Вестник Томского государственного университета. Биология. 2016. № 4 (36). С. 21–33

УДК 631.4
doi: 10.17223/19988591/36/2

SS Ryazanov, IA Sahabiev

*Research Institute for Problems of Ecology and Mineral Wealth Use,
Tatarstan Academy of Sciences, Kazan, Russian Federation*

Comparison of terrain-based drift models to improve the quality of soil predictive mapping at a field scale

The ecological, economic, and agricultural benefits of accurate interpolation of spatial distribution patterns of soil properties are well recognized. In the present study different approaches to build the drift model for the regression kriging are analyzed and compared for estimating the spatial variation of humus and physical clay at soil depth (0-20 cm) in Tatarstan, Russian Federation. The soil sampling was performed according to an agrochemical sampling design: the field was divided into 60 sections; within each section 12-15 sampling points were taken using a hand auger at the depth of 10-20 cm to produce one mixed sample. Three terrain-based drift models: principal component regression (PCR), partial least squares (PLS), and random forest were used to predict the spatial distribution of humus and physical clay. Cross-validation was applied to evaluate the accuracy of interpolation methods through mean error (ME), root mean square error (RMSE), root mean square standardized error (RMSSE), and ratio of the observed and the predicted variances (RVar). The results indicate that ordinary kriging (OK) is superior when the data have strong spatial dependence. But in other cases, the PLS approach had the best prediction performance.

The article contains 4 Figures, 5 Tables, 29 References.

Keywords: spatial interpolation; prediction; geostatistics; regression kriging; humus; physical clay.

Introduction

Spatial variability of soil properties is an important indicator of soil quality, and it is important in ecological modeling, environmental prediction, precision agriculture, and natural resource management [1]. Revealing the characteristics of spatial patterns will provide the basis for evaluating soil fertility, and assist in the development of sound agricultural management policies. So, there is a need for adequate information about spatio-temporal behavior of soil properties over a region and accurate interpolation at unsampled locations is needed for better planning and management.

In general, there are two major approaches to predict soil properties at unsampled location. Methods of “classic” statistics use linear and non-linear regression models to predict dependent variable using auxiliary data. Remote sensing data, topographic and morphologic attributes, climate, land-use and

geology are auxiliary parameters commonly used for the calibration of predictive models. For example, Rodriquez-Lado and Martinez-Cortizas used multiple linear regression, e.g. principal component regression and partial least squares, for modeling and mapping organic carbon content of topsoil using climatic and geological data as independent variables [2].

The second approach is geostatistics, which has been rapidly developing for last decades [3; 4]. Geostatistics is an efficient method for studying spatial allocation of soil characteristics and their inconsistency and reducing the variance of assessment error and execution costs [5]. Geostatistical methods model the local uncertainty about the attribute value at any particular location through the set of possible realizations of the random variable at that location [6]. Earlier researchers, who applied geospatial techniques to evaluate geographical changeability of soil characteristics, reported that ordinary kriging in most cases was the best method for prediction of the spatial distribution of soil properties [7; 8].

And there is the third, hybrid approach that uses advantages of the first two. And the typical example is regression kriging (RK) that uses regression models to explain deterministic part of spatial variation using auxiliary data and kriging technique to interpolate the residuary, stochastic part of spatial variation. In RK, the deterministic part can be explained using various statistical techniques. Many authors suggest the relative accuracy advantage of the RK compared to OK, and this prediction performance depends on the relationship between the target variable and the explanatory co-variables [9; 10]. The present study was undertaken to compare the accuracy of various approaches to model the deterministic part of regression kriging.

Materials and methods

2.1. Study area and sampling design

The study was carried out in the national crop testing field (CTF) that is located in the southeastern part of the Republic of Tatarstan (Russian Federation, 55°05’56.0”N 52°02’24.0”E). The relief of the field is flat in the northern part and changes to the gentle slope in the southern and south-eastern parts. The soil cover is represented by leached, silt loamy chernozems with varying rates of erosion. The most eroded chernozem is located in the eastern and south-eastern parts of the field. The soft eroded soils are located in the northern part of the CTF. Parent rocks are represented by fine loamy and clayey calcareous deluvium, underlined by the ancient alluvial deposits in the eastern part. Particle size distribution, according to the Russian classification by NA Kachinsky, is fine loamy [11].

