International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 24 (2016) pp. 11631-11638

Working-Out of the Geostatistical Model of Mass Cadastral Valuation of Urban Lands Evidence from the City Vsevolozhsk (Russia)

Alina Mikhailovna Rybkina, Polina Mikhailovna Demidova*, Vladimir Alekseevich Kiselev

Saint-Petersburg Mining University, Saint-Petersburg, 22nd line, d.2, 199106, Russia.

Abstract
This work analyzes the experience of using of the method applied in the Russian Federation for mass cadastral valuation of urban lands permitted use for private housing projects (PHP). According to the results of use of this method it has been found that the accuracy of obtained results does not satisfy the stated requirements. In consequence of the researches the existence of spatial autocorrelation in the base data was revealed and a new and more correct method for cadastral valuation in these conditions was offered – the method of geostatistical interpolation (kriging). The model received on application of kriging has high accuracy which is achieved only when sufficient selection criteria. The study developed a method that enables to carry out the cadastral valuation in areas with insufficient amount of information about transactions by the multivariable spatial modeling (cokriging), which increases the quality of the results of valuation.

Keywords: Real property assessment, cadastral valuation, regression, geostatistics, interpolation, kriging, cokriging.

1. INTRODUCTION

The problems of tax assessment traditionally come under more critical and debatable topics being a major focus of interest of the government and society [1]. This problem is the subject of many papers by various authors, which consider the structure and content of the tax assessment system. In particular, an international expert in the field of value systems of real property tax R. Almy in the property tax system three components were allocated:

- identification of subjects (taxpayers) and taxable objects, and linkage between them;
- identification of the cost of taxable objects;
- tax administration [2].

The component "definition of value objects" includes conducting mass appraisal of real property for tax purposes. In Russia the analogue of the international concept of "mass appraisal of real property for tax purposes" is the concept of "state cadastral valuation of real property".
At the present moment in Russia the mass appraisal of urban lands is carried out on the basis of methods for constructing regressive dependences of the cadastral value of the dependences of the factors affecting the value of real property. According to the requirements of the methodology in 2007 the composition of value factors is determined for each type of permitted use of land property as the part of urban lands on the basis of the indicative list of value factors and information analysis about the real property market of the Russian Federation [3]. In other words, the exact list of factors that should be included in the model is not listed, which means it is selected by an expert. This approach leads to the subjective component in the appraisal, which is one of the problems with the existing method. The second one is that for the application of regression analysis the basic data (values market prices of land properties) are independent. However, within the accepted method the data analysis to verify this requirement is not made.

The problems with the present method are that the detailed list of value factors is not approved, it is based on the opinion of an appraiser (expert), i.e. this approach leads to a subjective component in the valuation. For use of the regression analysis it is also necessary that the input data - the value of market prices of land properties are independent, but within the accepted method the data analysis for autocorrelation is not carried out.

The aim of the research is to assess the accuracy of the regression model, compiled according to the requirements of the methodology in 2007, the basic data research for the presence of spatial autocorrelation and search of the most appropriate method for mass cadastral appraisal of urban lands.

THE TERRITORY UNDER THE INVESTIGATION

As the object of the investigation the city Vsevolozhsk was chosen, which is the administrative center of the Leningrad region (located 7 km from Saint- Petersburg, the population is 62.1 thousand people).

The highest cost has the lands that are located near the park area "Dostoevsky" and in the territory of the exclusive villa development, which are to the south of it, called the "Golden Square". In the park the sports complex and tennis courts function. The famous Zhdanovsky lakes are located to the south, 600 meters from the land. The pointed territory is the most prestigious one provided for PHP in Vsevolozhsk.
The map of cadastral units in Vsevolozhsk is shown in Fig. 1.

Figure 1. The map of Vsevolozhsk

For the aims of the research the selection from 61 land properties with a known market value was composed. The permitted use is a private housing project (PHP).

ANALYSIS OF THE APPLICATION OF MATHEMATICAL STATISTICS for cadastral valuation of urban lands

During the analysis of the information about Vsevolozhsk, a list of factors that influence the value of land properties for PHP were made:

1. Availability of mainstream gas-supply;
2. Availability of water piping;
3. Availability of canalization;
4. Distance to Saint-Petersburg;
5. Distance to the nearest railway station;
6. Closeness to recreation zones;
7. Distance to a water pond;
8. Distance to route A128;
9. Distance to the nearest transit shelter;
10. Distance to a supermarket;
11. Distance to a school;
12. Distance to a kindergarten;
13. Remoteness from industrial area;
14. Status value of a territory;

The basic data are presented in table 1.

