Radosław Cellmer*

The Use of Selected Spatial Interpolation Methods for Analyzing the Worth of Land Zoned for Housing Development

1. Introduction

Most phenomena observed on the real estate market, in particular transactions and prices, are related to space. Location plays a special role in the group of factors determining the market value of property and, consequently, property prices on local markets. Spatial analyses determining the correlation between property prices and the property's location are, therefore, a useful tool in the process of evaluating the situation and the current trends on the property market.

The determination of a property's market value as the most probable price that can be achieved on that market requires knowledge about the prices of similar properties, where one of the main similarity criteria, in addition to property type, its zoning destination and legal status, is the property's location. One of the key principles in geography states that elements situated in the vicinity share more similarities than objects separated by a large distance. For that reason, the determination of a property's estimated value, which can be unambiguously identified in space, may rely on the averaged prices of sites with a similar location. Since the above approach is a mere approximation of the property's market value, the concept of property worth has been adopted in this study.

The issue of property similarity in a spatial approach has been analyzed by various authors. Basu and Thibodeau [2] and Gillen et al. [5] point to spatial correlations between property prices that can be expressed through autocorrelation. Tu, Sun and Yu [16] relied on the spatial autocorrelation phenomenon to identify the principles for determining the neighborhood boundaries of property, based mostly on spatial factors. They have argued that the same principle supports the spatial segmentation of the housing market. Spatial correlations on the property market may constitute a basis for building geostatistical models where price is a regionalized variable [9].

* Department of Land Management and Regional Development, University of Warmia and Mazury, Olsztyn

Various solutions for applying geostatistical principles in property market analyses have also been discussed by Anselin [1], Haining [7] and Leuanghton [12].

In Poland, property market surveys place special emphasis on the mutual correlation between location and transaction prices. Ligas and Kulczycki [10, 11] have proposed algorithms for selected geostatistical methods of interpolating geospatial data generated by the property market. The presented algorithms can be used in property market analyses. Cichociński [3] has reviewed various methods for the cartographic presentation of the results of analyses investigating spatial correlations between property location and property prices. Those methods can be deployed to develop property worth maps, and they could prove highly useful in a situation when the data on the distribution of property values and prices are the key factor supporting a rational decision-making process in property management.

2. Selected Spatial Interpolation Methods

One of the methods for the graphic presentation of economic phenomena in space relies on an analysis of a given phenomenon's spatial range, and it indicates that phenomenon's intensity in a given area. As regards land prices, this range may be illustrated by maps of separate zones where the determinants of the price and, consequently, the value of land property are relatively uniform, or by maps of isolines of identical worth [19].

In zone maps, the covered area is divided into uniform zones based on the set of adopted criteria. Unit land worth is determined for each zone by way of analysis. The determination of area boundaries is preceded by a market analysis which identifies market parameters that contribute to land value. Zone boundaries are set in view of the identified market parameters, while the value of land in the zone is estimated with the use of a value model developed by way of analysis. A map is developed on the assumption that the identified zones group properties with similar attributes, therefore, their unit value should be similar.

Isoline (isarithmic) maps are developed by expressing the spatial variability of a phenomenon with the use of non-intersecting curves (isolines) that are assigned successive increasing or decreasing values from identical value points which are connected by those curves [3]. In land worth maps, measurement points (transactions) should be situated at characteristic points for the investigated area and, if possible, they should be evenly distributed. The course of an isoline may be difficult to determine in practice. The acquisition of a sufficiently large dataset of transactions which supports reliable mapping of the analyzed phenomenon is a significant problem. Despite this difficulty, the method delivers a high level of transparency and an ease of interpretation [19].

Isoline interpolation methods applied in the development of geodesic and cartographic documents are not very useful for generating property worth maps whose main objective is not to accurately reflect spatial variation in prices, but to present a hypothetical, anticipated or probable price, i.e. a price which reflects market value in a given location. If a property's market value is represented by its most probable market price, determined in view of transaction prices quoted for similar properties, it can be assumed that interpolation methods relying on the price averaging principle can be applied in this case. Location can be assumed to be the main criterion of property similarity because most market attributes can be related to space (e.g. infrastructure, neighborhood, access, safety, etc.).

Spatial interpolation relies on the phenomenon of spatial autocorrelation to estimate value at a given point based on measurement data. Interpolation methods may use various algorithms for determining the value of the interpolated parameters. Analyses and comparisons of various interpolation methods in surface modeling have been performed by Weber et al. [17], Zimmerman et al. [18] and Goldsztejn and Skrzypek [6]. The existing studies focus mostly on the modeling of geological surfaces, therefore the resulting conclusions are not always applicable to economic phenomena comprising transaction prices as regionalized variables.

