Value-Driven Values:

Mass Cadastral Valuation Based on Spontaneous Data

Intro

In the Russian Federation, the public demand for high-quality cadastral valuation has sharply increased, due to the transition from calculating real estate tax on the basis of inventory, to the calculation on the basis of cadastral value. This being said, the calculation system that is currently used comes in for a fair share of criticism. The flaws that are brought up the most are: the systematic problems in collecting information on cadastral registration of real estate, the frequent loss of the data during the transition from one storage media to another, and the poor condition of the calculation part (in relation to methodological approaches) that is incapable of evaluating the impact on the price formation in the complex social systems of the modern city — all have led to a large gap which has emerged between the cadastral value of the property and the prevailing levels of market prices.
So-called spontaneous data generated by society as a by-product of daily life activities can be used in order to bridge that gap. Using this data allows us to take into account the non-material value of the location, which is expressed through its visual shape, flow of people passing by, prevailing stereotypes, and many other factors that play a major role when assessing the value of real estate.

Along with that, thanks to more accurate coordinate binding, it has become possible to pass on to evaluating the objects not on a quarterly, but on a case-by-case basis; and the fact that collecting spontaneous data generally does not require deploying any additional infrastructure makes it possible to reduce the number of staff required to carry out cadastral valuation of real estate. The use of spontaneous data also provides for more frequent re-valuation of the property, which really speeds up the process of responding to changes in the urban environment. From this perspective, the creation of a calculation system that includes indicators based on spontaneous data compares favorably with the calculations carried out today.

International Context

The major limitations of the current methods of real estate valuation have to do with the underutilization of non-economic factors, such as the importance of the location for the community, as well as the lack of spatial and temporal detail, particularly when monitoring the building's functional load.

Spontaneous data can be of good service when using hedonic regressions and indexes for evaluating real estate, i.e. for the assessment of some certain characteristics or just quality adjustment. Hedonic pricing is the most appropriate type of pricing in the space continuum. For instance, high crime rates in the area lead to a decrease in real estate prices, as people are paying not only for the property itself, but also for its inherent characteristics as well as for the characteristics of the surrounding space.

The most significant result of the analysis of spontaneous data is the evaluation of the perception of space shared by the community. Qualitative indicators can be used on a par with the quantitative at different scales — from a single building to a city, or to even a whole country.

According to some researchers, in the modern world the subjective assessment of well-being can say much more about the society than its economic indicators[1]. American psychologists have noticed that "economic indicators were extremely important in the early stages of economic development, when the fulfillment of basic needs was the main issue. As societies grow wealthy, however, differences in well-being are less frequently due to income, and are more frequently due to factors such as social relationships and enjoyment at work." Therefore, there have been cases where, despite a growth in income, life satisfaction declined among the population.

Another study has shown that a sentiment analysis strongly correlates with the evolution of people's satisfaction[2]. Based on this conclusion, researchers from the University of Oregon have proposed to calculate the level of "gross national happiness" based on the activity of 100 million Facebook users. Scholars from Great Britain managed to increase the scope of such developments and revealed a correlation between the tones of posts on Twitter and economic performance in different boroughs of London[3].

American economist Tran Quang suggested "using information collected from Yelp.com, to measure the effects of both quantity and quality of local consumption amenities, specifically from restaurants, on compensating differentials in the values of surrounding properties". He discovered that both the quantity and quality aspects of consumption amenities matter and that consumption amenity is capitalized in the value of nearby housing[4].

In the New England Institute, posts on Twitter were analyzed on the subject of people's emotions evoked by certain places. According to the survey, the most positive experiences are associated with parks, while the most negative ones are caused by transport interchange nodes and junctions[5]. On top of that, considerable emotional variations have been detected in health centers, cemeteries, and near waste disposal facilities. Traditionally, the proximity to such areas affects real estate value.

As a consequence, it can be assumed that the spontaneous data derived from social networks is representative of the actual processes that are taking place in the real estate market.

