Possibility of Using GIS and Geostatistic for Modeling the Influence of Location on the Value of Residential Properties

Milica COLAKOVIC and Darko VUCETIC, Serbia

SUMMARY

This paper is an attempt to present one of the possibilities of using spatial data for modeling the influence of the location factor on the value of residential units in Belgrade. Location factor was selected as the factor that determines the property in spatial sense with its components that are spatially dependent sub-factors. Modeling of location factor influence on the value of real estates is based on Multi Criteria Evaluation (MCE) method as a method of spatial analysis with GIS and regression kriging as geostatistical methods. At the end of the paper comparison of the obtained results from these two methods and conclusions on the possibilities of their use are done.

1. INTRODUCTION

Location takes a special place among the attributes that characterize residential properties, arguably being the most influential factor affecting property price. Location of a property is its unique spatial characteristic, determined by many spatially related attributes. If we have a model that in a good way represents location factor, we can say that the main factor determining property value is known. This paper studies the impact of the location factor on the value of residential properties in the urban area of the City of Belgrade. Under the term 'property location' various geographical, social, infrastructural and neighborhood quality characteristics, here called sub-factors, are implied. These sub-factors are chosen as the most typical for the study area. For modeling the value of the residential property as a function of its location, two techniques are used; Multi Criteria Evaluation and regression kriging. The possibilities of applying these techniques and their advantages and disadvantage have been examined.

2. STUDY AREA, REFERENCE AND FACTOR DATA

Due to the fact that this analysis is limited on investigation of location factor influence on the value of residential properties excluding family houses, as study area the urban territory of Belgrade is chosen (*Figures 1, 2*).

Input or reference data for this analysis are selling prices of residential properties on the area of interest. Selling prices of real estate are used as alternative solution due to the fact that real market data, i.e. sold prices, could not be collected. Data have been collected from the internet and have been selected in a way to ensure uniform distribution all over the area of interest. At

TS04G - Valuation - Mass Appraisal Techniques, 5943   1/9
Milica Colakovic and Darko Vucetic
Possibility of Using GIS and Geostatistic for Modeling the Influence of Location on the Value of Residential Properties

FIG Working Week 2012
Knowing to manage the territory, protect the environment, evaluate the cultural heritage
Rome, Italy, 6-10 May 2012

the same time real estate legality and conditions of property were checked as inevitable part of the procedure. Data for 747 apartments for sale were collected for reference (*Figure 3).*
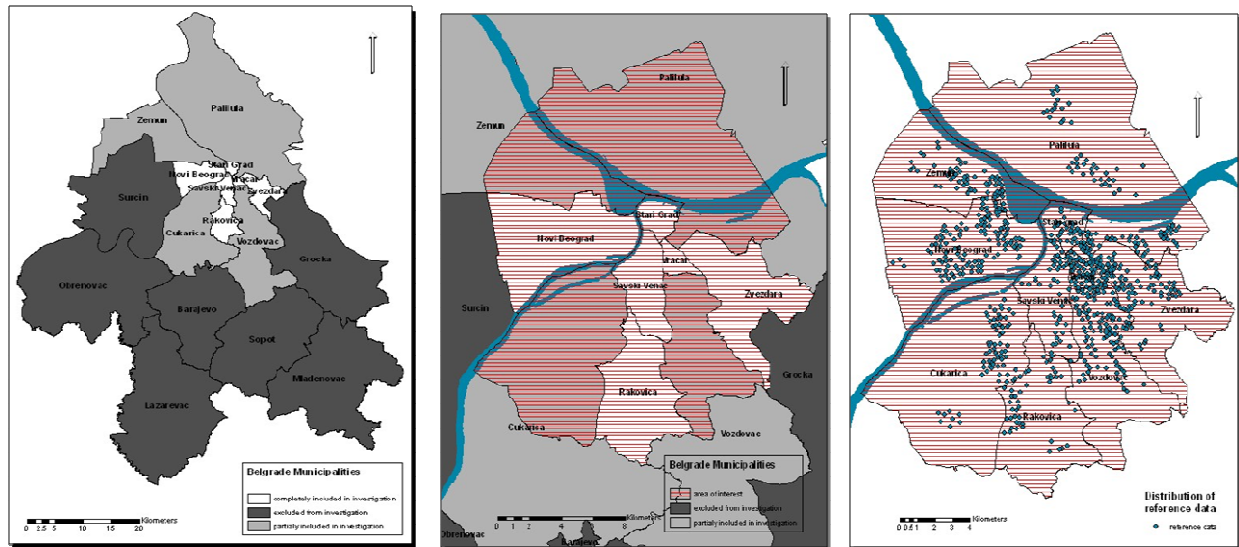
Figures 1, 2: Study Area   *Figure 3: Reference Data*