The choice of interpolation method for unevenly distributed measurement data should be dictated by the following features characterizing the analyzed dataset: the degree of homogeneity of data distribution, the number of points per surface unit, population variance (degree of data variance) and the type of surface represented by the data [6]. This study analyzes three methods of spatial interpolation of transaction prices: inverse distance weighting (IDW), local polynomial interpolation (LPI) and ordinary kriging. The above methods support the development of land worth maps, where value at a given point in space is determined based on the prices of property in the vicinity, thus satisfying the location-based similarity condition.

The inverse distance weighting method is based on the assumption that spatial objects that are close to one another are more similar than objects which are father apart [8]. The general formula for calculating value at a given point in space is as follows [18]:

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i),$$

where:

$\hat{Z}(s_0)$ – the searched value for location s_0,

n – the number of measured points (prices) which constitute the basis for prediction,

λ_i – the weight assigned to each measured point (price).

Weights are calculated with the use of the following formula [8]:

$$\lambda_i = d_{i0}^{-p} / \sum_{i=1}^{n} d_{i0}^{-p}, \quad \text{where} \sum_{i=1}^{n} \lambda_i = 1.$$

The weight of observations decreases with an increase in distance d_{i0} between the point where value s_0 is predicted and reference point s_i. The value of exponent p is selected so as to minimize the prediction error.

Local polynomial interpolation (also known as Lagrange interpolation) involves the approximation of the searched surface with the use of a polynomial, usually a first-degree polynomial:

$$Z(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \varepsilon(x_i, y_i),$$

where $Z(x_i, y_i)$ denotes the searched value in a given location, and $\varepsilon(x_i, y_i)$ is the random error.

Whereas global interpolation involves all data, local polynomial interpolation relies only on points in the direct vicinity of the point at which value is determined. The polynomial surface is fitted to the immediate vicinity of the interpolated point, and unknown parameters β_i are calculated locally, separately for each point, by the least squares method.

Therefore, global interpolation describes local trends, and it is calculated based on the minimization of the following expression:

$$\sum_{i=1} w_i \left(Z(x_i, y_i) - \mu_0(x_i, y_i) \right)^2,$$

where n is the number of points in the vicinity, $\mu_0(x_i, y_i)$ is the value at the measurement point, and weight wi is calculated using the following formula [8]:

$$w_i = \exp(-3d_{i0} / a),$$

where d_{i0} is the distance between the interpolated point and the measurement point from the immediate vicinity of the interpolated point, while a describes the rate at which the value of the weight decreases with an increase in distance. The higher the value of parameter a, the more local the polynomial becomes.

Kriging is a geostatistical method that provides the best linear unbiased estimators of the value of the analyzed regionalized variable. Sample data (measurement points) inside the area of estimation (area of the sample search) are assigned kriging coefficients (weights) to minimize the mean-square estimation error (kriging variance). Similarly to IDW, this technique assigns greater weight to points located closer to the analyzed point, but unlike in IDW, weights are determined based on a semivariogram calculated with the use of a classical semivariance estimator [2, 15]:

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}\left(z(s_i + h) - z(s_i)\right)^2,$$

where $z(s_i + h)$ and $z(s_i)$ represent the value of the analyzed parameter at points separated by distance h, while n is the number of pairs of points separated by distance h.

The semivariogram function, developed by mapping the semivariance value determined by distance h on the diagram, synthetically describes the structure of variance in the investigated set. The weighed average is approximated with the use of the following estimator:

$$\hat{Z}_v = \sum_{i=1}^{N} w_i Z(x_i).$$

The weights, determined with the use of a semivariogram, are assigned in a way ensuring that the estimator is unbiased and has a minimal variance, referred to as the kriging variance.

The most popular kriging estimators include ordinary kriging, used when average value is unknown, and simple kriging, applied in non-stationary conditions when the average estimated for the entire sample population is known or as a result of a local estimation. According to many experts [4, 6, 15, 17, 18], kriging is the most effective technique for interpolating non-uniform natural phenomena because its algorithm contains procedures supporting the minimization and random distribution of errors (the estimator is unbiased, implying that negative and positive errors are balanced out). Kriging accounts for general data trends, and is especially useful when data are not evenly distributed.

3. Results

This study evaluated the usefulness of three spatial interpolation methods for the cartographic visualization of land worth: inverse distance weighting (IDW), local polynomial interpolation (LPI) and ordinary kriging. The option of estimating land worth individually with the use of kriging was also investigated.

The study was carried out in 2007–2009 in the city of Olsztyn. It analyzed data from 188 transactions involving undeveloped land plots zoned for housing construction. The data were acquired from the register of property prices and values kept by the Property and Geodesy Department at the Olsztyn City Office. They were verified in the part relating to land zoning, type of land use and terms of transaction. All properties evaluated in this study were privately owned. The spatial distribution of the investigated transactions is presented in figure 1.

