ROAD NETWORK ACCOUNTING WHEN ESTIMATING SETTLEMENT FIELD POTENTIAL

P. Kuzmin [1], M. Karpenko [2], E. Panidi [1] *, A. Sebentsov [2]

[1] Saint Petersburg State University, Department of Cartography and Geoinformatics, Saint Petersburg, Russia - panidi@ya.ru, evgeny.panidi@spbu.ru
[2] Institute of Geography, Russian Academy of Sciences, Moscow, Russia

Commission IV, WG IV/3

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

ABSTRACT:

Population density is one of valuable factors of economic development and ecological stability in inhabited areas. Classic approach to direct estimation of the population density is the application of gravity model to the settlement. The distance in the model is estimated usually as a straight-line distance between settlements. Our study is devoted to implementation of gravitational model able to account distances in transportation network. This model is demanded when investigating transborder regions, where people migrations through the state border are possible at checkpoint locations only. We use the ArcGIS software to ensure settlement field potential computations and mapping, and Valhalla open source routing engine to build routes and compute distances. Current results of our study incorporates the data conversion and processing methodology, a set of algorithms (program code) that implements these techniques, and a map series produced for the Russia-Kazakhstan transborder region that illustrates performance of elaborated methodology.

1. INTRODUCTION

Population density is one of valuable factors of economic development and ecological stability in inhabited areas. It appears to be a key point of interest when studying and forecasting humanitarian risks of any nature. Population density study practice in human geography domain assumes application of Geographic Information Systems (GISs) to ensure mapping and geospatial data analytics (Sidorina et al., 2019; Vorobiev, 2019). GIS application in this case, can be denoted as a traditional instrument. However, used modelling approach can be different. In recent studies for instance, machine learning implementation in GIS helps to automate the settlement gravity model computations and ensures indirect estimation of population density field (Guzman et al., 2022). Some authors propose different mathematical concepts to be applied to population density modelling, like fractal computations (Yanguang, 2009). A number of authors investigates the idea proposed by Colin Clark (Clark, 1951) that pose the inverse dependence between population density and distance to the city center(s) (Clark, 1951; Martori, Suriñach, 2002). Also a classic approach to direct estimation (description, modelling) of the population density is the application of gravity model (Sen, Smith, 1995) to the settlement.

The gravity model derives physics idea of interaction between spatial objects (between settlements in our case). The approach assumes interpolation of a geographical variable called settlement field potential. GISs are applied as a tool for computations and mapping automation. The term used to entitle the variable and computational formula are varying from author to author (Dong et al., 2022; Sozdaev, Teslenok, 2019), while the general idea remains common, - the variable estimates the degree of mutual impact of the objects (settlements) onto each other according to their scale and distance. The population value is used in many cases as scale parameter, while distance is estimated as a straight-line distance between settlements (at least in the studies known to the us).

Bearing in mind that we are operating in geographic coordinate space, we can conclude that the straight-line distance have no unified interpretation. Especially in wide area regions, where its estimation will wary depending on map projection. Moreover, when we are dealing with population, transportation or economical interactions between settlements, we cannot operate in Cartesian coordinate space. All the interactions will be realized through nonstraight/nonlinear connections, for instance through the road network. In such a case, closely located objects can be split by geographic barrier of any nature, and asses distance in the transportation network can be extremely longer than nominal Cartesian distance. In connection to this, our study is devoted to implementation of gravity model able to account distances in transportation network distances. In other words, we replace the direct distances in computational model to the road network distances. This modified model is demanded particularly when investigating transborder regions (Golovina et al., 2015), where people migrations through the state border are possible in checkpoint locations only.

2. DATA AND METHODS

We use the Russia-Kazakhstan transborder region as the ground test area. Common state length is ~6900 km, while the study area measurements are ~3400 by ~1600 km. Currently, our study incorporates methodology elaborating that assumes integration of road graph computations (to estimate distances), orchestration of geospatial computational tools (to automate settlement field potential computations and mapping), and spatial analytics (to assess mapping results in the meaning of geographic correctness).

