

PAPER • OPEN ACCESS

Estimation of Single-Housing Areas Development Using Artificial Neural Network

To cite this article: Jan Janusz and Robert Lucka 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **471** 102064

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

Estimation of Single-Housing Areas Development Using Artificial Neural Network

Jan Janusz¹, Robert Lucka¹

¹ UTP University of Science and Technology in Bydgoszcz, Al. prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland

jan.janusz@utp.edu.pl

Abstract. The subject of the research is the analytical model of single-housing areas development covered with local spatial plans in the suburban zone of Poznań agglomeration outside the central city. The research is based on the regression analysis with selected techniques of artificial neural networks using SPSS 22 software environment. It is conducted on 61 local spatial plans, with a single-family housing area from 2 to 90 hectares with an average of 15 hectares. Since it was assumed to focus on the growth rate from 5 to 10 years from the implementation of the plan, the research samples come from 1993-2007. In conclusion, it can be stated that a relatively simple network, based only on a few variables, allows for an effective estimation of development in a given time. The model can support the decision-making process for the purposes of sustainable development.

1. Introduction

One of the most basic tools for controlling the processes of suburbanisation and urban growth are local spatial development plans. In Poland, their form and implementation procedure is regulated in detail by law. [1,2] The aforementioned acts state that the creation of such a plan is based on analyses, including forecasts, but they do not specify nor even propose what are the methods and techniques. Therefore, the subject of the research is focused on the creation of such an analytic model of single-housing areas development, dedicated to the local planning scale. The researched is focused on the impact of local conditions and factors on the development of single-family housing areas, therefore the regression analysis is used to establish the model. The research material comes from the 61 housing areas with an average area of 15 hectares during next 10 years from the local spatial plan implementation. All of them are located in the suburban zone of Poznań agglomeration, outside the central city. The specific task can be accomplished in many ways and with use of many techniques. In the paper below the regression model is based on selected techniques of artificial neural networks with use SPSS 22 software environment.

2. The research problem

The Poznań agglomeration is struggling with problems related to uncontrolled development. The urban sprawl is related to the significant growth of housing areas on the periphery of the agglomeration, [3]. It leads to the number of consequences, including landscape degradation, increase of the financial and social costs, degradation of the natural environment, loss of attractiveness of the recreation zones, and finally, the loss of the agricultural and investment sites and deterioration of housing conditions, [4]. Similar negative phenomena can be observed in numerous examples of other



agglomerations. [5,6]. There are many reasons for this situation, including, among others, economic, political and social aspects. There are several programs and movements aimed at comprehensive resolution to those problems, including the new urbanism [7], the intelligent development [8], and the compact cities [9]. At the same time, they require complex changes, time and awareness in various aspects of development, and using the planning and political solutions to help with counteraction of the negative processes effectively. The presented elaboration is based on existing instruments, especially local spatial development plans. It means a specialized approach to anticipate the urban sprawl by the implementation of analytical models. Irwing and Bockstael picture a low density of newly emerging development areas as the cause of city sprawl. [10] Their case study shows the advantage of the emergence of new development areas over the filling of existing ones. As a result, the housing settlements are not used effectively and extensive areas on the periphery are wasted. In addition to the negative impact mentioned above, the low density housing zones are deprived of adequate infrastructure, because in such an area its implementation is extremely expensive, remote from workplaces and service facilities. This causes a heavy load on communication and consequently, traffic jams. In addition, the report of the Supreme Audit Office [11] states that the space is ineffectively managed and this is the result of the overestimation of the future number of houses and the allocation of too large areas for housing development in inappropriate places. Therefore, it can be stated that the paper is aimed at finding the analytical model for the effective use of available legal planning instruments through forecasting and evaluation in order to control suburbanisation processes in the Poznań agglomeration.

