Digital soil mapping as a basis for climatically oriented agriculture a thematic on the territory of the national crop testing fields of the Republic of Tatarstan, Russia

I A Sahabiev[1], K G Giniyatullin[1], S S Ryazanov[2]

[1]Institute of Environmental Sciences, Kazan Federal University, Kremlevskay str., 18, 420008, Kazan, Russia
[2]Institute for Problems of Ecology and Mineral Wealth Use, Tatarstan Academy of Sciences, Kazan, Russia
E-mail: ilnassoil@yandex.ru

Abstract. The concept of climate-optimized agriculture (COA) of the UN FAO implies the transformation of agriculture techniques in conditions of changing climate. It is important to implement a timely transition to the concept of COA and sustainable development of soil resources, accurate digital maps of spatial distribution of soils and soil properties are needed. Digital mapping of soil humus content was carried out on the territory of the national crop testing fields (NCTF) of the Republic of Tatarstan (Russian Federation) and the accuracy of the maps obtained was estimated.

1. Introduction

The future of the world's agrifood system is linked to the concept of climate-smart agriculture that is based on an integrated approach to transforming agricultural systems in order to effectively support sustainable development and ensure food security in a changing climate [1]. To achieve the goals of ensuring food security and agricultural development, adaptation to climate change and reduction of emission intensity for each yield output will be required. Such a transformation must be carried out without depletion of the natural resource base, which in turn is possible with the improvement of approaches and agrotechnologies. Such technologies aimed at implementing the concept of CSA are called smart technologies.

The Food and Agriculture Organization of the United Nations (FAO) developed the concept of CSA in 2013 [2]. Since then, smart technologies have been gaining increasing popularity in agroecology, agriculture and soil science around the world. The role of "smart" technologies in the ecologization of farming systems, agroecological monitoring, improving the quality of products and utilization of agricultural waste is emphasized in Russia [3].

In the concept of CSA, the management of soil resources is presented as strategies that conserve carbon in the soil, reduce greenhouse gas emissions and promote intensification of production, while strengthening the natural resource base [1]. In this regard, knowledge of the state of soils is fundamental for making decisions about management practices that contribute to the rational land use in view of climate change. The FAO concept emphasizes that "... ideally, soil information should be available in the form of continuous maps ..." [3]. Nevertheless, obtaining information about soils and their properties in the form of maps is a non-trivial task, since it involves a number of difficulties, among which the key one is the high cost of soil surveys. At the same time, it should be understood

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

that with the help of single data not realized in the form of cartographic material, it is practically impossible to make managerial decisions, both in traditional agriculture and in CSA.

Along with traditional soil maps, digital soil maps based on the concept of digital soil mapping (DSM) can be used for CSA purposes. In the last decade, the DSM has received rapid development in soil science and has been singled out as a separate sub-discipline [4]. The essence of the DSM comprises the prediction of soils and their properties on the basis of spatially distributed quantitative characteristics of soil formation factors [5]. Maps based on the DSM concept allow to estimate the accuracy of mapping, which can be improved by using various statistical methods (variants of regression analysis, regression trees, random forests, etc.) and approaches (geostatistics, hybrid methods, linear mixed modeling, etc.).

2. Material and methods

Taking into account the foregoing, digital maps of humus content were created for the territory of two national crop testing fields (Arsky and Zainsky NCTF) of the Republic of Tatarstan (Russian Federation). The territory of the NCTF was chosen as an object of research for several reasons: 1) for a long time the territory of the NCTFs was subjected to intensive agricultural use; 2) NCTFs use scientifically grounded agrotechniques, thanks to which sorts and hybrids of crops studies, and evaluates for their cultivation stability.