Factor data represent sets of sub-factors that define the properties in terms of location or geographical position. For this data set the most important geographic, infrastructural and socio-economic data that describe the quality of a neighborhood are selected. They are divided into four sub-factor types depending on its characteristics. The geographical sub-factors are proximity to the city centre, proximity to the business centre, proximity to the green areas and water areas. The infrastructural sub-factors are proximity to the main roads and railways, proximity to the public transport stations and proximity to the primary schools and kindergartens. Sub-factors that determine quality of neighborhood are proximity to the parks and playgrounds, proximity to the culture and entertainment areas, proximity to the green markets and shopping malls, level of noise and air pollution. The average income per municipality is chosen as social-economic sub-factor.

3. MULTI CRITERIA EVALUATION (MCE) ANALYSIS

As GIS solution for spatial analysis, MCE method is chosen. This analysis is performed using one of the most contemporary and wide used software ArcGIS. The Multi Criteria Evaluation process diagram applied for the purpose of analysis of influence of the location factor on prices of residential properties in Belgrade is shown in *Figure 4*. Definition of the main goal and criteria that will be used in the analysis is first step in the MCE process. In case of this analysis the goal is production of the unique map that shows dependency of the residential property values on selected criteria. Selected criteria should be presented as raster maps and standardized with the possibility of their combination. On the base of standardized maps and calculated weights for each sub-factor, model producing result map has been created.
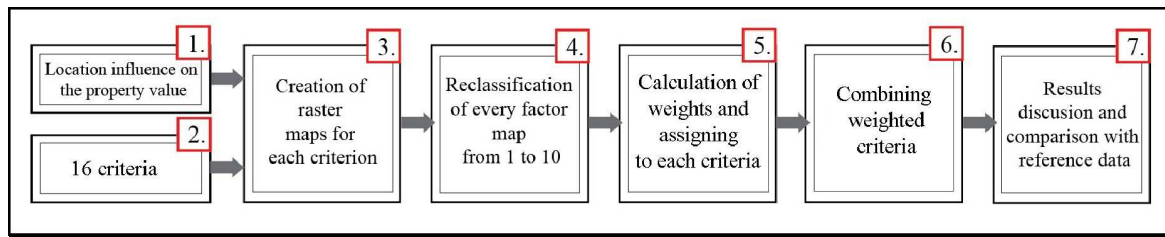
Figure 4: MCE process

Input data for the MCE method are raster maps created using various tools of spatial analysis. Beside those raster maps, on the basis of reference data-selling prices of real estates, as a dependent variable, and factor data as independent variables, weights are calculated.

Weights are used to emphasize that not all sub-factors included in the research have equal impact on the prices of residential apartments, i.e. to rank them in order to demonstrate which sub-factors have the biggest impact on the residential apartment market value. Weights are calculated on the base of output coefficients of regression analysis that shows change of dependant variable by increasing or decreasing single independent variable. According to regression analysis results 7 of 16 factors are excluded from further analysis.
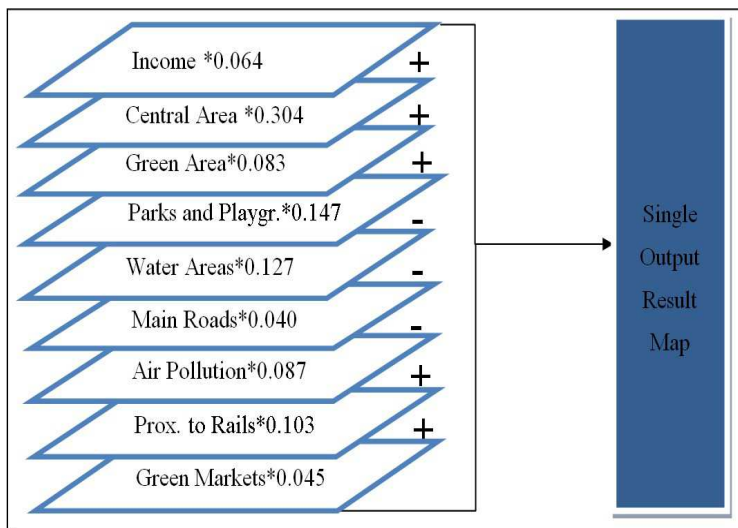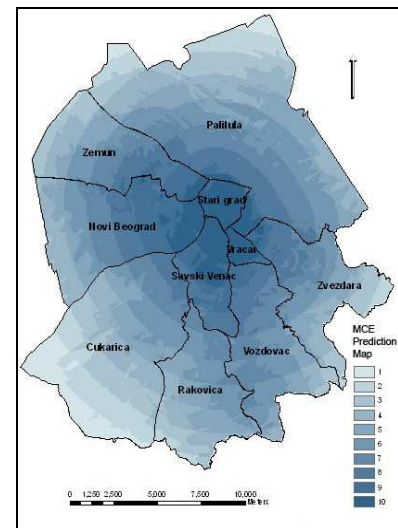