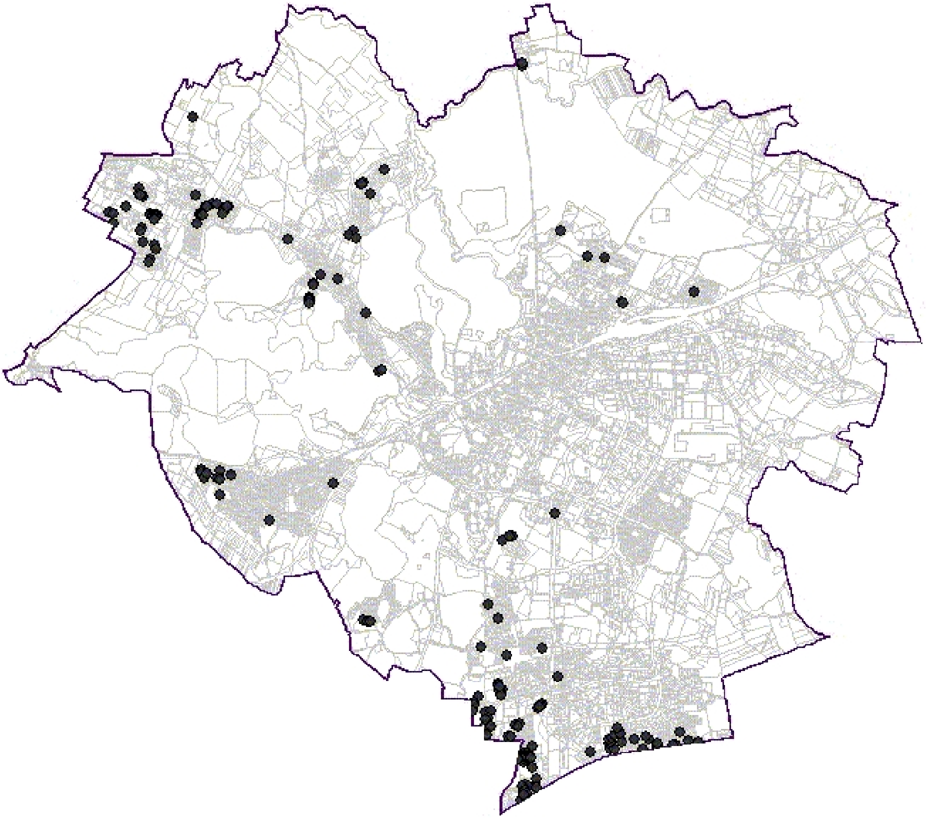


Fig. 1. Spatial distribution of transactions of land property zoned for housing construction in 2007–2009

Although the data were interpolated throughout the entire city of Olsztyn, the results can be sensibly interpreted only in locations where land was zoned for housing purposes. Discontinuity will be taken into account at a further stage of the study. A relatively high price increase (Fig. 2), reaching around 18% per annum, was observed in the analyzed period. An expensive time span of 36 months necessitated price adjustments, therefore the applied prices were updated as of the analysis date (beginning of 2010).

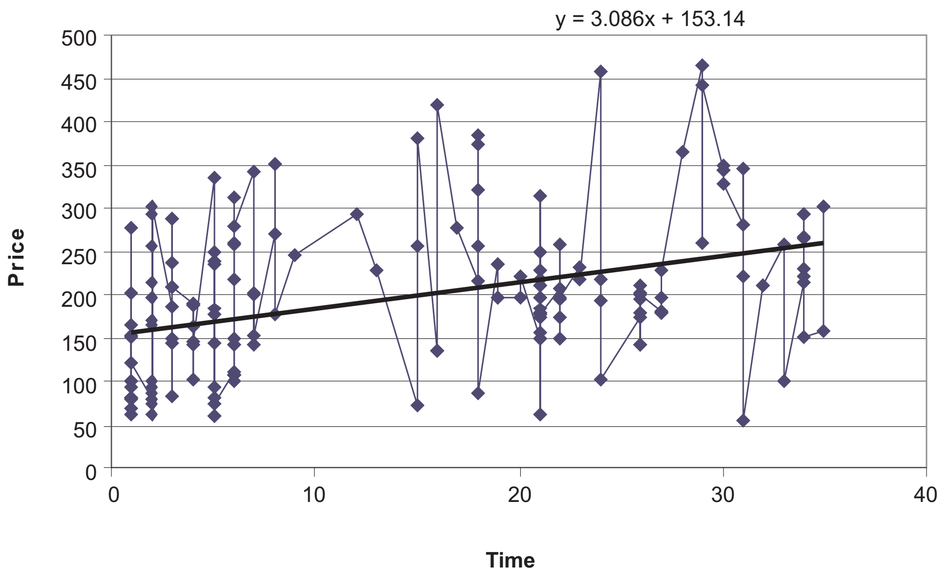
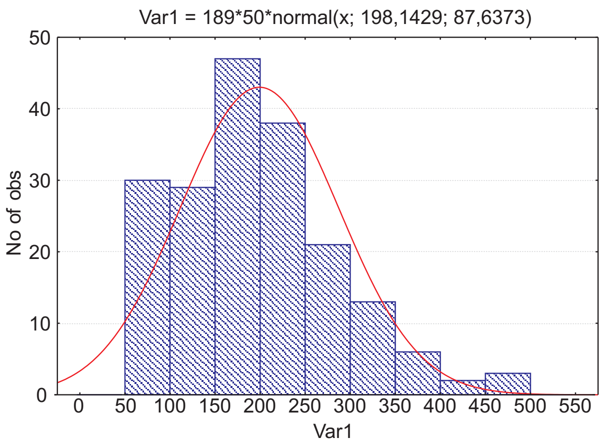


