

SPATIO-TEMPORAL ANALYSIS OF THE REAL ESTATE MARKET USING GEOGRAPHIC INFORMATION SYSTEMS*

Piotr Cichociński, PhD

*AGH University of Science and Technology
Department of Geomatics
e-mail: Piotr.Cichocinski@agh.edu.pl*

Janusz Dąbrowski, PhD

*AGH University of Science and Technology
Department of Geomatics
e-mail: geo-staszic@wp.pl*

Abstract

The paper proposes the use of geographic information system tools for the analysis of spatial and temporal aspects of the real estate market. In particular, it focuses on the graphical presentation of the spatial distribution of price and its variability over time. The possibility of presenting an image of the spatial distribution of prices in the form of a 3D model is studied. A topographic surface is proposed as an alternative to traditional methods of spatial interpolation. Visual verification and numerical comparison have shown its superiority over other previously used methods.

The best method of presenting four-dimensional data – the variation in time of the spatial distribution of house prices – was sought. The possibility of taking time into account as one of the attributes of the analyzed and presented objects, available in advanced GIS software, was used for this purpose. The undertaken activities were based on formal guidelines for the registration of time set out in the ISO 19100 series of standards dedicated to geographic information.

Potential sources of data for this kind of analysis were identified and their availability was examined. The paper also presents how to build a spatial database on the basis of the available information, which is a starting material for further analysis.

The carried out research demonstrated the benefits of the spatial approach to trends of changes in real estate prices, which can be used, among others, for mass appraisal.

Keywords: geocoding, spatial interpolation, topographic surface.

JEL Classification: C33, C81, R31.

Citation: Cichociński P., Dąbrowski J., (2013), "Spatio-temporal analysis of real estate market using geographic information systems", *Real Estate Management and Valuation*, vol. 21, no. 2 pp. 73-82.

DOI: 10.2478/remav-2013-0019.

1. Introduction

The colloquial meaning of the market is a meeting place for buyers and sellers, at which the buying and selling of goods takes place. In economics, the market is defined as a trade zone, which is characterized by the exertion of mutual influence by potential buyers and sellers, resulting in the formation of supply and demand and the determination of the price level (HOPFER, CELLMER 1997). On this basis, it can be stated that the basic elements of the market are: supply, demand and the resulting

* This work has been prepared as a part of scientific research carried out in Department of Geomatics.

price (REKOWSKI 1995).

Real estate is a special kind of commodity. The essence of real estate is that it cannot be moved from place to place. The result is that the location of the real estate is of utmost importance, since, basically, this feature cannot be changed. Furthermore, it allows people, especially those not oriented in the specifics of the market, to identify areas with lower prices.

The extraction, processing and sharing of data regarding the location, geometric properties and spatial relationships between objects that can be identified in relation to the Earth are the basic tasks performed by geographic information systems (GIS). These systems provide a number of tools for analyzing the characteristics (attributes) and relationships (resulting from location) of spatially defined objects (features).

Any action related to the real estate market (such as, for example, purchase or valuation) should be preceded by an analysis of this market. The determination of prices in a given area is closely related to the spatial distribution of properties and, therefore, is a task well accomplished by using GIS tools.

2. The issue of time in geographic information systems

Notarial deeds are a valuable source of information about properties that were the subject of a trade. Such data collected over an extended period of time may have particular value. The availability of large amounts of constantly updated data opens up the possibility of systematizing them and obtaining useful knowledge, which would otherwise have been lost.

Currently, the majority of geospatial data analyses are performed on a static image of reality registered at a certain point in time and describing only the spatial characteristics of the phenomena (Yao 2003). However, because all geographic phenomena evolve over time, both the spatial and temporal characteristics are key to the understanding of geographic processes and events. Furthermore, knowledge extracted from spatio-temporal data will help to better predict the spatial processes and events. Therefore, it is important to carry out analyses on the spatio-temporal data sets. This introduces the aspect of time to such analysis and thus, gives it spatio-temporal character.

Many geoinformatic systems can collect temporal data. To systematize time related issues, works were carried out within the International Organization for Standardization Technical Committee No. 211 "Geographic Information/Geomatics" (ISO/TC 211), which resulted in the development of the International Standard ISO 19108 Geographic Information - Temporal Schema, subsequently adopted as the Polish Standard. The authors of this standard assumed that the standardized conceptual schema for the temporal characteristics would increase the possibility of using such information in various areas of application. For the purpose of exchanging temporal information, it builds on existing standards in the field of information technology. The standard defines the concepts necessary to describe the temporal characteristics of geographic information observed in the real world. Temporal characteristics include feature attributes, operations performed on the features, feature relationships, and temporal metadata elements.

However, although the ISO/TC 211 defines the temporal model, from the point of view of information technology, this model does not clearly define temporal aspects that the data model should associate with the types of geographical features and their attributes. Therefore, in practical applications, it is convenient to continue using the simple and proven solution (Peuquet 2001) - time-related data can be stored in existing GIS databases using the "snapshot" data model. This type of representation consists of a different use of one of the two traditional data models used in geographic information systems - a sequence of raster images. However, instead of a single file containing a complete thematic layer, as is the case in a static (not time related) spatial database, each image represents the state of reality in a certain thematic range, registered at a particular point in time. Pixels in the individual "snapshots" contain values corresponding to the respective places in space for a given moment in time. Although it introduces a high degree of data redundancy, when the described phenomenon does not change continually and throughout the whole area, this approach is conceptually simple and it is easy to restore a complete state of reality for each of the stored spatio-temporal locations. Furthermore, this reflects a way of collecting data from popular sources, such as aerial photographs and satellite images, or census.

