




Territorial Extrapolation of Basic Data as a Solution of the Problem of Its Deficiency during Mass Appraisal

Jana Volkova ¹, Elena Bykova ² , Maria Heldak ^{3,*} , Katarzyna Przybyła ³  and Sebastian Pawlak ³

¹ Saint Petersburg State University of Architecture and Civil Engineering, 4 Vtoraya Krasnoarmeiskaya, 190005 Saint Petersburg, Russia; yavolkova@lan.spbgasu.ru

² Department of Engineering Geodesy, Saint Petersburg Mining University, 21-Line, 2, 199106 Saint Petersburg, Russia; Bykova_EN@pers.spmi.ru

³ Department of Spatial Management, Wrocław University of Environmental and Life Sciences, ul. Grunwaldzka 55, 50-357 Wrocław, Poland; katarzyna.przybyla@upwr.edu.pl (K.P.); 115625@student.upwr.edu.pl (S.P.)

* Correspondence: maria.heldak@upwr.edu.pl

Abstract: The article is devoted to the application of the territorial extrapolation of basic data method during a mass (cadastral) assessment of a territory that is characterized by an acute lack of market information. In the framework of the study, an acute lack is understood as the conditions when for the assessing territory there are less than five transaction (offer) prices suitable for regression models. The idea of the method is to use market information of territories that are comparable in a composition of pricing factors and the nature of their influence on the cost, as well as in terms of price levels. The developed method includes such stages as collection of basic data, creation of thematic maps, grouping of estimated territories by price level and composition of pricing factors and modeling. The method was applied to assess land plots that have the type of permitted use “for individual housing construction” and belong to the mass appraisal segment “gardening and horticulture, low-rise residential buildings” in the settlements of the Republic of Udmurtia. The results of approbation shown that the method of territorial extrapolation helps to overcome an acute shortage of market information and build statistically significant models of the cadastral values of land plots.

Keywords: cadastral valuation; method of territorial-temporal extrapolation; individual residential development; land plot; modeling; cadastral value



Citation: Volkova, J.; Bykova, E.; Heldak, M.; Przybyła, K.; Pawlak, S. Territorial Extrapolation of Basic Data as a Solution of the Problem of Its Deficiency during Mass Appraisal. *Land* **2021**, *10*, 750. <https://doi.org/10.3390/land10070750>

Academic Editors: Izabela Rącka, Mirosław Belej and Małgorzata Krajewska

Received: 15 June 2021

Accepted: 15 July 2021

Published: 17 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, the term “cadastral value” has become firmly included in the active vocabulary of Russian real estate owners: this is evidenced by both regular mention in the media and Internet search statistics. The reason for this popularity is not the desire to acquire new knowledge, but rather landowner dissatisfaction with the excessive tax burden, which is a consequence of the overestimated cadastral value.

European countries represent a variety of solutions in the field of real estate taxation. All systems can be summarized in two general groups: valuable systems, where the tax base is the value of the property obtained by the mass valuation, and systems for surface based on the surface of taxable property [1].

In almost all countries with modern tax systems, the basis for real estate tax is the property value. This is defined in very different ways in relation to particular social and historical circumstances [2,3]. According to Jasińska and Preweda, in the EU member states the most popular is cadastral system, which is based on the capital value of the property or rental value [1].

According to Grover et al., the benefits of property taxation are also recognized in those countries which do not have a comprehensive list of taxable properties and adequate

data on transaction prices [4]. In many countries with a valuable system, calculating the amount of tax is determined by the cadastral value of a property. The issues of determining the cadastral value of a real estate, which has local names in different countries (market, tax, regulatory and other), are being dealt with by researchers around the world, as disclosed in detail by Wang [5].

Russia also belongs to the countries with a valuable system. The modern approach to determining the amount of real estate tax is based on the cadastral value, which is determined mainly by mass valuation methods. The cadastral value is the most probable price of a real estate object at which it can be acquired, based on the possibility of continuing the actual type of its use. The main properties of mass appraisal and cadastral value in Russia are shown in Table 1.

Table 1. Main properties of mass appraisal and cadastral value in Russia.

Properties	Substance
Appraisal objects	Land plots, buildings, structures, construction in progress, premises, parking spaces, single immovable complexes—all objects classified as immovable property
Type of appraisal	Mass appraisal (mainly)
Approaches	The cost approach, the sales comparison approach, and the income approach
Source of market data for mass valuation	(1) Databases of organizations providing services in support of transactions with real estate, (2) information about the real estate objects sold at auctions, (3) periodicals, (4) sites with advertisements for the sale of real estate, (5) state information system “Monitoring of the Real Estate Market”.
Pricing factors	There is a list of pricing factors in [6]; it contains most common factors for different types of real estate. E.g., for land plots, pricing factors are size, allowed land use, encumbrances (restrictions), etc. Assessors should review the entire list. During market analyses in case of refusal to consider some factor or addition of the list, such actions must be justified.
Source of pricing factors	(1) Federal Service for Registration, Cadastre, and Cartography, (2) local administrations, (3) digital thematic maps, (4) archives of technical inventory organizations, (5) other cadastres, registers, information systems.