Figure 5: MCE modeling

Figure 6: MCE result map

Remaining 9 factors and corresponding weights create model shown in Figure 5. Combining of factor maps 'pixel per pixel' using algebraic operators is done by Map algebra method, and function Single Output Map Algebra for obtaining of unique output map *(Figure 6)*.

4. REGRESSION KRIGING

The starting idea of geostatistics is application of spatial correlation of the variable being analyzed, in order to improve its spatial prediction or interpolation. Not only correlation within the variable itself, but also correlation of variable with other spatial variables. Kriging method is synonym for geostatistical methods of interpolation.

TS04G - Valuation - Mass Appraisal Techniques, 5943   3/9
Milica Colakovic and Darko Vucetic
Possibility of Using GIS and Geostatistic for Modeling the Influence of Location on the Value of Residential Properties

FIG Working Week 2012
Knowing to manage the territory, protect the environment, evaluate the cultural heritage
Rome, Italy, 6-10 May 2012

Modification of basic formulas and their adjustment to specific applications has resulted in increasingly wide application in different spatial analysis. Especially attractive method for spatial analysis is regression kriging. Regression kriging could be successfully implemented in R software environment using gstat package for modeling, prediction and simulation of spatial data and sp package that enables manipulation of spatial data. Results obtained in R software can be easily converted to any standard GIS formats that enable additional manipulation and analysis of the results inside commercial GIS environments.

Reference data are loaded as ASCII files containing X and Y coordinates and selling prices of properties per square meter. Preliminary analysis of reference data includes statistical calculations in order to show their distribution. Normal distribution of input data is a prerequisite for the regression kriging prediction. The results of this analysis show minimum and maximum values of samples, their mean, median, QQ-plot and Shapiro-Vilk test results.

Min.	1st Quart.	Median	Mean	3rd Quart.	Max.
700	1259	1545	1633	1905	4235
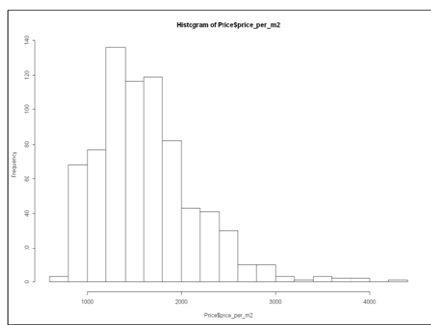
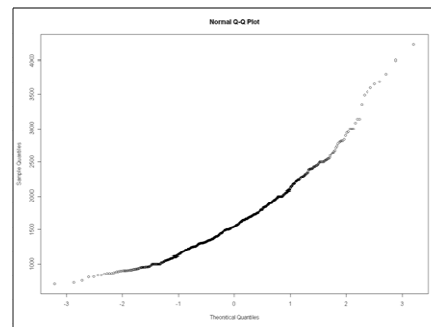*Figure 7: Histogram*

Figure 8: QQ - plot

Shapiro-Wilk test	
W=0.9359	p-value=2.2e-16

Histogram obtained shows no central distribution of sample data that indicates the need for transformation of input data. After logarithmic transformation, the sample shows normal distribution:

Min.	1st Quart.	Median	Mean	3rd Quart.	Max.
700	1259	1545	1633	1905	4235
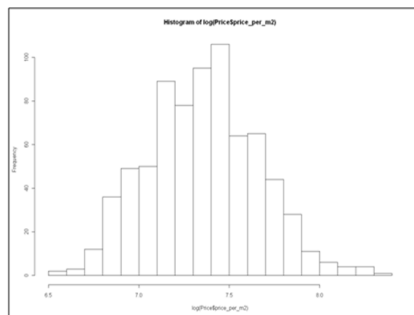
Figure 9: Histogram after log transformation
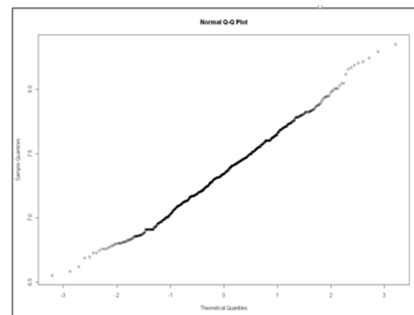
Figure 10: QQ plot after log transformation

TS04G - Valuation - Mass Appraisal Techniques, 5943   4/9
Milica Colakovic and Darko Vucetic
Possibility of Using GIS and Geostatistic for Modeling the Influence of Location on the Value of Residential Properties

FIG Working Week 2012
Knowing to manage the territory, protect the environment, evaluate the cultural heritage
Rome, Italy, 6-10 May 2012

Shapiro-Wilk test	
W=0.9953	p-value=0.02118

After preliminary analysis and preparation of sample data, through multiple regression analysis, examination whether additional predictors show correlation with variable of interest and whether additional predictors can be used to improve the spatial prediction, is done. Multiple regression analysis results are presented in *Table 1*:

Regression Statistics	
Multiple R-squared	0.637
Adjusted R Square	0.627
Residual Standard Error	0.187
Observations	747

	Coefficients	Standard Error	t Stat	P-value	Signif. code
(Intercept)	7.30E+00	1.50E-01	48.689	2.00E-16	***
Income	4.70E-06	1.64E-06	2.856	0.004411	**
MainRoads	2.12E-05	1.17E-05	1.803	0.071848	*
NoisePollution	-1.07E-04	6.85E-03	-0.016	0.987514	
Soot	-2.97E-02	1.27E-02	-2.349	0.019085	*
no2	3.96E-02	8.62E-03	4.594	5.11E-06	***
so2	2.24E-02	7.98E-03	2.81	0.00509	**
PowerPlants	1.37E-03	9.13E-03	0.15	0.881177	
Kindergartens	2.28E-05	3.03E-05	0.754	0.451364	
BusinessArea	-2.72E-05	1.19E-05	-2.284	0.022652	*
CentralArea	-6.61E-05	1.09E-05	-6.052	2.29E-09	***
Culture	1.27E-05	1.69E-05	0.75	0.453557	
Parks	-2.99E-05	1.18E-05	-2.535	0.011466	*
GreenMarkets	-9.22E-05	1.08E-05	-8.511	2.00E-16	***
Railways	-4.20E-05	8.09E-06	-5.199	2.61E-07	***
WaterAreas	2.85E-05	6.95E-06	4.096	4.68E-05	***
Schools	7.20E-05	2.82E-05	2.558	0.010742	*
ShoppingMalls	2.92E-05	1.10E-05	2.648	0.008271	**
GreenAreas	-4.61E-05	1.27E-05	-3.645	0.000287	***
PublicStations	-5.24E-05	5.69E-05	-0.921	0.357124	

Table 1: Multiple regression analysis outputs

From the results, it can be concluded that the variable of interest shows significant correlation with 11 of 19. Data marked with *** have the biggest impact, while data designated with ** and * have less influence and data without mark do not affect the dependent variable. Variables that do not show an effect on the explanatory variable are excluded from further analysis.

Milica Colakovic and Darko Vucetic
Possibility of Using GIS and Geostatistic for Modeling the Influence of Location on the Value of Residential Properties

FIG Working Week 2012
Knowing to manage the territory, protect the environment, evaluate the cultural heritage
Rome, Italy, 6-10 May 2012

After analysis of the dependent variable properties, spatial prediction starts. The experimental variogram is calculated in order to show the existence of spatial autocorrelation between samples *(Figure 11)*.