The inclusion of time as an additional dimension also causes challenges for the optimal visualization of spatio-temporal data. In fact, all the methods traditionally used to present data on conventional spatio-temporal maps continue to apply in the age of computer cartography

(ANDRIENKO et al. 2003).

One of them is a “strip map” (KRAAK, MACEACHREN 1994), involving the juxtaposition of a chronological series of maps (in a style similar to that of a comic book), where each map shows the state of the phenomenon at a different point in time. Differences arising from the passage of time are shown by the differences between the various maps depicting the successive “snapshots” (HILL, DONALD 2003). This way, the temporal sequence is converted into a spatial sequence (the order in which the individual maps are presented to the recipient). It is a versatile technique, suitable for use with any type of spatio-temporal data. Of course, the number of visible images which can be represented simultaneously on the computer screen or a sheet of paper is reduced, so that long series have to be presented with a low temporal resolution.

An extension of this concept are animated maps which, however, require media other than paper. They consist of either a suitably rapid display of sequences composed of static maps arranged in chronological order or the presentation of a map, which changes dynamically. In this case, the changes resulting from the temporal component are presented on a single map. The variability in time does not result from the spatial sequence, but from real changes of the map image, which can be regarded as a movement. This type of animation shows the changes in position and/or attribute values, observed from a fixed point of view. A classic example of an animation is a film prepared using a computer, showing the development of cities in the region of Detroit (TOBLER 1970).

In today's advanced GIS software (ESRI 2011) it is sometimes possible (although recently, this has become widely and easily available through the popular and not even addressed to professionals Google Earth software) to use a slider that allows one to move along a timeline and select an interesting point in time.

In addition to the above described methods, change maps are also used to estimate changes in individual objects or locations, allowing the locations in which the changes occurred to be shown precisely. They are created by calculating the mathematical difference between two images (“snapshots”) registered at certain moments (DENG et al. 2009).

3. Address geocoding

One should be aware that the location of a given piece of real estate in notarial acts is presented in the form of an address rather than coordinates. Therefore, geocoding is an action which must precede spatial analysis. Geocoding is the process of determining the location, usually expressed in terms of (point) coordinates, of the object described by the address. This is done by comparing the relevant elements of address information with source material. Addresses can be written in many different forms. In Poland, for towns in which the streets have names, it is usually assumed that the address includes the name of the street, followed by the sequence number of the building (and possibly the number of the apartment). In addition, the zip code and the name of the city, town or village are given. This set of information can uniquely identify a particular location anywhere in the whole country.

Geocoding is a process which consists of three major steps (CHARIF et al. 2010):

- 1) The structuring and standardization of searched addresses in order to identify individual components in the entered complete address (house number, street name, postcode, city, etc.),
- 2) Finding a matching address in the source database by comparing it with the corresponding elements of the searched address,
- 3) Proper geocoding, which determines the coordinates of the identified address.

There are three basic methods for determining the location of an address. The most accurate one consists of getting the coordinates of the point object. Another popular method is coordinate interpolation performed along a linear object, assuming a range of numbers assigned to each of these objects. The basis for performing the third method is determining, by any means, a point within a polygon or line.

The oldest and best-known method of geocoding requires a set of line segments representing the streets to which a name and numbering range are assigned on their left and right sides in the form of attributes. Geocoding, in this case, consists of first finding a matching street, then determining the segment the numbering range of which includes the relevant building number and finally, interpolating the position on the line on the basis of number range limits. This method is called street geocoding. Its application involves a lot of problems, widely reported in literature (ZANDBERGEN 2007, ZANDBERGEN 2011, ZIMMERMAN, LI 2010). Most objections are raised by linear interpolation which is

carried out to determine the coordinates. Some authors report (CAYO, TALBOT 2003) that the error of this operation can be as high as 3 km (calculated as the distance from the actual position of the address to the location determined by geocoding).

As to the advantages of street geocoding, one can mention the lower sensitivity to the incompleteness of a database, since the sought address does not have to be found directly and thus, stored in the database, as it is sufficient when the range it is included in is made available. Additionally, cases where the searched address is outside of any available range have also been considered.

Geocoding on the basis of points is rated more positively (ZANDBERGEN 2008, VIEIRA et al. 2010) due to its excellent positional accuracy. It also provides an additional check of the entered address data as it requires finding the real address in the source database and is not based on the numbering range that may be incomplete in the field.

Considering the availability of appropriate software and despite the above comments (KARIMI et al. 2004), the stated comparability of algorithms used, the basis for successfully carrying out the geocoding process, is to have an appropriate source material which will determine the location of the address.