At the initial study stage, we use most simple and well-known settlement field potential formula (Dong et al., 2022; Sozdaev, Teslenok, 2019):

---

* Corresponding author

$$V_i = H_i + \sum_{i=1}^{i=n} \frac{H_j}{R_{ij}}$$

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

Developed methodology assumes next steps:

1. Geocoding of population data
2. Building settlement connecting routes in the road graph with distance estimation
3. Computing settlement field potential settlement field potential values for settlement locations
4. Spatial interpolation of settlement field potential
5. Map visualization and consequent analysis of interpolated settlement field potential

We use the *ArcGIS* software to ensure settlement field potential computations and mapping (obviously the software selection in this case is based upon the researcher habit, one of many other desktop GISs can be used, including open source *QGIS*). To build routes and compute distances, we attract *Valhalla* open source routing engine (*https://github.com/valhalla/valhalla*) that uses

Open Street Map road graph in computations (*https://www.openstreetmap.org*). The road graph composed as a full graph and incorporates all levels of roads presented in the study area. However, being the open source data, it is not guaranteed to be free of data gaps (El-Ashmawy, 2016; Quinn, Bull, 2019).

Population amount data for town/city type settlements were retrieved from the open statistical sources and official publications. The data were geocoded manually, a set of 422 settlements was formed. The dataset was composed of point geometries and population amount values estimated annually.

Currently we cannot estimate the uncertainty of initial data directly (both for population data and roads data), as we have no alternative reference data sources to compare. In the future work however, we plan to produce test computations using derivative sparse datasets to estimate uncertainties statistically.

