

## **EFFICIENCY OF INTERPOLATION METHODS BASED ON GIS FOR ESTIMATING OF SPATIAL DISTRIBUTION OF PH IN SOIL**

**Myslyva Tamara, Kutsaeva Olesya, Krundzikava Natallia**  
Belarusian State Agricultural Academy

### **Abstract**

The main objective of this study is to review and evaluate three common interpolation methods namely: Inverse Distance Weighting (IDW), Radial Basis Function (RBF) and Ordinary Kriging (OK), and generate maps of soil pH using these methods. The accuracy and efficiency of the generated maps have been examined as well as the most fitting technique for estimating spatial distribution of soil pH in the study area is identified. Studies were conducted within the limits of land use of RUP “Uchkhov BGSMA” (Republic of Belarus, Mogilev region, Goretzky district). The total area of the surveyed territory is 3197.89 hectares. For the analysis data is used about  $pH_{KCl}$  of soil solution obtained from materials of an agrochemical survey executed in 2014. Forecasting and visualization of the spatial distribution of  $pH_{KCl}$  was carried out using the Geostatistical Analyst module of the ArcGIS software. The experimental anisotropic variograms were calculated to determine the possible spatial structure of soil pH. Based on cross-validation results, a polynomial function was identified as the best variogram model. The model created by the method of radial basis functions turned out to be the most suitable for forecasting purposes (the value of the root-mean-square error was 0.763). In terms of interpolation accuracy, the investigated deterministic and geostatistical methods are located in the next descending row: RBF > IDW > OK.

Key words: geospatial analysis,  $pH_{KCl}$  of soil solution, interpolation, spatial distribution.

### **Introduction**

Rational use of land resources is one of the most important factors in the modern development of the Republic of Belarus. Monitoring the state of land becomes a reference point for public authorities for developing regulatory legal acts regarding their use, conducting territorial planning, implementing measures to protect land and reproduce soil fertility. The economic efficiency of land use and the efficiency of agriculture generally largely depend on the quality of land resources the quality of which, in turn, is determined by the fertility of their soil cover.

Acid-baseline state is one of the most important factors of soil fertility and causes numerous features of the behavior of chemical elements in the soil ecosystem. Regimes of organic matter and elements of mineral nutrition are associated with it, as well as the mobility of various compounds, including those that are toxic to plants. Indicators-constants, to which the pH of the soil solution belongs, are the main source of information on basic ecological permanent soil. Therefore, it is extremely important to predict the spatial distribution of soil pH for assessing the state of the soil system and planning measures for the rational use of land resources and the reproduction of soil fertility.