As part of building the infrastructure for spatial information in Poland (USTAWA 2010), a database of address points is yet to be developed (ROZPORZĄDZENIE 2012). It is based on the address of a building which consists of the name(s) of the state, county, municipality, town, street (and their respective identifiers), ordinal number and postcode. Every ordinal number will be assigned to the address point determined by the approximate center of the wall of the building facing the street or square, which is related to the ordinal number of the building, or the approximate center of the entrance to the building. The point address will be described by a pair of coordinates, specifying the position of the address point.

Although at present address points are basic map elements (GŁÓWNY GEODETA KRAJU 1998), the drawback that prevents their direct application in geocoding is that information associated with them relates only to the number of the building, whilst the second required component, the name of the street, is missing.

4. Data preparation

The primary source of data for the analyses presented below was recorded in a spreadsheet list of notarial deeds of real estate sales and purchases from the years 2003-2011 in the area of Jarosław – a medium-sized town in south-eastern Poland, with a population of just over thirty-eight thousand inhabitants. It should be noted that the database includes all transactions that took place on the real estate market, and so the complete population was analyzed, not just a statistically representative sample. From this list of dwelling transactions described by a set of attributes, those relevant for further actions, i.e., the date of the transaction, the address and area of the property and the transaction price were selected. By dividing the price by the area of the dwelling, the unit price was determined. In some cases, information about the floor on which the dwelling is located and the number of rooms was also available.

A map of the land and buildings of Jarosław, divided into five sheets corresponding to five precincts with each stored in a separate *dwg* file (AutoCAD drawing format), was used as the data source for geocoding. Since the numbering of buildings saved there was not directly suitable for geocoding, new address points were drawn on this background, described by the street name and building number. To reduce time-consumption of this stage of works, it was limited only to those addresses appearing on the list of transactions. This allowed the correctness of addresses stored in notarial acts to be verified. It turned out that in two cases they pointed to industrial buildings, and in some other cases made reference to buildings not shown on the map. In such situations the addresses were corrected, assigning them the most probable new value.

Geocoding was conducted in Esri's ArcGIS, which has a suitable tool for this purpose. It required the creation of a so-called address locator, which determined the roles played by the individual fields (columns) from the attribute table of the source data set, that is in which field the numbers, street names, postal codes, etc. were stored.

A list of 1,499 dwellings from the area of Jarosław traded in the years 2003-2011 was subjected to geocoding. The address locator that was built used two pieces of information, namely, the street name

and building number, which means that the point method was used, with accuracy to individual buildings. Matching options were left unchanged at the default levels suggested by the program, that is, with a spelling sensitivity of 60%, a minimum score of candidates at 10% and minimum score of the match at 90%. The preparation of own address points made it possible to avoid potential problems arising from the diversity in the ways of storing street names derived from names. Two variants could be distinguished in descriptions of the sold dwellings: forename and surname or surname only. In comparison, only the surname usually occurred on the cadastral map, although there were exceptions. Therefore, it was decided that in this experiment, short street names (consisting only of a surname) would be used and they were adjusted to this form both in the real estate database and in the reference data set. After such data preparation, it was possible to achieve an address match of one hundred percent.

5. Estimation of the distribution of unit real estate prices

Spatial interpolation methods previously described and analyzed by Cichociński (2011) were used to generate models of the spatial distribution of selected dwellings. The availability of a new, comprehensive data set provided the opportunity to verify the previously obtained results. The following interpolation methods were applied: inverse distance weighted (IDW), radial basis function (RBF), and kriging.

The first issue appeared during data loading. Transactions of various dwellings located in the same building were linked to the same address point and thus had the same position. However, most of the tools for performing interpolation require only one value subjected to interpolation, associated with a single point. Therefore, the program makes it possible to average a number of values to obtain a single, representative one. The second solution, which the authors of the present article also decided to investigate, was to keep the original data diversity. Geostatistical Wizard, a part of the Geostatistical Analyst extension of Esri's ArcGIS software, was used and the available ability was applied to automatically select the interpolation parameters in such a way that cross-validation errors were minimized. The resulting final precision parameters of the examined methods for data from the year 2011 are presented in Table 1.

Table 1

The result of the cross-validation of individual interpolation methods.

Method	Original values		Average values	
	Mean error	RMSE	Mean error	RMSE
RBF - inverse multiquadratic function	-	-	2	669
RBF - spline with tension	-	-	11	697
RBF - completely regularized spline	-	-	13	704
RBF - multiquadratic function	-	-	22	792
RBF - thin plate spline	-	-	22	951
Kriging	2	743	6	685
IDW	42	779	21	677

Source: own work

The obtained results confirmed the expectation that models built on the basis of average values would be characterized by smaller errors. Methods characterized by the smallest deviations at the data points were selected for further studies. If these methods were used exclusively for presentation in the form of the traditional isoline maps (CICHOCIŃSKI 2007), the resulting models would probably be sufficient. However, it was decided to perform three-dimensional visualization, which exposed the weaknesses of the chosen methods. Especially in the case of the IDW method, rapidly increasing values near the data points and generally, a rather flat surface between them can be seen (Fig. 1). In addition, it turned out that kriging and radial basis functions tend to flatten the model: the original dwelling unit prices in 2011 ranged from 976 to 4,417, while the model obtained on the basis of these prices determined by the kriging method had a minimum value of 2,139 and a maximum value of 3,108.