Spontaneous data can be used not only for analyzing the already existing areas, but also for planning future ones. For the sake of controlling and designing the urban environment, it is essential to have some knowledge of the functional use of land and the identification of key

[1] Diener, E., Seligman, M. E. P. 2004. "Beyond money: Toward an Economy of Well-Being" http://internal.psychology.illinois.edu/~ediener/Documents/Diener-Seligman_2004.pdf
[2] Kramer A. D. I. 2010. "An Unobtrusive Behavioral Model of "Gross National Happiness" http://www.uvm.edu/~pdodds/files/papers/others/2010/kramer2010a.pdf
[3] Quercia, D., Ellis, J., Capra, L., Crowcroft, J. 2012. "Tracking "Gross Community Happiness" from Tweets". https://www.cl.cam.ac.uk/research/srg/netos/papers/p965-quercia.pdf
[4] Chun Kuang. 2015. "Does Quality Matter in Local Consumption Amenities? An Empirical Investigation with Yelp" http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2606775
[5] Bertrand K.Z., Bialik M., Virdee K., Andreas Gros, Bar-Yam Y. 2013. "Sentiment in New York City: A High Resolution Spatial and Temporal View" http://necsi.edu/research/social/newyork/newyorksentiment.pdf

objects. Undertaking surveys based on direct observations and questionnaires is a method that is traditionally used for collecting this kind of information. These techniques have certain limitations caused by their high cost, as well as the reluctance of residents to provide such information.

A number of researchers suggest employing social networking that use geo-tags to determine the type of land use. Such studies were carried out using data from a social network called Foursquare, which has pre-defined categories of place (restaurant, movie theater, etc.)[6]. A similar study based on photos from Flickr defined attractions as places that have the highest concentration of photos[7]. Direct observations on the ground were conducted in order to assess the accuracy of this method. Bringing these results together confirmed the validity of using spontaneous data.

A similar study based on Twitter posts showed an equivalent list, which justifies using this type of data[8]. The authors suggest using this technique to identify new points of attraction and landmarks appearing in real time. The second part of this study defines the type of usage of a particular territory. Based on the information of the location and time of the posts, 4 types of land usage have been identified: residential, commercial, recreational (parks, public gardens etc), and those having to do with nightlife. The first 3 clusters coincided/tallied with the official zoning on 62, 77 and 79%, respectively. The fourth type, "night life", is not included in the city regulations, although nightlife does have a considerable effect on the noise performance, safety, and cleanliness of the streets. Thus, the usage of spontaneous data can serve as a very effective addition to the information on urban zoning.

Modeling the Cadastral Value of Commercial Real Estate in Russian Cities

In the first place, the spectrum of spontaneous data that can be used for mass valuation of this type of real estate was determined within the framework of the project.

Among others, we have selected the data derived from: 1) Google Routes, that allowed to estimate the parameters of social mobility indirectly, according to the congestion index 2) the social network Vkontakte, which acted as a proxy for the volume and dynamics of a certain part of people's activity, defining the general attitude towards the sites through the key statements in the posts, as well as a part of the overall characteristics of the online real estate offers, and the demand for a property through the posts with leasing deals.

Data from Google Places and Open Street Map was used to obtain information on the number of service facilities in particular areas, the distances between the objects and transport networks connecting them, work schedules, and the level of prices as well as visitor reviews.

---

[6] Noulas A., Scellato S., Mascolo C., Pontil M. 2011. "Exploiting semantic annotations for clustering geographic areas and users in location-based social networks" https://www.cl.cam.ac.uk/~cm542/papers/SMW11.pdf

[7] Crandall D. J., Backstrom L., Huttenlocher D., Kleinberg J. 2009. "Mapping the world's photos" https://www.cs.cornell.edu/home/kleinber/www09-photos.pdf

[8] Frias-Martinez V., Soto V., Hohwald H., Frias-Martinez E. 2012. "Characterizing Urban Landscapes using Geolocated Tweets" https://pdfs.semanticscholar.org/5062/feceb81c8f8e250bfbe67224dde2e434bfd1.pdf

AIRSTATE's databases (air pollution index, that includes the concentration of SO2, CO, dust etc) and thermal satellite shots made by USGS EROS made it possible to provide an overview of weather conditions and environmental pollution.