3. RESULTS

The gravity model itself is just one of possible metrics, which can be applied to settlement description. In our current study, we are not discussing its pros and contras, but focusing onto the model modifications needed to fit it into the transborder region studies. Figures 1 and 2 show map visualizations of settlement field potential mapped using road network distances computation (Fig. 1) and settlement field potential mapped using direct distances computation.
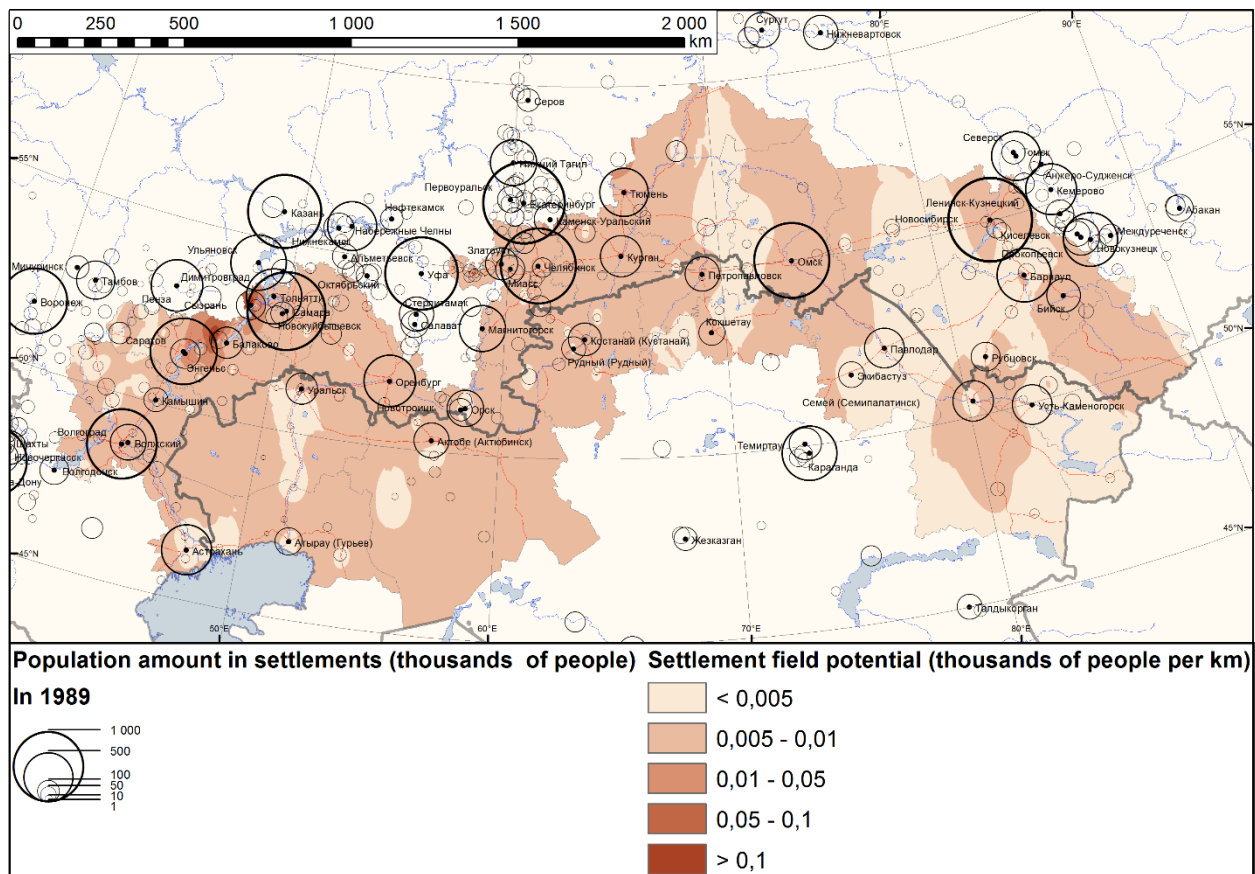
Figure 1. Draft map visualization of the road-network-based settlement field potential for the Russia-Kazakhstan transborder region. WGS-84/Alberts conic projection. Inverse distance weighting interpolation was used for spatial interpolation of settlement field potential.
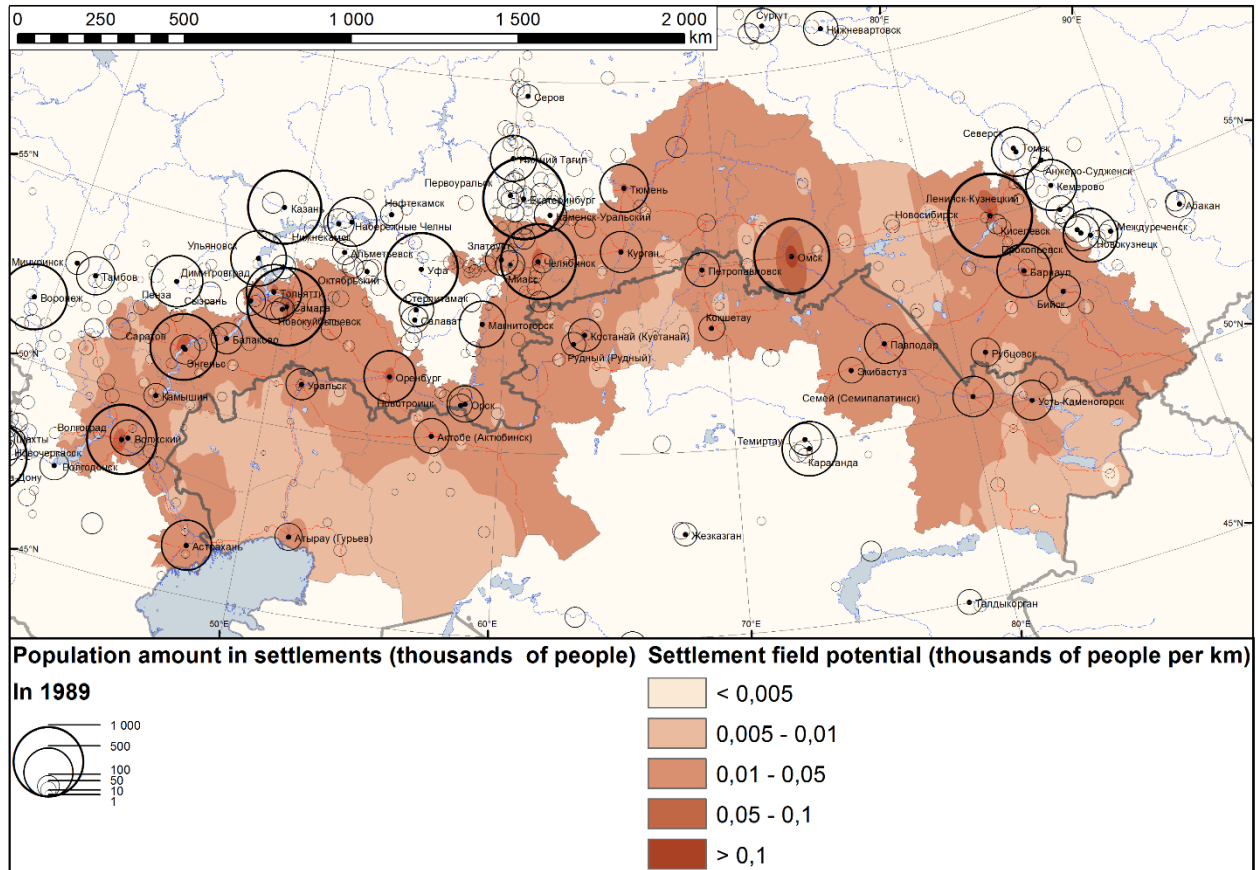
Figure 2. Draft map visualization of the direct-distance-based settlement field potential for the Russia-Kazakhstan transborder region. WGS-84/Alberts conic projection. Inverse distance weighting interpolation was used for spatial interpolation of settlement field potential.