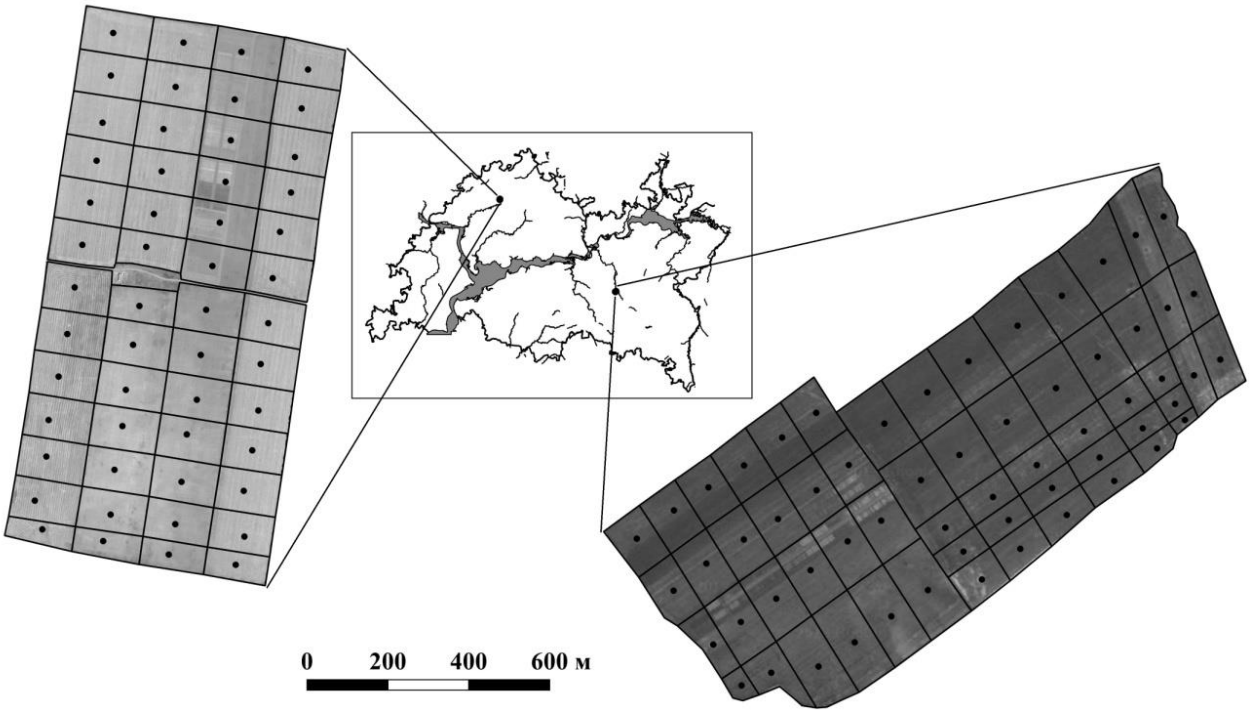


Figure 1. Location of the NCTFs and sampling design.

Zainsky NCTF is located in the south-eastern part of the Republic of Tatarstan (figure. 1). The territory of the test plot occupies an area of 80 hectares. The relief of the field is flat in the northern part that changes to the gentle slope in the southern and south-eastern parts. The soil cover is represented by leached, silt loamy chernozems (Luvic Chernozems (Pachic)) 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 heavy loamy and clayey calcareous deluvium, underlined by the ancient alluvial deposits in the eastern part. Texture is heavy loamy.

Arsky NCTF is located in the northwestern part of the Republic of Tatarstan (figure 1). The testing plot occupies an area of 83 hectares. The soil cover consists of sod-podzolic (Glossic Albic Dystric Retisols (Loamic, Cutanic, Ochric)) and light gray forest soils (Eutric Retisols (Loamic, Cutanic, Ochric)). The granulometric composition of the soils is also heavy loamy. The relief of the testing field is represented by a weak gentle slope in a watershed area. The northern part of the NCTF is on an elevated part of relief, which gradually decreases in the southern direction.

Digital maps of the humus content were created using geostatistical approach. For the purpose of spatial interpolation the regression kriging (RK) was used. The deterministic part of spatial variation was modeled using LASSO- and ridge-regression. The variogram analysis was used for specification of the stochastic part of spatial variability. The ordinary kriging results were used as a baseline to assess the performance of other interpolation techniques.

The relief attributes that have potential impact to spatial humus distribution were used as explanatory variables. The relief attributes were calculated from SRTM1 (spatial resolution – 30 m. (1 arc second)) after preliminary smoothing filtration. The averaged values of the relief attributes within elementary sampling section were assigned to the middle point of the corresponding section.

