

ESTIMATION AND MAPPING OF THE SETTLEMENT FIELD POTENTIAL BASING ON REAL TRANSPORTATION CONNECTIONS BETWEEN SETTLEMENTS

P. Kuzmin ¹, M. Karpenko ², E. Panidi ^{1,*}, A. Sebentsov ²

¹ Saint Petersburg State University, St. Petersburg, Russia – panidi@ya.ru, evgeny.panidi@spbu.ru

² Institute of Geography, Russian Academy of Sciences, Moscow, Russia

KEY WORDS: Gravity Model of Settlement, Population Density, GIS-based mapping, Automated Mapping, Road Graph.

ABSTRACT:

In the article, we describe an approach to computational estimation of the settlement field potential, and mapping its spatial distribution in the desktop geographic information system software. The settlement field potential is a quantitative geographical variable that reflects mutual impact of settlements according to the population amount in the settlements and distances between settlements (two or many). This geographical variable can be used as an extra metric in human geography alongside with population density metric. We propose a methodology for automation of the settlement field potential mapping, and propose to estimate the settlement field potential value using distances measured in real transportation network instead of traditionally used straight-line measurement of the distances. To test and verify proposed approach and methodology we produced the maps of settlement field potential for the study area that covers Russia-Kazakhstan transborder area.

1. INTRODUCTION

Geographic science in whole and human geography in particular operate wide range of computational facilities and techniques to ensure quantitative description, analysis, estimation and forecasting of discovered object of interest. Over the time, new technologies like fractal computations (Yanguang, 2009), remote sensing (Kubanev et al., 2014), machine learning (Guzman et al., 2022), and other are involved in human geography studies, while the application of automated computation, modelling and mapping (Bashirov, 2017; Sidorina et al., 2019; Vorobiev, 2019; Sumping et al., 2021) in geographic information systems appears to be classical approach applied to study social processes.

Analysis elaborated in geographic information systems (GISs) assumes generally a quantitative description of explored phenomenon and its spatial characteristics, like coordinates, areas and lengths (distances). Accounting of distance between settlements alongside with the population amount at settlements is a widely used approach in human geography applied when studying spatial distribution of population (Clark, 1951; Martori, Suriñach, 2002). As an extra to other spatial approaches like population density and more general gravity model (Sen, Smith, 1995), the distance-based settlement field potential (Sozdaev, Teslenok, 2019; Dong et al., 2022; Kuzmin et al., 2023) involves and ensures an ability to select and apply different metrics when evaluating population and its spatial distribution.

Settlement field potential (SFP) being a geographic variable reflects the impact of settlements onto each other and surrounding area. It reflects the inverse dependence between population density and distance to the city center(s), and appears to be directly proportional to population amount and inversely proportional to mutual remoteness of the settlements.

However, GIS-based implementation of SFP computation and mapping faces a number of ambiguities that lead to unavailability of common technique for SFP modelling in GISs. In our study we resolve two most significant of such ambiguities, tending to ensure SFP modelling in transborder region located in southern Russia and northern Kazakhstan across the Russia-Kazakhstan border.

The first one ambiguity is the variability of SFP essence and availability of different computational formulas in publications of different authors (Kushnyr, 2015; Sozdaev, Teslenok, 2019; Vorobiev, 2019; Mandyt, 2021; Dong et al., 2022). At least three most popular formulas can be easily found in scientific publications, while SFP values estimated using any of the found can differ by an order of magnitude. In addition, most of published approaches assume application of these formulas to SFP computation at the location of selected settlement:

$$V_i = H_i + \sum_{j=1}^n \frac{H_j}{R_{ij}} \quad (1)$$

$$V_i = H_i + \sum_{j=1}^n \frac{H_j}{2 \cdot R_{ij}} \quad (2)$$

$$V_i = \sum_{j=1}^n \frac{H_j}{R_{ij}} \quad (3)$$

where V_i = the settlement field potential value estimated for the location of some given settlement
 H_i = is a population value in the given settlement
 H_j = are population values in all other settlements
 R_{ij} = is a distance between settlements i and j , respectively

Due to this, the issue of spatial interpolation of SFP estimated at irregular nodes (settlement locations) is involved in modelling process. As the interpolation in general cannot be conducted universally, it leads to incompatibility of modelling results made by different authors upon a different time. Some authors however mention the possibility to estimate SFP at regular grid nodes when mapping it over some area (Kolosov et al., 2014) that may exclude the interpolation issue.

