried out, digital soil models were constructed (using the example of the soils of the Chelyabinsk region). The obtained models, which clearly reflect the content of heavy metals, are recommended for use in making managerial decisions in the field of the agro-industrial complex. The use of these models contributes to the transition to digital crop technologies.

1. Introduction
Currently, agricultural producers are increasingly using precision farming technologies, that is, technologies based on digital data that are used to manage and optimize production [1-6]. Digital crop technologies include the development of "digital maps" of agricultural land, which are compiled from the analysis of land resources. Such maps contribute to the transition to rational fertilization, taking into account the needs of different parts of the field in micro-, macroelements, etc. taking into account the state of the soil, recommendations are issued on agricultural turnover and plant protection products.

Digital maps and information models of soils can be one of the ways to present a field passport - a visual database that reflects reliable information about the state of a particular land plot [7, 8]. The introduction of digital crop production technologies can increase yields up to 20% and reduce the cost of seeds and fertilizers up to 25% [9].

The solution to the issue of digital land modeling, which makes it possible to make optimal management decisions in the field of the agro-industrial complex, taking into account the state of land resources, is also relevant for the Chelyabinsk region. In this regard, the purpose of the research was to build digital land models based on spectral analysis data. The research objectives included:

- spectral (energy dispersive X-ray fluorescence) analysis of soils in the Chelyabinsk region;
- mathematical processing of the analysis results;
- development of digital models.
The processing of experimental data of physical analysis of soils is an applied problem. The solution to any applied problem begins with the construction of its model. In practice, it is possible to build a 3D soil model.

Structured data, associated, in particular, with the physical analysis of lands, can be used in making effective management decisions in the field of the agro-industrial complex (Figure 1).

![Figure 1. An example of an information model [7].](image)

Important information when choosing such solutions is reliable information on the content of various microelements in soils, in particular, lead, cadmium, zinc, manganese, cobalt, chromium, etc.

When describing the model, all the essential characteristics of an object or phenomenon are analyzed, significant properties are taken into account and secondary ones are discarded. Thus, we obtain an information-logical model that reflects all the interdependencies of interdependence. Let's structure the initial data and write them using a table, getting a relational model of soil composition. This data can be visualized using charts, which makes the model more visual (Figure 2).

![Figure 2. Stages of creating a digital model.](image)

2. Materials and methods

Compilation of a digital model begins with monitoring the state of land resources. To control the state of the environment, classical methods of chemical analysis and modern methods of instrumental analysis, including spectrometry, are used [10, 11] (Figure 3). Many spectroscopic methods are non-destructive, allowing multiple analyzes of the same sample to be performed multiple times.
With the help of spectroscopic methods, it is possible to solve the following problems: finding and extracting raw materials; development of new products and technologies; design and optimization of production processes; ensuring the required quality of manufactured products.

For successful monitoring of production and technological processes (including environmental monitoring), it is necessary to know the initial, background content of elements in natural objects. In

| METHODS FOR DETERMINING POLLUTANTS IN THE BIOSPHERE |
|-----------------------------------------------------|
| Gravimetric                                         | Titrimetric                                     |
| Visible spectrophotometry                           | Ultraviolet spectrophotometry                   |
| Flame Emission Spectroscopy                        | Atomic absorption spectroscopy                   |
| Gas chromatography                                  | Voltammetry                                     |
| Spectrofluorimetry                                  | X-ray fluorescence spectrometry                 |
| Liquid chromatography                               | Polarography                                    |
| Infrared spectrophotometry                          | Microbiological                                 |

**Figure 3.** Methods for determining pollutants in the biosphere:
- solid,
- liquid,
- gaseous samples

| Gravimetric                                         |
|-----------------------------------------------------|
| Titrimetric                                         |
| Visible spectrophotometry                           |
| Ultraviolet spectrophotometry                       |
| Flame Emission Spectroscopy                        |
| Atomic absorption spectroscopy                      |
| Gas chromatography                                  |
| Voltammetry                                         |
| Spectrofluorimetry                                  |
| X-ray fluorescence spectrometry                     |
| Liquid chromatography                               |
| Polarography                                        |
| Infrared spectrophotometry                          |
| Microbiological                                    |
In this case, it is often required to determine the concentration of a number of elements at the level of 10^{-4} ... 10^{-7}\% and lower in samples of complex composition; reliably distinguish the background content of elements in environmental objects from technogenic components [12].

When determining the composition of samples, various methods of spectral analysis are used, which differ in their main characteristics: sensitivity, selectivity, reproducibility, productivity and economy.

The most promising method is the X-ray fluorescence analysis method, which is widely used in the work of research laboratories, since it has a number of advantages over the known methods [9]. For spectral analysis of land resources, a Rayny EDX-720 Shimadzu spectrometer was used.

To study the relationship between the content of elements in the soil, the methods of correlation and regression analysis were used. On the basis of these methods, the presence of an e relationship between the elements was determined, the tightness of the relationship between two quantities, linearly related to each other, was studied using the correlation coefficient. Finding the parameters of the equation expressing the relationship between the mean values of the dependent variable and the values of the independent variable was performed using the least squares method. The calculations were performed using the statistical analysis package of the Ms Excel program [13].