The soil sampling was performed according to an agrochemical sampling design: the field was divided into 60 sections; within each section 12-15 sampling points were taken using a hand auger at the depth of 10-20 cm to produce one mixed sample (Fig. 1).

For geostatistical analysis the mixed samples were georeferenced into centers of the corresponding sections. The following soil properties were measured: humus

content by the Tyurin method and the particle size distribution by the Kachinsky-Robinson-Kehl pipet method [12]. As an indicator of the particle size distribution, the sum of particles <0.01 mm was used. In Russian classification, this range of particle sizes is called "physical clay", and this term is used henceforward.

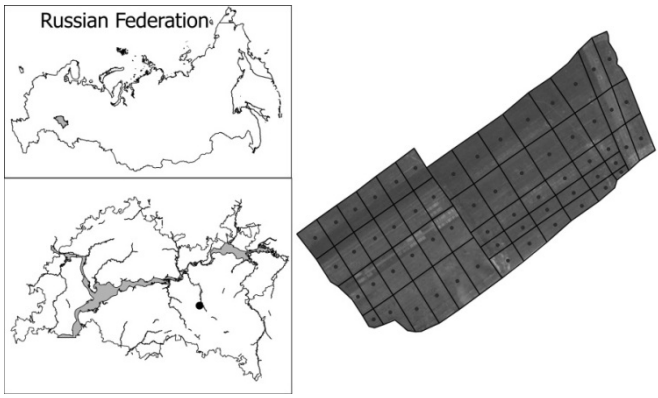


Fig.1. Study area showing the location of the sampling points

2.2. Terrain variables

Terrain analysis was based on a 30-m grid digital elevation model (DEM). In order to model the distribution of soil properties, we considered a set of 34 GIS-based geographic covariates in the form of raster maps. As many terrain indices as possible were calculated because a large set of predictors can compensate unaccounted variables [13]. A full set of used terrain indices can be found in Fig 3. All terrain variables were averaged within the section of the mixed sample. The terrain analysis was performed using SAGA GIS software [14].

2.3. Interpolation techniques

2.3.1. Ordinary kriging

OK is a geostatistical interpolator of the kriging family. Here the predictions are based on the model:

$$Z(s) = \mu + \varepsilon'(s), \tag{1}$$

where μ is the stationary mean of the local neighborhood; ε'(s) is the spatially correlated stochastic part of variation [6].

The spatial structure is quantified by the experimental variogram, which is the plot of the semivariance γ (2) against the distance between points. Semivariance is an important concept in geostatistics; it represents the differences between the neighboring values:

$$\gamma(h) = \frac{1}{2} E[(z(s_i) - z(s_i + h))^2], \tag{2}$$

where $z(s_i)$ is the value of a target variable at a sampled location, $z(s_i+h)$ is the value of the neighbor at distance h [15].

The kriging procedure gives weightings to sample data based on the relationship between distance and semivariance [16]. The ordinary kriging results were used as a baseline to assess the performance of other interpolation techniques.

2.3.2. RK with Principal Component Regression

The regression kriging is similar to OK and the main difference is that the stationary mean is replaced with a linear function of secondary variables. These secondary variables should be available at all primary data points and at all points of the region of interest [17].

$$Z(s) = m(s) + \varepsilon'(s) + \varepsilon'' \quad (3)$$

In the regression kriging the drift function and residuals can be fitted separately and then summed afterwards [18; 19]. Because of this appears a big advantage of the RK over other methods of spatial interpolation: it can be easily combined with different statistical modeling techniques [20].

Hengl et al. proposed a generic framework for spatial interpolation based on regression kriging [21]. This approach predicts a soil property at unvisited location by summing the predicted drift $m(s)$ and residuals that are interpolated using ordinary kriging. In this article the different drift modeling approaches were used.

During the calibration of linear models, multicollinearity often occurs when using a large set of auxiliary data, leading to problems of overfitting. PCR was proposed as an effective technique when the evidence of multicollinearity exists. The principal component regression approach is based on the principal component analysis. It involves the summarizing of auxiliary data into new uncorrelated principal components, and then using these components as the predictors in a linear regression model that is fit using least squares [22]. These components are linear combinations of the original covariates, and retain the maximum amount of their variability. Then, the scores from the PCA decomposition are regressed onto the response data for prediction purposes.