The use of the ridge and lasso regression was due to the fact that a large set of multicollinear auxiliary variables in the form of relief attributes (32 parameters) was used to construct a digital maps of humus content.

Ridge regression is the implementation of the classical L2-regularization, where into the RSS loss function, that minimizes using the least squares to estimate the model coefficients, the additional term λ is added [6]. This term depends on the estimated coefficients. Thus, instead of RSS, the following expression is minimized:

$$L_2 = RSS + \lambda \sum_{i=1}^{n} \beta_i^2 \quad (1)$$

This introduces a penalty for excessive values of β_i, and the value of this penalty is proportional to the value of the λ. The λ term is used to adjust the model fitting algorithm. If $\lambda \to 0$, the regularized solution tends to the solution of least squares, that is a simple linear regression. If $\lambda \to \infty$, excessive regularization leads to a degenerate solution of $\beta \to 0$. To find the optimal value of the λ the generalized cross-validation (GCV) is used:

$$GCV = \frac{1}{n} \sum \left(\frac{y_i - \hat{y_i}}{1 - \frac{tr(H)}{n}} \right)^2 \quad (2)$$

where tr(H) is the effective degrees of freedom.

LASSO stands for least absolute shrinkage and selection operator. The method is similar to ridge regression. The model coefficients is also estimated on the condition of minimizing the loss function taking into account the penalty, but the penalty parameter is calculated somewhat differently:

$$L_1 = RSS + \lambda \sum_{i=1}^{n} |\beta_i| \quad (3)$$

This leads to qualitatively different estimation of the coefficient vector during the optimization. In addition to the uniform adjustment of all set of β, some of the coefficients become equal to 0, which allows to perform a selection of significant predictors [7].

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

$$ME = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i) \quad (4)$$

Root mean square error is given by

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2\right]^{\frac{1}{2}} \quad (5)$$

Ratio of the observed and the predicted variances is given by

$$RVar = \frac{Var[p]}{Var[o]} \quad (6)$$

The model is better if ME is closer to zero and RMSE is smaller. RMSSE should be close to 1. The closer RVar is to 1, the better is the ability of a spatial interpolation method to preserve the observed variance.

The calculation and analysis of the relief attributes were conducted using SAGA GIS [8]. All statistical analyses presented in the article were performed within the statistical environment R [9]. Variogram analysis and spatial interpolation were made using the "gstat" package [10]. The final maps were formed using QGIS [11].

3. Results and discussion

Variogram analyses shows that the crop testing fields have strong differences in the spatial structures of humus content. The presence of spatial anisotropy was assessed using the variogram maps and directional variograms with a horizontal tolerance of ±20º. Parameters of the fitted variogram models are presented in table 1. The spatial structure of the humus content shows the presence of a geometric anisotropy, the direction of which corresponds to the sampling grid orientation (45º). The autocorrelation range in a minor direction (315º) is 0.6 times less.

The spatial structure of humus values of the Arsky NCTF is described by linear isotropic variogram. The autocorrelation range of the values is larger than the size of the field and the semivariogram graph does not reached "plateau" within the field (figure 2). According to Cambardella et al. the ratio nugget/sill makes it possible to characterize the power of the spatial dependence of the studied soil values [12]. Proceeding from this, the values of humus content of the Zainsky and Arsky NCTFs have a strong spatial dependence (table 1).

Table 1. Parameters of the fitted variogram models.

Model	Range	Nugget (C_0)	Partial sill (C_1)	Sill (C_0+C_1)	$C_0/(C_0+C_1)$	Anisotropy main dir.	Anisotropy coef.	**Method**
				Zainsky NCTF				
Sph	800.5	0.03	0.59	0.62	0.05	45^0	0.6	OK
Sph	333.7	0.16	0.13	0.28	0.55	-	-	RK+Ridge
Sph	285.5	0.16	0.09	0.25	0.64	-	-	RK+Lasso
				Arsky NCTF				
Lin	1246.14	0.06	0.22	0.3	0.20	-	-	OK
Lin	1187.689	0.07	0.01	0.09	0.86	-	-	RK+Ridge
Nug	-	0.078	0	0.08	1	-	-	RK+Lasso

Sph – spherical model, Lin – linear model, Nug – nugget effect.