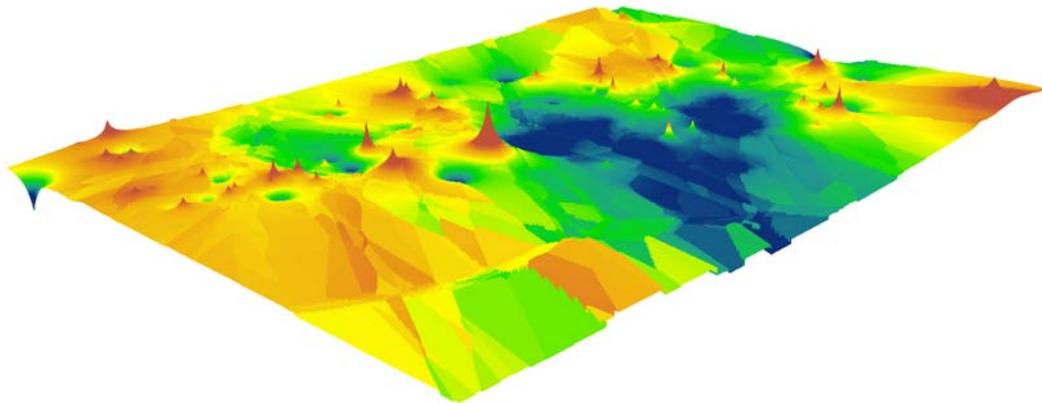


Fig. 1. The values of a model obtained by the IDW method, rapidly increasing around the data points.
Source: own work

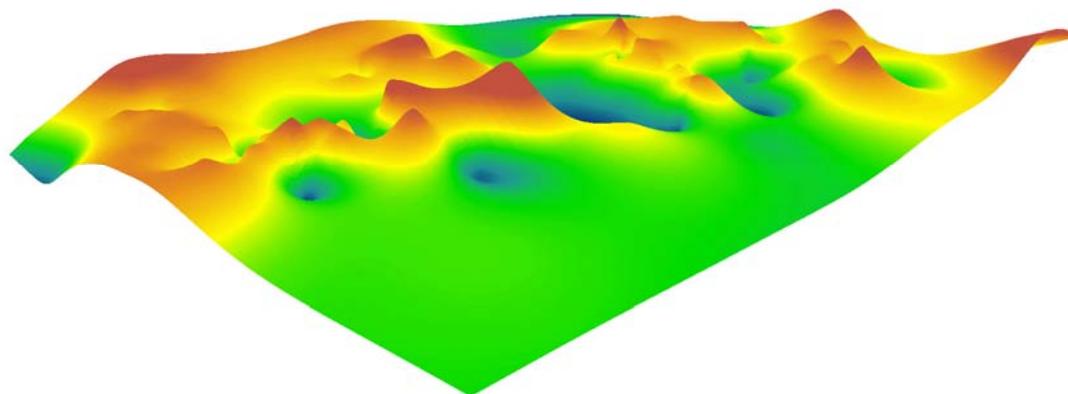


Fig. 2. Natural-looking shape of the surface of a model obtained by the topographic surface method.
Source: own work

These not fully satisfactory results led the authors to look for another solution. The spatial interpolation method, which fulfilled the requirements of both precision and visualization (Fig. 2), proved to be available in the 3D Analyst extension of Esri's ArcGIS software topographic surface method. It is based on the ANUDEM program, developed by Michael Hutchinson (ESRI 2011) specifically to build hydrologically correct digital elevation models (DEM). It uses the technique of splines with constraints resulting from the knowledge of surface topography imposed on the process of interpolation and, as a result, gives the structure of combined drainage areas and a correct representation of ridge and drainage lines. While it does not have a handy tool to automatically make an accuracy assessment of model fitting by cross-validation, such an operation can also be performed by comparing data point values with the interpolated value at the same points. Since this function cannot average values located in the same points, it was necessary to properly prepare the input dataset. For data on transactions that took place in 2011, the resulting standard deviation of differences between the real unit price and the price determined from the model was at a level of 16 zloty, which is a result better by at least an order of magnitude than those obtained by other methods of interpolation.