2.3.3. RK with Partial Least Squares

Partial Least Squares is a technique similar to PCR. Like PCR, PLS combines features from principal component analysis and multiple linear regression. But PLS uses the response in order to identify new components that not only approximate the auxiliary data well, but are also related to the response [23].

2.3.4. RK with Random Forest

Random Forest (RF) is the special case of bootstrap aggregation of regression trees. In RF, a number of regression trees are built on bootstrapped training samples. When building these decision trees, each time a split in a tree is considered, a

random sample of *m* predictors is chosen as split candidates from the full set of *p* predictors. The split is allowed to use only one of those sampled predictors [22]. This tweak decorrelates the bootstrapped regression trees and makes the average of the resulting trees less variable and more reliable. The optimal value of *m* was chosen with respect to the Out-of-Bag error estimate [24].

2.4. Accuracy assessment

The spatial interpolation models were compared using leave-one-out cross-validation (LOOCV). In LOOCV each sampling point o_i is removed sequentially. The spatial interpolation model is fit on n-1 observation, and a prediction p_i is made for the excluded observation, using its X values. Several error measurements were calculated using the difference $o_i - p_i$:

Mean error is given by

$$ME = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i) \tag{7}$$

Root mean square error is given by

$$RMSE = [\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2]^{1/2} \tag{8}$$

Root mean square standardized error is given by

$$RMSSE = [\frac{1}{n}\sum_{i=1}^{n}(p_{si} - o_{si})^2]^{1/2} \tag{9}$$

Ratio of the observed and the predicted variances is given by

$$RVar = \frac{Var[p]}{Var[o]} \tag{10}$$

Li and Heap made a review of several criteria for using error measurements to judge the performance of the spatial interpolation methods [25]. The model is better if ME is closer to zero and RMSE is smaller. RMSSE should be close to 1. If RMSSE>1, the method underestimates the depended variable, and if RMSSE<1, it overestimates the depended variable. The closer RVar is to 1, the better the ability of a spatial interpolation method to preserve the observed variance.

2.5. Software

All statistical analyses here presented were performed within the statistical environment R [26]. The PCR and PLS approaches were performed using the "pls" package [27]. The random forest models were fitted using the "randomForest" package [28].

Results and discussion

3.1. Spatial structure of the raw data

The investigated field was characterized by a high content of humus with high variability. The density function skewed toward higher values. The particle size distribution of the topsoil was fine loamy with a low coefficient of variation (Table 1). Both, humus and physical clay content, had nonnormal distribution of values, but, because of no effect on the experimental variograms, no data transformation were performed.

Table 1

Summary statistics of raw data

Variable	min	Mean	Median	max	var	skewness	kurtosis
Humus	3.95	6.86	6.87	8.05	0.73	-0.84	3.72
Ph. clay	35.47	50.72	52.93	58.17	28.37	-1.21	3.53

The presence of spatial anisotropy was assessed using the variogram maps and directional variograms with a horizontal tolerance of ±20°. Parameters of the fitted variogram models are presented in Table 2.