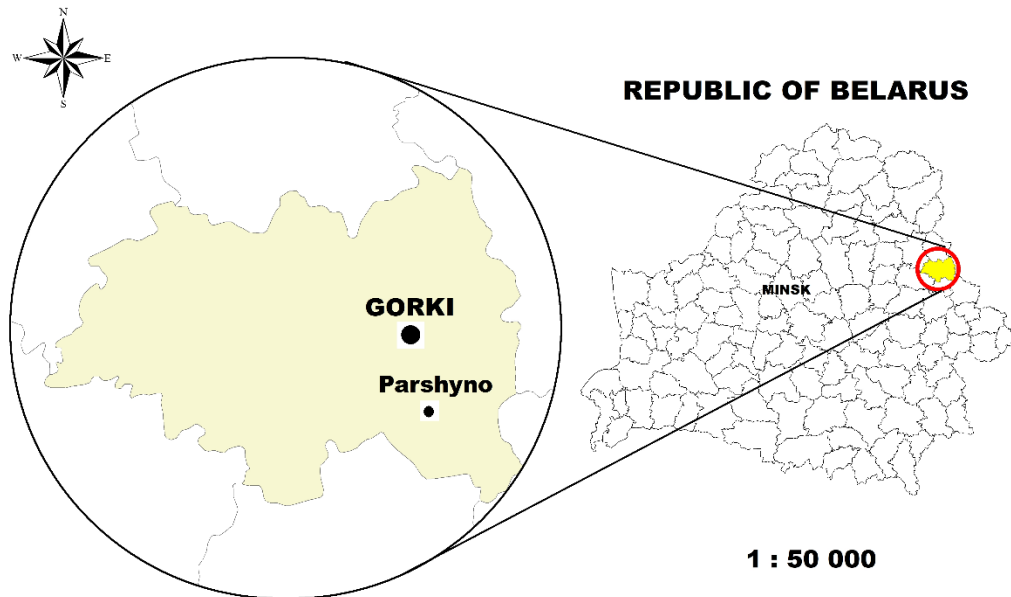
Geostatistics is an effective method for studying the spatial distribution of soil characteristics and their inconsistencies. A number of researchers have used Geospatial methods and the comparison of their effectiveness in assessing the spatial relationship of the agrochemical properties of soils and the geographic variability of soil characteristics (Zandi et al., 2011; Liu et al., 2014; Behera, 2015). In particular, the research of Indian scientists is devoted to the problems of assessing the suitability of various interpolation models for studying the spatial distribution of organic matter in soil depth of 0–20 cm (Gouri et al., 2016). Another study compared the effectiveness of different interpolation methods to estimate the spatial distribution of the upper pH and EC in Hamadan Province, Western Iran (Attaeian et al., 2016). Scientists from Turkey evaluated interpolation approaches to characterize spatial variability of soil organic matter, phosphorus, lime and boron in soil in the Tuz lake basin (Taha et al., 2017). Scientists from Lithuania tested several geo-statistical and interpolation methods for monitoring the spatio-temporal evolution of ashes in peat soil after a fire (Martin and Jord, 2013). The problems of assessing the suitability of various interpolation models for studying the spatial distribution of soil properties are devoted to studies of Russian (Simbatova et al., 2016; Samsonova et al., 2017) and Ukrainian (Moskalenko, 2012; Kohan et al., 2013) scientists. However, in the scientific literature there is not enough information about the development of this area of research in the territory of Belarus.

In this paper, both deterministic and geostatistical methods were investigated. The choice of methods was based on two requirements: 1) ease of use in practical activities of land management

organizations; 2) accuracy in predicting the spatial distribution of the pH of the soil solution. Among the determined methods, the following were investigated: Inverse Distance Weighting (IDW) and Radial base Function (RBF) while among geostatistical methods Ordinary Kriging (OK) was studied. The performances of interpolation methods were evaluated and compared using the cross-validation. Cross-validation model statistics were compared to determine the model with the lowest values of the average error (ME) and the mean square error (RMSE).

### Methodology of research and materials

The purpose of this study was to compare the effectiveness of various interpolation methods – the Inverse Weighted Distance method (IDW), the Radial Basis Function method (RBF) and Ordinary Kriging (OK) to estimate the spatial distribution of soil pH within the land use of RUP “Uchkhov BGSHA” (Fig. 1), and estimate the use of cross-validation to assess the accuracy of spatial modeling and identification the most suitable method.



**Fig. 1.** The location of the studied territory

Geospatial analysis of data was carried out using the Geostatistical Analyst module of the ArcGIS software version 10.2. The data obtained from the materials of the agrochemical survey of the territory of RUP “Uchkhov BGSHA” (Republic of Belarus, Mogilev region, Goretzky district), executed in 2014 by the Mogilev Regional Design and Exploration Station of Agrochemicalization were used for the analysis. The total area of the surveyed territory is 3197.89 hectares. The soil cover of the study area is represented mainly by sod-podzolic sandy loams on water-glacial sandy loams soils and sod-podzolic loamy on loesslike loams soils.

In the present study, deterministic (create surfaces from measured points) and geostatistical (utilize the statistical properties of the measured points) interpolation methods were used. We used the method of Inverse Distance Weighting (IDW) and the method of Radial Basis Functions (RBF) among the deterministic interpolation methods, as well as Ordinary Kriging (OK) among the geostatistical methods.

The IDW is one of the most applied and deterministic interpolation methods in the field of soil science. IDW estimates were made based on nearby known locations. The weights assigned to the interpolating points are the inverse of its distance from the interpolation point. Consequently, the closest points are made-up to have more weights (so, more impact) than distant points and vice versa. The known sample points are implicit to be self-governing from each other (Robinson and Metternicht, 2006; Gouri et al., 2016).

$$Z(x_0) = \frac{\sum_{i=1}^n \frac{x_i}{h_{ij}^\beta}}{\sum_{i=1}^n \frac{1}{h_{ij}^\beta}} \dots\dots\dots(1)$$

where  $Z(x_0)$  is the interpolated value,

$n$  representing the total number of sample data values,

$x_i$  is the  $i$ th data value,

$h_{ij}$  is the separation distance between interpolated value and the sample data value,

$\beta$  denotes the weighting power.