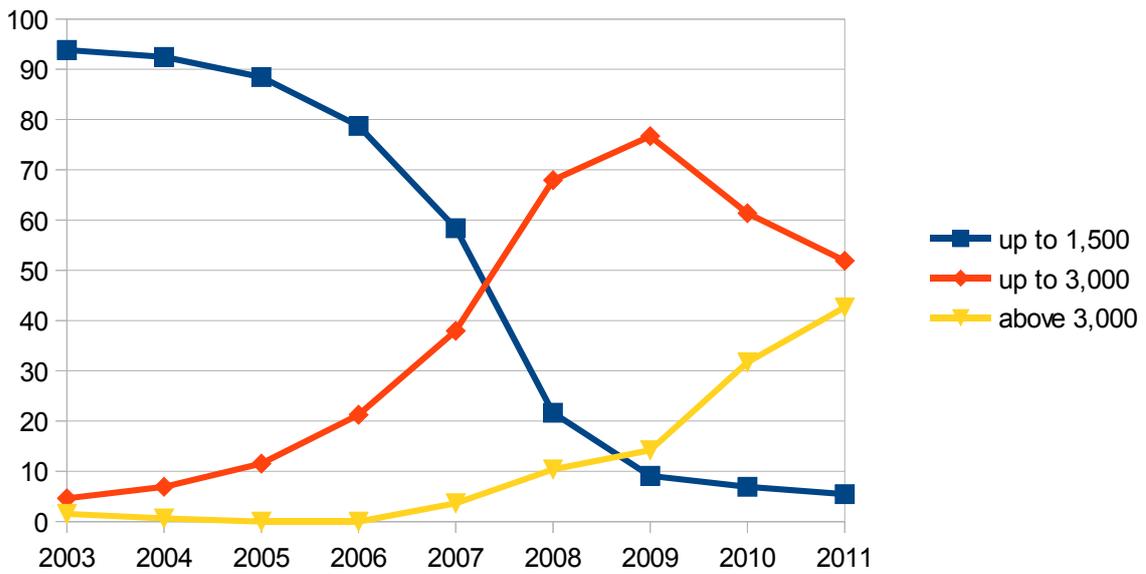


Fig. 3. Percentage of the number of transactions in each range of unit prices in the individual years.
Source: own work

Additionally, the usefulness of this method of interpolation may be proven by the great similarity of graphs presented in Figures 3 and 4, respectively, showing the percentage of the number of transactions in each range of unit prices in the individual years as well as the percentage of the area of city parts in each range of unit prices in the individual years determined on the basis of the models.

Maps of the distribution of dwelling unit prices in the years 2003-2011 interpolated using the topographic surface method are presented in Figure 5. Possessing information on the distribution of real estate prices in the subsequent years also allows spatio-temporal analysis to be conducted. As an example of such possibilities, the authors present maps of the growth of dwelling values throughout the 2003-2011 period covered by the study (Fig. 6) as well as during the short, but characterized by a significant price increase, period of 2007-2008 (Fig. 7).

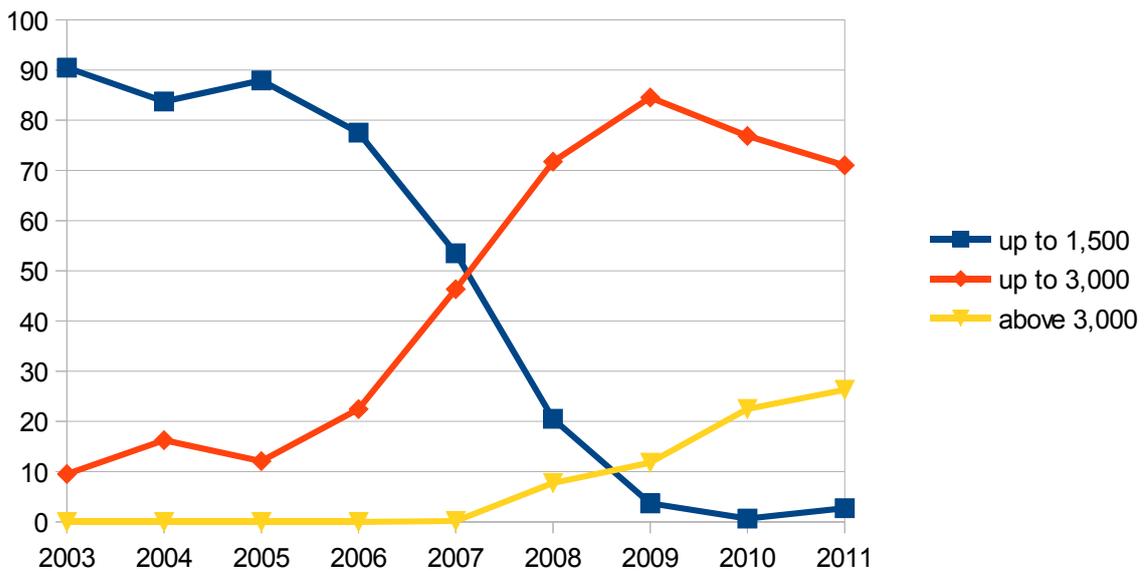


Fig. 4. Percentage of the area of city parts in each range of unit prices in the individual years.
Source: own work