The data collected by real estate agencies was also put into use when making calculations related to the cadastral value of the property.

It is important to formulate a hedonic regression model using independent variables and to avoid spurious regression which can reduce the quality of the analysis.

Following that, the value model of real estate has been designed, in which in addition to traditional indicators, indicators calculated on the basis of spontaneous data were used as dependent variables.

The accuracy of the model was evaluated by comparing the results of the calculations with real market prices. The model has been tested in five Russian cities (Nizhny Novgorod, Ufa, Khabarovsk, Kazan, Yaroslavl), and as the testing has proved that the performance of the indicators varies from city to city, the final models had a different set of standards.

The resulting calculations have demonstrated that the information obtained with the use of Internet resources, such as social networks and navigation applications, performs relatively better than the traditional statistics.

Thus, the calculations showed that the cadastral value of non-residential real estate is underestimated throughout the city of Khabarovsk. The results of the simulation that involved spontaneous data carried out a non-residential real estate evaluation that turned out to be very similar to results of the traditional cadastral valuation.
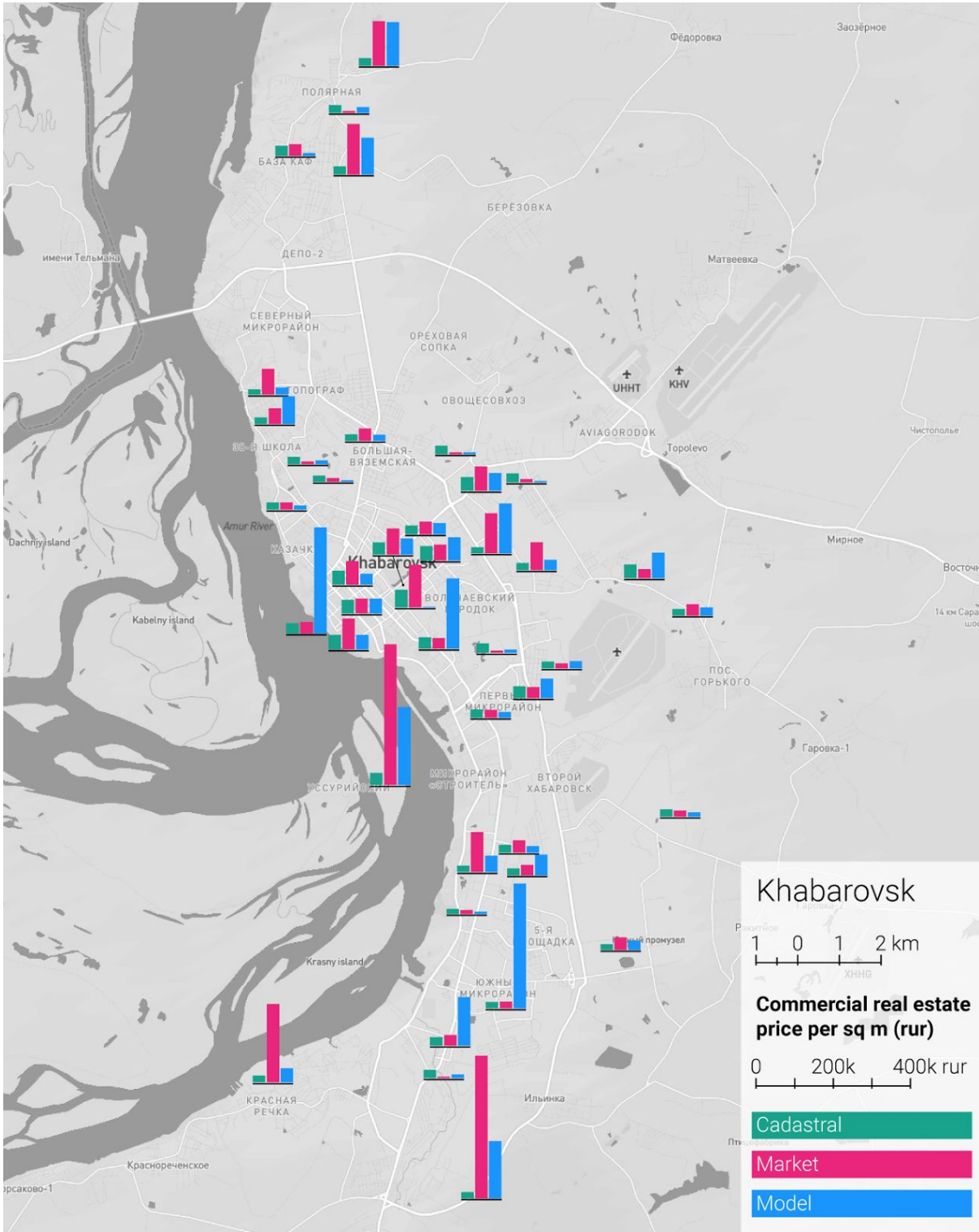
Figure 1. Commercial real estate price per sq m (rub.), Khabarovsk