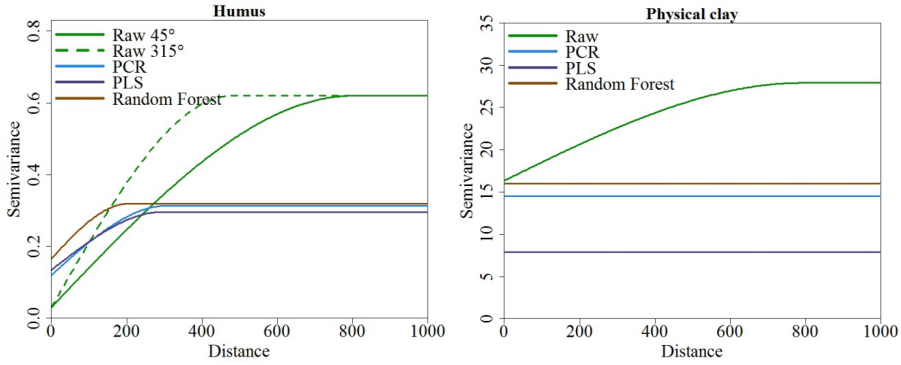
Table 2

Parameters of the fitted variogram models

Variable	Model	SSErr	Range	C0	C1	C0+C1	C0/(C0+C1)
Raw data							
Humus	Sph	4.6E-05	800.5*	0.03	0.59	0.6	0.05
Ph. clay	Sph	3.9E-02	791.9	16.36	11.55	27.9	0.59
PCR residuals							
Humus	Sph	1.7E-07	307.0	0.12	0.19	0.3	0.38
Ph. clay	Nug	-	-	14.50	-	-	1
PLS residuals							
Humus	Sph	2.0E-07	296.5	0.13	0.16	0.3	0.45
Ph. clay	Nug	-	-	7.90	-	-	1
Random Forest residuals							
Humus	Sph	3.1E-06	205.6	0.17	0.14	0.3	0.54
Ph. clay	Nug	-	-	16.00	-	-	1

** The range for the main direction* (45°).

The spatial structure of the humus content showed the presence of a geometric anisotropy, the direction of which corresponds to the sampling grid orientation (45°). The autocorrelation range in a minor direction (315°) is 0.6 times less. The nugget/sill ratio showed the high spatial dependence of the humus values [29]. The physical clay content had an isotropic spatial structure with a medium spatial dependence. The fitted variograms were used to perform the ordinary kriging interpolation (Fig. 4, *a, b*).



a *b*

Fig. 2. The fitted variogram models. Green line - Raw data; blue - Residuals after the PCR; purple - Residuals after the PLS; brown - Residuals after the random forest

3.2. RK with Principal Component Regression

The principal component analysis effectively deals with high collinearity in high dimensional data. After the PCA, seven principal components (PC) with the cumulative proportion of variance of 93% were selected for the following regression modeling. The number of PCs was selected using the screeplot. The regression models were fitted on the principal components using the forward stepwise selection approach, therefore only significant PCs were included. As shown in Table 3, the first two PCs explained 54% of the humus variation; and the first, the second and the fourth PC explained 47% of the physical clay variation.

Table 3

Parameters of the PCR models

Response	Explanatory	Coefficient	Adj. R-sq.
Humus	Intercept	6.86	0.54
	PC1	0.09	
	PC2	0.18	
Ph. Clay	Intercept	50.72	0.47
	PC1	0.40	
	PC2	0.99	
	PC4	-1.14	

Inclusion of the terrain variables into the model of regionalization explained the part of spatial variation of the response, which was reflected on the experimental variograms of the model residuals (Fig. 2, *a*). The PCR drift model eliminated the geometric anisotropy and reduced the autocorrelation range and the overall spatial dependence of the humus values (Table 2). In the case of physical clay, relief-based PCR model fully explained the spatially dependent variation. The resultant variogram had a pure "nugget-effect" form (Fig. 2, *b*). The final humus

and physical clay maps were produced as the sum of the PCR function and the interpolated residuals (Fig. 4, *c, d*).

3.3. RK with Partial Least Squares

Like PCR, partial least squares compute the latent vectors (LV) which are linear combinations of the original predictors. The number of LV was determined by the RMSE, obtained by leave-one-out cross-validation. The lowest cross-validation error occurred when only one LV was used to predict the humus content and when six LVs was used to predict the physical clay content. In comparison with PCR, the PLS models explained higher amount of the response variations (Table 4). This is because the PLS, in contrast to PCR, searches for directions that explain variance in both the predictors and the response.