Radial basis function (RBF) predicts values identical with those measured at the same point and the generated surface requires passing through each measured point. The predicted values can vary above the maximum or below the minimum of the measured values (Li et al., 2011; Gouri et al., 2016). RBF method is a family of five deterministic exact interpolation techniques: thin-plate spline (TPS), spline with tension (SPT), completely regularized spline (CRS), multi-quadratic function (MQ) and inverse multi-quadratic function (IMQ). RBF fits a surface through the measured sample values while minimizing the total curvature of the surface (Johnston et al., 2001; Gouri et al., 2016). RBF is ineffective when there is a dramatic change in the surface values within short distances (Cheng and Xie, 2009; Gouri et al., 2016). In this study spline with tension (SPT) was selected.

Ordinary kriging method incorporates statistical properties of the measured data (spatial autocorrelation). The kriging approach uses the semivariogram to express the spatial continuity (autocorrelation). The semivariogram measures the strength of the statistical correlation as a function of distance. The range is the distance at which the spatial correlation vanishes, and the sill corresponds to the maximum variability in the absence of spatial dependence (Gouri et al., 2016). Kriging estimate  $Z(x_0)$  and error estimation variance  $\sigma_k^2(x_0)$  at any point  $x_0$  were, respectively, calculated as follows (2) and (3):

$$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (2)$$

$$\sigma_k^2(x_0) = \mu + \sum_{i=1}^n \lambda_i \gamma(x_0 - x_i) \quad (3)$$

where  $\lambda_i$  are the weights,

$\mu$  is the lagrange constant,

$\gamma(x_0 - x_i)$  is the semivariogram value corresponding to the distance between  $x_0$  and  $x_i$  (Gouri et al., 2016).

Semivariograms were used as the main tool for studying the structure of the spatial distribution of agrochemical indicators. Based on the regional theory of variations and internal hypotheses (Gouri et al., 2016), the semivariogram is expressed as follows (4) (Wang, Shao, 2013):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (4)$$

where  $\gamma(h)$  denotes semi-variant,

$h$  denotes the lag interval,

$Z$  denotes soil property parameter,

$N(h)$  denotes the number of pairs of places separated by the distance lag  $h$ ,

$z(x_i)$  and  $z(x_i+h)$  denotes the values of  $Z$  at the positions  $x_i$  and  $x_i + h$ .

The interpolation accuracy was determined from the mean error (ME), mean square error (RMSE), average standard error (ASE) and standard error RMSS (5), (6), (7) (8) (Myslyva et al., 2018):

$$ME = \frac{\sum_{i=1}^N (O_i - S_i)}{N} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - S_i)^2}{N}} \quad (6)$$

$$ASE = \sqrt{\frac{\sum_{i=1}^N \sigma^2(S_i)}{N}} \quad (7)$$

$$RMSSE = \frac{RMSE}{\Delta} \quad (8)$$

where  $O_i$  denotes the observed value,

$S_i$  denotes the predicted value,

$N$  denotes the sample size,

$\sigma$  denotes the dispersion,

$\Delta$  denotes the range equal to the difference between the maximum and minimum observable values.

## Discussions and results

The application of the module “Geostatistical Analyst” for spatial modeling of distribution  $pH_{KCl}$  of soil solution provides a preliminary assessment of the initial data for their suitability for modeling purposes. As a result of the use of tools for exploratory analysis of spatial data, a histogram of the

distribution of initial data is created and the form of their distribution is investigated, as well as the basic statistical characteristics of the sample are calculated (Table 1).