* Corresponding author

Another one ambiguity that is extremely meaningful for our study is a selection of distance metric. Conventional approach assumes straight-line Cartesian distances accounting when estimating SFP, while all the transportation operated along nonlinear corridors. Moreover, as the distance computations in almost all studies are realised using road maps presented in different cartographic projections, produced results appears to be distorted and (again) incompatible one to another.

Assuming the possibility to exclude map projection distortions when estimate distances in GISs as geodetic distances, we have to underline that real transportation routes can differ significantly from Cartesian distances between settlements due to different geographic barriers. In transborder regions (Golovina et al., 2015) these differences are greater due to the limitation of border crossing locations (Fig. 1).

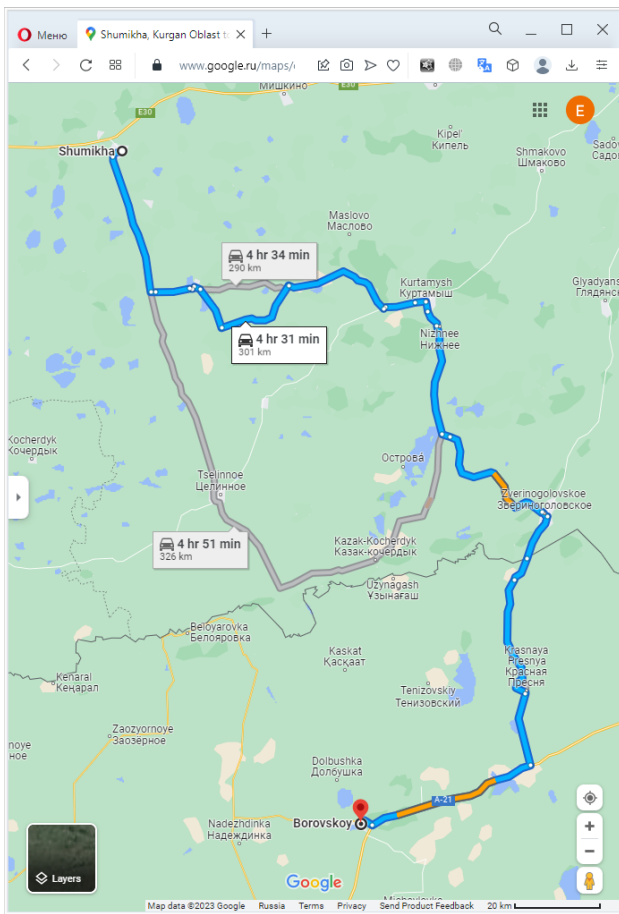


Figure 1. Route between Shumikha (Russia) and Borovskoy (Kazakhstan) through the Zverinogolovskoye border crossing point built in the road network using GoogleMaps routing engine.

2. DATA AND METHODS

The area of our study incorporates all Russia and Kazakhstan regions that assimilate Russia-Kazakhstan state border as an administrative border of a region. Study area dimensions are ~3400 kilometers from west to east and ~1600 kilometers from south to north, while state border length is ~6900 kilometers. A set of 422 settlements with highest population amounts was formed and used in the study. Population amount in the settlements was accounted according to the 2017.

Basing on analysis of preliminary modelling results conducted for the area (Kuzmin et al., 2023), we excluded the need of SFP spatial interpolation between settlement locations. Currently we implement all the computations in the regular grid nodes to exclude interpolation distortions and ambiguity of interpolation technique selection.

To ensure computations in regular grid nodes we applied formula 3 modified according to the study area dimensions. To modify the formula we establish the hypothesis of gravity buffer zone. We assumed that settlements located far then 300 kilometres from estimated location (grid node) have no social impact onto the location, while settlements located far then 50 kilometres (up to 300 kilometres) have a half-cut impact. So generally the distance weighting coefficient was added to the formula (all the designations remain equal to the formula 3):

$$V_i = \sum_{j=1}^n \begin{cases} \frac{H_j}{R_{ij}}, R_{ij} \leq 50 \\ \frac{H_j}{2 \cdot R_{ij}}, 50 < R_{ij} \leq 300 \\ 0, R_{ij} > 300 \end{cases} \quad (4)$$

To build and estimate transportation routes we used OpenStreetMap (www.openstreetmap.org) road graph. OpenStreetMap (OSM) database was downloaded from official repository (<https://planet.openstreetmap.org>). Dataset size built to cover our study area was ~4.5 GB. Being an open sources data OSM is well applicable in exploratory research, while is not free of data gaps and errors (El-Ashmawy, 2016; Quinn, Bull, 2019). As at the preliminary stage of the study (Kuzmin et al., 2023) we applied SFP computations at the settlement locations, the distance matrix was composed of 422 rows and 422 columns (according to the list of 422 accounted settlements). Consequently, 88,831 routes were built and measured according to this approach taking into account duplication of direct and inverse distances in the matrix and presence of 422 zero distanced at the matrix diagonal.