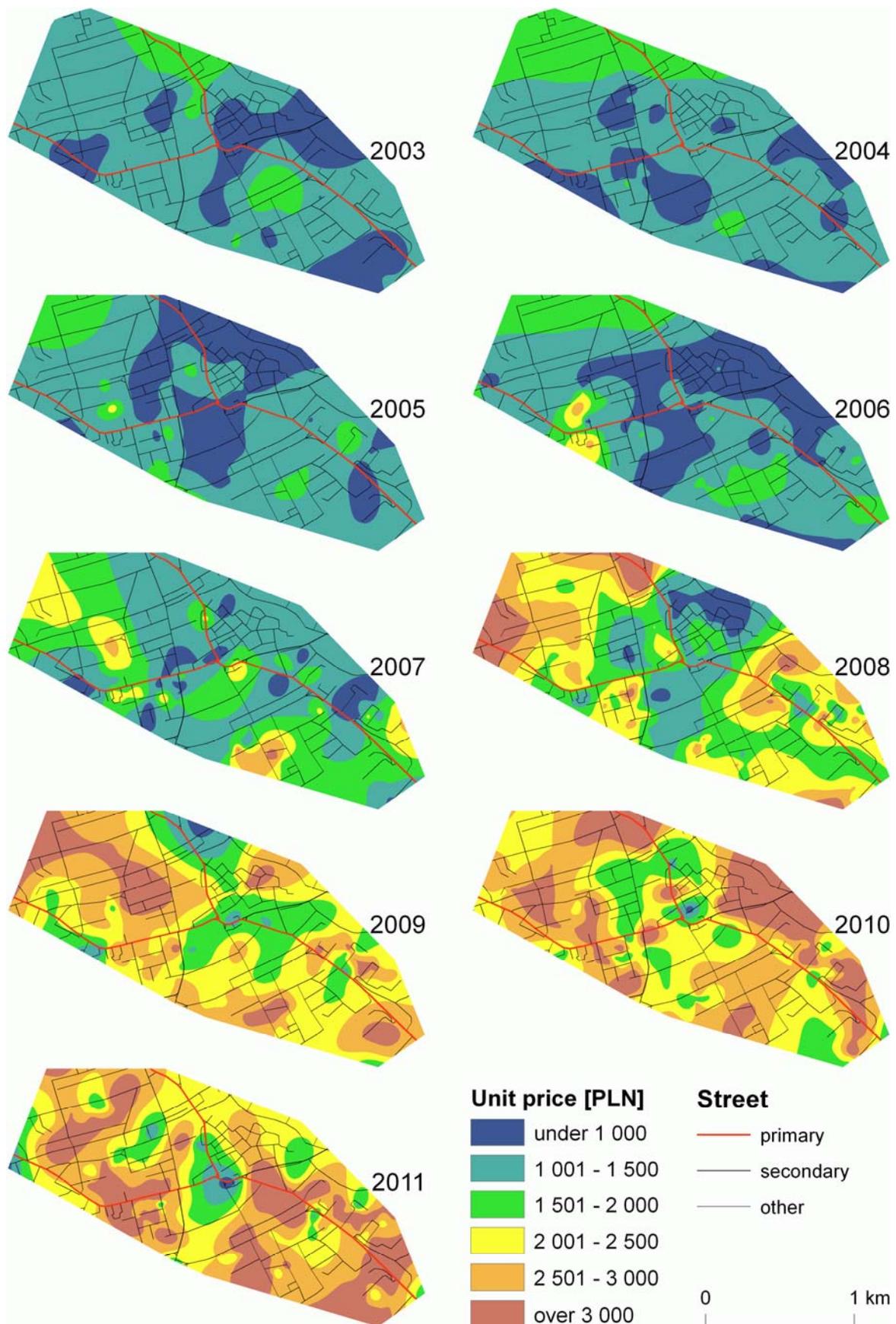


Fig. 5. Maps of the distribution of dwelling unit prices in the years 2003-2011. Source: own work