We discovered that commercial real estate is dramatically undervalued in Nizhny Novgorod. In the historical center it significantly exceeded the market. And just like in all central locations, there are diverse services and art and cultural institutions attracting people and influencing the perception of centrality. Therefore both local authorities and property owners may take advantage of this underestimated potential.
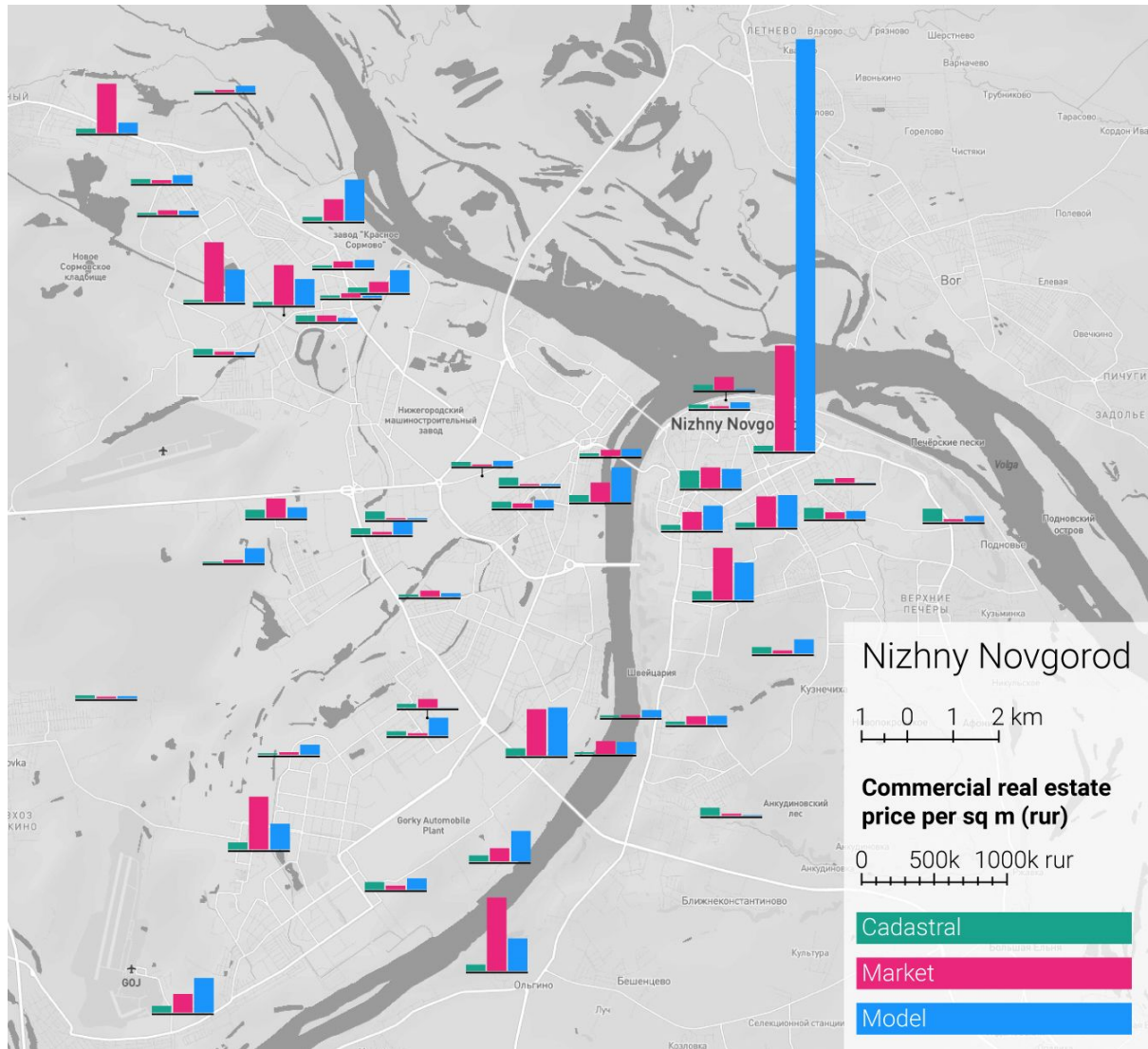
Figure 2. Commercial real estate price per sq m (rub.), Nizhny Novgorod