Computations at the regular grid nodes lead us to the need of 45,375,550 routes building, as the distances to all the 422 settlements have to be estimated for every grid node. This amount of computations can be classified as a massive, or at least quasimassive geospatial computations. Due to this previously used Valhalla open source routing engine (<https://github.com/valhalla/valhalla>) was replaced to Open Source Routing Machine (OSRM - <https://project-osrm.org>). Both routing engines are free and open source and demonstrate high computational efficiency, also in comparison to commercial routing engines (Saki, Hagen, 2022; Fu et al., 2023). However, at the current stage of our study, we found that massive routing computations in prebuilt OSRM container (<https://hub.docker.com/r/osrm/osrm-backend>) can be ensured with smaller amount of engine tuning in comparison to the prebuilt Valhalla container (<https://hub.docker.com/r/gisops/valhalla>).

All the data processing and computations were conducted on desktop computer with 16 GB of random access memory (RAM) and Intel Core i5-4460 processor.

3. RESULTS AND DISCUSSION

Downloaded OSM database dump composed of ~65 GB of the map data was clipped using Osmium open source program library (<https://osmiumcode.org>) according to the bounds of study area to the 4.5 GB dataset in ~1 hour. Use of clipped original full size database in this case helped to exclude the routing errors possible when composing road graph from previously clipped map

segments available for download in different formats at the side sources like Geofabrik (<https://download.geofabrik.de>) and other.

Between-settlement distance matrix was built originally using online OSRM service provided by FOSSGIS (<https://routing.openstreetmap.de>). This technique assumed execution of 88,831 routing Web queries to the service. Taking into account the needed of interrupts for Web queries execution and repeated queries in the connection lost cases, the data processing time took ~6 hours. As the data processing demanded some corrections and iterative reprocessing, this approach was discovered as not successful in the meaning of time spending and scalability.

To gain higher productivity we applied Valhalla routing engine deployed in local container alongside with OSM road graph data for the study area. Time spending in this case were composed of ~336 minutes spent for preprocessing and building of the road graph dataset (on the basis of clipped OSM dataset) and ~15 minutes spent for routes measurement and distance matrix filling. At the stage of rectangular distance matrix building for the regular-grid-based computations local Valhalla-based routing was failed however, due to the great size of processed array and consequent computational resources consumptions.

Rectangular distance matrix was built using OSRM deployed similarly in a local container. Road graph unpack was conducted in three computational flows and took ~115 minutes. Additionally, graph restructuring to ensure OSRM routing took ~216 minutes. So the whole preprocessing time spending were ~331 minutes and were equal to the ~336 minutes of Valhalla preprocessing. Rectangular matrix with 107,526 by 422 size was built in ~5 minutes using three-flows computations.

The possibility of locally deployed OSRM application to 107,526 by 422 matrix building in opposite to Valhalla inapplicability is explained by the computational logic. Valhalla container used in the study assumes a one-step distance matrix filling using function execution in Python programming language. The used OSRM container is operated through local Web request equipped with distance matrix parameters and coordinates of routes' ending nodes, while the sequential building of the routes is operated by routing engine itself.

After building of the distance matrix it is used to compute the SFP values for the settlements – in the case of square (at-the-settlement-locations) distance matrix building, or to produce an array of SFP values at the nodes of regular grid – in the case of rectangular (regular-nodes-based) distance matrix building.

In all mentioned cases routes were built and measured as shortest routes in the meaning of geometric distance. As the beginning and ending nodes of the routes generally were allocated not at the road graph segments, both nodes were mapped to the nearest road graph element, and these increments of the route length were summarized with the route length measured between beginning and ending nodes projections mapped onto road graph elements.