Approaches and methods are selected depending on the objects of assessment and available data. For example, the income approach is recommended when there is reliable data on income and expenses for real estate properties, on the total capitalization rate and/or discount rate, while the comparative approach is generally preferred over other approaches to assessing. The main and preferred method for determining the cadastral value is the method of statistical (regression) modeling. It is used when assessors have a sufficient amount of information, and in case of its deficiency, they are forced to resort to the method of the reference object, the method of specific indicators of the cadastral value, etc.

Scientists and cadastral assessors, such as Leifer [7], Korostelev [8,9], Dubovik and Pavlova [10], Boyko [11] and others highlight a number of problems typical of the cadastral valuation in Russia, but they all agree that the main disadvantage of the Russian mass appraisal system is the insufficient quantity and low quality of basic data. As for quality of the initial data, it can be improved by in the following ways.

Firstly, using only reliable sources of information. For example, real estate listing in periodicals and on sites may contain false information (according to statistics, in St.

Petersburg, on average 30% of advertisings for the sale of real estate objects contain false information [12]). Also, according to the experts, tenders for land plots are held with a large number of violations, which affects the resulting price [13].

Secondly, obtaining the values of some pricing factors using spatial interpolation [12–14], which makes it possible to abandon the multi-format information (often conflicting) coming from various databases.

The problem of insufficient amount of initial data is more acute because:

- (1) Firstly, regression allows building statistically significant value models correlated with the market situation but requires a large amount of basic data.
- (2) Secondly, other assessment methods are either laborious (the method of the reference object) or have poor accuracy (the method of specific indicators of the cadastral value).
- (3) Thirdly, methods proposed by researchers (based on neural networks, fuzzy sets), also require large amounts of information [15–18].

There is no consensus between researchers about the required ratio of the number of dependent variables (market information) and independent variables (pricing factors) for a reliable regression model.

Kacman, Kosorukova and Rodin in their textbook [19] set the ratio of the amount of market information and pricing factors by the ratio described by Formula (1):

$$(n + m) \leq (n - m)^2 \quad (1)$$

where n is the amount of market information (number of transaction prices (offers)) and m is the number of pricing factors.

Smith and Draper say that the number of dependent variables should be 5–10 times greater than the number of factor signs [20]. Russian authors Gribovsky et al. proposed a formula to determine the minimum number of units of market information (Formula (2)) [21,22].

$$n = 2(m + 2) \quad (2)$$

Tabachnick and Fidell in their textbook [23] advise a minimum of 50 pieces of dependent variables plus eight for each independent variable.

The views on the quantity of dependent variables of some other authors can be read in the works [24–26].

The required ratio of a number of transaction prices (offers) and pricing factors proposed by the authors Draper and Smith, and Gribovsky et al. [20–23] is shown in Table 2.

Table 2. Required ratio of the number of pricing factors and units of market information.

Quantity of Pricing Factors	Draper N.R., Smith H.		Tabachnick B.G. and Fidell L.S.
	Number of Transaction Prices (Offers)		
	Fivefold	Tenfold	50 + 8 for Each PF
1	5	10	58
2	10	20	66
3	15	30	74
4	20	40	82
5	25	50	90

Source: Own study.

In addition to the table above, it is worth referring to the study “Classification of inhabited localities by the level of development of individual residential land market” [27]. This study is devoted to the classification of territories according to the level of development of the market for land plots for individual residential construction. According to the study, in 96% of Russian territories, the number of transactions varies from 0 to 10 per year.

This is especially typical of rural settlements, namely land plots for individual housing construction. Thus, we can conclude that, firstly, the methods proposed by Draper and Smith, and Tabachnick and Fidell are not applicable in mass evaluation; secondly, for the predominant part of territories, the use of regression analysis becomes impossible, that forces appraisers to use inaccurate methods, e.g., method of specific indicators of the cadastral value.

Both the use of low-quality data to build the regression and low accurate methods, lead to lower income to a budget (if a cadastral value is undervalued) and landowner dissatisfaction with the excessive tax burden, which is a consequence of the overestimated cadastral value. Every year, courts and special commissions consider a large number of disputes about the results of determining the cadastral value, most of which are devoted to land plots (Figure 1). Annually, the total value of the cadastral value (of all real estate in Russia) according to the results of the work of commissions and courts is offset by an average of 30%.