3. Urban growth models

The assumed goal requires the research on the literature and case studies concerned with the method of predicting of development and urban growth. This problem can be addressed in many ways. The several studies below are based on the classification of the land use changes over the time. Another approach, the one used in the paper, is the estimation of the unknown quantitative parameter, future development rate, based on known local parameters. The available research material is the observation of development of housing areas in the past, so the research can be classified as correlational research. [13] The numerous studies aimed at the establishment of a model for estimating the future development that use a variety of regression techniques and can be found in the literature. Their abundant taxonomy is represented by Triantakou and Mountrakis. [13] They established that artificial neural networks (ANN), a general linear model (GLM) and a generalized linear and nonlinear model (GLNM), fractal modelling, agent-based model (ABM) and systems dynamics (SD) are most commonly applied. Other techniques which are also used, include regression trees and expert based models. Simultaneously, the original study introduces a novelty, since the subject of analysis are small areas of homogeneous housing function. This approach is not widespread in the literature, hence the references to such case studies can only be considered at the level of techniques, but not at the level of results. ANN analyses are used on various scales, ranging from the scale of the region [14], through metropolitan areas [15], to the scales of a single city [16]. Within the framework of ANN, various methods of organizing and processing spatial information are used. They can be combined with a cellular automaton or fuzzy clusters, [17]. Beim presents the research based on the artificial neural networks and cellular automaton for the entire Poznań agglomeration area. [18] What is more, the classification can be made on image analysis, using the Tensorflow library. However, these techniques are used on a larger scale and in that form they cannot be applied in the presented study. At the same time, they point to the high efficiency of SNN in the study of spatial development.

4. Methodology

The studies above illustrate the key advantages and disadvantages of these techniques in the context of other methods. The main advantages include the ability to find complex, non-linear relationships and the ability to rely on different variable scales. The methods based on multiple regression in urban planning encounter a problem of correlation between variables or spatial dependencies, including

autocorrelation and heterogeneity. [19] In the context of the study, the basic disadvantage of ANN is black-box style approach to the resulting model. It means the impossibility of full intuitive explanation of the relation based on the calculation model. Although there are some techniques, different depending on the type of network, to estimate the impact of a specific variable, it cannot be read directly from the model. [19] The another disadvantage is the risk of overfitting, especially dangerous in the situation of too few variables. The problem of overfitting means the excessively precise fitting of the model to a specific, local situation. It is associated with the loss of the ability to broaden the investigated phenomenon. [20] This problem is a relative, because it depends on the purpose of the study. A different degree of generalization and abstraction is desired in different elaborations. In the presented study, the aim is to create a model that describes the overall spatial relationship, therefore overfilling is a significant risk. It leads to the conclusion that the simple model based on a small number of variables is required.

5. Research process

The search for the original model is based on the ANN techniques. They are biologically inspired mathematical structures used for many purposes, including classification, data processing and process-control. However, in the study, they are used to create a regression model. Unfortunately, the obtained databases of buildings, named EGiB (Records of land and buildings) at the selected areas were neither complete, nor accurate, [21, 22]. It was necessary to verify them using archival satellite images, [23]. It slowed down the work and limited the possible range to 9 counties: Czerwone, Komorniki, Murowana Goślina, Oborniki, Rokietnica, Suchy Las, Szamotuły and Tarnowo Podórne. The archival orthophoto maps were available there. As a result, the measurement includes 61 areas covered by local plans in the years 1993-2007, observed over 10 years, with 2404 constructed buildings of the selected destination. Variables were measured in two scales; continuous and dichotomous based on this material. The independent variable can be defined as the percentage of the number of buildings completed within 5 or 10 years to the total number defined in the local plan, with full use of the land dedicated to the housing function. In some plans this number is strictly written, while in others, it can be estimated by the plan regulations with acceptable accuracy. Independent variables are related to many thematic areas:

- Independent variables are related to many thematic areas:
- Nuisance
- Access to specific objects
- Access to natural resources
- Distances from the center
- The housing area in the local plan
- Planned organization of the construction process (for own needs, or by a large entity for sale or rent)
- Typology

At first, the multiple regression analysis was performed for the variable sets to eliminate the majority of the 28 variables. Next, 10,000 artificial neural networks were tested, including radial basis function network and multilayer perceptron with a separate learning and testing group in a 7: 3 ratio. As a result, a network for the development analysis was selected. The study is primarily focused on determining whether the development of such a model is possible using artificial neural networks, then on its description and finally, the assessment of the possibility of applying it in spatial planning.