Experimental variogram shows significant autocorrelation that leads to conclusion about spatial dependency of sample. When samples show a spatial dependency, variogram modeling samples based on mathematical function is created. For the model of fitted variogram exponential model was chosen because it best fits the experimental variogram. The values of the most important terms that describe the variogram are:

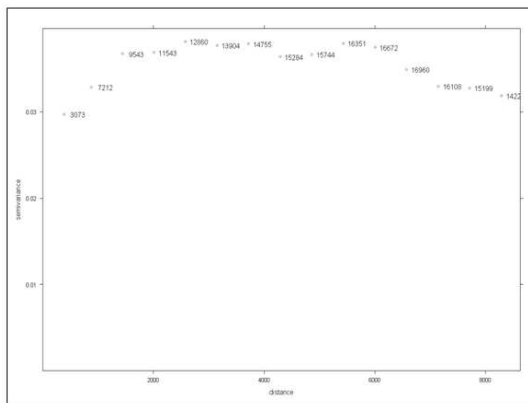| **Range=173m**	**Nugget=0.0026**	**Partial sill=0.0347**

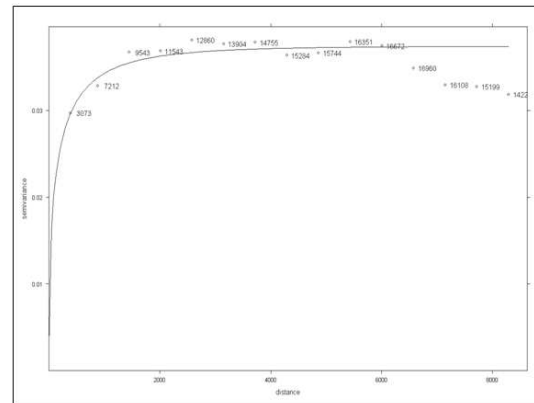
Figure 11: Experimental variogram

Figure 12: Fitted variogram

Defining of the semivariogram is important because it allows the spatial prediction is performed using regression kriging. The results obtained are presented in *Figure 13*.
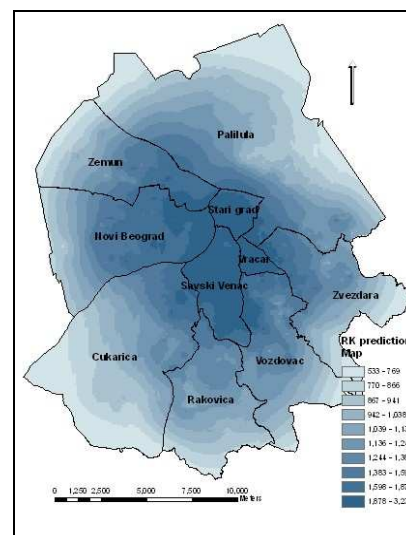

Figure 13: RK result map

5. RESULTS

Estimation of results is done through comparison of the map for comparison (MC) and the result maps obtained from the MCE process and the RK process respectively. The reference map is obtained by interpolation of reference data. Inverse Distance Weighted (IDW) method was used.