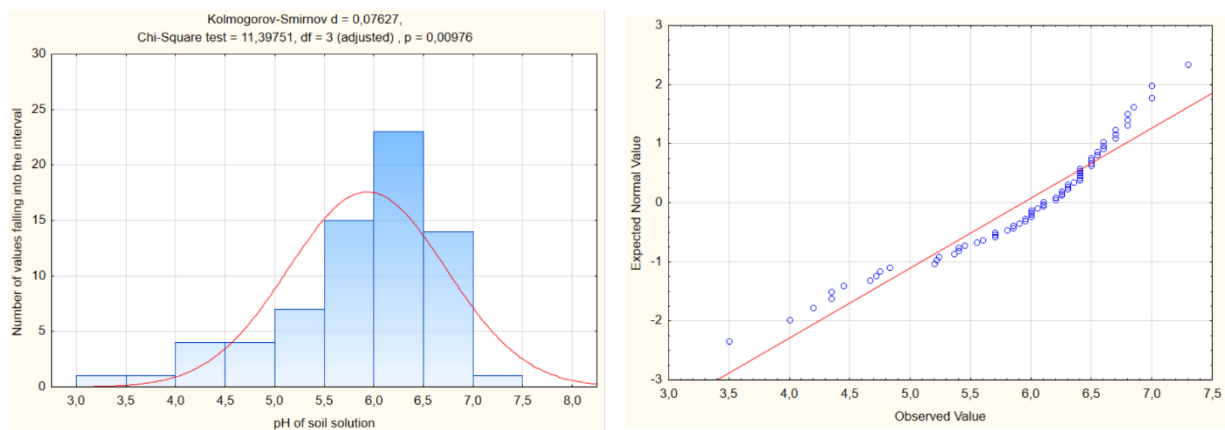
**Table 1**

Statistical characteristics of a sample of data on pH of the soil solution used to construct interpolation models, n = 70

Indicator name	Indicator value			Sd	Cv, %	Med	Kurtosis	Skewness
	min	max	mid					
pH <sub>KCl</sub>	3.50	7.30	5.93	0.79	13.4	6.10	0.68	-1.03

Note: Sd is the standard deviation; Cv is the coefficient of variation; Med is the median.

Preliminary evaluation of the data makes it possible to establish the necessity of carrying out their transformation with the subsequent modeling of the distribution surface. If the data distribution differs significantly from the normal one, you need to convert them. In particular, if the data is asymmetrically distributed, the logarithmic transformation that approximates the distribution to normal is applied to such data. In our case, although the variation coefficient of 13.4 % indicates that the data sample is quantitatively homogeneous, the distribution of data is not normal and requires a logarithmic transformation. In particular, this is evidenced by the values of the Kolmogorov-Smirnov coefficient and the Pearson's acceptance criterion (Fig. 2). Moreover, the magnitude of the coefficient of kurtosis less than zero indicates the flatness of the actual distribution of data compared to normal as well as the value of the coefficient of asymmetry -1.03 indicates the skewness of the distribution of data to the left.



**Fig. 2.** Frequency histogram and normal probability plot of the pH of the soil solution, soil layer 0-20 cm

The Trend Analysis tool of the Geostatistical Analysis module allows to display data in a three-dimensional perspective. The locations of the reference points, which in our case are the locations of the selection of soil samples for agrochemical analysis, are plotted on the x, y plane. A unique feature of this tool is that the values are projected onto the perpendicular planes x-z and y-z in the shape of dispersion diagrams. Then, polynomials are fitted by using scattering diagrams on the projected planes. The line of the best fit (polynomial), drawn through the projected points, shows the trends of data changes in certain directions. In our case, a certain trend for soil pH level is observed both in the direction of the west-east and in the direction of the north-south. Since the trend is U-shaped, it is advisable to use a second-order polynomial as a global trend model for performing interpolation as well as apply the trend removal option for constructing models using the ordinary kriging method.

The experimental anisotropic variograms were calculated to determine the possible spatial structure of pH of the soil solution. The polynomial kernel function and spherical semivariogram model were identified as the best, the type of the circle was standard, the type and the number of sectors was 4 with a displacement of 45°, and the lag was 100 meters. The results of estimating the predictive models generated by the ordinary kriging method are presented in Table. 2.

The best model was chosen based on five criteria: mean error (ME), mean square error (RMSE), mean square normalized error (RMSS), mean standard error (ASE) and the difference between RMSE and ASE. In this study, the spherical variogram model in most cases meets the requirements for the best model.