Table 1. The basic data

№	Cadastral number of real estate	Market value, rub./sq. m	Availability of mainstream gas-supply (1-yes, 0-no)	Availability of water piping (1-yes, 0-no)	Availability of canalization (1-yes, 0-no)	Distance to Saint-Petersburg, km	Distance to the nearest railway station, km	Closeness to recreation zones, km	Distance to a water pond, km	Distance to route A128, km	Distance to the nearest transit shelter, km	Distance to a supermarket, km	Distance to a school, km	Distance to a kindergarten, km	Remoteness from industrial area, km	Status value of a territory (1- yes, 0-no)
		y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
1	47:07:1302017:29	3444	1	1	0	11.0	0.85	1.05	0.52	3.48	0.35	1.79	1.62	1.13	6.29	0
…																
61	47:07:1302108:6	3100	1	1	0	15.4	3.01	0.49	1.51	6.41	0.55	2.31	2.51	1.31	5.15	0

The formation procedure of market value selection can not guarantee its homogeneity, so testing of the hypothesis of normal distribution of the sample data was conducted at the initial stage. The confirmation of the hypothesis of normality of the sample data on prices for analogues is required for the correct application of correlation and regression methods in determining the value of the property assessment to reflect its differences from prototypes on one or more influencing factors. It is known that the presence of the optimal properties in the least squares method, used in the construction of regression dependences, which is closely related to the normal distribution of the selection [4].

The research of distribution normality was performed using the criterion χ^2 ("chi-squared"), due to a sufficient selection size (more than 50 items). When using the pointed criterion χ^2 the hypothesis H_0 is forged: the selection data obtained from the general population with a known distribution law, the alternative hypothesis is H1: the selection data obtained from the general population with the different distribution law [4].

When testing the hypothesis H_0 using the built-in functions Excel, it was found that $\chi^2 = 6.886$, when $\chi^2_{critical} = 7.815$. The obtained results demonstrate that the hypothesis H_0 can not be rejected. Consequently, the hypothesis of normal distribution of value selection data of land properties for PHP in Vsevolozhsk is confirmed (Fig. 2).

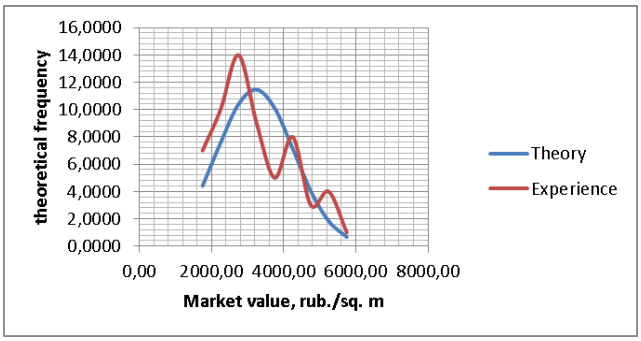


Figure 2. Hypothesis testing about the normal distribution law

Another condition for the correct construction of the regression model is the condition of linear independence of factors. If this condition is violated, i.e. if one of the factors can be expressed in a few others, then they say that there is a complete collinearity. In practice, full collinearity is rare, the situation when a high degree of correlation is between the factors is more often appeared, and then it indicates the presence of multicollinearity factors [6].

To check the data on the presence of multicollinearity the matrix of paired linear correlation coefficients was built, for its compilation the superstructure Excel "Data Analysis – Correlation" was used (Fig. 3).