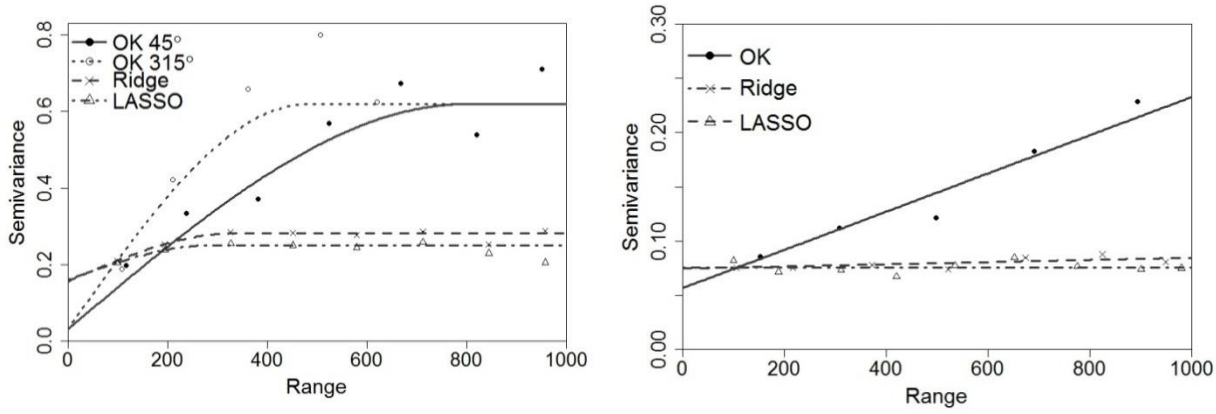
Ridge regression requires an accurate selection of the tuning parameter λ. The leave-one-out cross validation was used to tackle this problem. The ridge regression was fitted with the tuning parameter value for which the cross-validation error was the smallest. In the case of humus, the λ was relatively

large (λ=1.43), indicating that the optimal fit involves a high amount of shrinkage relative to the least squares solution.

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 (figure 2). The ridge drift model eliminated the geometric anisotropy and reduced the autocorrelation range and the overall spatial dependence of the humus values (table 1).

The performance of the LASSO regression also depends on the selection of the parameter λ. The method of the parameter selection was the same as for the ridge regression. As opposed to the ridge regression, the LASSO acts as the feature selection method. Because of the small shrinkage, the coefficients of the humus model are very similar to those we can obtain with OLS regression. The results of LASSO modeling indicate the terrain indices that influence the spatial distribution of humus.

In the case of the LASSO regression in the Arsky NCTF, the variogram has the form of a pure nugget effect - the spatial structure of humus is completely described by the regression model (table 1, figure 2). In the case of Zainsky NCTF, the spatial structure of humus is preserved after applying regression (figure2), although the spatial dependence weakens (table 1).



Zainsky NCTF (a) Arsky NCTF (b)

Figure 2. Sampling variograms with the fitted variogram models.

The final maps of the interpolated humus content are shown in figure 3. Digital maps based on the DSM concept have a continuous smoothed outlook, unlike traditional maps and electronic maps created in various GIS systems in which the soil areas are expressed as discrete units, and the values of soil properties are enclosed within the limits of the elementary sampling sites [13].