At the current stage of the study we have modified slightly the earlier proposed (Kuzmin et al., 2023) methodology of data processing. Currently operated methodology incorporates next steps:

1. Geocoding of population data
2. Building a distance matrix
 - Obtaining a copy of the OSM dataset for the studied area (clipping)
 - Preprocessing of a road graph to be used by selected routing engine
 - Tuning parameters (coordinate system and spatial resolution) for the regular grid that is used for computing and mapping of the SFP values
 - Reprojecting the regular grid nodes to the coordinate system used by the routing engine (generally, to the WGS 84)
3. Building a distance matrix
3. Postprocessing of the distance matrix, calculation of the SFP values
 - Distance increments accounting, distances correction
 - Filtering and transformation of the distance matrix
 - Calculation of SFP values array
4. Cartographic visualization of the SFP spatial distribution
 - Export of SFP values array to the raster (raster format is optional, GeoTIFF is used generally)
 - Integration the SFP raster layer with the basic map in GIS software, and layer styling

Applying the elaborated methodology and tools we can produce a SFP map for the study area (Fig. 4) and compare it with the SFP map produced in GIS environment using straight-line estimations for distance measurements, and formula 4 for the convolution of distance matrix to the SFP array (Fig. 2). It is clearly observed that graphical pattern of SFP produced in the case of straight-line distance estimations appears to be rough and geometric artefacts incorporating in comparison to the SFP map produced using real distance estimations made in the road graph. Also the areas of higher SFP values are observed on the map in the Figure 4 along the primary roads, while this feature reflects real phenomenon of settlements grouping along primary roads. Such an effect is almost not observed on the map in the Figure 2.

In addition, we can compare both maps with the map produced using square distance matrix (matrix of the distances between settlements only) and GIS-based spatial interpolation of the SFP estimated at the settlement locations (Fig. 3). This mapping technique have to be denoted as most rough and distorting the real character of the studied parameter (SFP). In this map we observe poor local differentiation in the SFP and false local minimums of the SFP at the locations of some settlements.

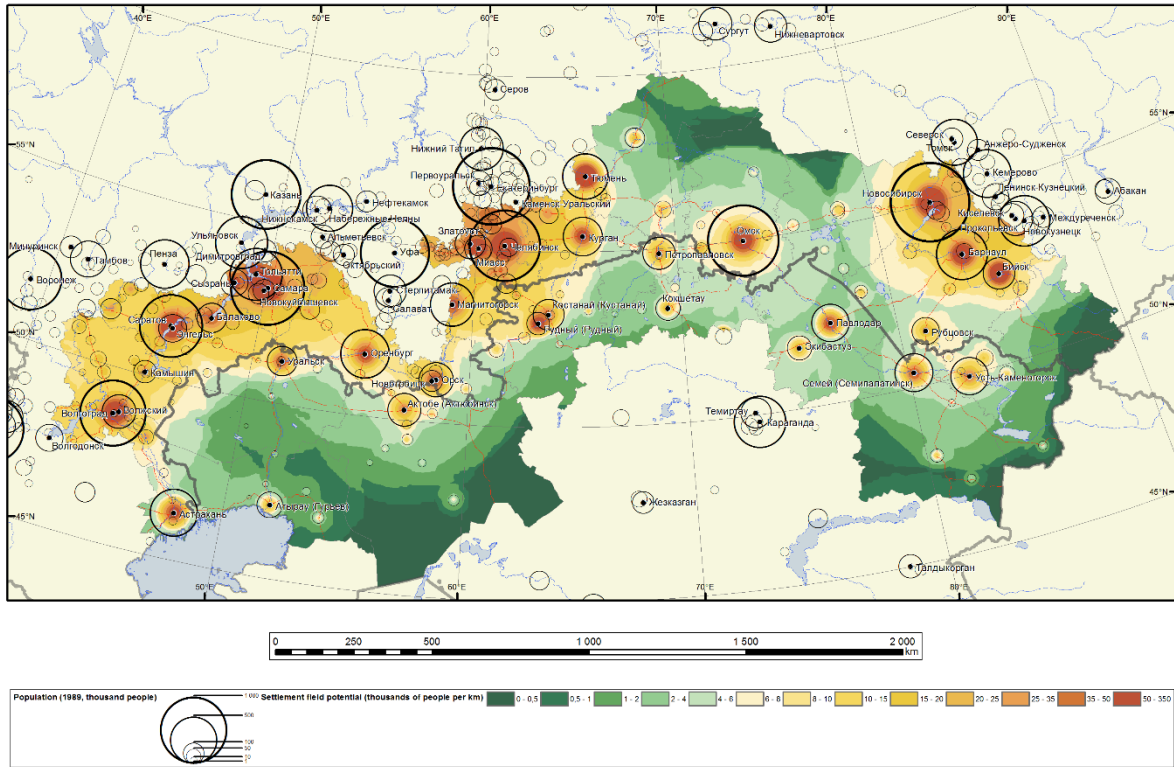


Figure 2. Map visualization of the direct-distance-based settlement field potential for the Russia-Kazakhstan transborder region. Settlement field potential estimated at the regular grid nodes in the 10 km/pix spatial resolution grid. Basic map data courtesy of OpenStreetMap.