	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
y	1														
x1	0,45	1													
x2	0,79	0,32	1												
x3	0,78	0,17	0,53	1											
x4	0,27	0,26	0,43	0,19	1										
x5	-0,46	-0,46	-0,19	-0,26	0,24	1									
x6	-0,70	-0,31	-0,68	-0,52	-0,40	0,19	1								
x7	-0,33	-0,09	-0,29	-0,20	0,34	0,46	0,28	1							
x8	0,26	0,40	0,42	0,17	0,92	0,20	-0,46	0,28	1						
x9	-0,48	-0,12	-0,36	-0,31	0,31	0,49	0,28	0,36	0,35	1					
x10	-0,22	0,07	0,04	-0,19	0,69	0,63	0,00	0,51	0,73	0,71	1				
x11	-0,41	-0,02	-0,19	-0,38	0,51	0,51	0,08	0,50	0,49	0,62	0,73	1			
x12	-0,53	-0,12	-0,32	-0,42	0,30	0,43	0,38	0,49	0,29	0,61	0,69	0,63	1		
x13	-0,32	-0,31	-0,40	-0,25	-0,94	-0,07	0,30	-0,29	-0,87	-0,29	-0,62	-0,41	-0,26	1	
x14	0,80	0,18	0,55	0,95	0,19	-0,28	-0,51	-0,26	0,17	-0,31	-0,21	-0,43	-0,46	-0,26	1

Figure 3. The matrix of paired linear correlation coefficients

Coefficients of linear pair correlation of some factors are high (more than 0.7), that is why these factors are collinear (are in a linear dependence among themselves). It is recommended to exclude one from two clearly collinear factors regression equation. Preference is given to the possible factor that with a sufficiently close link with the result has the least crowded due to other factors [5].

As a result of the analysis it has been found that the data has multicollinearity. One of the ways to eliminate multicollinearity is the exclusion of the variables of the model [5], and thus to build a model, you must perform step-by-step method of variable selection. For these purposes the program SPSS Statistics 17.0 was selected, in the result a model has been built and the regression coefficients for the five variables have been obtained (table 2).

Table 2. The regression coefficients for the five variables

Variables	Coefficients
Y – crossing	2741.013
x_1	445.323
x_2	779.831
x_5	-161.152
x_6	-223.877
x_{14}	1118.801

The design equation is represented by the formula (1):

$$y = 2741.013 + 445.323x_1 + 779.831x_2 - 161.152x_5 - 223.877x_6 + 1118.801x_{14}, \quad (1)$$

where y - the rate of the cadastral value, rub./sq.m.; x_1 - availability of gas pipeline (1 – yes, 2 – no); x_2 – availability of water piping (1 – yes, 2 – no); x_5 – distance to the nearest railway station, km; x_6- closeness to recreation zones, km; x_{14} - status value of a territory (1 – yes, 2 – no).

Equation (1) describes the dependence of the land value from the factors affecting it, and can be used to determine the specific indicator of the cadastral value of the real estates of Vsevolozhsk with permitted use - for private housing project. All the coefficients of the unknowns in it are significant and the coefficient of determination R^2 is equal to 0.88.

The minimum value of the error of approximation for this equation is equal to 0.4%; maximum – 44.1%; average - 8.9%. According to [7] it is considered, that sufficient quality of the obtained model is achieved when the error of approximation is 10% to 12%.

The obtained results lead to the understanding that it is necessary to use a more accurate method of cadastral appraisal of urban lands free from shortcomings.

DEVELOPMENT OF MODEL AND METHOD OF MASS CADASTRAL VALUATION OF URBAN LANDS GEOSTATISTICAL METHODS

At the next stage of the research the assumption of a mutual dependence of values of market price for land plots, located near each other was formulated. In the article "The argumentation the use of geostatistical method of interpolation of initial data for mass cadastral valuation of urban lands by the example of Vsevolozhsk" the existence of a spatial auto-correlation in the test data has been proved, on the basis of that it is offered to use more acceptable, in these conditions, geostatistical interpolation method: ordinary kriging with the use of spherical model of semivariograms [9].

The minimum value of the approximation errors for this method is equal to 0%; maximum - 3.2%; the average - 0.7%. In this case, the average error of the approximation is 13 times less than the average errors of approximation obtained in the result of the construction of equation (1).

The model obtained with the application of the kriging has the precision that can be achieved only when a sufficient selection. To determine the amount of initial data when the chosen model ceases to conform to the stated requirements the situation of lack of market price values of real estates was simulated. For this the entire array of initial data incrementally decreased to 10 plots, and in each case the accuracy of the obtained data was evaluated. In the result the graph of dependence of the middle error of approximation of the sample volume was built (Fig.4).