Fig. 2. Change trends in land property prices in 2007–2009

Relatively high variations in the updated prices were also noted in the range of 55 to 466 PLN/m^2 which, in line with the adopted hypothesis, could indicate that location significantly affects unit prices. The empirical distribution of unit prices and their basic statistics are presented in figure 3.



	Basic statistics
Valid n	188
Mean	198.14
Median	196.00
Minimum	55.00
Maximum	466.00
Variance	7680.30
Stat.Dev.	87.64

Fig. 3. Empirical distribution of updated unit prices of land property

The usefulness of interpolation methods should be evaluated mainly in view of the size of interpolation errors. An interpolation error is calculated from the difference of a parameter's interpolated value at the tested point and the parameter's field value [6]. Error estimation may pose certain problems in the interpolation of transaction prices because the price, regarded as a fact, is, in principle, an errorless value, and it does not always reflect a given location's market attributes. The appearance, legibility and realistic content of the map is also significant in the evaluation of an interpolation method.

The obtained data were analyzed with the use of the ArcGIS v. 9.3 application with Geostatistical Analyst and Spatial Analyst extensions as well as SAGA GIS (System of Automated Geoscientific Analysis) software. The results are presented in the form of isoline maps, and the contour interval was determined with the use of „natural intervals", i.e. an algorithm that automatically identifies clusters and „intervals" in attribute values.

Interpolation results obtained with the use of the IDW method are presented in figure 4. In the analyzed case, prediction errors generated by this technique may reach up to 70 PLN/m^2, i.e. more than 30% of the property's average unit price.

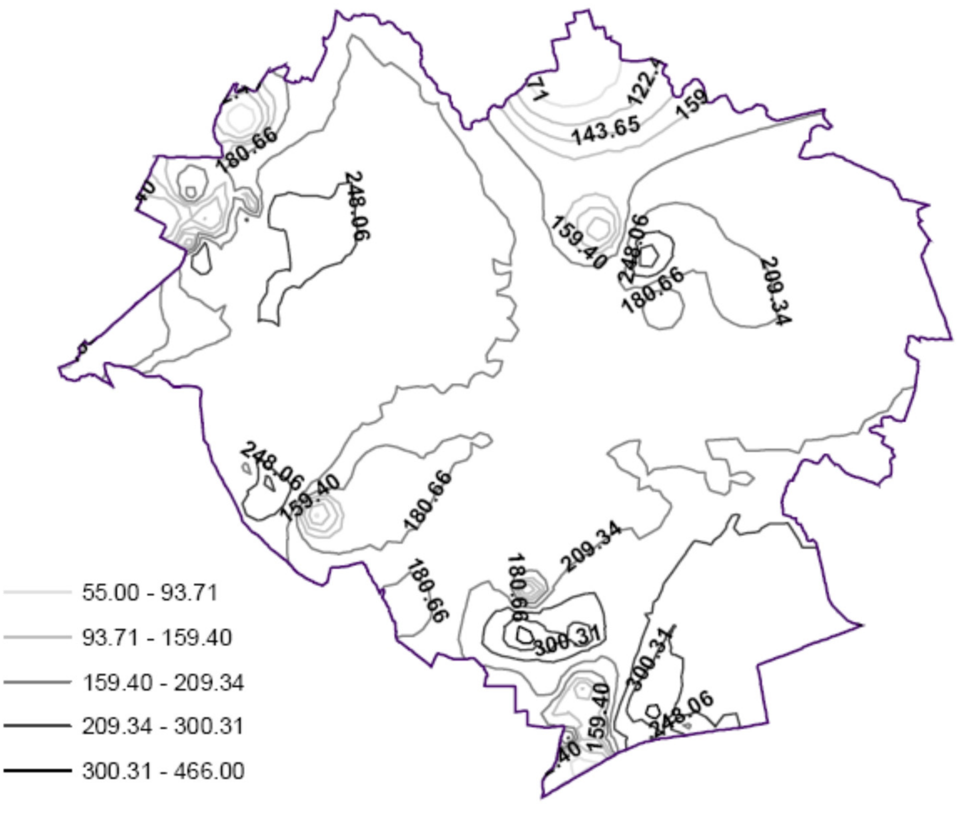


Fig. 4. Interpolation by inverse distance weighting (IDW)