Current cadastral valuation of real estate in Yaroslavl is also seriously undervalued comparing to market prices. In the case of Yaroslavl, the most significant variables were those that marked the city's habitable periphery (which negatively affect the value of commercial real estate). In addition, there are three anomalous points in Yaroslavl that deserve some closer attention. In these points both the model calculation and the cadastral value differed considerably from the cost of the transaction. It can be assumed that these transactions, for whatever reason, have been concluded on a non-market basis.
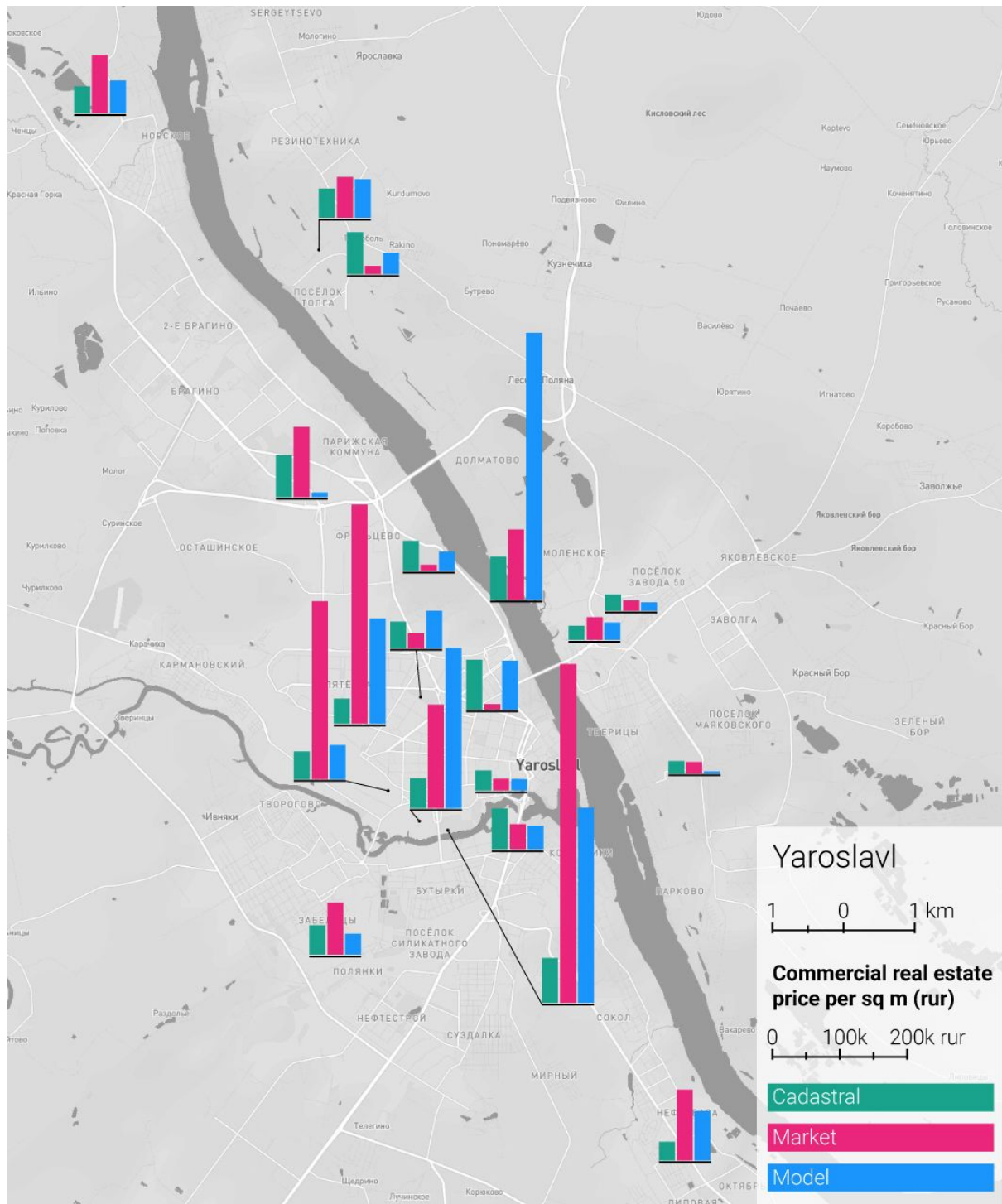
Figure 3. Commercial real estate price per sq m (rub.), Yaroslavl