Both models are highly statistically significant, but adjusted R² points to a relatively weak fit of the model. It shows correlation, but it is not satisfying for prognostic purposes with such research sample. In addition, the variables "Library1000m" and "PublicGreenery500m" due to Pearson's test are correlated at the level of 0.416. They are included in both models. In the second model, "NeighbIndustry" and "SportFacilities1000m" are correlated according to this test in 0.473. Apart from further analysis, it should be stated that such a correlation influences the assessment of the model negatively, so it is rejected as a prognostic tool.

Table 1. The primary set of variables

n.	Gr.	Description	S/U	Short name
A		Dependent: constructed in 5 years building divided by the number specified in the local plan	%	Growth5Years
B		Dependent: constructed in 10 years building divided by the number specified in the local plan	%	Growth10Years
1	Nuisance	The neighbourhood of railway in the distance of 400m	%	Railway400m
2		The noisy road in the distance of 200m (Cat: A,S,GP)	%	Road200m
3		The neighbourhood of industry facilities	D	NeighbIndustry
4		Sewage treatment plant in the distance of 2000 m	D	SewageTreatmentP2000
5		The neighbourhood of multi-housing	D	NeighbMultihousingN
6	Access to facilities	Access to pitch in the distance of 1000m	D	Pitch1000m
7		Access to sport kindergarten in the distance of 1000m	D	Kindergarten1000m
8		Access to medical facilities in the distance of 1000m	D	MedicalHelp1000m
9		Access to primary school in the distance of 500m	D	PrimarySchool500m
10		Access to primary school in the distance of 1000m	D	PrimarySchool1000m
11		Access to primary school in the distance of 2500m	D	PrimarySchool2500m
12		Access to high school in the distance of 1000m	D	HighSchool1000m
13		Access to high school in the distance of 2500m	D	HighSchool2500m
14		Access to sport facilities in the distance of 1000m	D	SportFacilities1000m
15		Access to culture facilities in the distance of 1000m	D	CultureFacilities1000m
16		Access to library in the distance of 1000m	D	Library1000m
17		Access to post office in the distance of 1000m	D	PostOffice1000m
18		Access to railway station in the distance of 1000m	D	RailwayStation1500m
19	Natural resources	Access to public greenery in the distance of 500m	D	PublicGreenery500m
20		Access to open forest in the distance of 1000m	D	OpenForest1000m
21		Access to lake in the distance of 1000m	D	Lake1000m
22		Access to river in the distance of 1000m	D	River1000m
23	Others	possibility of connection to the sewerage network	D	SewageNetwork
24		Percentage of attached houses typology	%	AttachedHousing
25		Average size of the plot in the spatial plan	m ²	AvaragePlotArea
26		distance from the center of the Poznań agglomeration	km	DistanceFromCent@km
27		The size of the housing area designated in the plan	ha	HousingArea@ha
28		the percentage of buildings constructed for selling	%	ByDeveloper

S/U - scale or unit, D - Dichotomous Variable, % - percentage, ha - hectares

The variables above were tested in the study. Two sets were selected based on multiple stepwise regression analysis with 9 and 7 independent variables.