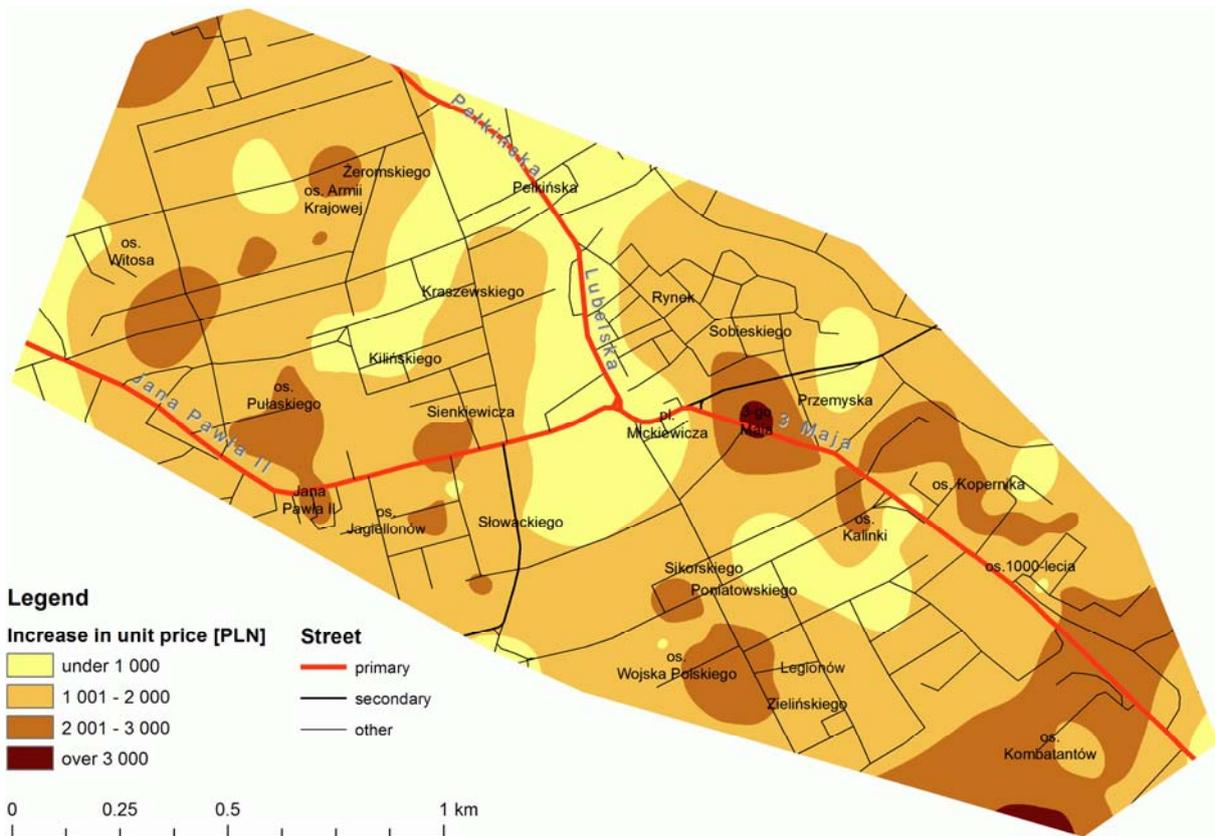


Fig. 6. Map of the increase in dwelling unit prices between the years 2003 and 2011. Source: own work

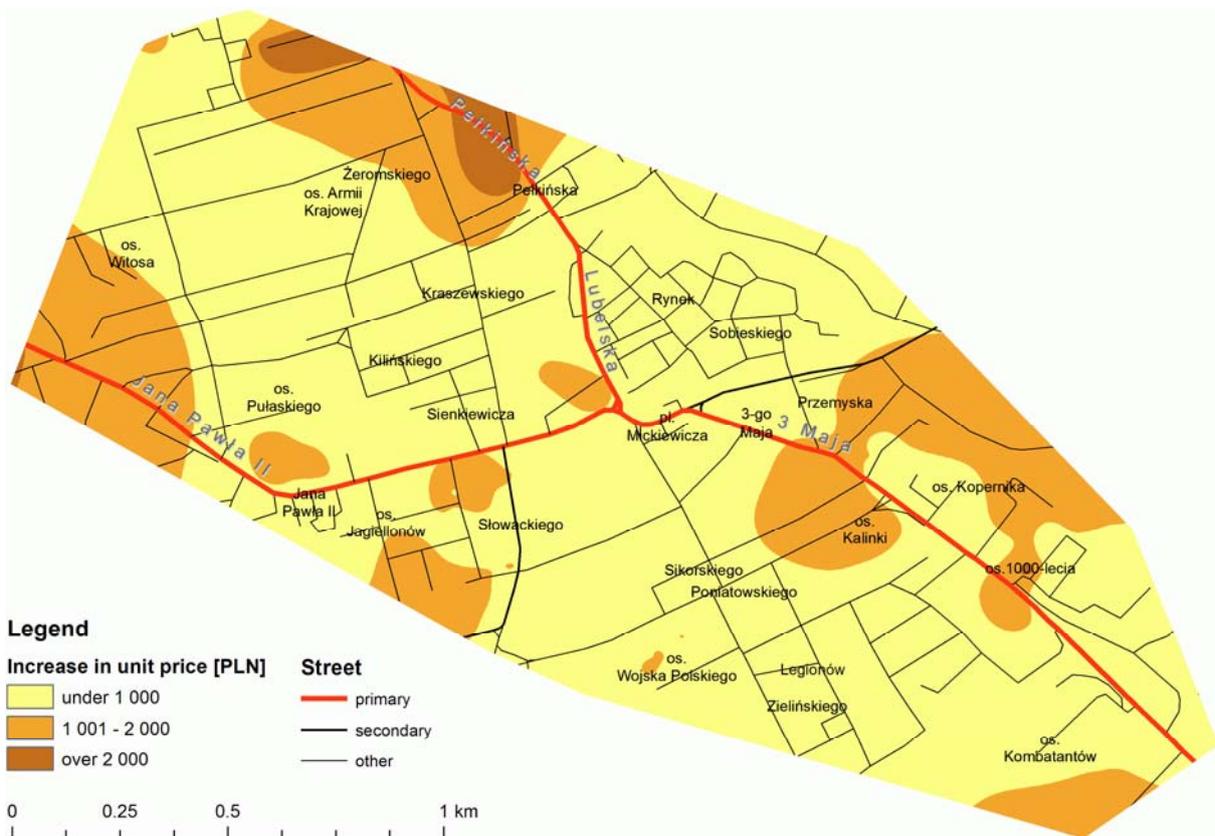


Fig. 7. Map of the increase in dwelling unit prices between the years 2007 and 2008. Source: own work

6. Summary and conclusions

Data collected over an extended period of time is always valuable research material. It allows us to capture not only the current state of the phenomenon, but also some trends that may even allow a cautious attempt at prediction, although in the case of real estate, experience has shown this not to be a safe operation. This is because there is no guarantee, that the observed trends will actually persist, which may be due to macroeconomic factors.

It is even better when one has a complete set of data that describes the whole phenomenon and not just a representative sample. This type of data was collected by the authors of this article. Unfortunately, as of now, information about transactions is not collected in the form of spatial databases, making the preparation phase of data which is to be used for analysis difficult and prolonged. It also requires the availability of the necessary address details for the process of geocoding. The existing maps are not best suited for this purpose but it is hoped that the implementation of the Act on Spatial Information Infrastructure will significantly improve the situation in this regard. However, given the existing legal solutions, one has to take into account that access to such data for all those who are interested (not just authorities) may be limited.