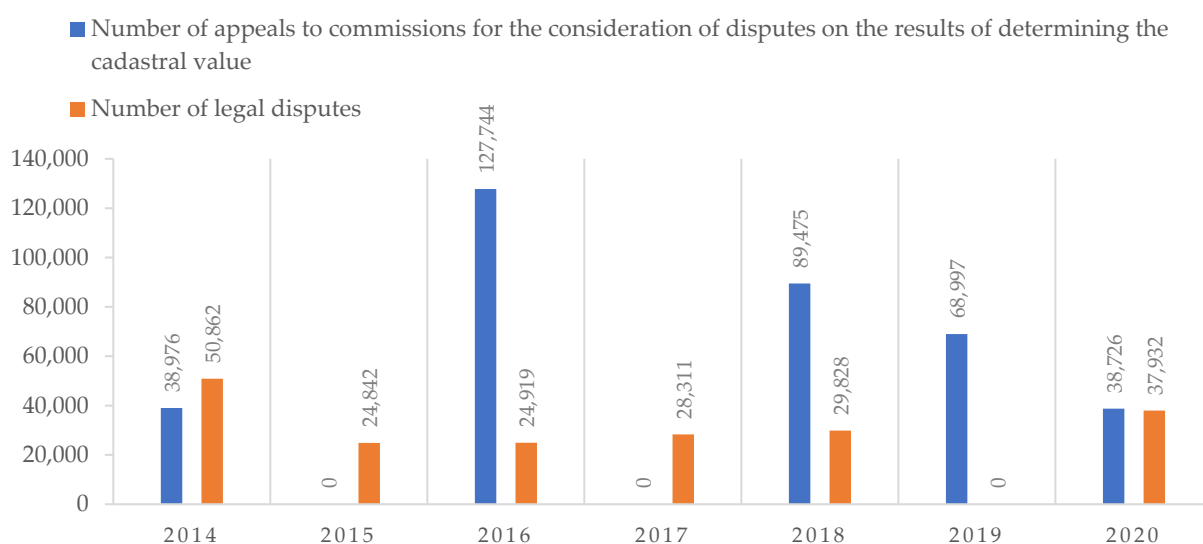


Figure 1. Statistics of appeals to commissions for the consideration of disputes on the results of determining the cadastral value and legal disputes. Source: Federal Service for Registration, Cadastre, and Cartography.

The purpose of the research is to assess the application of the territorial extrapolation of basic data method during a mass (cadastral) assessment of land or territory that is characterized by an acute lack of market information.

2. Materials and Methods

For all of the above reasons, a territorial extrapolation method (TEM) was developed. The idea of TEM is to consolidate settlements (territories) that are comparable in a composition of pricing factors and the nature of their impact on value, as well as price levels, and to build a united model of a cadastral value. A territory (grouping unit) can act as a set of nearly located land plots with one land use, similar pricing factors and the nature of their impact on value, as well as price levels. In rural settlements, grouping units can be equal to the entire territory of village. Figure 2 shows the steps of this method. The method of territorial extrapolation can be effectively implemented in the current methodology of cadastral valuation, because the 1st, 2nd and 4th stages overlap with the regulatory framework of mass appraisal in Russia, but has a number of peculiarities.

1
Analysis of the market and its segments
2
Creation of thematic maps of the subject for comparison of territories
3
Grouping of territories by spatial and price characteristics
4
Creation of a list of comparable territories of the subject
5
Creation of a cadastral value model based on the extrapolated in the space data

Figure 2. Stages of territorial extrapolation of market data method. Source: Own study.

Market data should be obtained from reliable resources as databases of organizations providing services for supporting transactions with real estate and state information system “Monitoring of the Real Estate Market” for two years. Analysis of a market and its segments allows making a preliminary conclusion about pricing factors, the nature of their impact on value and required minimum of dependent variables for regression. It is a stage when assessors should identify the need to integrate the method of territorial extrapolation.

First of all, it is necessary to determine factors that influence value of land plots. For this, the territory is analyzed for the influence of factors characterizing:

1. environment;
2. immediate surroundings of the land;
3. land plot.

For example, such information could include: the position of the administrative center, water bodies; availability of water supply, sewerage, heat supply; the possibility of flooding and soil erosion; information about the road network; position of land plots and (or) restrictions and others.

Particular attention should be paid to the nature of the influence of factors on the value. For example, in most cases, the proximity of a reservoir has a positive effect on the value, but in case of identified coastal erosion, it acts as a factor in value reduction; in the cadastral valuation of small rural settlements, as a rule, one of the pricing factors is the proximity to regional centers or large settlements.

This is followed by a grouping by price level among those territories that are comparable in a composition of pricing factors and the nature of their influence on the cost. It is advisable to use clustering methods and the method of principal components. Clustering methods are divided into hierarchical and sequential. Hierarchical methods show good results working with small samples (up to 150 objects), otherwise, it is better to use one of the sequential clustering methods [27]. K-means is one of the most popular methods of sequential clustering. Its advantages are the transparency of the algorithm, high linear speed of operation, and efficient processing of large amounts of data. In the absence of market data, the comparison by the price level is not made and the assignment to a certain group can be carried out by the expert method. This is followed by grouping territories that are comparable in price levels and pricing factors. For this, various classification methods or algorithms can be used; for example, a graph of nearest neighbors.

The next stage of TEM is devoted to collecting the actual values of pricing factors. Information about pricing factors can be obtained only from official resources that could contain up-to-date and complete data. The pricing factors that can be represented graphically are determined from digital thematic maps [6] created by assessors.

The last and the most important step is creation of a united cadastral value model based on the extrapolated in the space data. The unified cadastral value model is a regression model built using data from territories that are comparable. Various mass valuation methods as well construction of multiple regression models is widely covered in the scientific literature, therefore, we will focus only on the key points of the proposed method. All factors used to build a model are subject to the following requirements:

- Firstly, the factors must be quantifiable. The diverse nature of the pricing factors makes it necessary to take into account both quantitative and qualitative factors. In this regard, before the specification of the model, it is necessary to transform basic data. There are different methods of transformation such as binary variables (if the factor can take two values), coding by choosing a basic property, etc.
- Secondly, factors should not be correlated with each other. In the case of a strong correlation between factors, it is impossible to determine their isolated effect on cost, and the parameters of the regression equation are not interpretable. To overcome a strong inter-factor relationship the method of principal components, the transition to combined regression equations, etc. are used. These methods make it possible to express a set of pricing factors, but inevitably lead to the loss of some information; therefore, in conditions where there is a lack of information, it is more expedient to exclude one of the correlating factors.