Comparison of maps produced using original and our modified gravity models gives possibility to conclude that due to the higher mean distance value, settlement field potential values appear to be more homogeneous. Obviously, it is a result of using distance as a divisor in settlement field potential formula. In addition, the extremum settlement field potential values are allocated out of settlement locations due to that the road network coordinate space appears to be non-Euclidean.

Designed computational methodology composed of distances computation stage and settlement field potential computation/interpolation stage. *Valhalla* engine was selected as the in fact only open source tool available on the market able to conduct massive multiple routing computations without any fees and time/quantity limitations. Another one OSM-based open source routing tool was tested and shown its applicability, which is *Open Source Routing Machine* (*OSRM*) (*https://routing.openstreetmap.de*). However, being a Web service it have limitations in the meaning of Web requests number processed without blocking. Due to this feature, *OSMR* appears to be less used friendly when we have to fill the 422 by 422 distance matrix. To implement distance computations basing on initial *Microsoft Excel* table containing settlement coordinates and distance matrix filling in *Microsoft Excel* format, we applied *Python* scripting (*https://www.python.org*), and implemented an algorithm that automates formation of Web requests to *Valhalla* engine and harvesting of *Valhalla* Web responses. Here we have to point that *Valhalla* engine is operated ordinary on local operation system through the Web requests to local Web server (localhost).

After distance matrix filling, it was split by columns and joined to the settlements point dataset. To implement settlement field potential computations we applied *Field Calculator* and *Summary Statistics* tools in *ArcGIS*. The first one was applied to divide population volume values by correspondent distances, while the second was used to summarize division results corresponding to the same settlement. Additionally, the *Point Distance* toll was used to compute direct distances matrix for comparison purposes.

At the mapping stage, we applied Inverse Distance Weighting Interpolation (IDW) basing on idea that the settlement field potential is a distance-weighted parameter. Our experiments shown that Natural Neighbor interpolation produces comparable results and can be used for detailed analysis of settlement field potential distribution, while Spline interpolation produces unexpected extremums in the discovered area and is hardly applicable in this case.

In the meaning of results achieved, we can pose that our experiments demonstrate low complexity and high efficiency of automation road-distance-based settlement field potential computations. We discover current study stage as a prototyping stage.