When applied in price analyses, the IDW method has a number of shortcomings. Instead of interpolating values for recreating „pits", the IDW algorithm maps local „dips" where peaks should be determined, and local „elevations" where pits should be mapped [8, 14]. Goldsztejn and Skrzypek [6] also noted that the method creates bull's eye type of structures, an unrealistic accumulation of isolines around extreme values, thus generating relatively improbable surfaces. The results of this study validate the above shortcomings, in particular in north-western and southern parts of Olsztyn (Gutkowo and Jaroty estates). It can also be postulated with a very high degree of probability that the IDW technique will produce incorrect results in price analyses covering areas not described by measurement data, i.e. in estates where no transactions involving undeveloped land property took place.

Local polynomial interpolation produces clearly different results than IDW (Fig. 5), not only in areas marked by a high concentration of measurement points (transactions), but also in areas with dispersed transactions. Although the mapped surface seems to reflect market attributes more accurately, prediction errors proved to be higher than in the IDW technique.

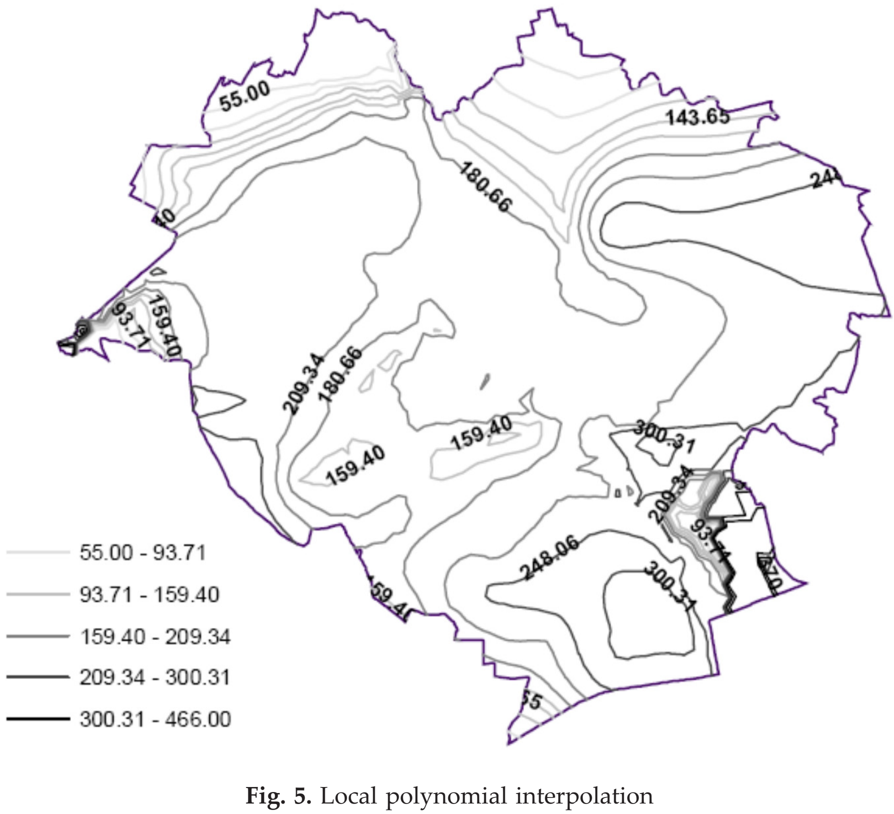


Fig. 5. Local polynomial interpolation

Local polynomial interpolation generalizes local trends and smoothes the surface even with use of fourth- and fifth-degree polynomials. Significant surface smoothing has been particularly observed in the northern part of the analyzed area where land property zoned for housing construction is scarce. A non-standard course of the isoline in the south-eastern part of the city (Pieczewo estate) was also noted, pointing to sudden fluctuations in worth and suggesting that this method generates unreliable results when extrapolation is required outside the area containing measurement data (transaction prices).

Interpolation involving ordinary kriging produced the most varied surface, representative of land worth (Fig. 6). The isoline was also characterized by a non-standard course, pointing to anisotropy in the eastern part of Olsztyn. The average land worth determined in this area accurately reflects, at least intuitively, the trends on Olsztyn's real estate market (around 200 PLN/m^2). Empirical semivariogram is presented in figure 7.