To determine the cadastral value, linear, multiplicative and regression models are specified. The choice of the type of model can be carried out expertly or as a result of building models of all types and comparison of their quality indicators.

A quality indicators analysis of a statistical model includes a set of procedures:

- taking into account all potentially influencing pricing factors for which the objects of comparison differ and whose changes are capable of influencing the change in value;
- the validity of the signs with the coefficients of the statistical equation, that is, their compliance with the nature of the influence of pricing factors;
- correspondence of the type of influence function of each pricing factor to the nature of such influence on the real estate market;
- by Student's *t*-test (the coefficients of the model are recognized as reliable if the calculated value of the criterion exceeds the tabular value for a given level of reliability);
- by the coefficient of determination;
- by the calculated value of F—Fisher's criterion;
- by the average approximation error.

The actual values of the effective indicator differ from the theoretical values calculated using the regression equation. The smaller these differences are, the closer the theoretical values are to empirical data, and the better the quality of the model. Lewis' interpretation of average approximation error is given in Table 3.

Table 3. Interpretation of average approximation error.

Average Approximation Error	Interpretation
<10	Highly accurate forecasting
10–20	Good forecasting
20–50	Reasonable forecasting
>50	Inaccurate forecasting

Source: Own study according to Lewis [28].

In the case of united model of cadastral value, the maximum allowable value of the average approximation error should not exceed 20%. This is due to the facts that:

1. TEM is only applied in conditions of lack of market information (the larger the sample size, the more sensitive the model);
2. during grouping territories, it is not always possible to take into account some hidden factors [29];
3. according to Lewis, an approximation error of 20% allows good forecasting.

On the subject of mass valuation using the model, Trawiński et al. [30], Kokot [31], Lis [32], Doszyń [33] and other authors [34–37] wrote about research done in Poland.

Currently, these operations are performed automatically in special software products, such as Microsoft Excel, Statistica, R Studio, SPSS Statistics, etc. The use of software tools allows avoidance of technical errors and reduces the time of work.

In the case of satisfactory results of assessing the quality of the united model, they begin testing on the objects of the control sample. In the 2017 Methodological guidelines on cadastral valuation [6] and Standard on Mass Appraisal of Real Property [38], the number of objects in the control sample is not methodically fixed, therefore the method of temporary extrapolation of market data provides for such a minimum number of items in the control sample, which allows making unambiguous conclusions about the quality of the model.

To test the method of territorial extrapolation, the settlements of the Republic of Udmurtia were selected. As a result of the analysis of the market of land plots for individual housing construction (IHC), a list of pricing factors was prepared. It included: the area of the land plot, the presence of encumbrances (mortgage), the distance to the nearest reservoir, shopping center, school, kindergarten, public transport stop, conglomeration center, availability of land plot of electricity, gas supply, water supply and sewerage.

3. Results and Discussion

Due to the fact that a cadastral valuation is carried out within a constituent entity of the Russian Federation, it is advisable to consider the cadastral valuation in the context of one constituent entity of the Russian Federation. Rural settlements of the Republic of Udmurtia were chosen as a pilot object of the research. There are 184 settlements in the republic, six of which are cities (Votkinsk, Glazov, Izhevsk, Mozhga, Sarapul, Kambarka), and the rest are rural settlements. On average, the land plots of individual housing construction occupy about 40% of the territory of settlements of the subject.

During the analysis of the real estate market and collection of market data, it was revealed that:

- in 19 settlements in two years there was not a single purchase and sale transaction in relation to land plots for individual housing construction (e.g., in villages Azino and d. Ershovka);
- there are settlements where about 100 transactions take place annually (cities Izhevsk, Mozhga);
- in 20 settlements, the number of transactions did not exceed 9 units.

As a result of the market analysis, lists of pricing factors for land plots for individual development in settlements of Udmurtia were formed. This is presented in a generalized form in Table 4.

Table 4. List of pricing factors.

No.	Pricing Factor	Unit of Measurement	Data Source
1.	Space of land plot	m ²	Federal Service for Registration, Cadastre, and Cartography data base
2.	Mortgage	yes\no	
4.	Distance to a nearest water body	km	
5.	Distance to a nearest mall		
6.	Distance to a nearest school		
7.	Distance to a nearest kindergarten		
8.	Distance to a nearest public transport stop		
9.	Distance to a nearest city (for villages)		
10.	Distance to core enterprise		yes/no
11.	Availability of electricity		
12.	Gas supply availability		
13.	Availability of sewage		

Source: Own study.

To calculate the values of the factors, thematic maps of settlements were created. The example of a thematic map is shown in Figure 3.