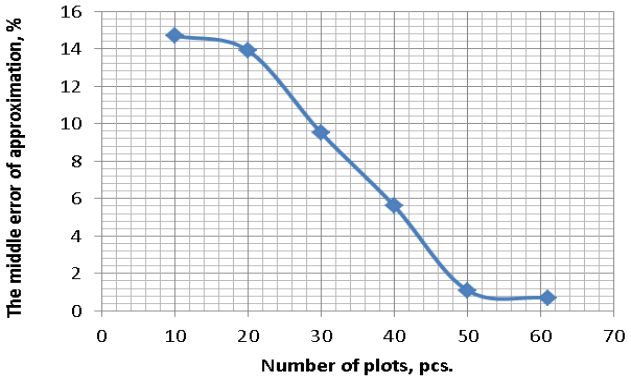


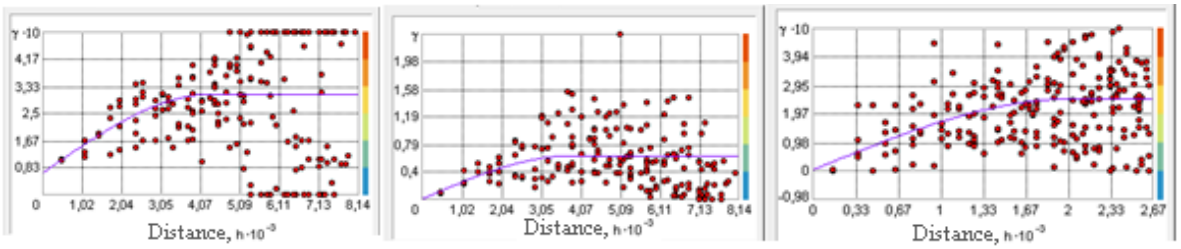
Figure 4. The graph of dependence of the middle error of approximation of the selection

To improve the quality of the model in the case of lack of initial data the method of multivariate spatial modeling, namely cokriging was considered.

ANALYSIS OF APPLICATION OF MULTIVARIATE SPATIAL MODELING FOR MASS CADASTRAL VALUATION OF URBAN LANDS

The assessment measuring of the principal variable (market price) takes place during the use of kriging, but it also can be accompanied by the additional information provided in the measurements of other variables. In this case, a need arises to exam multivariate spatial modeling, namely cokriging.

The main condition for the possibility and usefulness of additional information is its correlation with the main evaluated variable. Cokriging requires much larger number of estimates that include both an assessment of the autocorrelation for each variable and mutual correlation for additional variables to the main [9]. Earlier in the work a number of factors was highlighted affecting the value of the cadastral appraisal of lands in Vsevolozhsk with the permitted use - for PHP, and a regression equation was built, in accordance with the data of the conducted regression analyses the most significant factors - additional variables were chosen. As the program ArcGIS gives the opportunity to include in the model only 3 additional variables, from the equation (1) the factors were selected having the greatest value of the coefficient of correlation with the main variable (Fig. 3): availability of water piping, closeness to recreation zones and status value of a territory. For the quantitative description of the spatial continuity and modeling of spatial autocorrelation in the data semivariograms for the all selected additional variables were constructed (Fig. 5).



a) b) c)

Figure 5. The semivariogram for the additional variable: a – availability of water piping; b – closeness to recreation zones; c - status value of a territory

Cokriging is a natural generalization of the kriging in case of multivariate data when between variables are a spatial correlation [10]. It should be noted that the value of all variables (basic and advanced) in all points are not required, for example, for the ordinary kriging it is necessary to have at least one dimension of the main variable, for the simple knowledge it is its average value, and the rest information is entered at the expense of additional variables [10].

The application of cokriging for the purposes of cadastral valuation of urban lands within the simulated situation of lack of market price values, was also carried out step by step (with the inclusion of 10 additional plots on each stage).

In the cokriging the procedure of maps construction of interpreted values is similar to the procedure which was carried out using kriging, i.e. it is necessary to delete a trend ,to include the anisotropy in the model, and then to choose the optimal method and semivariogram model. Due to the repeated realization of the operations on the maps construction of the interpolated values for cokriging in ArcGIS environment a table of error for the selection, consisting of 10 plots was created (table 3).