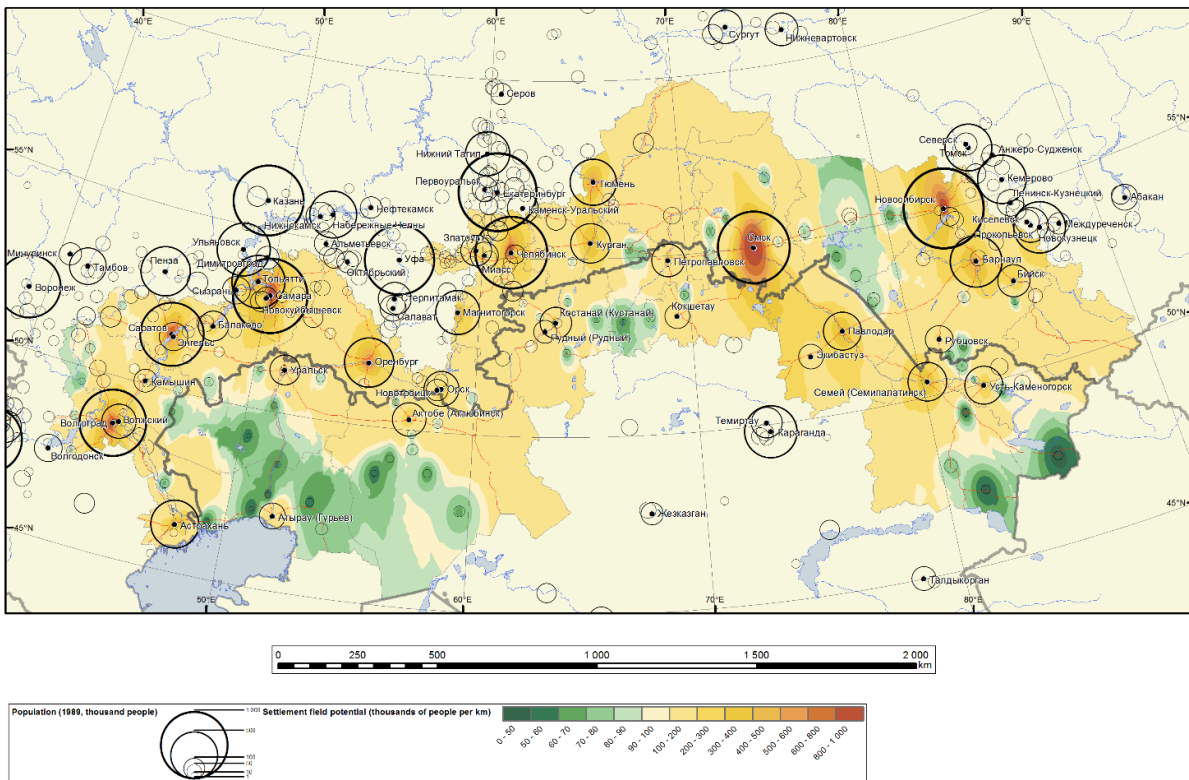


Figure 3. Map visualization of the road-network-based settlement field potential for the Russia-Kazakhstan transborder region. Road-network-based settlement field potential estimated at the settlement locations and interpolated using Inverse Distance Weighting algorithm. Basic map data courtesy of OpenStreetMap.