**Table 2**

Results of cross-validation of semivariogram models

Kernel function	ME	RMSE	RMSS	ASE	RMSE-ASE
Semivariogram model – Spherical					
Exponential	0.0072	0.812	0.955	0.872	-0.060
Polynomial	-0.0012	0.772	1.051	0.747	0.025
Gaussian	-0.0013	0.775	1.076	0.738	0.037
Semivariogram model – Exponential					
Exponential	0.0098	0.814	0.905	0.909	-0.095
Polynomial	0.0113	0.822	0.928	0.899	-0.077
Gaussian	0.0130	0.827	0.942	0.896	-0.069
Semivariogram model – Gaussian					
Exponential	-0.0023	0.831	1.021	0.842	-0.011
Polynomial	0.0030	0.822	0.997	0.857	-0.035
Gaussian	0.0046	0.826	1.006	0.855	-0.029

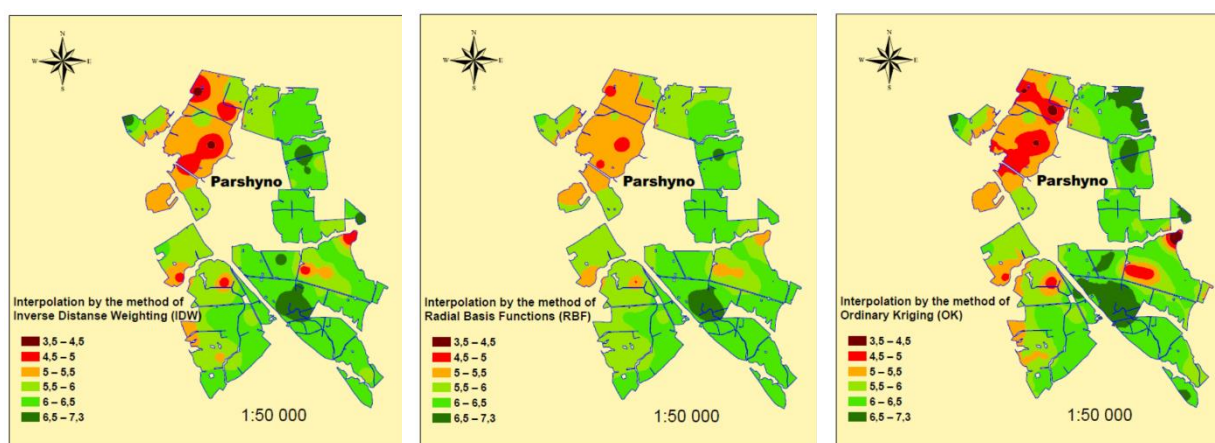
Note: a semivariogram kernel function that allows selecting a function for selection a plane when using kriging

The interpolation methods under study were implemented to estimate the values of the unmeasured data on the pH of the soil solution and to create surfaces from the predicted points. The parameters used in the interpolation models and used to create soil pH prediction maps are presented in Table 3, and the visualization of the prediction results is shown in Fig. 3.

**Table 3**

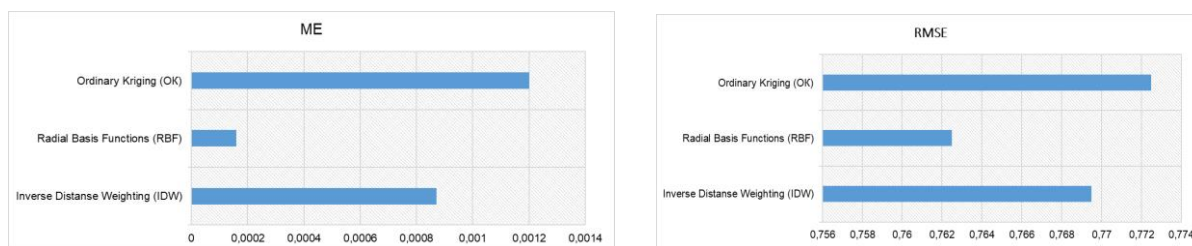
Parameters of interpolation methods used to create soil pH forecast maps  
(type of search area – standard; sector type – 4 sectors with a displacement of 45°; angle – 90°)

Inverse Distance Weighting (IDW)	Radial Basis Functions (RBF)	Ordinary Kriging (OK)
Power parameter: 2.5335 Search area: min = 10; max = 15	Search area: min = 10; max = 15 Kernel function: spline with tension; kernel parameter: 0.0806	Search area: min = 2; max = 15 Kernel function: polynomial; semivariogram model – spherical



**Fig. 3.** Prediction maps of soil pH distribution

Because of the interpolation methods of RBF and IDW during the cross-validation procedure only two criteria are defined – the average error (ME) and the mean square error (RMSE), the comparison of the effectiveness of the methods for soil pH prediction was carried out by these indicators. The most suitable for forecasting purposes was the model created by the method of radial basis functions, the value of the root-mean-square error of which was 0.763 (Fig. 4).