In Ufa, the cadastral value of non-residential real estate is too high throughout the city, especially in the central areas. Interestingly, an unusual way of measuring the periphery – through the distance to sports facilities and the number of different functions, was discovered in Ufa.
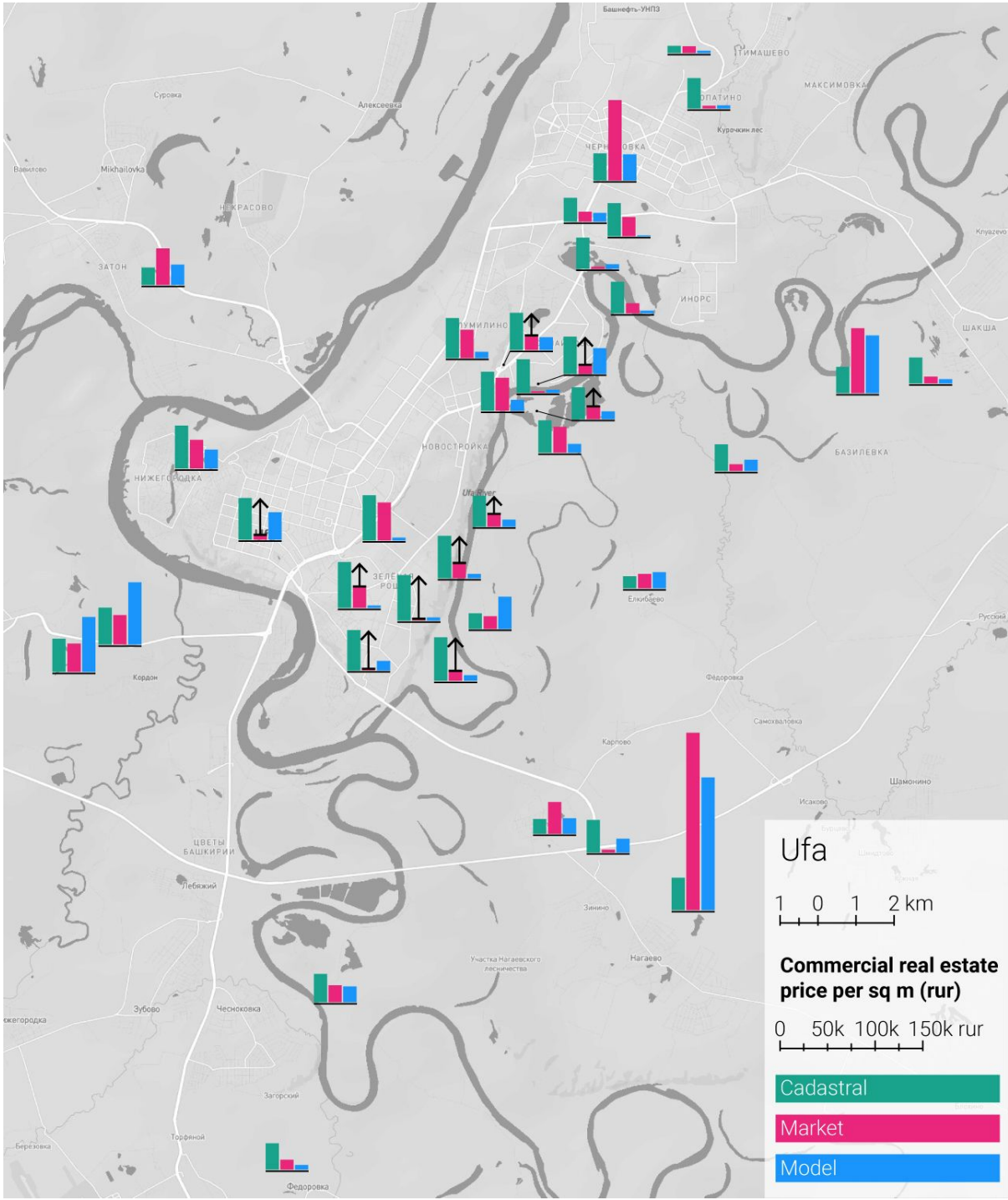
Figure 4. Commercial real estate price per sq m (rub.), Ufa