Table 2. Multiple stepwise regression model summary with 7 and 9 independent variables

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Statistics			Durbin-Watson
					F test	df1	Sig. F	
1	0.784	0.614	0.546	11.8341449	9.022	9	0.000	2.205
2	0.754	0.568	0.511	12.2787741	9.971	7	0.000	2.285

Table 3. Multiple stepwise regression model coefficients with 7 and 9 independent variables

	Model 1					Model 2				
	Unstandardized Coefficients		Stand Coef.	t	Sig.	Unstandardized Coefficients		Stand Coef.	t	Sig.
	B	Std. Error	Beta			B	Std. Error	Beta		
(Constant)	60.165	5.801		10.372	0.000	56.052	5.717		9.805	0.000
DistanceFromCent@km	-0.887	0.210	-0.414	-4.219	0.000	-0.799	0.206	-0.373	-3.880	0.000
HousingArea@ha	-0.228	0.095	-0.218	-2.406	0.020	-0.209	0.098	-0.199	-2.137	0.037
RailwayIn400m	-0.100	0.058	-0.177	-1.732	0.089					
NeighblIndustry	-14.367	4.045	-0.382	-3.552	0.001	-10.449	3.637	-0.278	-2.873	0.006
SewageTreatmentP2000	-19.141	7.976	-0.238	-2.400	0.020	-18.136	8.203	-0.225	-2.211	0.031
SportFacilities1000m	9.648	5.702	0.187	1.692	0.097					
Library1000m	-11.911	5.580	-0.253	-2.135	0.038	-15.200	4.925	-0.323	-3.087	0.003
PublicGreenery500m	9.912	4.172	0.255	2.376	0.021	12.944	4.121	0.333	3.141	0.003
SewageNetwork	12.317	3.779	0.336	3.259	0.002	12.621	3.910	0.344	3.228	0.002

6. The result of the ANN model

The next step is to perform the analysis using artificial neural networks for growth in 10 and 5 years.

Table 4. The parameters of the ANN for growth analyses in 10 and 5 years

	Dependent Variables	Growth10Years	Growth5Years			N	Percent
	Rescaling Method for Covariates (9 units)	Standardized	Standardized	Sample	Training	43	70.5%
Hidden Layer(s)	Number of Hidden Layers	1	1		Testing	18	29.5%
	Number of Units in Hidden Layer 1 ^a	8	10	Valid		61	100.0%
	Activation Function	Hyperbolic tangent	Hyperbolic tangent			10 years	5 YEARS
Output Layer	Type:	MLP Multilayer perceptron	MLP Multilayer perceptron	Training	Sum of Squares Error	1.743	17.265
	Number of Units	1	1		Relative Error	0.083	0.822
	Rescaling Method for Scale Dependents	Standardized	Standardized	Testing	Sum of Squares Error	3.588	2.155
	Activation Function	Identity	Identity		Relative Error	0.543	0.900
	Error Function	Sum of Squares	Sum of Squares				

The model for 10 years' development has almost a two-time smaller relative error in testing group and approximately ten times smaller in learning group, therefore the elaboration will focus on the better model. In the case of regression for the variable " Growth10Years " regression with artificial neural networks, 10,000 models of RBF and MLP networks were tested, according to the adopted methodology, limiting network complexity to one hidden layer with maximum number of 18 hidden neurons. The MLP 9-8-1 network was selected based on the lowest relative error value in the test group. It is a multilayer perceptron with one hidden layer with 8 neurons. In the automatically selected model, the learning type was applied to the whole set by using the scaled conjugate gradient algorithm.

Table 5. Multilayer perceptron network structure with weights

Predictor		Predicted							
		Hidden Layer 1							
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)
Input Layer	(Bias)	1.019	0.186	0.761	-0.081	1.831	0.730	0.720	-0.059
	DistanceFromC@km	-1.225	1.203	-0.059	-0.375	-0.140	0.589	0.387	0.726
	HousingArea@ha	0.547	0.078	0.157	0.138	1.453	-0.074	-0.384	-0.107
	RailwayIn400m	0.387	-0.193	0.995	0.093	0.785	0.323	1.790	0.059
	NeighbIndustry	0.189	0.646	0.966	0.290	0.679	-0.515	-0.722	0.152
	SportFacilities1000m	0.313	-0.180	0.260	-0.116	-0.016	0.174	-0.506	-0.182
	Library1000m	-0.403	-0.457	-0.337	-0.152	0.356	0.238	0.259	-0.238
	PublicGreenery500m	-1.288	-0.099	-0.884	0.044	-1.513	-0.284	0.548	0.110
	SewageNetwork	-0.918	-1.487	-0.146	0.367	-1.043	-0.727	0.847	0.150
	SewageTreatP2000	-0.376	-0.113	-0.008	0.120	0.148	-0.205	0.145	-0.078
Hidden Layer 1	(Bias)								0.381
	H(1:1)								0.924
	H(1:2)								-0.990
	H(1:3)								-0.611
	H(1:4)								0.030
	H(1:5)								-1.061
	H(1:6)								0.478
	H(1:7)								-0.565
	H(1:8)								0.493