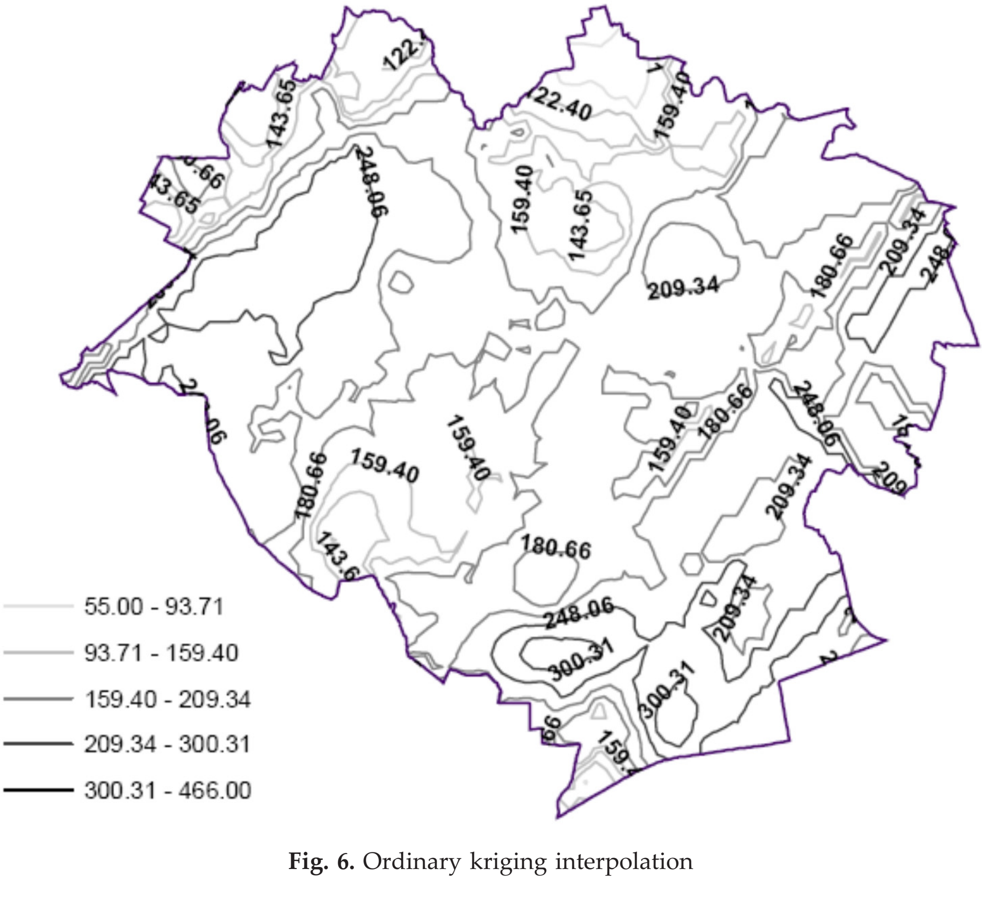


Fig. 6. Ordinary kriging interpolation

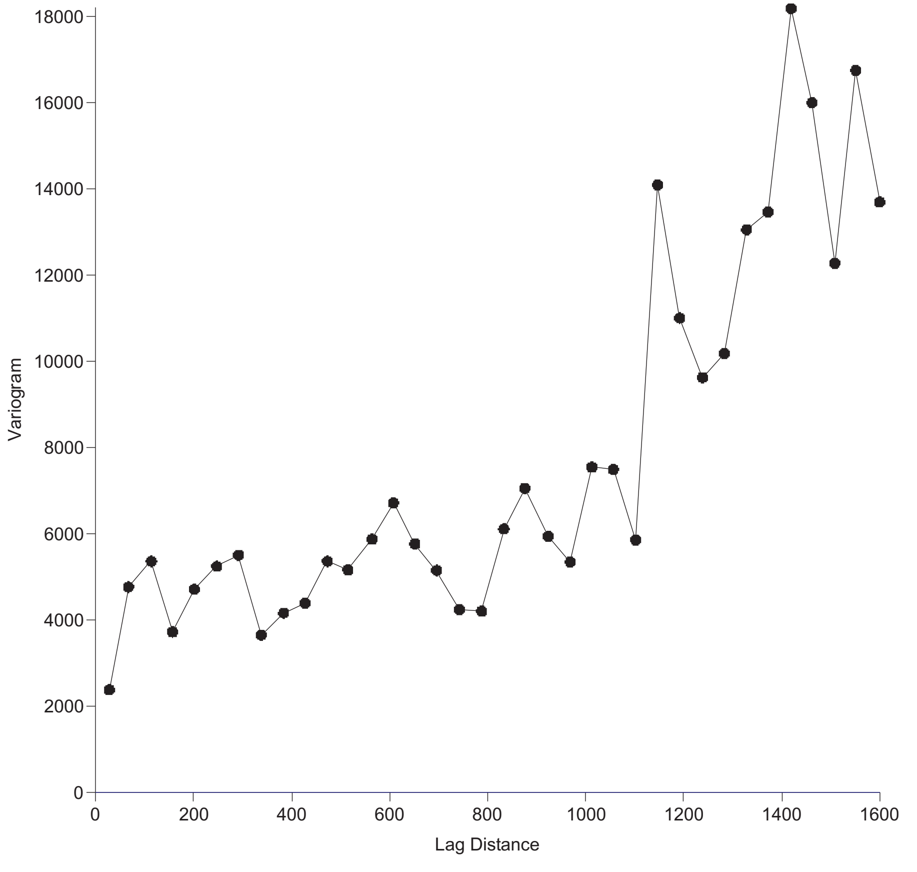


Fig. 7. Empirical semivariogram

Prediction errors in ordinary kriging were significantly lower than in the two previous techniques, and they were clearly determined by the density of measurement points, i.e. land property transactions (Fig. 8).