**Fig. 4.** Results of cross-validation of predictive models of the pH of the soil solution

In terms of interpolation accuracy, the investigated deterministic and geostatistical methods are located in the next descending series RBF> IDW> OK (Tab. 4).

**Table 4**

Differences between the actual and predicted spatial distribution of the pH of the soil solution

Range of values	Land area falling in the interval, ha					
	Interpolation method IDW		Interpolation method RBF		Interpolation method OK	
	1	2	1	2	1	2
< 4.5	28.94	26.12	2.11	-0.71	29.45	26.63
4.51–5.0	159.59	117.34	47.54	5.29	246.22	203.97
5.01–5.50	480.89	-91.69	586.21	13.63	515.36	-57.22
5.51–6.0	887.22	-156.99	1021.48	-22.73	797.26	-246.95
6.01–6.50	1488.79	47.21	1449.10	7.52	1278.35	-163.23
6.51–7.0	152.46	58.01	91.45	-3	331.25	236.8

Note: 1 – predicted value; 2 – difference between predicted and actual value.

The IDW method interpolates the surface with sufficient accuracy if the sampling points are relatively normally distributed and the surface is uniform. In our case, none of these conditions is fulfilled: the sampling points are unevenly distributed, clustering takes place, and the steepness of the slopes of the study area varies from 0.5 to 10.0 degrees. Therefore, this method is not suitable enough to predict the spatial distribution of the pH of the soil solution.

Kriging is the optimal interpolation method if the data meets certain conditions: normally distributed, stationary and no trends. Since, in our case, the data are not normally distributed and their transformation is required as well as the presence of a trend in their spatial distribution is established, the use of the ordinary kriging method to predict the spatial distribution of the pH of the soil solution is also impractical.

The RBF method approaches the construction of a surface through measured values while minimizing the total curvature of the surface and is ineffective when an abrupt change in values occurs over short distances. This phenomenon is absent in our case, and the change in the studied indicator is rather smooth, which causes the high efficiency of the RBF method for interpolation. The possibility of using this method to predict the spatial distribution of soil acidity is also indicated in studies carried out by Iranian scientists (Attaeian et al., 2016).

### Conclusions and proposals

Among the interpolation methods that have been studied, the most optimal method for predicting and mapping the spatial distribution of soil pH is the radial basis function method, which provides the root-mean-square error of the average predicted values at 0.763. Both the interpolation method IDW and the RBF method require adjusting the power parameter and determining the optimal search radius to improve the prediction accuracy. Although the IDW method is relatively easy to use and does not provide for the calculation of standard interpolation errors, in our case it provided higher accuracy in predicting the spatial distribution of soil pH compared to ordinary kriging.



The choice and use of the most optimal interpolation method for predicting the spatial distribution of soil pH largely depends on the characteristics of a particular set of experimental data as well as on the spatial characteristics of the territory for which the prediction is performed. Consequently, in each particular case, the selection of the optimal interpolation method should be performed according to the specific conditions and characteristics of the data, and the definition of a single universal interpolation method is not possible.

Further studies should be concentrated in the direction of comparing the results of interpolation obtained for the pH of arable and subsurface soil horizons.