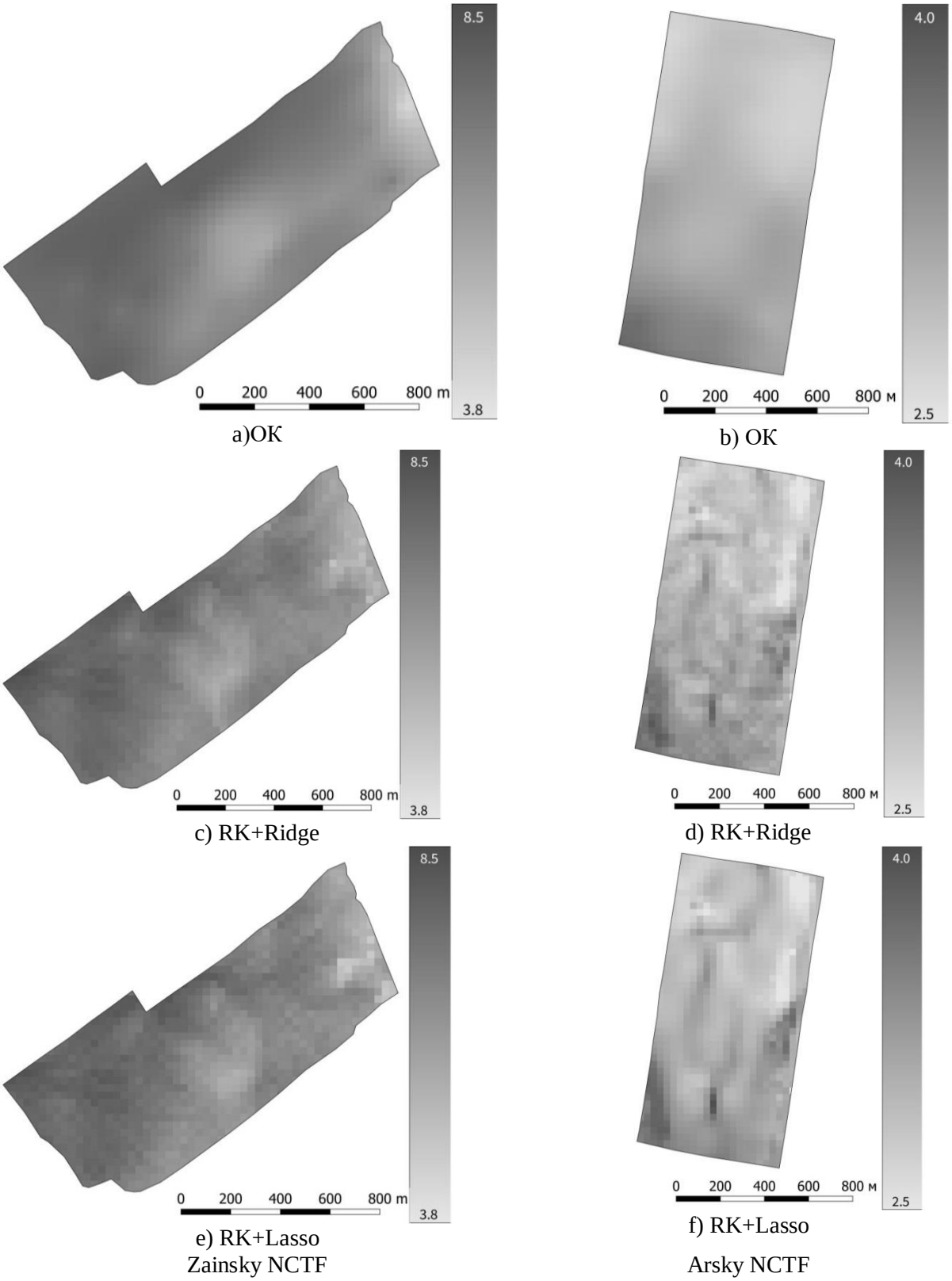


Figure 3. Interpolated maps of the humus content in the Arsky and Zainsky national crop testing fields.

The geomorphological features of the fields, as well as the inhomogeneity of the humus distribution, have caused the differences in the accuracy of the spatial interpolation methods. The results of cross-validation are shown in table 2. The method with the highest performance is the one with the lowest ME and RMSE values. The ability of spatial interpolation to keep the variance of the target variable is estimated from the value of RVar [14].

Table 2. Accuracy assessment.