Table 4

Parameters of the PLS models

Response	Explanatory	Coefficient	Adj. R-sq.
Humus	intercept	6.86	0.57
	LV1	0.21	
Ph. Clay	intercept	50.72	0.69
	LV1	1.17	
	LV2	1.11	
	LV3	0.38	
	LV4	0.98	
	LV5	0.95	
	LV6	1.94	

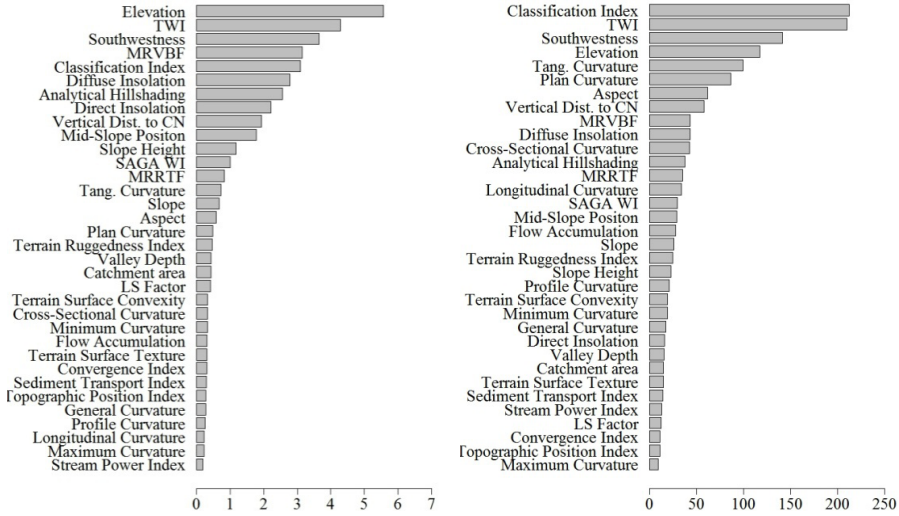
The PLS drift models was also reflected on the experimental variograms of the residuals. When compared with the variograms of the PCR residuals a strong decrease in the random spatial variability was observed.

3.4. RK with Random Forest

The random forest regression was performed with 10000 bagging sampling iterations, to ensure that every input row got predicted at least a few times. The number of variables randomly sampled as candidates at each split was set to 11.

The results of the random forest technique is less interpretable than the results of ordinary regression trees, nevertheless it is possible to obtain an overall summary of the importance of each predictor (Fig. 3). The variable importance graph shows the total decrease in node impurities from splitting on the variable, averaged over all trees. Impurity was measured as the total amount that the RSS was decreased due to splits over a given predictor, averaged over all 10000 trees.

The residuals showed large variances relatively to the over drift modeling approaches (Table 2). Random Forest drift model effectively eliminated the spatial dependence of the physical clay data. It also resulted in a great explanation of the spatial dependence of the humus (Fig 2).