**Figure 3.** Thematic map of Novaya Kazmaska village. Source: Own study.

Qualitative factors (the availability of sewerage, gasification and electricity) have a binary form; therefore, they were coded (1 if yes, 0 if no). Correlation analysis and analysis of statistical significance (using stepwise regression) showed that the proximity to the cities—Izhevsk, Votkinsk, Glazov, Kambarka, Mozhga, Sarapul—is one of the main pricing factors for most rural settlements, and settlements form a kind of conglomeration. It was decided to group settlements according to proximity to cities and price levels.

The process of grouping settlements according to spatial data consisted in collecting information about the coordinates of settlement centers in Udmurtia, and for grouping by price level, information on transaction prices was analyzed and the price level was calculated (Table 5).

Table 5. Basic data for grouping settlements by spatial and price characteristics (fragment).

Settlement	Coordinates of Settlement Centers		The Price Level for Land Plots of Individual Housing Construction (Rub. per m ²)
	<i>x</i>	<i>y</i>	
Azino	56.96630	52.76360	N/A
Alnashi	56.18200	52.48040	301
Bogatirevo	56.25542	51.63474	1437.5
Bezmeshur	56.45410	51.33120	N/A
Bolshaya Veniya	56.72810	53.19720	254

Note: the designation N/A is used for settlements, there is no information on the price level of sales (offers). Source: Own study.

During the grouping of settlements by spatial and price characteristics, it was revealed that: First, conglomerations, on average, include 6–7 settlements. The largest conglomerate of settlements is formed around Izhevsk, including 37 settlements, more than half of the total sample size.

Grouping according to the price level was made using k-means. Algorithm of the k-means method consists of the following steps:

1. k cluster centers are chosen randomly;
2. each sample is attributed to the closest center;
3. the calculation of the cluster centers is done as the average parameter value of all objects belonging to the certain cluster;
4. each sample is attributed to the closest cluster center.

Steps 3 and 4 are repeated until the cluster centers stop changing during the following calculations or the maximum number of operations is reached. As a result, k clusters with the minimum value of the objective function should be received.

The sum of square distances within the clusters is usually used as the objective function, being described by the Formula (3).

$$s(k) = \sum_{i=1}^n \sum_{j=0}^p (x_{ij} - x_{kj})^2 \quad (3)$$

where *k* is the cluster,

- is the value of the *j* variable for *i* review (of the sample),
- is the average value of the *j* variable in the *k* cluster
- *p* is the number of the variables [39].

The basic data for the clusterization using the k-means are the analyzed data massive, the metrics, the number of the clusters and the decomposition method.

The definition of the number of clusters has been made with the use of specialized NbClust software package in R Studio [40]. It was developed on the basis of the research by Milligan and Cooper [41] dedicated to the analysis of the stopping rules—approaches defining the best suitable number of the clusters. A total of 30 stopping rules are used in the total package, such as CH (Calinski and Harabasz) index, Duda index, Pseudo t2, C-index, Gamma-index, Beale index, Cubic clustering criterion (CCC), Point-biserial correlation coefficient, Gplus-index and others. Several features were used as definition criteria for the number of clusters, such as the dispersion matrix within the clusters, the dispersion matrix between the clusters, the total data amount in the selection, the total data amount within the cluster, the sum of the distances between the clusters and the sum of the distances within the clusters. Some of stopping rules have their own specialization, which was discussed in detail in the works [27,40]. To determine the optimal number of clusters (intervals), KL, CH, Hartigan, CCC, Scott, Marriot stopping rules were used. Among all: 1

proposed 2 as the best number of clusters, 3 proposed 3 as the best number of clusters, 2 proposed 4 as the best number of clusters. According to the majority rule, the best number of clusters is 3. As a result, the settlements were divided into three intervals (Table 6).

Table 6. Intervals of price levels for land plots for IHC.

Number of Intervals	The Minimum Value (Rub)	The Maximum Value (Rub per m ²)
Interval 1	165	577
Interval 2	635	1006
Interval 3	1161	2005
No available information	—	—

Source: Own study.

As a method of grouping, the graph of nearest neighbors was used. The essence is one of the main metric algorithms for finding similar elements of a set. The idea of its construction is in seeking the most similar object in the set X to a given query q . Under the conditions of the current task, the main settlements act as q , and the rest of the settlements act as the set X . However, it should be noted that the grouping process can be automated if all the algorithms used are formalized in the form of a program or macro.

Then, settlements with the same set of pricing factors and the nature of their influence on the value were identified using the expert method. This made it possible to form a conglomeration of settlements comparable in terms of price levels, a set of pricing factors and the nature of their impact on value. For example, one conglomeration was formed with villages located near Izhevsk—Golyany, Bolshaya Venya, Shudya, Malaya Venya, Novaya Kazmaska, Old Mikhailovskoye. The amount of market information in these settlements is shown in Table 7 (collected on all transactions for the sale and purchase of IHC land for the period from 5 January 2015 to 5 January 2017).

Table 7. The amount of market information.

Settlement	Sample Volume, Pieces
Golyany	2
Bolshaya Venya	2
Shudya	27
Malaya Venya	1
Novaya Kazmaska	1
Old Mikhailovskoye	36

It is obvious that the application of the traditional approach to cadastral valuation using regression for these settlements is impossible, therefore, TEM was applied.