Table 3. The table of errors

Method \ Model	Errors	Circular	Spherical	Tetraspherical	Pentaspherical	Exponential	Gaussian	Rational Quadratic	Hole Effect	K-Bessel	J-Bessel	Stable
Ordinary kriging	1	0.078	0.003	0.004	0.010	-0.057	0.135	0.033	0.119	0.106	0.087	0.103
	2	489.5	464.1	468.3	468.6	511.9	585.3	497.7	863.1	701.7	963.1	679.0
	3	94.4	79.4	55.1	45.4	163.1	191.7	245.4	305.9	501.3	300.9	511.0
	4	1.06	1.04	1.05	1.06	1.50	0.85	0.81	0.77	0.60	0.78	0.59
Simple kriging	1	-0.05	-0.07	-0.07	-0.06	0.11	0.02	-0.13	-0.05	0.02	-0.06	0.01
	2	499.4	519.4	541.5	563.8	546.1	578.5	529.4	787.1	554.4	1035.0	615.3
	3	30.4	42.4	96.3	143.9	187.5	185.8	138.5	258.9	210.9	241.6	260.6
	4	1.05	1.15	1.26	1.36	1.47	0.82	1.49	0.77	0.79	0.78	0.77
Universal kriging	1	0.096	0.011	-0.005	0.010	-0.065	0.165	0.026	0.121	0.118	0.092	0.115
	2	509.5	471.4	471.2	470.7	511.7	621.7	503.7	865.5	715.6	970.3	692.4
	3	84.4	75.1	53.8	44.0	162.4	167.5	241.1	304.5	448.4	294.7	498.6
	4	1.08	1.06	1.05	1.06	1.50	0.88	0.82	0.77	0.61	0.78	0.60
Disjunctive kriging	1	-0.05	-0.07	-0.07	-0.06	0.05	0.02	-0.13	-0.05	0.02	-0.07	0.01
	2	498.6	519.1	541.5	563.8	511.7	579.1	530.0	787.9	554.8	794.4	614.8
	3	31.3	42.1	96.3	143.9	174.1	185.2	139.2	258.1	210.5	240.6	260.9
	4	1.05	1.15	1.26	1.36	1.41	0.82	1.49	0.77	0.79	0.78	0.77

Note to table: 3 1- Mean Standardized; 2- Root-Mean-Square, rub./sq. m; 3- Deference between root-mean-square and average standard error, rub./sq. m; 4- Root-Mean-Square Standardized.

The analysis of the obtained results was performed according to the following criteria:

- the closest to zero mean standardized;
- the lowest root-mean-square;
- the closest to the root-mean-square average standard error;
- the closest to one the root mean square standardized.

To build a map of interpreted values of the specific indicator of the cadastral value in Vsevolozhsk the best combination of: spherical model semivariogram and ordinary kriging method was chosen (Fig. 6).

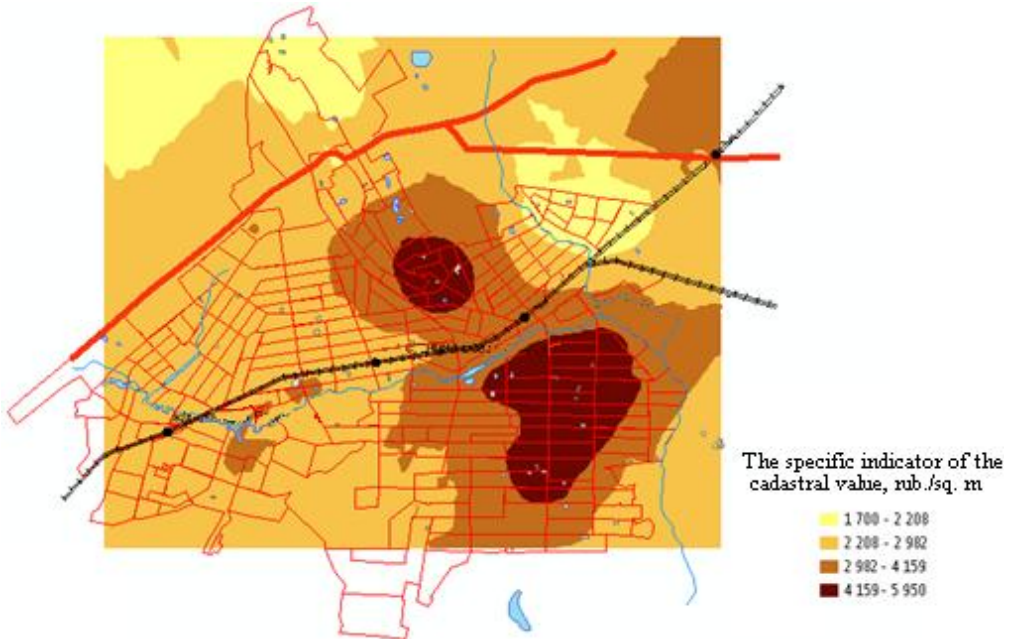


Figure 6. The map interpreted values of the specific indicator of the cadastral value in Vsevolozhsk (cokriging)

After applying the cokriging method at all samples mean errors of approximation of the obtained surfaces were evaluated.