a. Humus *b*. Physical clay

Fig. 3. Variable importance as measured by a random forest

3.8. Performance of spatial interpolation approaches

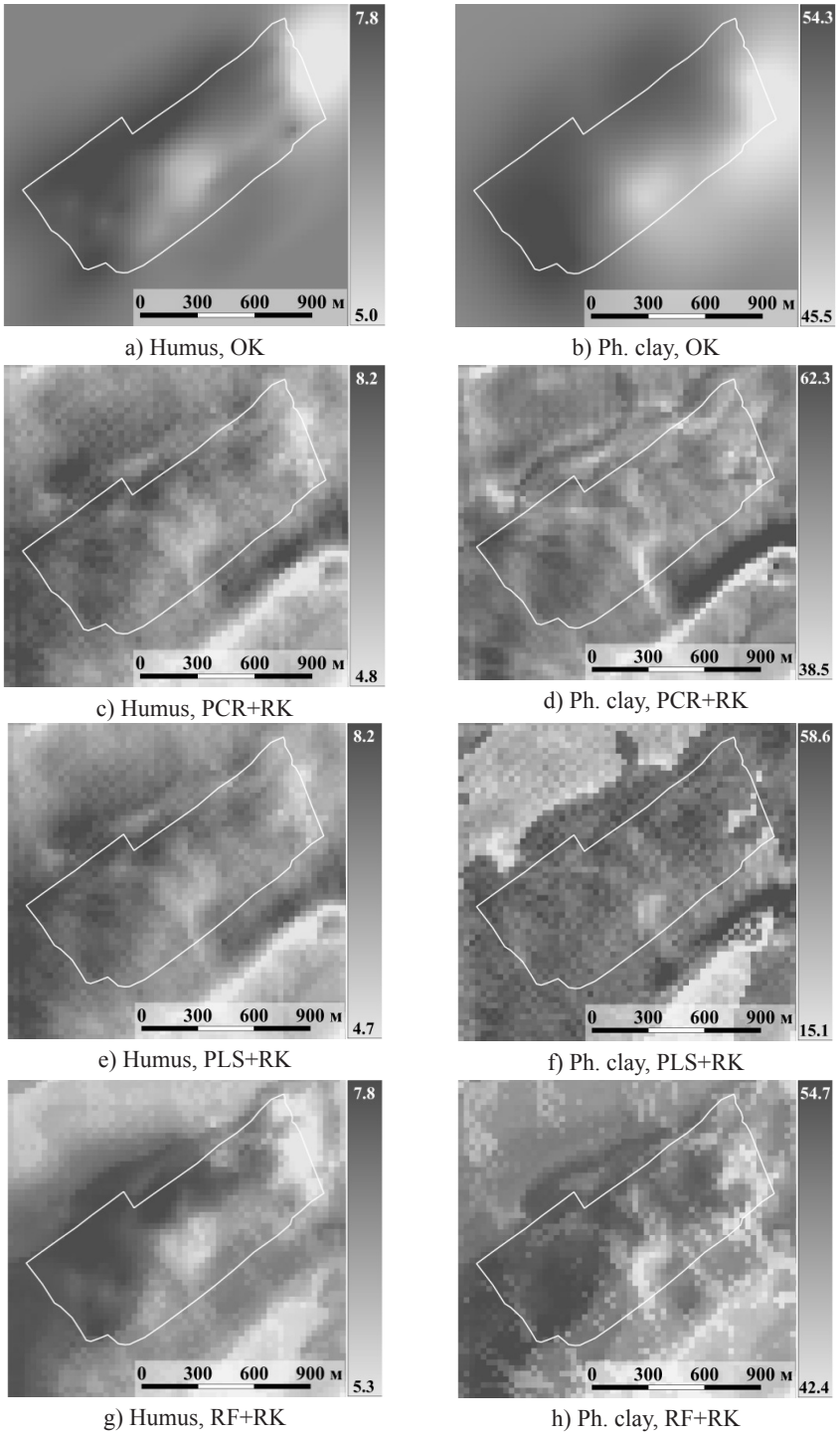
Because of the lack of additional sampling points, the accuracy of the interpolation approaches was assessed using internal cross-validation. By the ratio of the predicted and the observed variance of the physical clay, there was a doubling of the preserved variability in the case of PCR, relative to the ordinary kriging. The PLS drift model with the regression kriging showed the best ability to preserve the observed variance (Table 5). Inclusion of the relief information conversely smoothed the variability of the humus values.

Table 5

Cross-validation results

Variable	ME	RMSE	RMSSE	RVar
OK				
Humus	-0.015	0.476	0.579	0.626
Ph. Clay	0.007	4.104	0.837	0.257
PCR+RK				
Humus	-0.005	0.513	0.632	0.622
Ph. Clay	-0.024	4.017	0.828	0.494
PLS+RK				
Humus	-0.006	0.543	0.674	0.607
Ph. Clay	0.024	3.994	0.778	0.841
randomforest+RK				
Humus	-0.004	0.574	0.719	0.586
Ph. Clay	0.062	3.981	0.818	0.368

In the case of humus all methods overestimated the observed values, and the bias reduces in the following order: OK>PLS>PCR>RF. But all the drift-models increased the RMSE of the predictions.



a) Humus, OK b) Ph. clay, OK

c) Humus, PCR+RK d) Ph. clay, PCR+RK

e) Humus, PLS+RK f) Ph. clay, PLS+RK

g) Humus, RF+RK h) Ph. clay, RF+RK

Fig. 4. Interpolated maps with different drift models

In the case of the physical clay, the highest bias observed for the random forest drift model, and the lowest for the OK model. Root mean square error reduced in the order: OK>PCR >PLS>RF, although the OK showed the lowest overestimation rate.