Multilayer perceptron network structure with weights.

The error function was the sum of error squares in the test group, abbreviated as: SOS. The function of activation of the hidden layer is hyperbolic tangent and the function of activating the output neuron is identity, so the sum of the hidden layer's neurons and the error of this layer was equal to the standardized prediction value for a given sample (theoretical value), because all variables were standardized. The test group was preliminarily selected and it remained the same for each of 10,000 tests. The relative error in the training group is 0.083, while in the test group is equal 0.543. The table represents the specific network structure with its weights.

The comparison of presented models is shown in scatterplots for multiple regression and artificial neural networks for 5 and 10 years.

7. Conclusion

The comparison of the obtained models proves the possibility of creating models estimating the development of single-housing area covered by the local spatial development plan. Moreover, there is an improvement in the model's "fit" with the amount of time passed from the plan implementation in the observed sample. Models analysis indicates that artificial neural networks allow for more accurate mapping of the process than multiple regression. At the same time, the model created by the MLR does not meet all assumptions, including the correlation of independent variables. Finally, ANN for 10 years was chosen as adequate for analysis. However, it is not sufficient for a reliable forecast due to the size of the research sample, the value of relative error and the changing durability. This model can be successfully used as an evaluation analysis allowing to determine whether a given local plan design has prospects for the desired high development or to diagnose the danger of low growth. R2 value in multiple regression, as well as relative error in ANN and analysis of scatterplots allow to state that regression studies of local plans, dedicated to residential development, offer promising perspectives. It is possible to obtain a model that describes effectively the development phenomenon in a way useful in spatial planning and urban design. In addition, such models allow to deduce about the importance of individual variables, which provides knowledge about the phenomenon of development.