Subtraction of the map for comparison and MCE result maps is performed using the 'Single Output Map Algebra' *(Figure 14, 1st picture)*. The absolute values of the difference range from 0 to 7. Areas in the difference map with value 0 represent areas of absolute agreement between the two maps, while we can say that in areas with value 7 the MCE result map shows the biggest deviation from the map for comparison. In Figure 14 different levels of agreement are shown, depending on the tolerance specified. The absolute agreement map represents areas with '0' difference between reference and MCE result map. Agreement with tolerance +/- 1 shows areas where absolute difference of classes on reference and MCE result maps is 1, while agreement with tolerance +/- 2 allows absolute difference of 2 classes between the compared maps. The results are shown in *Table 2*.
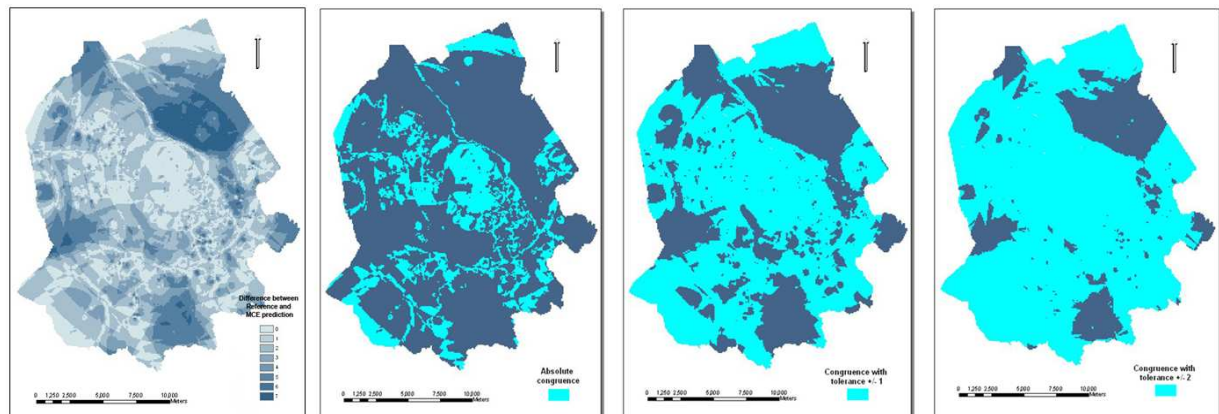

Figure 14: Comparison of MCE result map and map for comparison

Besides this comparison between the MCE result map and the map for comparison, calculation of agreement of 747 input data with the corresponding data point from the MCE analysis is done. For each of the 747 input selling properties its respective value in the classified map for comparison and the MCE result map is calculated. The values are subtracted. The results are presented in *Table 3*.

Tolerance	Agreement
0	26%
0 or 1	60%
0, 1 or 2	78%

Table 2: Agremeent between MC and MCE

Tolerance	Number of congruence
0	220 of 747 (29%)
0 or 1	503 of 747 (67%)
0, 1 or 2	629 of 747 (84%)

Table 3: Agremeent between MC and MCE on sample points

Subtraction of the map for comparison and RK result maps is performed using the 'Single Output Map Algebra' *(Figure 15, 1st picture)*. Absolute values of the subtraction map are from 0 to 7. Areas with subtraction map value 0 represent areas of absolute agreement

Milica Colakovic and Darko Vucetic
Possibility of Using GIS and Geostatistic for Modeling the Influence of Location on the Value of Residential Properties

FIG Working Week 2012
Knowing to manage the territory, protect the environment, evaluate the cultural heritage
Rome, Italy, 6-10 May 2012

between the two maps, while we can say that in areas with value 7 the RK result map shows the biggest difference between the map for comparison and the RK result map. In Figure 15 the different levels of agreement are shown, depending on the tolerance specified. The absolute agreement map represents areas with '0' difference between reference and RK result maps.
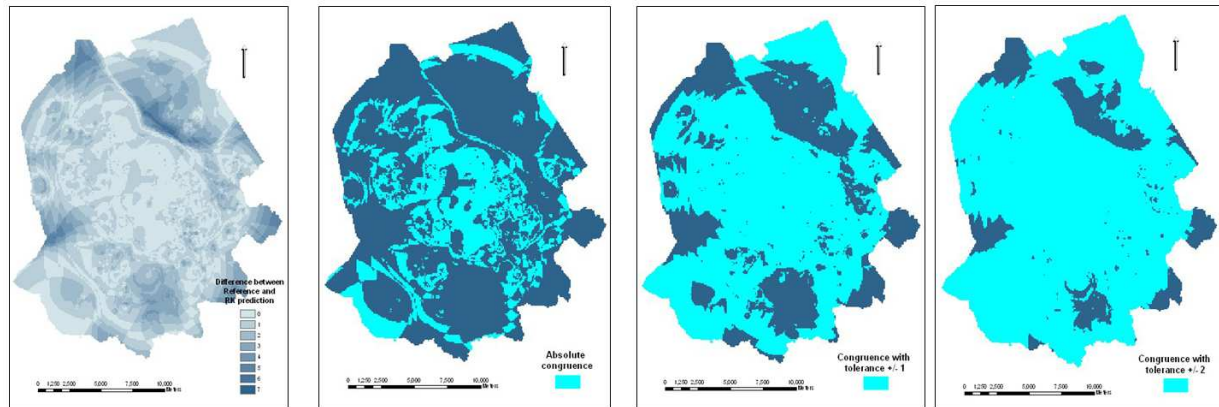