Thus, although data adaptation to the needs of spatial analysis required some effort, it can now be used for complex spatio-temporal analysis of the real estate market. In particular, the use of a single color scale on the "strip map" (Fig. 6) allowed for the sharp increase in the unit prices of dwellings in late 2007 and 2008 to be captured, as also evidenced by graphs (Fig. 3, Fig. 4). Simultaneously, areas where the increase was found to be the highest and places that do not respond to trends in the market were indicated. Maps of real estate unit price increase played an auxiliary role in this regard (Fig. 6, Fig. 7) and facilitated the identification of areas with the highest growth rates, although in this particular case, making classifications in the same intervals (for the sake of visualization comparability) probably resulted in some blurring of details, given the relatively small increase in price in one year.

Examples of the visualization of phenomena and object variability presented in this paper do not cover all the possibilities. The authors are planning further research in this area.

7. Bibliography

- ANDRIENKO N., ANDRIENKO G., GATALSKY P., 2003, *Exploratory spatio-temporal visualization: an analytical review*, Journal of Visual Languages & Computing, 14(6), 503-541.
- CAYO M.R., TALBOT T.O., 2003., *Positional error in automated geocoding of residential addresses*, International journal of health geographics, 2(1):10.
- CHARIF O., OMRANI H., KLEIN O., SCHNEIDER M., TRIGANO P., 2010, *A method and a tool for geocoding and record linkage*, CEPS/INSTEAD Working Paper Series 2010-17.
- CICHOCIŃSKI P., 2007, *Zastosowanie metod kartograficznych i geostatystycznych do wstępnej analizy rynku nieruchomości*, Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 15, nr. 3-4, s. 155-166.
- CICHOCIŃSKI P., 2011, *Porównanie metod interpolacji przestrzennej w odniesieniu do wartości nieruchomości*, Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 19, nr 3, s. 120-132.
- DENG J. S., WANG K., HONG Y., QI J. G., 2009, *Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization*, Landscape and Urban Planning, 92(3), 187-198.
- ESRI, 2011. ArcGIS Desktop 10 Help.
- GŁÓWNY GEODETA KRAJU, 1998, *Instrukcja techniczna K-1. Mapa zasadnicza*, Główny Urząd Geodezji i Kartografii.
- HILL M. J., DONALD G. E., 2003, *Estimating spatio-temporal patterns of agricultural productivity in fragmented landscapes using AVHRR NDVI time series*, Remote Sensing of Environment, 84(3), 367-384.
- HOPFER A., CELLMER R., 1997, *Rynek Nieruchomości*, ART Olsztyn.
- KARIMI H.A., DURCIK M., RASDORF W., 2004, *Evaluation of uncertainties associated with geocoding techniques*, Computer-Aided Civil and Infrastructure Engineering 19, pp.170-185.
- KRAAK M.-J., MACEACHREN A.M., 1994, *Visualization of spatial data's temporal component*, Proceedings, Spatial Data Handling, Advances in GIS Research, Edinburgh, Scotland, 5-9, September, 1994, IGU
- PEUQUET D.J., 2001, *Making space for time: Issues in space-time data representation*, GeoInformatica, 5(1), 11-32.

- PN-EN ISO 19108:2010, *Informacja geograficzna -- Schemat czasowy*.
- REKOWSKI M., 1995, *Wprowadzenie do mikroekonomii*, Polsoft.
- ROZPORZĄDZENIE MINISTRA ADMINISTRACJI I CYFRYZACJI z 9 stycznia 2012 r. w sprawie ewidencji miejscowości, ulic i adresów, Dz. U. z 2012 r., poz 125.
- TOBLER W., 1970, *A Computer Movie Simulating Urban Growth in the Detroit Region*, *Economic Geography*, 46, 2 (1970), pp. 234-240.
- Ustawa z 4 marca 2010 r. o infrastrukturze informacji przestrzennej, Dz. U. nr 76, poz. 489.
- VIEIRA V.M., HOWARD G.J., GALLAGHER L.G., TONY FLETCHER T., 2010, *Geocoding rural addresses in a community contaminated by PFOA: a comparison of methods*, *Environmental Health*, 9:18.
- YAO X., 2003, *Research issues in spatio-temporal data mining*, In White paper submitted to the University Consortium for Geographic Information Science (UCGIS) workshop on Geospatial Visualization and Knowledge Discovery, Lansdowne, Virginia, Nov (pp. 18-20).
- ZANDBERGEN P.A., 2007, *Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads*, *BMC Public Health* 2007, 7:37.
- ZANDBERGEN P.A., 2008, *A comparison of address point, parcel and street geocoding techniques*, *Computers, Environment and Urban Systems* 32 (2008) 214–232.
- ZANDBERGEN P.A., 2011, *Influence of street reference data on geocoding quality*, *Geocarto International* vol. 26, no. 1, February 2011, 35–47.
- ZIMMERMAN D.L., Li J., 2010, *The effects of local street network characteristics on the positional accuracy of automated geocoding for geographic health studies*, *International Journal of Health Geographics* 2010, 9:10.