## References

1. Attaeian B. Farokhzadeh, Akhzari D., Artimani M. M. (2016) Comparing interpolation methods for estimating spatial distribution of topsoil pH and EC. *Ecopersia*. Vol. 3(4), p. 1145–1159.
2. Behera S. K., Shukla A. K. (2015) Spatial distribution of surface soil acidity, electrical Conductivity, soil organic carbon content and exchangeable Potassium, calcium and magnesium in some cropped acid Soils of India. *Land Degrad. Deviation*. Vol. 16, p. 71–79.
3. Gouri S.B., Pravat K.S., Ramkrishna M. (2016) Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *Journal of the Saudi Society of Agricultural Sciences*, Vol. 2, p. 1–13.
4. Johnston, K., Ver, Hoef J.M., Krivoruchko, K., Lucas, N., 2001. Using ArcGIS Geostatistical Analyst. ESRI Press, Redlands, CA.
5. Kohan S. S., Moskalenko A. A., Shylo L. G. (2013) Geoinformation providing of qualitative estimation of soils. *East European Magazine of Advanced Technology*. Vol. 6, p. 18–25.
6. Li, X.F., Chen, Z.B., Chen, H.B., Chen, Z.Q. (2011) Spatial distribution of soil nutrients and their response to land use in eroded area of South China. *Proc. Environ. Sci.*, Vol. 10, p. 14–19.
7. Liu L., Wang H., Dai W. (2014) Spatial variability of soil organic carbon in the forestlands of northeast China. *J. Forest. Researches*. Vol. 25(4), p. 867–876.
8. Martin D., Jord A. (2013) Spatial models for monitoring the spatio-temporal evolution of ashes after fire – a case study of a burnt grassland in Lithuania. *Solid Earth*. Vol. 4, p. 153–165.
9. Moskalenko A. A. (2012) Justification of geoinformation system on land soils quality monitoring. *Land system, cadastre and land monitoring*. Vol. 3–4, p. 108–112.
10. Myslyva T.N., Sheluto B.V., Kutsaeva O.A., Naskova S.A. (2018) Use of geospatial analysis methods in land management and cadaster. *Baltic Surveying*. Vol 9(2), p. 56–62.
11. Robinson, T.P., Metternicht, G.M. (2006) Testing the performance of spatial interpolation techniques for mapping soil properties. *Comput. Electeron. Agric.*, Vol. 50, p. 97–108.
12. Samsonova V. P., Blagoveshchenskii Y. N., Meshalkina Y. L. (2017) Use of empirical bayesian kriging for revealing heterogeneities in the distribution of organic carbon on agricultural lands. *Eurasian Soil Science*. Vol. 50(3), p. 305–311.
13. Simbatova A. N., Ryazanov S. S., Sahabiev I. A. (2016) Modeling of spatial distribution of soil organic matter: an overview of modern approaches. *Russian Journal of Applied Ecology*. Vol. 2, p. 48–54.
14. Taha G., Elif S., Aysegul T. (2017) Interpolation Approaches for Characterizing Spatial Variability of Soil Properties in Tuz Lake Basin of Turkey IOP Conf. Ser.: Earth Environ. Sci. 95 062010
15. Wang Y.Q., Shao M.A. (2013) Spatial variability of soil physical properties in a region of the Loess Plateau of PR China subject to wind and water erosion. *Land Degrad. Dev.*, Vol. 24 (3), p. 296–304.
16. Zandi S., Ghobakhlou A., Sallis P. (2011) Evaluation of spatial interpolation techniques for mapping soil pH. Perth, Australia, p. 1153-1159.

## Information about authors

**Tamara Myslyva** is a Doctor of Agricultural Sciences, Professor, Head of the Department of Geodesy and Photogrammetry of Belarusian State Agricultural Academy. Her current area of study surrounds the topics of GIS-analysis, GIS-modelling and forecasting the yield of grain and forage crops using GIS technologies and remote sensing data., e-mail: [byrty41@yahoo.com](mailto:byrty41@yahoo.com) , tel: +375 2992 16839.

**Olesya Kutsaeva** is a Senior lecturer of the Department of Geodesy and Photogrammetry of Belarusian State Agricultural Academy. Her current area of study surrounds the topics of GIS-analysis, photogrammetry and remote sensing. e-mail: [alexa-1982@bk.ru](mailto:alexa-1982@bk.ru) , tel: +375 2933 21631.

**Natallia Krundzikava** is a Senior lecturer of the Department of Cadastre and Land Law of Belarusian State Agricultural Academy. Her current area of study surrounds the topics of GIS-analysis, photogrammetry and remote sensing. e-mail: [krundikova.nata@gmail.com](mailto:krundikova.nata@gmail.com) , tel: +375 2934 53710.