Although, the RF approach showed the lowest RMSE, the interpolated map obtained by this method was very noisy (Fig. 4, *h*). In contrast, the OK produced a very smoothed surface with the lowest ratio of preserved variance. This is a well-known feature of the OK interpolator in the presence of strong "nugget effect". Apparently, the PLS was the best approach to predict the physical clay content. With only slight increase in bias, the PLS prediction had the highest RVar and the second lowest error rate. Withal, the PLS residuals showed the lowest random variation among the five methods (Fig. 2, *b*). The map, obtained by the PLS was not so noisy as the RF case and contained a well-detailed spatial variation related to the terrain.

The raw humus data had very low "nugget effect" and very strong spatial dependence (Fig. 2, *a*). Inclusion of the drift models, cleared the spatial dependence and removed the geometric anisotropy, but, at the same time, the residuals of the all three drift models showed the higher amount of random, unexplained variation. This explained the highest performance of the OK method, which showed the highest amount of the preserved variance, and the lowest RMSE. Because of accurate variogram modeling, the humus map, interpolated by the OK, contained information about local variation (Fig. 4, *a*). The maps, produced by the regression kriging with terrain-based drift model, can still be useful, as they well reflected the humus erosion path in the central part of the field (Fig. 4, *c, e, g*).

Conclusion

The study shows that OK interpolator is superior than regression kriging with various drift models if the original data have a strong spatial dependence with low rate of "nugget effect". And vice versa, if the data have high amount of random variation, the inclusion of auxiliary data can increase the prediction performance. Among the three drift models, the PLS method is the most optimal to use within regression kriging. A set of significant auxiliary variables and corresponding parameters of the relief models is unique for different mapping sites and depends on the content and spatial structure of the target soil property, terrain heterogeneity of the field, etc. Nevertheless, our findings on the performance of the interpolation techniques are applicable for other sites. Finally, the results guide to the amplification of trustworthy maps of soil properties which can significantly contribute to proper application of agricultural modeling.

References

1. Wei JB, Xiao DN, Zeng H, Fu YK. Spatial variability of soil properties in relation to land use and topography in a typical small watershed of the black soil region, northern China. *Environ Geol*. 2008;53:1663-1672. doi: 10.1007/s00254-007-0773-z

2. Rodriques-Lado L, Martinez-Cortizas A. Modelling and mapping organic carbon content of topsoil in an Atlantic area of southwestern Europe (Galicia, NW-Spain). *Geoderma*. 2015;245-246:65-73. doi: 10.1016/j.geoderma.2015.01.015
3. McBrantey AB, Mendoca Santos ML, Minasny B. On digital soil mapping. *Geoderma*. 2003;117:3-52. doi: 10.1016/s0016-7061(03)00223-4
4. Minasny B, McBratney AB. Methodologies for global soil mapping. In: *Digital Soil Mapping. Bridging Research, Environmental Application and Operation.* Boettinger JL, Howel DW, Moore AC, Hartemink AE, Kienast-Brown S, editors. Dordrecht: Springer Netherlands; 2010. pp. 429-436. doi: 10.1007/978-90-481-8863-5
5. Behera SK, Shukla AK. 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. Develop*. 2015;26:71-79. doi: 10.1002/ldr.2306
6. Goovaerts P. Geostatistics for Natural Resources Evaluation (Applied Geostatistics). New York: Oxford University Press; 1997. 497 p.
7. Zare-Mehrjardi M, Taghizadeh-Mehjardi R, Akbarzadeh A. Evaluation of geostatistical techniques for mapping spatial distribution of soil pH, salinity and plant cover affected by environmental factors in Southern Iran. *Not Sci Biol*. 2010;2:92-103.
8. Gouri SB, Pravat K, Ramkrishna M. Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *Journal of the Saudi Society of Agricultural Sciences*. 2016. In press. doi: 10.1016/j.jssas.2016.02.001
9. Zhang S, Huang Y, Shen C, Ye H, Du Y. Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. *Geoderma*. 2012;171-172:35-43. doi: 10.1016/j.geoderma.2011.07.012
10. Li J, Heap A. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance an impact factor. *Ecological Informatics*. 2011;6(3-4):228-241. doi: 10.1016/j.ecoinf.2010.12.003
11. Shein EV. Kurs fiziki pochv [Soil physics]. Moscow: Moscow State University Publ.; 2005. 432 p. In Russian
12. Yagodin BA, Deryugin IP, Zhukov YuP, Demin VA, Peterburgskiy AV, Kidin VV, Slipchik AF, Kulyukin AI, Sablina SM. Praktikum po agrokhimii [Manual on agrochemistry]. Moscow: Agropromizdat Publ.; 1987. 512 p. In Russian
13. Miller BA, Koszinski S, Wehrhan M, Sommer M. Impact of multi-scale predictor selection for modeling soil properties. *Geoderma*. 2015:239-240:97-106. doi: 10.1016/j.geoderma.2014.09.018
14. Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Boehner J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev*. 2015;8:1991-2007. doi: 10.5194/gmd-8-1991-2015
15. Webster R, Oliver M. Geostatistics for Environmental Scientists. Chichester: John Wiley & Sons, Ltd; 2001. 271 p.
16. Robinson TP, Metternicht G. Comparing the performance of techniques to improve the quality of yield maps. *Agricultural Systems*. 2005;85:19-41. doi: 10.1016/j.agsy.2004.07.010
17. Hengl T. A Practical Guide to Geostatistical Mapping. Amsterdam: University of Amsterdam Publ.; 2009. 293 p.
18. Odeh I, McBratney A, Chittleborough D. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*. 1994;63:197-214. doi: 10.1016/0016-7061(94)90063-9
19. Odeh I, McBratney A, Chittleborough D. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*. 1995;67:215-226. doi: 10.1016/0016-7061(95)00007-B
20. McBratney A, Odeh I, Bishop T, Dunbar M, Shatar T. An overview of pedometric techniques of use in soil survey. *Geoderma*. 2000;97:293-327. doi: 10.1016/S0016-7061(00)00043-4