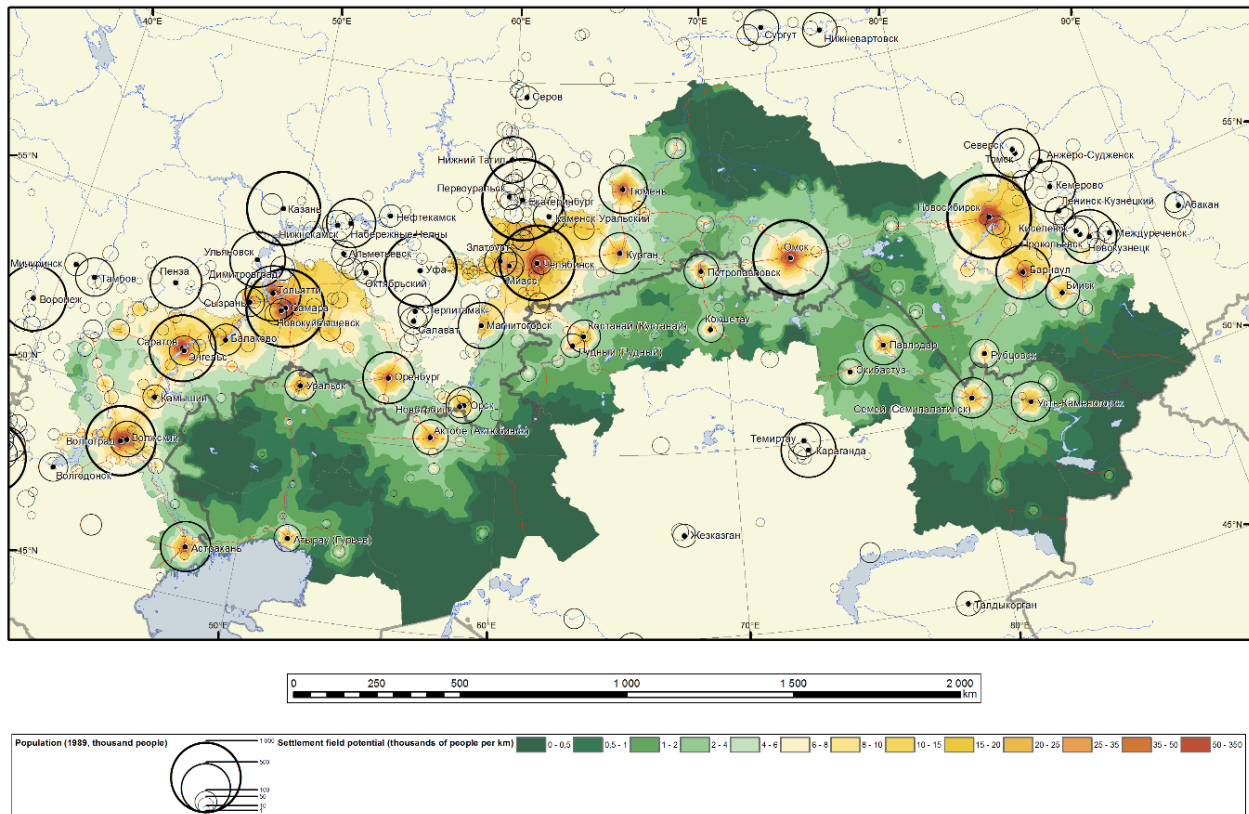


Figure 4. Map visualization of the road-network-based settlement field potential for the Russia-Kazakhstan transborder region. Settlement field potential estimated at the regular grid nodes in the 10 km/pix spatial resolution grid. Basic map data courtesy of OpenStreetMap.

4. CONCLUSIONS

In our study, we estimate SFP in the regular grid nodes. It excludes the need of spatial interpolation and enhances quality of map visualisation. Current results of our study incorporate the data conversion and processing methodology designed to estimate and map SFP for the studied areas, a set of program code algorithms that implements the methodology, and a set of maps (Fig. 2-4) produced for the Russia-Kazakhstan transborder region that illustrates performance of elaborated methodology, and makes it possible to ensure comparative analysis of the road-graph-based and straight-line-based modeling.

ACKNOWLEDGEMENTS

The study was carried out as part of the work on the RSF project No 22-27-00661 "Assessment of the potential for regionalization at the internal and external borders of the EAEU".

Basic map data courtesy of OpenStreetMap.

REFERENCES

Bashirov, V.R., 2017. Enhancement of methodology for mapping of Russian settlement system. PhD thesis, Moscow, MIIGAiK, 123. (in Russian)

Clark, C., 1951. Urban population densities. *Journal of the Royal Statistical Society*, 114, 490-496.