In Kazan, where the cadastral value of non-residential real estate is underestimated throughout the city, especially in the central areas, the model performed best in the Vakhitovskoe area, which is the most problematic area for traditional cadastral valuation and which has the highest concentration of trading-entertainment and office centers, that have a complex pricing structure .
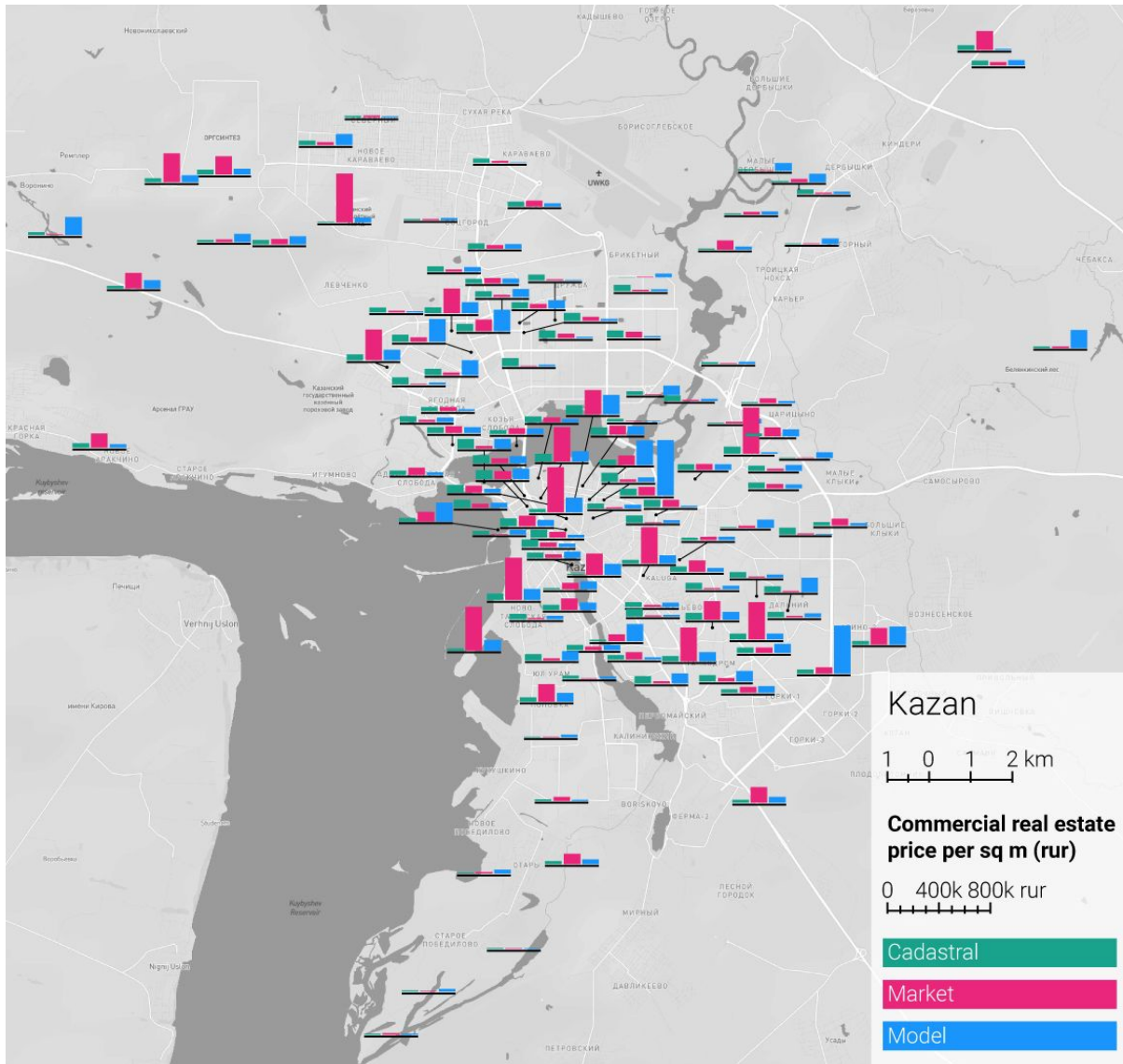
Figure 5. Commercial real estate price per sq m (rub.), Kazan

Summary

This study has confirmed that when being verified, maintained and updated by the urban community, spontaneous data is far more relevant and has a greater spatial resolution than traditional statistics. This fact led to a natural result: the accuracy of the calculations in most of the cities has increased. The estimated value has got closer to market indicators.

At the same time, indicators showing the behavior of communities, such as the level of activity on social networks (reflecting the locations preferred by the population), the temperature level (indicating the level of activity), traffic jams, and data on transport availability have proved their relevance in the pricing of commercial real estate in the cities studied.

Cadastral price modeling is part of the "Alternative Statistics for Better City Management" prepared by Habidatum in cooperation with the Institute for Urban Economics by the request of FCPF. The project also includes research focused on public transportation accessibility

and urban agglomerations. Our analysis is tailored to help urban decision-makers measure, manage and adapt to dynamic processes in a constantly changing urban environment. For more information please ask@habidatum.com