To determine the size of the selection, in which the inclusion of additional variables into the model increases its accuracy, was built the graph of dependence of the mean error of approximation of the selection size for two methods: kriging and cokriging (Fig. 7).

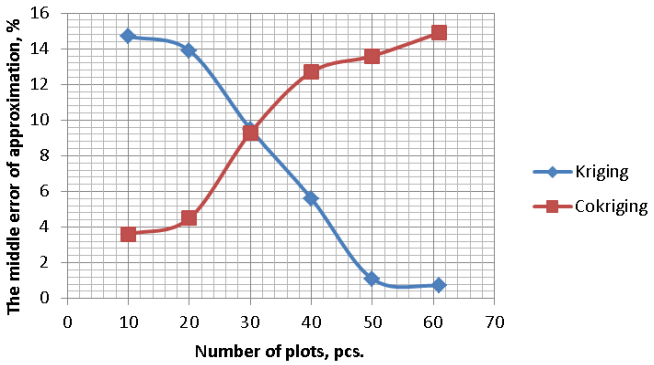


Figure 7. The graph of the dependence of the mean error of approximation of the selection size

The graph shows that the introduction of additional variables leads to improvement of the model quality if the number of sectors in the sample does not exceed 30 pcs. Otherwise it is necessary to apply kriging for mass cadastral valuation of urban lands in Vsevolozhsk.

The additional information contributes to the assessment of the principal variable in the areas with lack of information about the market price of plots, where there are data of available additional variables, as their use allows to assess areas which have been zones of extrapolation for the main variable, but when using the additional variables they become a zone of interpolation.

The obtained result is explained by the fact that in the market prices of plots the values of the factors affecting their cost are already taken into account, so if the selection consists of more than 30 plots, there is a situation of excessive amounts of information, which leads to an increase in evaluation error.

When using the cokriging the most accurate is surface which is built for a selection consisting of 10 plots (Fig. 6). The minimum value of the approximation error for a given surface is 0%, maximum - 8%; the average of 3.6%. It is possible to speak about the good quality of the obtained model when the error of approximation is 10 to 12% [7].

ANALYSIS OF THE QUALITY OF THE CONSTRUCTED MODELS

In accordance with the method of 2007 the quality analysis of the constructed model is held at the training and the control selection.

Under the training selection the market information is understood, on the basis of which the calculation models are built. Under the control selection the market information is understood, on the basis of which the quality of the constructed models of calculation are checked, not including the training selection [3].

To check the quality of models on the control selection the plots were selected which not involved in the creating of models (table 4).

Table 4. The control selection

№	Cadastral number of a real estate	Market price, rub./sq. m	Availability of mainstream gas-supply (1-yes, 0-no)	Availability of water piping (1-yes, 0-no)	Availability of canalization (1-yes, 0-no)	Distance to Saint-Petersburg, km	Distance to the nearest railway station, km	Closeness to recreation zones, km	Distance to a water pond, km	Distance to route A128, km	Distance to the nearest transit shelter, km	Distance to a supermarket, km	Distance to a school, km	Distance to a kindergarten, km	remoteness from industrial area, km	Status value of a territory (1- yes
		y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
1	47:07:1302065:7	3000	1	1	0	11.3	0.83	0.24	0.19	2.64	0.62	0.73	1.82	0.97	7.56	0
…																
4	47:07:1302102:21	4545	1	1	1	14.1	1.66	0.05	1.09	5.01	0.51	1.48	1.27	1.21	4.62	1
…																
7	47:07:1301075:5	4700	1	1	1	10.6	2.13	1.08	0.16	1.47	0.51	2.34	2.40	1.01	7.37	1

Using the control selection were tested: the additive model (table. 5), the kriging model for 60 plots (table. 6), the cokriging model to 10 plots (table 7).