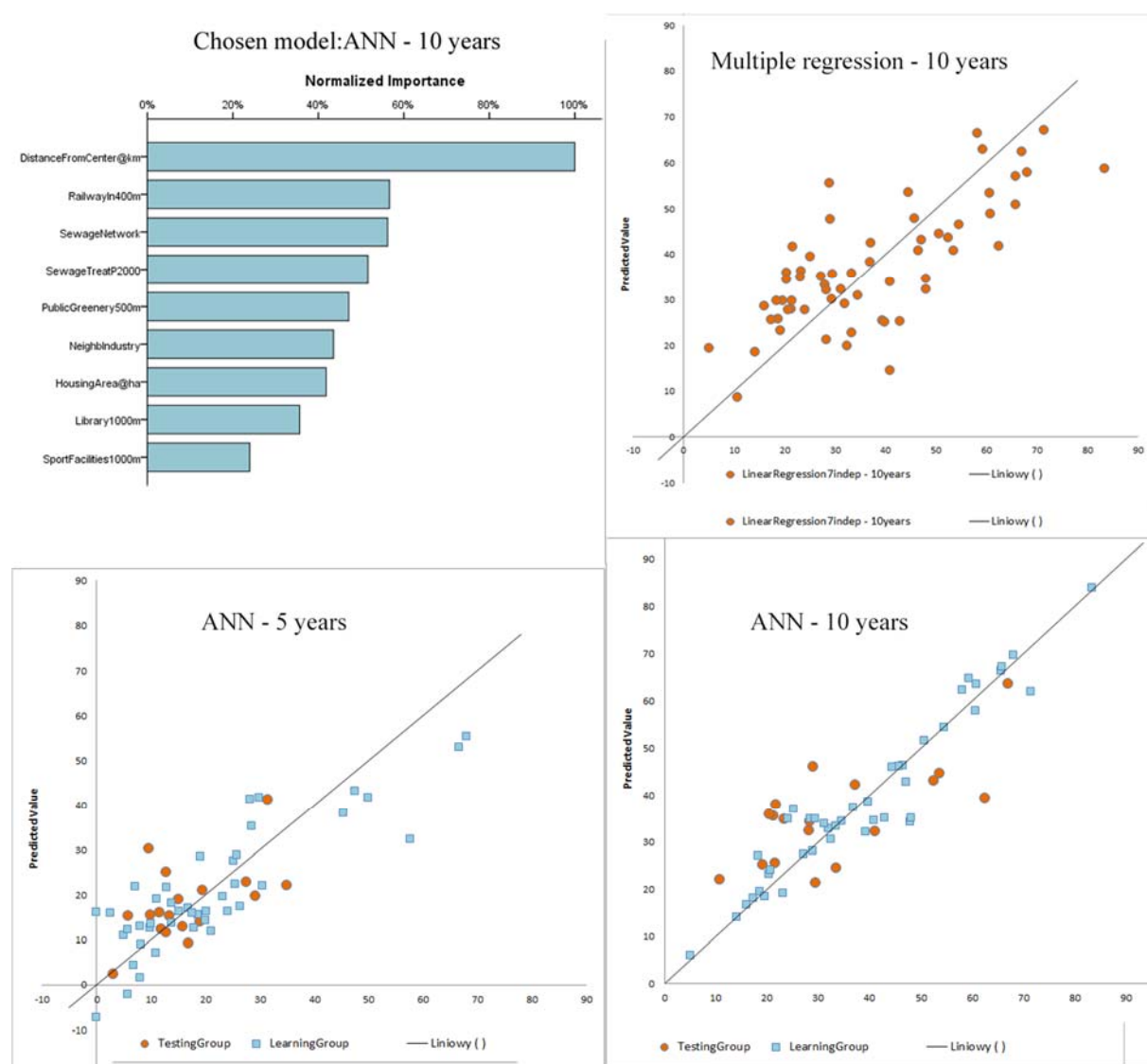


Figure 1. The scatterplots for prepared models and the importance graph of Growth10Years ANN model of development.

References

- [1] The Act on Revitalization, 9 October 2015. Dz.U. 2015 poz. 1777, Art. 4 ust. 1., Art. 15. Ust. 1.
- [2] The Act on Planning and Spatial Development, 27 March, 2003. Dz.U. 2003 nr 80 poz. 717, art. 1, ust. 3.
- [3] T. Kaczmarek, and M. Mikula, "Study of conditions for spatial development of the Poznań Agglomeration - Studium uwarunkowań rozwoju przestrzennego Aglomeracji Poznańskiej", *Adam Mickiewicz University in Poznań*, pp. 82, 2012.
- [4] Kowalewski, J. Mordasewicz, J. Osiatyński, J. Regulski, J. Stępień, and P. Śleszyński, "Report on economic losses and social costs of uncontrolled urbanization in Poland - Raport o ekonomicznych stratach i społecznych kosztach niekontrolowanej urbanizacji w Polsce", *Foundation for the Development of Local Democracy - Fundacja Rozwoju Demokracji Lokalnej*, Warsaw, pp 3-7, 2013.
- [5] N. A. Salingaros, "Compact City Replaces Sprawl, Rozdział monografii: Crossover: Architecture, Urbanism, Technology", editor: *Arie Graafland i Leslie Kavanaugh*,