Figure15: Comparison of RK result map and map for comparison

Agreement with tolerance +/- 1 shows areas where the absolute difference of the classes in the map for comparison and the RK result maps is 1, while agreement with tolerance +/- 2 allows absolute difference of 2 classes between compared maps. These results are shown in *Table 4*. Besides this comparison between the RK result map and the map for comparison, calculation of agreement of 747 input data with the corresponding data point from the RK analysis is done. For each of the 747 input selling properties its respective value in the classified map for comparison and the RK result map is calculated. These values are subtracted. The results are presented in *Table 5*.

Tolerance	Agreement
0	31%
0 or 1	67%
0, 1 or 2	86%

Tolerance	Number of congruence
0	294 of 747 (39%)
0 or 1	577 of 747 (77%)
0, 1 or 2	691 of 747 (93%)

Table 4: Agremeent between MC and MCE   *Table 5: Agremeent between MC and MCE on sample points*

6. CONCLUSION

The results obtained lead to conclusion that both methods give good result for modeling the impact of factors on the value of the location of residential units on the territory of Belgrade. Due to the fact that the research conducted relies on the real estate valuation, tested models could find application in this field. Specifically, for the application in the field of valuation, proposed methods should extend with internal characteristics of properties, which would result in more complex, but more complete and more accurate model. For the purpose of creating prediction model, regression kriging method is a more reliable solution, because it

Milica Colakovic and Darko Vucetic
Possibility of Using GIS and Geostatistic for Modeling the Influence of Location on the Value of Residential Properties

FIG Working Week 2012
Knowing to manage the territory, protect the environment, evaluate the cultural heritage
Rome, Italy, 6-10 May 2012

allows the integration of internal characteristics data in a simple manner and provides predicted values and accuracy assessment of the model used.

REFERENCES

[1] Colakovic M., Dabic S., 2011, *The value of residential properties as a function of location - comparison of two modeling techniques, multi criteria evaluation and regression kriging –* master thesis
[2] Hengl T.; *A Practical Guide to Geostatistical Mapping*, 2009
[3] Maquire D., Batty M., Goodchold M., 2008, *GIS, Spatial Analysis and Modeling,* ESRI Press, Redlands, California
[4] Vucetic D., 2009, *Aplication of Regression kriging for mapping of humus and sludge in the soil –* final graduate thesis

BIOGRAPHICAL NOTES

Milica Colakovic

Career-to-date:
October 2009 - up to date, Republic Geodetic Authority of Serbia, Head of Unit for topographic-cartographic database,
November 2004 - October 2009, Map publishing company 'Geokarta' Ltd - engineer in division for cartography, department for planning and development,
June 2003 - November 2004, Geomatics company 'MapSoft', assistant engineer for geographic information systems, digital processing and mapping.

Publications:
'3D city modeling methods' - Mitrovic M.; Nedeljkovic I.; Colakovic M.;
First Mediterranean Conference on Earth Observation (Remote Sensing); Session II:
Processing of remote sensing data, Belgrade, April 2004.
'Cartography base for navigation and transport' - Colakovic M.,
GEOInfo Conference; Divcibare, October 2009.
'Modeling the Influence of Location Factor on the Value of Residential Properties in Belgrade' - Colakovic M., Vucetic D., Dabic S.,
First Serbian Geodetic Congress, Belgrade, December 2011.

CONTACTS

Ms. Milica Colakovic
Republic Geodetic Authority
Bulevar vojvode Misica 39
Belgrade
SERBIA
Tel. + 381 11 2651 245
Fax + 381 11 2651 064
Email: milica.colakovic@rgz.gov.rs
Web site: www.rgz.gov.rs

TS04G - Valuation - Mass Appraisal Techniques, 5943   9/9
Milica Colakovic and Darko Vucetic
Possibility of Using GIS and Geostatistic for Modeling the Influence of Location on the Value of Residential Properties

FIG Working Week 2012
Knowing to manage the territory, protect the environment, evaluate the cultural heritage
Rome, Italy, 6-10 May 2012