Table 5. The approximation error for the additive model

№	Market price, rub./sq. m	The specific indicator of the cadastral value (the equation of regression), rub./sq. m	The difference between the values, rub./sq. m	The approximation error, %
1	3000	3779	779	26,0
…				
4	4545	4806	261	5,4
…				
7	4700	3720	980	26,3

The minimum value of the approximation error for the additive model is equal to 0.9%; maximum – 26.3%; average – 8.31%.

Table 6. The approximation error for kriging

№	Market price, rub./sq. m	The specific indicator of the cadastral value (when using kriging), rub./sq. m	The difference between the values, rub./sq. m	The approximation error, %
1	3000	3063	63	2.1
…				
4	4545	4552	7	0.2
…				
7	4700	4717	17	0.4

The minimum value of the approximation error to the surface obtained in the result of the application of the kriging is equivalent to 0.1%; maximum – 2.7%, average – 1.0%.

Table 7. The approximation error for cokriging

№	Market price, rub./sq. m	The specific indicator of the cadastral value (when using cokriging), rub./sq. m	The difference between the values, rub./sq. m	The approximation error, %
1	3000	3029	29	1.0
…				
4	4545	4606	61	1.3
…				
7	4700	4869	169	3.6

The minimum value of the approximation error to the surface obtained in the result of application of the cokriging is 1.0%; maximum – 4.7%; average - 2.5%.
As a result of the analysis it emerged that the most accurate model is a model created for a selection of 60 plots, when using the geostatistical method of interpolation - ordinary kriging using a spherical model semivariogram.

CONCLUSIONS

The analysis of methods of mathematical statistics for mass cadastral valuation of urban lands showed that the built additive model has a low quality that has led to the discovery of another more accurate valuation method.
In the course of work the presence of spatial autocorrelation in the initial data was proved, therefore, for example, Vsevolozhsk a method of mass cadastral valuation of urban lands was developed with the use of geostatistical methods.
For cadastral valuation of urban lands, where there is a sufficient number of initial data it is encouraged to use more correct method - ordinary kriging with using a spherical

model. This approach is the most accurate and provides an opportunity to increase the objectivity of the problem and its solution by taking account of the interdependence between the values of the initial data, waive the need to establish significant influencing factors and determine their values.

Also, this technique gives an opportunity to spend the cadastral valuation in areas with insufficient information regarding transactions through multivariate spatial modeling (cokriging), thus increasing the quality of the valuation results.

At present the existence and availability of powerful geographic information systems opens good prospects for application of the proposed method in many organizations carrying out the mass cadastral valuation of urban lands.

REFERENCES

[1] Bezrukov, V. B., Dmitriev, M. and Pylaeva, A. 2011. *Tax and cadastral valuation of real estates*. N.Novgorod: NNGASU.

[2] Almy, R. 2002. Real Property Assessment Systems. *choices*, 6 p. 11.

[3] Guidelines for state cadastral valuation of urban lands. 2007. *Prikaz minekonomrazvitiya Rossiiskoi Federacii №39*, 15.02.2007.

[4] Greenwood, P. E. and Nikulin, M. S. 1996. *A guide to chi-squared testing*. New York: Wiley.

[5] Babeshko, L. O. 2010. *The basis of econometric modeling*. Moskva: ComKniga.

[6] Magnus, I., Katyshev, P. K. and Peresetskii, A. A. 2000. . *konometrika*. Moskva: Delo.

[7] Eliseeva, I. I. 2001. *Econometrics*. Moskva: Finansy i statistika.

[8] Snytko, A. M. and Kiselev, V. A. 2013. The argumentation the use of geostatistical method of interpolation of initial data for mass cadastral valuation of urban lands by the example of Vsevolozhsk. *Ingenernyi vestnik Dona*, (3), Available at: http://ivdon.ru/magazine/archive/n3y2013/1797 [Accessed: 21 Feb 2014].

[9] Resources.arcgis.com. 2014. *ArcGIS Help (10.2 and 10.2.1)*. [online] Available at: http://resources.arcgis.com/en/help/main/10.2/index.html#//003100000053000000 [Accessed: 21 Feb 2014].

[10] Demyanov, V. V. and Saveleva, E. A. 2010. *Geostatic: theory and practice*. Moskva: Nauka.