21. Hengl T, Heuvelink G, Stein A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*. 2004;120:75-93. doi: 10.1016/j.geoderma.2003.08.018
22. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning with applications in R. New York: Springer-Verlag; 2013. 440 p. doi: 10.1007/978-1-4614-7138-7
23. Mevik B, Wehrens R. The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software*. 2007;18:2:1-23. doi: 10.18637/jss.v018.i02
24. Breiman L. Random Forests. *Machine Learning*. 2001;45:5-32. doi: 10.1023/A:1010933404324
25. Li J, Heap AD. A Review of spatial interpolation methods for environmental scientists. *Geoscience Australia, Record 2008/23*. 137 p.
26. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. [Electronic resource]. Available at: http://www.R-project.org/ (accessed 05.10.2016)
27. Mevik B-H, Wehrens R, Liland KH. Partial Least Squares and Principal Component Regression. R package version 2.5-0.; 2015. [Electronic resource]. Available at: https://cran.r-project.org/web/packages/pls/index.html (accessed 05.10.2016)
28. Liaw A, Wiener M. Classification and Regression by random Forest. *R News*. 2002;2(3):18-22.
29. Cambardella C, Moorman T, Novak J, Parkin T, Karlen D, Turco R, Konopka A. Field-Scale Variability of Soil Properties in Central Iowa Soils. *Soil Sci. Soc. Am. J*. 1994;58:1501-1511. doi: 10.2136/sssaj1994.03615995005800050033x

Received 3 August 2016; Revised 7 October 2016;
Accepted 19 October 2016; Published 13 December 2016

Authors info:
Ryazanov Stanislav S, Research Fellow, Research Institute for Problems of Ecology and Mineral Wealth Use, Tatarstan Academy of Sciences, Laboratory of Soil Ecology, 28 Dayrskaya Str., Kazan 420087, Russian Federation.
E-mail: erydit@yandex.ru
Sahabiev Ilnaz A, Research Fellow, Research Institute for Problems of Ecology and Mineral Wealth Use, Tatarstan Academy of Sciences, Laboratory of Soil Ecology, 28 Dayrskaya Str., Kazan 420087, Russian Federation.
E-mail: ilnassoil@yandex.ru

Ryazanov SS, Sahabiev IA. Comparison of terrain-based drift models to improve the quality of soil predictive mapping at a field scale. *Vestnik Tomskogo gosudarstvennogo universiteta. Biologiya – Tomsk State University Journal of Biology*. 2016;4(36):21-33. doi: 10.17223/19988591/36/2