Kriging supports an independent selection of semivariogram functions, and it accounts for anisotropy (directions of spatial dependencies), which is highly significant for determining the worth of land in the vicinity of linear spatial objects (mainly roads, forest boundaries, lake shores, etc.).

The presented methods for the spatial interpolation of transaction prices of land are a valuable source of information on local market trends. Spatial interpolation facilitates the automation of the process of determining worth at any point whose coordinates are known, e.g. at points situated at the center of land plots, provided that they satisfy the criteria of similarity to the analyzed property. The results of the study were used to produce a land worth map for selected parts of Olsztyn (Fig. 9).

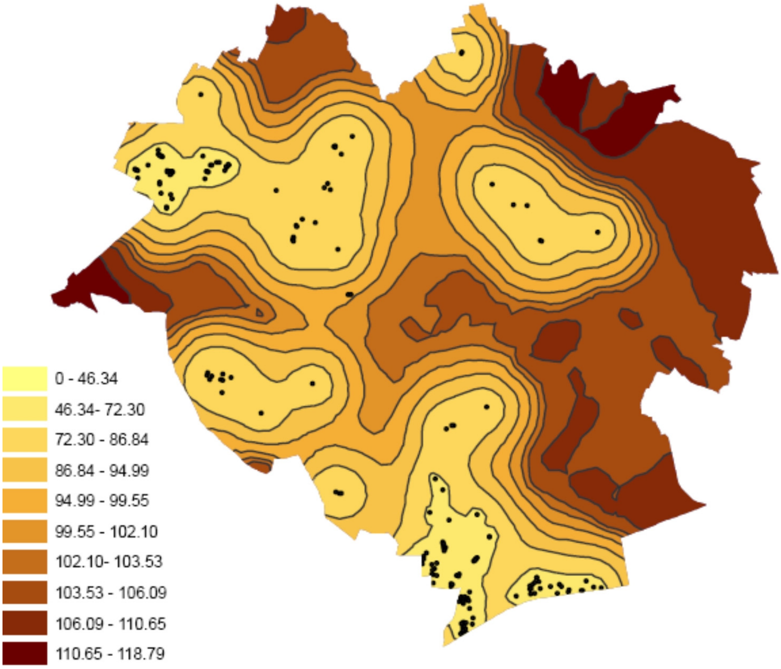


Fig. 8. Spatial distribution of standard prediction errors using ordinary kriging

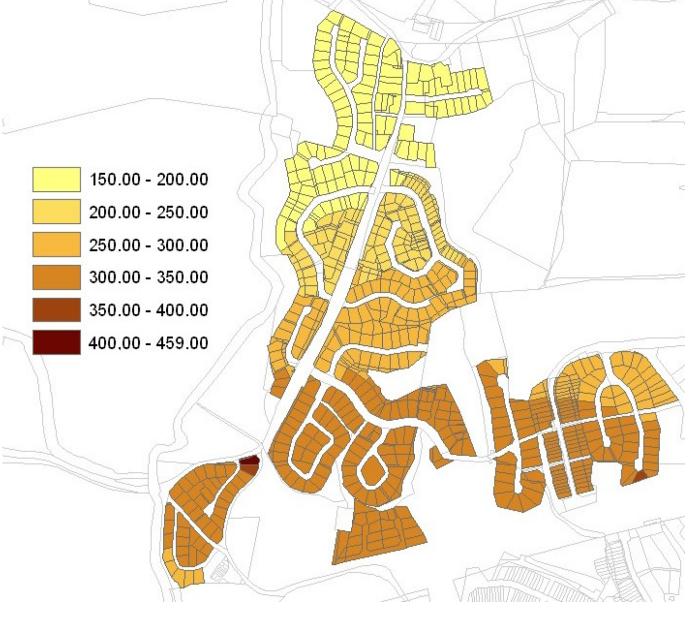


Fig. 9. Fragment of a map presenting the worth of land zoned for housing development (Olsztyn, Brzeziny estate)

The study revealed that various factors distort the shape of the interpolated surface, including high density of transactions in a relatively small area (sale of neighboring properties), significant variations in the prices of neighboring properties as well as non-spatial factors (e.g. surface area and shape of the land plot). The above does not undermine the usefulness of the reported results which may be applied to evaluate spatial phenomena on the real estate market and support a rational decision-making process in property management.