Leave-one-out cross-validation			Method
ME	RMSE	RVar	
Zainsky NCTF			
-0.015	0.476	0.626	OK
-0.0008	0.536	0.546	RK+Ridge
-0.001	0.583	0.694	RK+Lasso
Arsky NCTF			
-0.0008	0.333	0.391	OK
0.00214	0.343	0.313	RK+Ridge
-0.00004	0.329	0.333	RK+Lasso

Ordinary kriging in the case of Zainsky NCTF shows the lowest RMSE value. A similar picture is consistent with the previous work of the authors, when the OK showed the greatest accuracy in the spatial prediction of the humus content [15]. In the case of the Arsky NCTF, regression kriging with the lasso regression turned out to be the most accurate (table 2). Nevertheless, in some works it is noted that not always lasso and ridge regression have the greatest accuracy of prediction in comparison with other prediction methods [16, 17]. In this case, much depends on the choice of the regularizer and there must be a trade-off with respect to regularization.

In the digital soil mapping, the use of various statistical methods does not guarantee that with the complexity of the method, the productivity and accuracy of the spatial prediction increases. To obtain adequate estimates, several methods must be used. Nevertheless, in comparison with traditional soil mapping, the DSM allows more rapid creation of soil maps. Maps of soil properties, which in the traditional version were not created and were practically replaced by cartograms, can be created with high accuracy in the DSM. At the same time, the concept of the DSM does not violate the principle of the landscape approach in the CSA, since the DSM uses a large set of available environmental variables.

Acknowledgement
This work was supported in part by the Russian Foundation for Basic Research, research project №17-04-00846.

References

[1] *Climate-Smart Agriculture* http://www.fao.org/climate-smart-agriculture/ru/
[2] *Climate smart agriculture* 2013 (Food and Agriculture Organization of the United Nations) p 570
[3] *Russia looks to smart technologies for the future of agroecology and farming.* http://www.fao.org/partnerships/container/news-article/ru/c/881418/
[4] Minasny B and McBratney A B 2016 Digital soil mapping: A brief history and some lessons *Geoderma* **(264)** pp 301–311
[5] Mc Bratney A B, Mendonça S M L and Minasny B 2003 On digital soil mapping *Geoderma* **(117)** pp 3–52

[6] Masticki S E and Shitikov V K 2014 *Statistical analysis and data visualization using R* (Moscow: DMK-Press) p 496 (In Russian)
[7] James G, Witten D, Hastie T and Tibshirani R 2013 *An introduction to Statistical Learning with Applications in R* (Springer New York Heidelberg Dordrecht London) p 440
[8] Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V and Boehner J 2015 System for Automated Geoscientific Analyses (SAGA) **v. 2.1.4** *Geosci. Model Dev.* (8) pp 1991–2007
[9] *R Core Team. R: A Language and Environment for Statistical Computing* 2016 (Austria: R Foundation for Statistical Computing) http://www.R-project.org/
[10] Pebesma E J 2004 Multivariable geostatistics in S: the gstat package *Computers & Geosciences* (30) pp 683–691
[11] *QGIS Development Team QGIS Geographic Information System 2016 Open Source Geospatial Foundation Project* http://qgis.osgeo.org
[12] Cambardella C, Moorman T, Novak J, Parkin T, Karlen D, Turco R and Konopka A 1994 Field-Scale Variability of Soil Properties in Central Iowa Soils *Soil Sci. Soc. Am. J.* **(58)** pp 1501–1511
[13] Sahabiev I A and Ryazanov S S 2015 Investigation of the spatial variability of soil properties using the geostatistical approach *Russian Journal of applied ecology* **(2)** pp 32–37 (In Russian)
[14] Li J and Heap A D 2011 A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors *Ecological Informatics* **(6)** pp 228–241.
[15] Ryazanov S S and Sahabiev I A 2016 Comparison of terrain-based drift models to improve the quality of soil predictive mapping at a field scale *Tomsk State University Journal of Biology* **4(36)** pp 21–33.
[16] Nussbaum M, Spiess K, Baltensweiler A, Grob U, Keller A, Greiner L, Schaepman M E and Papritz A 2017 Evaluation of digital soil mapping approaches with large sets of environmental covariates *SOIL Discuss.*
[17] Song X D, Brus D J, Liu F, Li D C, Zhao Y G, Yang J L and Zhang C L 2016 Mapping soil organic carbon content by geographically weighted regression: A case study in the Heihe River Basin, China *Geoderma* **(261)** pp 1–22