The determination of the bond strength between the content of chemical elements: Cd-Pb, Cd-Zn, Pb-Zn was carried out for the soils of the Chelyabinsk region. On the basis of empirical values, a correlation field is constructed, according to the form of which a linear relationship between random variables can be assumed. Based on this, the tightness of the linear correlation was determined.

Taking into account the assumption that the quantities are distributed according to the normal law, the value of the correlation coefficient was calculated [14]

\[
\rho_{xy} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right) \cdot \left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}}
\]

where \( x_i, y_i \) is the content of a chemical element (Cd-Pb, Cd-Zn, Pb-Zn).

The correlation coefficient was tested for significance using the hypothesis that the correlation coefficient is significantly different from zero. To test the hypothesis, the Student’s t test was used [14]:

\[
t_{calc} = \frac{|\rho_{xy}|}{\sqrt{\frac{1}{n-2}}}
\]

where \( n \) is the sample size.

Compliance with the condition, \( t_{calc} > t_{cr} \) meant that the correlation coefficient is significantly different from zero, therefore, the relationship is statistically significant [14].

When developing a digital model, the presentation of the obtained data in the form of diagrams, graphs or diagrams was used (to visualize the results) (Figure 4)
3. Results and discussion

Based on the results of spectral analysis of land resources, it was found that the content of a number of elements (cadmium, lead, zinc) exceeded the permissible concentration. For these elements, the relationship was calculated (Table 1). Based on the corresponding calculations, it was concluded that the relationship between the content of elements Cd-Zn is very weak (r_{xy} \approx 0.13); of Cd-Pb elements, the connection is direct moderate (r_{xy} \approx 0.33); of Pb-Zn elements, the bond is direct moderate (r_{xy} \approx 0.56).

| №  | The elements | Correlation coefficient | t_{calc} |
|----|--------------|-------------------------|----------|
| 1  | Cd-Pb        | 0.1338                  | 0.59     |
| 2  | Cd-Zn        | 0.3302                  | 1.52     |
| 3  | Pb-Zn        | 0.5611                  | 2.95     |

At a 5% significance level, the critical value of the statistic was 2.08. Since the condition t_{calculated} > t_{cr} was observed for the Pb-Zn elements, the correlation coefficient was significantly different from zero. Hence, it was concluded that the relationship was significant. The results obtained indicate that there is a moderate direct relationship between the values of the content of Pb-Zn elements in the kidney.

Next, a regression equation was drawn up, expressing the dependence of the zinc content on the lead content in the soil. The coefficients of the regression equation were found using the least squares method, the resulting regression equation has the form:

$$y = 22.05 + 2.9x$$

(3)

The correlation field and the empirical regression line for Pb-Zn are shown in Figure 5.
After the calculations, a digital model was built based on the results of spectral analysis of soils (Figure 6). This model visualizes the content of heavy metals (cadmium, lead, lead) in the studied land resources of the Chelyabinsk region. Analysis of the model revealed a significant content of one of the elements - lead.

4. Conclusions
Thus, as a result of the research carried out, digital soil models have been built. These models, which clearly reflect the content of heavy metals, are recommended for use in making managerial decisions in the field of the agro-industrial complex. The use of these models contributes to the transition to digital crop technologies.

References
[1] Analysis of Trace Elements in Water using Ultra Thin Film. Rayny EDX-700/800 2008 - 2009 Application Data No. 12 - 17 (Japan, Shimadzu)
[2] Basarygina E M, Pakhomova N A and Akulich O E 2021 E3S Web of Conferences 282 07016
[3] Onesimo C M. G; Dias D D and Vale Beirao M 2021 *RES EC* 29(1)
[4] Puno G. R; Marin R A and Puno R C C 2021 *GL JOUR OF ENV SC & MAN-GJESM* 7(1)
[5] Elizarov N V, Popov V V and Semendyaeva N V 2020 *EUR S SC* 53(12) 1701-1708
[6] dos Santos Teixeira A F, Godinho Silva S H, de Carvalho T S 2021 *CATENA* 199 105083
[7] Spatial information systems Internet resource: http://www.spsinf.ru/
[8] Tsvetkov V Ya 2016 *Ed res & tech* 4 (16) 114-120
[9] Rubaeva O D, Zubareva I A, Pakhomova N A and Malykhina E A 2020 *Jour of Env Man & Tour* 692-704 https://doi.org/10.14505/jemt 2020
[10] Becker Yu 2009 Spectroscopy (Moscow: Technosphere) p 527
[11] Schmidt V 2007 Optical spectroscopy for chemists and biologists (Moscow: Technosphere) p 378
[12] Chetyrkin Y B, Basarygina E M and Hiller V V 2009 Monitoring of production and technological processes at agricultural enterprises: the use of spectral methods of analysis. scientific. analyte. Overview (Chelyabinsk) p 160
[13] Vadzinsky R N 2008 Statistical calculations in the Excel environment (Saint Petersburg: Peter) p 608
[14] Dobrodets B S 2020 Computational probabilistic analysis: models and methods (Krasnoyarsk: Sib. Federal University) p 236