On its basis, a list of market information used for modeling was compiled. As a result of the stepwise regression, a model was built (2):

$$y = 1114.53 + 34.15x_1 + 685.55x_2 + 1097.58x_3 - 57.52x_4 \quad (4)$$

where x_1 is the distance to the school in kilometers, x_2 is the availability of sewerage (yes/no), x_3 is the availability of electricity (yes/no) and x_4 is the distance to Izhevsk in kilometers.

Table 8 presents the quality characteristics of the cadastral value model for the settlements of the Izhevsk conglomeration. Table 5 checks the predictive ability of the model on the control sample.

Table 8. Quality characteristics of the cadastral value assessment model.

No.	Criterion	Value (at the Level 0.05 of Significance.)	
		Calculated Value	Table Value
1	Determination coefficient R^2	0.66	—
2	Adjusted coefficient of determination R^2_{adj}	0.63	—
3	Fisher's F-test for the coefficient of determination	25.78	4.49
4	Average approximation error A	15.00%	—
5.	Student's t -test for model coefficients		
5.1	For x_1	2.74	2.73
5.2	For x_2	3.15	
5.3	For x_3	7.66	
5.4	For x_4	3.48	

Source: Own study.

The coefficient of determination (R^2) shows the percentage of deviation of the resulting indicator due to the variation of the factor. This indicator varies in the range from 0 to 1, and its value is better closer to 1. In this model, it is 0.66, which indicates the preliminary suitability of the model. The adjusted coefficient of determination corrects the number of degrees of freedom and the equality is 0.63.

Fisher's criterion allows evaluation of the obtained value of the criterion using a table. Under these conditions, $F_{crit} < F_{calculated}$, which proves the statistical significance of the equation. The average error is in the specified range of values ($A \leq 20\%$), which indicates sufficient predictive power. According to Student's criterion, the equation coefficient is considered reliable if its calculated value exceeds the critical value. In the context of the task, this condition is met. We can conclude that the model is suitable for forecasting.

Table 9 checks the predictive ability of the model on the control sample. In 2017 Methodological guidelines on cadastral valuation there is no criteria of mass value accuracy, therefore Standard on Ratio Studies [42] was used, which sets a ratio, described by Formula (5).

$$Ratio = \frac{\text{Appraised value}}{\text{Sales value}} = 0.9 - 1.1 \quad (5)$$

Table 9. The model checking on a control sample.

Coefficients				$Y_{\text{sales value, Rub Per m}^2}$	$Y_{\text{appraised value, Rub Per m}^2}$	Ratio = $Y_{\text{realkl}} / Y_{\text{calc}}$ [32]	Difference %
x_1	x_2	x_3	x_4				
11	0	1	17	2469	2160	0.88	12
7	0	0	17	1505	1292	0.85	15

Source: Own study.

The calculated ratios are slightly outside the specified interval, which can be considered acceptable, since by rounding it turns out to be 0.9, and the resulting value does not exceed the market value, i.e., the rights of citizens are not violated.

4. Conclusions

Universal real estate taxation is related to the valuation of many real estates. This can be performed at a mass scale or individually, yet the latter method requires a huge amount of funds, is time-consuming and is hard to execute and revalue at a later time [43,44].

The key problem of the cadastral valuation of land in Russia at the present stage is the insufficient quantity and low quality of the initial data [45,46]. This leads firstly to the impossibility of taking into account all pricing factors due to the lack of market

information, and secondly to unsatisfactory results of the cadastral valuation even if the appraiser observes all procedural requirements.

The developed method has a number of advantages: it could be effectively integrated into the sequence of works during regular mass appraisal, by including additional stages of territory analysis and increasing the volume of market information; allows reduction of the time for processing information and building models; and allows the use of mass appraisal methods in cases where the 2017 Methodological guidelines [6] recommends resorting to individual appraisal methods.

As a result of the approbation of the method on the example of settlements in the Republic of Udmurtia, the values of the specific indicators of the cadastral value are 54% closer to the market value of land plots than those that were obtained during the current cadastral valuation.

Author Contributions: Conceptualization, J.V., E.B. and M.H.; methodology, J.V. and E.B.; software, M.H., K.P. and S.P.; validation, E.B. and M.H.; formal analysis, J.V., E.B.; investigation, M.H., K.P.; resources, S.P., J.V. and M.H.; data curation, J.V. and E.B.; writing—original draft preparation, J.V., E.B. and M.H.; writing—review and editing, J.V., E.B., M.H., K.P. and S.P.; visualization, J.V.; supervision, M.H.; project administration, M.H. and K.P.; funding acquisition, E.B., M.H., K.P. and S.P. All authors have read and agreed to the published version of the manuscript.