- Rotterdam, pp. 100 – 115, 2006.
- [6] J.K. Brueckner, “Urban sprawl: diagnosis and remedies”, *International Regional Science Review* 23, 2, pp. 160–171, 2000.
 - [7] C. Ellis, “The New Urbanism: Critiques and Rebuttals”, *Journal of Urban Design*, Vol. 7, No. 3, pp. 261-291, 2002.
 - [8] A. Downs, “Smart Growth Why We Discuss It More than We Do It”, *Journal of the American Planning Association*, Vol. 71, No. 4, pp. 368, 2005.
 - [9] R. Alter, “Compact City Policies, A Comparative Assessment”, *OECD Green Growth Studies*, pp. 29-44, 2012.
 - [10] E. G. I. Irwin, and N. E. Bockstael, “The evolution of urban sprawl: Evidence of spatial heterogeneity and increasing land fragmentation”, *Proceedings of National Academy of Science of United States of America*, vol. 104 no. 52, editor: L. Turner, pp. 20673-20676, 2007.
 - [11] T. Emiljan, and M. Łuczak, K. “Spatial management system of the commune as a public good, Information on the results of the inspection - System gospodarowania przestrzenią gminy jako dobrem publicznym, Informacja o wynikach kontroli”, *Report number: 193/2016/KIN*, Department of Infrastructure, The Supreme Audit Office, Warsaw, pp. 10-16, 2017.
 - [12] L. Groat, and D. Wang, “Architectural Research Methods”, *John Wiley & Sons*, ISBN 0-471-33365-4, pp. 85-96, 2002.
 - [13] D. Triantakoustantis, and G. Mountrakis, “Urban Growth Prediction: A Review of Computational Models and Human Perceptions”, *Journal of Geographic Information System*, 2012, 4, pp. 555-587, 2012.
 - [14] M. L. Manna, P. Berck, M. A. Moritz, E. Batllori, J. G. Baldwin, C. K. Gately, and D. Richard, “Modeling residential development in California from 2000 to 2050: Integrating wildfire risk, wildland and agricultural encroachment”, *Land Use Policy* 41, pp. 439–451, 2014.
 - [15] R. B. Thapa, “Scenario based urban growth allocation in Kathmandu Valley, Nepa”, *Landscape and Urban Planning*, vo. 185, Issue 1-2, pp. 140-148, 2012.
 - [16] C. M. Almeida, J. M. Gleriani, E. F. I. Castejon, and B. S. Soares-Filho, “Using neural networks and cellular automata for modelling intra-urban land-use dynamics”, *International Journal of Geographical Information Science*, Vol. 22, No. 9, pp. 943–963, 2013.
 - [17] G. Grekousis, P. I. Manetos, Y. N. Photis, “Modeling urban evolution using neural networks, fuzzy logic and GIS: The case of the Athens metropolitan area”. *Cities* 30, pp. 194-202, 2013
 - [18] M. Beim, “Modeling of the suburbanisation process in the Poznań agglomeration with use of artificial neural networks and cellular automata - Modelowanie procesu suburbanizacji w aglomeracji poznańskiej z wykorzystaniem sztucznych sieci neuronowych i automatów komórkowych”, *Department of Spatial Econometrics, Institute of Socio-Economic Geography and Spatial Management, Adam Mickiewicz Uniwersytet in Poznań, PhD dissertation, Zakład Ekonometrii Przestrzennej Instytut Geografii Społeczno-Ekonomicznej i Gospodarki Przestrzennej Uniwersytet im. Adama Mickiewicza w Poznaniu, praca doktorska*, pp. 100- 130, 2007.
 - [19] D. Triantakoustantis, D. Stathakis, “Urban Growth Prediction in Athens, Greece, Using Artificial Neural Networks”, *International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering Vol:9, No:3*, pp. 268-273, 2015.
 - [20] B. C. Pijanowski, D. G. Brown, B. A. Shellito, and G. A. Manik, “Using neural networks and GIS to forecast land use changes: a Land Transformation Model”, *Computers, Environment and Urban Systems*, 26(6), pp. 553-575, 2002.
 - [21] EGiB license number GKG.4012.13299.2016_3021_N
 - [22] EGiB license number GK.6621.1625.2016
 - [23] ORTO license number DIO.7211.736.2016_PL_N