Dong, Y., Cheng, P., Kong, X., 2022. Spatially explicit restructuring of rural settlements: A dual-scale coupling

approach. *Journal of Rural Studies*, 94, 239-249. doi: 10.1016/j.jrurstud.2022.06.011

El-Ashmawy, K.I.A., 2016. Testing the positional accuracy of OpenStreetMap data for mapping applications. *Geodesy and Cartography*, 42(1), 25-30. doi:10.3846/20296991.2015.1160493

Fu, X., Kakkar, D., Chen, J., Moynihan, K. M., Hegland, T. A., Blossom, J., 2023. A comparative study of methods for drive time estimation on geospatial big data: a case study in USA. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-4/W7-2023, 53-60, doi:10.5194/isprs-archives-XLVIII-4-W7-2023-53-2023, 2023.

Golovina, E.D., Zotova, M.V., Sebentsov, A.B., Tikunov, V.S., 2015. Mapping of the barrier function of State borders. *Geodesy and Cartography*, 3, 29-38. (in Russian) doi:10.22389/0016-7126-2015-897-3-29-38

Guzman, L.A., Camacho, R., Herrera, A.R., Beltrán, C., 2022. Modeling population density guided by land use-cover change model: a case study of Bogotá. *Population and Environment*, 43(4), 553-575. doi:10.1007/s11111-022-00400-5

Kolosov, V.A., Rudenko, L.G., Tikunov, V.S., Gercen, A.A., Golovina, E.D., Zotova, M.V., Sebentsov, A.S., 2014. Atlas information system of Russian-Ukrainian borderland. *InterCarto, InterGIS*, 20, 24-44. (in Russian) doi:10.24057/2414-9179-2014-1-20-24-44

Kubanek, J., Nolte, E.M., Taubenböck, H., Wenzel, F., Kappas, M., 2014. Capacities of remote sensing for population estimation in urban areas. In: *Earthquake Hazard Impact and Urban Planning*, 45-66. Environmental Hazards. Springer, Dordrecht. doi:10.1007/978-94-007-7981-5_3

Kushnyr, O.V., 2015. Development of methodology for mapping of population concentration areals. PhD thesis, Moscow, MIIGAiK, 117. (in Russian)

Kuzmin, P., Karpenko, M., Panidi, E., and Sebentsov, A., 2023. Road network accounting when estimating settlement field potential. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-4/W2-2022, 65-70. doi:10.5194/isprs-archives-XLVIII-4-W2-2022-65-2023

Mandyt, M.K., 2021. Territorial features of the formation and distribution of the population in the Tyva republic. PhD thesis, Ulan-Ude, Buryat State University, 231. (in Russian)

Martori, J.C., Suriñach, J., 2002. Classical models of urban population density. The case of Barcelona Metropolitan Area. *Nederlandse Geografische Studies*, 303, 109-123.

Quinn, S., Bull, F., 2019. Understanding threats to crowdsourced geographic data quality through a study of open street map contributor bans. *Geospatial Information System Use in Public Organizations: How and Why GIS Should be Used in the Public Sector*, 80-96. doi:10.4324/9780429272851-6

Saki, S., Hagen, T., 2022. A practical guide to an open-source map-matching approach for big GPS data. *SN Computer Science*, 3, 415. doi:10.1007/s42979-022-01340-5

Sen, A., Smith, T.E., 1995. Gravity models of spatial interaction behavior. Springer. doi:10.1007/978-3-642-79880-1

Sidorina, I.E., Pozdnyakova, N.A., Panidi, E.A., Andreeva, T.A., Litvinova, M.V., 2019. Integration of traditional and modern methods in GIS-based mapping. *InterCarto. InterGIS*, 25, 35-46. (in Russian) doi:10.35595/2414-9179-2019-1-25-35-46

Sozdaev, A.A., Teslenok, K.S., 2019. Method of preparing materials for mapping settlement density. *Ogarev-online*, 3(124), 8 p. (in Russian)

Sumping, A., Rasam, A.R.A., Adnan, N.A., Md Saad, N., 2021. Mapping immigrant populations: a GIS-based public perception exploratory analysis. *IOP Conf. Ser.: Earth Environ. Sci.*, 620, 012018. doi:10.1088/1755-1315/620/1/012018

Vorobiev, A.N., 2019. GIS-based mapping of regional population dynamics. PhD thesis, Irkutsk, IG SO RAN, 165. (in Russian)

Yanguang, C., 2009. A new model of urban population density indicating latent fractal structure. *International Journal of Urban Sustainable Development*, 1(1-2), 89-110. doi:10.1080/19463131003661025