Funding: The research and the publication is financed under the Leading Research Groups support project from the subsidy increased for the period 2020–2025 in the amount of 2% of the subsidy referred to Art. 387 (3) of the Law of 20 July 2018 on Higher Education and Science, obtained in 2019.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data analyzed in this study is subject to the following licenses/restrictions: Some of datasets presented in this study are included in the article. Some of the data can be obtained by contacting the authors. Requests to access these datasets should be directed to Elena Bykova, Bykova_EN@pers.spmi.ru.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jasińska, E.; Preweda, E. Determining the cadastral-tax areas for the real estate premises based on the model of qualitative and quantitative. In Proceedings of the Environmental Engineering 10th International Conference, 27–28 April 2017. [CrossRef]
2. Manganelli, B.; Morano, P.; Rosato, P.; De Paola, P. The Effect of Taxation on Investment Demand in the Real Estate Market: The Italian Experience. *Buildings* **2020**, *10*, 115. [CrossRef]
3. Wołowicz, T.; Szybowski, D. Legal regulations of the European real estate taxation systems. *Int. J. Leg. Stud.* **2017**, *2*, 103–135. [CrossRef]
4. Grover, R.; Walacik, M.; Buzu, O.; Gunes, T.; Raskovic, M.; Yildiz, U. Barriers to the use of property taxation in municipal finance. *J. Financ. Manag. Prop. Constr.* **2019**, *24*, 166–183. [CrossRef]
5. Wang, D.; Li, V.J. Mass Appraisal Models of Real Estate in the 21st Century: A Systematic Literature Review. *Sustainability* **2019**, *11*, 7006. [CrossRef]
6. Prikaz Ministerstva Ekonomicheskogo Razvitiya Rossijskoj Federacii “Metodicheskie Ukazaniya o Gosudarstvennoj Kadastrovoj Ocenke”. 12.05.2017 No. 226. Available online: http://www.consultant.ru/document/cons_doc_LAW_217405/1cfba317e93c368b7e808fa9caa217b550814122/ (accessed on 10 May 2021).
7. Lejfer, L.A. Analiz Metodicheskogo i Programmnoho Obespecheniya Kadastrovoj Ocenki na Sootvetstvie Ocenочноj Metodologii i Sovremennym Statisticheskim Metodam Analiza Dannyh. *Imushchestvennyye Otnos. RF* **2010**, *6*, 52–64.
8. Korostelev, S.P. *Kadastrovaya Ocenka Nedvizhimosti*; Marosejka: Moscow, Russia, 2010; p. 366.
9. Korostelev, S.P. “O Edinoj Federal’noj Metodologii” Kadastrovoj Ocenki “Nedvizhimosti I Zemli”. Biblioteka Oshchenshchika. Available online: http://www.labrate.ru/articles/2017-1_korostelev.pdf (accessed on 15 May 2021).
10. Dubovik, B.I.; Pavlova, E.B. Nekotorye voprosy sovershenstvovaniya metodiki kadastrovoj ocenki zemli. *Ekon. I Upr. Nar. Hozyajstvom* **2015**, *8*, 67–72.
11. Bojko, A. YU. Kadastrovaya ocenka. Problemy i perspektivy. Available online: <http://spravks.ru/2018/05/10/kadastrovaya-ocenka-problemi-i-perspektivi/> (accessed on 10 May 2021).
12. Rodyna, S.M. Using the method of fuzzy logic in real estate appraisal (on the example of housing). Scientific and technical statements of the St. Petersburg State Polytechnic University. *Comput. Sci. Telecommun. Control* **2009**, *6*, 135–140.