4. Conclusions

Interpolation methods account for general price trends in different ways. Based on the findings of a comparative analysis, ordinary kriging emerges as a technique delivering the most reliable and realistic property worth maps in the process of spatial interpolation of market data. The results of this study validate the observations made by other authors [4, 6] who found kriging to be a useful method for interpolating relatively small datasets (188 transactions in this study) marked by uneven distribution in space. The study was performed on the assumption that the effect of distance on land worth is a regular, continuous and isotropic phenomenon (showing identical change in every direction). This assumption has not been verified in every situation on the real estate market. Although it can be intuitively assumed that spatial variation in prices and values has a continuous nature and it can be represented by isolines from the interpolation process, this assumption does not always hold true due to the spatial distribution of price determinants (e.g. infrastructure, proximity of noxious facilities). Owing to the presence of non-spatial attributes affecting transaction prices, the analyzed dataset cannot be regarded as homogenous. This problem can be solved by combining traditional methods for analyzing the property market and performing bulk sale valuations (e.g. multiple regression analysis) with interpolation techniques [13].

The knowledge of transaction prices and their spatial distribution in the form of cartographic representations facilitates effective decision-making in the area of urban planning and administration, it creates new directions for development and investment, it supports property taxation systems and creates better access to information on the prices and value of land property.

References

[1] Anselin L.: *Advances in spatial econometrics.* Springer, Berlin 2004.

[2] Basu S., Thibodeau T.G.: *Analysis of spatial autocorrelation in housing prices*. Journal of Real Estate Finance and Economics, 17, 1998, pp. 61–85.

[3] Cichociński P.: *Zastosowanie metod kartograficznych i geostatystycznych do wstępnej analizy rynku nieruchomości*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 15, nr 3–4, pp. 155–165.

[4] Davis J.C.: *Statistics and Data Analysis in Geology*. John Wiley & Sons, New York 1986.

[5] Gillen K., Thobodeau T., Wachter S.: *Anisotropic autocorrelation in house prices*. Journal of Real Estate Finance and Economics, 23(1), 2001, pp. 5–30.

[6] Goldsztejn P., Skrzypek G.: *Wykorzystanie metod interpolacji do numerycznego kreślenia map powierzchni, geologicznych na podstawie nieregularnie rozmieszczonych danych*. Przegląd Geologiczny, vol. 52, nr 3, 2004, pp. 233–236.

[7] Haining R.: *Spatial analysis of regional geostatistics data*. Cambridge University Press, 2003.

[8] Johnston K., Ver Hoef J.M., Krivoruchko, Lucas N.: *Using ArcGIS Geostatistical Analyst*, ESRI, New York 2003.

[9] Kelley Pace R., Barry R., Sirmans C.F.: *Spatial statistics and real estate*. Journal of Real Estate Finance and Economics, 17, 1998, pp. 5–13.

[10] Kulczycki M., Ligas M.: *Zastosowanie analizy przestrzennej do modelowania danych pochodzących z rynku nieruchomości*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 15, nr 3–4, 2007, pp. 145–154.

[11] Kulczycki M., Ligas M.: *Interpolacja danych geoprzestrzennych pochodzących z rynku nieruchomości*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 17, nr 2, 2009, pp. 77–78.

[12] Leuangthong O., Khan K.D., Deutsch C.V.: *Solved problems in geostatistics*. John Wiley & Sons, 2008.

[13] Ligas M.: *Zastosowanie modelu regresja-kriging do predykcji wartości nieruchomości*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 17, nr 1, 2009, pp. 7–16.

[14] Longley P., Goodchild M., Maguire D., Rhind D.: *Geographic Information Systems and Science*. John Wiley & Sons, The Atrium, Southern Gate, Chichester, 2005.

[15] Stach A.: *Analiza i modelowanie struktury przestrzennej*. [in:] Zwoliński Z. (Ed.), *GIS – platforma integracyjna geografii*. Bogucki Wydawnictwo Naukowe, Poznań 2009.

[16] Tu Y., Sun H., Yu S.: *Spatial autocorrelations and urban housing market segmentation*. Journal of Real Estate Financial and Economy, 34, Springer Science, LLC, 2007, pp. 385–406.

[17] Weber D., Englund E.J.: *Evaluation and comparison of spatial interpolators*. Mathematical Geology, 24(4), 1992, pp. 381–391.

[18] Zimmerman D., Pavlik C., Ruggles A., Armstrong M.P.: *An experimental comparison of Ordinary and Universal Kriging and Inverse Distance Weighting*. Mathematical Geology, 31(4), 1999, pp. 375–390.

[19] Źróbek S., Cellmer R., Kuryj J.: *Land value maps a source of information about local real estate market*. Geodezja (AGH-UST semi-annual), t. 11, z. 1/1, 2005, pp. 68–73.