13. Anisimov, A.P.; Kozlova, M.J.; Rizhenkov, A.J. Actual problems of bidding for the sale of land (lease) in the Russian Federation. *Kalmyk Univ. Bull.* **2013**, *1*, 94–100.
14. Danilov, A.; Pivovarov, I.; Krotova, S. Geostatistical analysis methods for estimation of environmental data homogeneity. *Sci. World J.* **2018**, *2018*, 7424818. [[CrossRef](#)] [[PubMed](#)]
15. Pashkevich, M.A.; Bech, J.; Matveeva, V.A.; Alekseenko, A.V. Biogeochemical assessment of soils and plants in industrial, residential and recreational areas of Saint Petersburg. *J. Min. Inst.* **2020**, *241*, 125–130. [[CrossRef](#)]
16. Rybkina, A.M.; Demidova, P.M.; Kiselev, V.A. Analysis of the application of deterministic interpolation methods for land cadastral valuation of low-rise residential development of localities. *Int. J. Appl. Eng. Res.* **2017**, *12*, 10834–10840.
17. Jia, Q.; Zhessakov, A. Study on ecological evaluation of urban land based on GIS and RS technology. *Arab. J. Geosci.* **2021**, *14*, 261. [[CrossRef](#)]
18. Vasileva, N.V.; Kadyrov, E.D. Membership function of process parameters formation based on fuzzy clustering of production data. *J. Min. Inst.* **2012**, *202*, 251–253.
19. Kacman, V.E.; Kosorukova, I.V.; Rodin, A.Y.U. *Osnovy Ocenочноj Deyatel'nosti*. M.: Moskovskaya Finansovo-Promyshlennaya Akademiya, 2010; p. 272.
20. Smith, H.; Draper, N. *Applied Regression Analysis*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2014.
21. Gribovskij, S.V.; Sivec, S.A.; Levykina, I.A. *Matematicheskie Metody Ocenki Stoimosti Imushchestva*; Marosejka: Moscow, Russia, 2014; p. 352.
22. Gribovskij, S.V.; Barinov, N.P.; Anisimova, I.N. *O Povyshenii Dostovernosti Ocenki Rynочноj Stoimosti Metodom Sravnitel'nogo Analiza*; Voprosy Ocenki; 2002. Available online: <https://www.elibrary.ru/item.asp?id=17291908> (accessed on 10 May 2021). (In Russian)
23. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*; California State University: Los Angeles, CA, USA, 2019; p. 815.
24. Cochran, W.G. *Sampling Techniques*, 3rd ed.; Wiley: Hoboken, NJ, USA, 1977; p. 448.
25. Snedecor, G.W.; Cochran, W.G. *Statistical Methods*; Wiley-Blackwell: Hoboken, NJ, USA, 1991; p. 524.
26. Altman, D.G. *Statistics in Medical Journals: Developments in the 1980s*; Wiley: Hoboken, NJ, USA, 1991. [[CrossRef](#)]
27. Bykova, E.N. *Classification of Inhabited Localities by the Level of Development of Individual Residential Land Market*; Bykova, E.N., Baltyzhakova, T.I., Volkova, Y.A., Eds.; Bulletin of the Tomsk Polytechnic University, Geo Assets Engineering: Tomsk, Russia, 2018; Volume 329, pp. 17–30.
28. Lewis, C.D. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*; Butterworth Scientific: London, UK; Boston, MA, USA, 1982.
29. Pivovarov, I.; Makhovikov, A. Statistical methods of ecological zoning. *Res. J. Appl. Sci.* **2016**, *11*, 321–326. [[CrossRef](#)]
30. Trawinski, B.; Smetek, M.; Lasota, T.; Trawinski, G. Evaluation of Fuzzy System Ensemble Approach to Predict from a Data Stream. In *Intelligent Information and Database Systems*; Springer: Bangkok, Thailand, 2014; pp. 137–146.
31. Kokot, S. Model wielu regresji pojedynczych w wycenie nieruchomości, w: *Analiza i modelowanie rynku nieruchomości na potrzeby wyceny*, S. Żróbek (red.). *Studia Mater. Tow. Nauk. Nieruchom.* **2004**, *12*, 106–122. (In Polish)
32. Lis, C. *Wykorzystanie Metod Ilościowych w Procesie Powszechnej Taksacji Nieruchomości w Polsce, w: Metody Matematyczne, Ekonometryczne i Informatyczne w Finansach i Ubezpieczeniach*, p. Chrzan; Wydawnictwo Akademii Ekonomicznej im; Oskara Langego we Wrocławiu: Wrocław, Polish, 2008. (In Polish)
33. Doszyń, M. Ekonometryczna wycena nieruchomości. *Studia I Pr. Wydziału Nauk. Ekon. I Zarządzania* **2012**, *26*, 41–52. (In Polish)
34. Sawiłow, E.; Akińcza, M. Zastosowanie teorii modelowania dla potrzeb powszechnej taksacji nieruchomości [The use of the theory of the modelling for needs of the general valuation of real estates]. *Infrastruktura i Ekologia Terenów Wiejskich* **2011**, *04*, 129–140.
35. D'Amato, M.; Siniak, N. Mass Appraisal Modelling in Minsk: Testing different Models Location sensitive. Available online: <https://doi.org/10.13128/Aestimium-13175> (accessed on 6 June 2021).
36. Borst, R. Artificial neural networks: The next modeling/calibration technology for the assessment community. *Prop. Tax J.* **1991**, *10*, 69–94.
37. Kopylova, N.S. Methods for displaying data using web technologies for the Arctic region and the continental shelf. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *913*, 042026. [[CrossRef](#)]
38. Standard on Mass Appraisal of Real Property. International Association of Assessing Officers. Available online: https://www.iaao.org/media/standards/MARP_2013.pdf (accessed on 10 May 2021).
39. Kabacoff, R. *R in Action*, 2nd ed.; Manning Publications Co.: Shelter Island, NY, USA, 2015.
40. Charrad, M.; Ghazzali, N.; Boiteau, V.; Niknafs, A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61(6). Available online: <http://www.jstatsoft.org/> (accessed on 15 May 2021).
41. Milligan, G.W.; Cooper, M.C. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **1985**, *50*, 159–179. [[CrossRef](#)]
42. Standard on Ratio Studies. International Association of Assessing Officers. Available online: https://www.iaao.org/media/standards/Standard_on_Ratio_Studies.pdf (accessed on 6 June 2021).
43. Heldak, M.; Stacherzak, A.; Baumane, V. Real estate value tax based on the Latvian experience. *Real Estate Manag. Valuat.* **2014**, *22*, 60–67. [[CrossRef](#)]
44. Heldak, M.; Baumane, V. The tax system of real property in Poland and in Latvia. *Balt. Surv.* **2014**, *1*, 109–115.

-
45. Kiselev, V.A.; Lepichina, O.Y. The analysis of sufficiency and reliability of market information in small and average settlements in the northwest district for estimation of the statistical method of ground area cadastral cost definition application possibility. *J. Min. Inst.* **2011**, *189*, 217–221.
 46. Lepihina, O.Y.U. Sovremennye problemy metodicheskogo obespecheniya kadaastrovoj ocenki zemel' naselyonnyh punktov Rossii. *Vseross. Zhurnal Nauchnyh Publ.* **2012**, *1*, 30–33.