Current results demonstrate applicability of modified gravity model for massive automated computations. Potentially, this model have to describe settlement field with higher truth, as real communication between settlements is provided in road network, but the strait distances. Model truthfulness have to be verified additionally in the future investigations using more test areas.
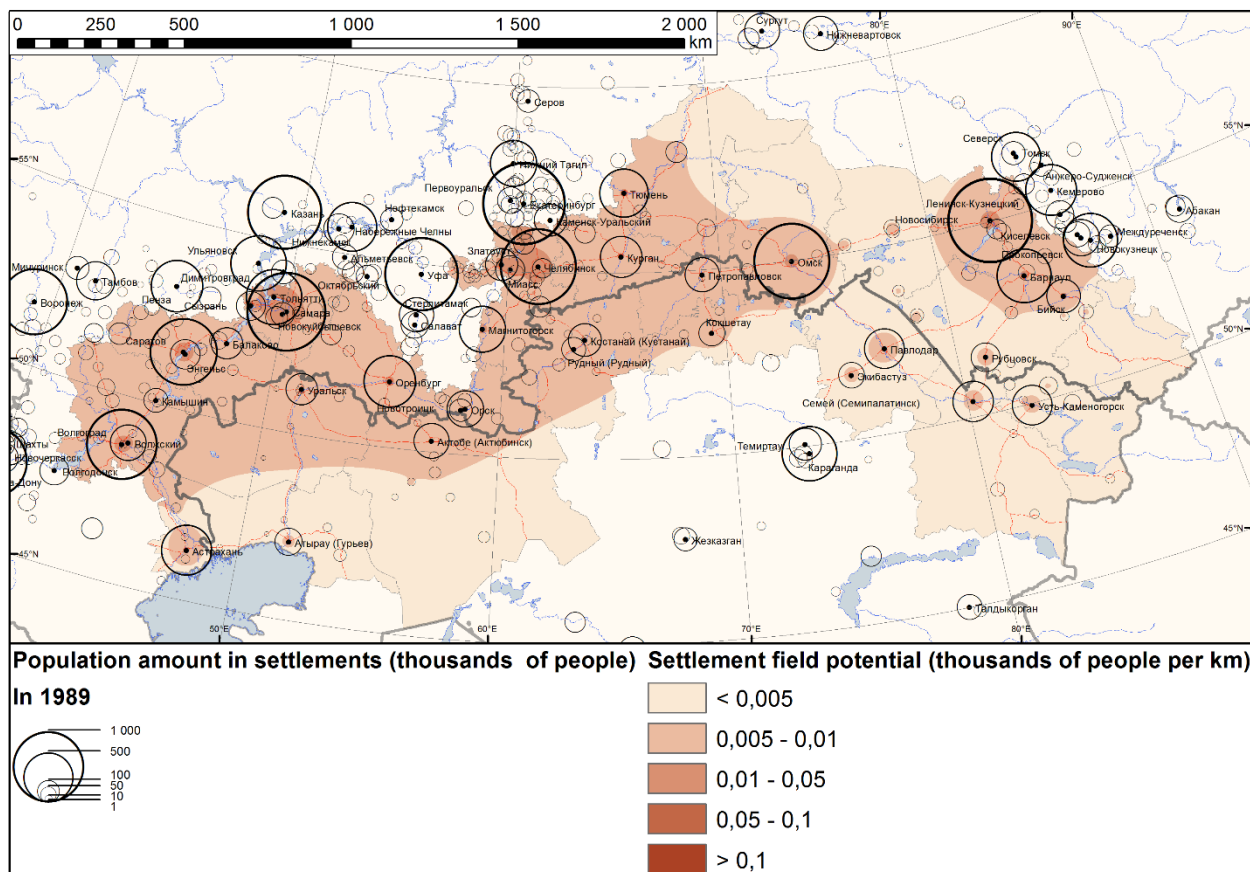
Figure 3. Draft map visualization of the direct-distance-based settlement field potential for the Russia-Kazakhstan transborder region. WGS-84/Alberts conic projection. The map was produced basing on pixel-by-pixel computations; no spatial interpolation techniques were used.

4. DISCUSSION

Gravity-model-based settlement field potential modeling and mapping is not a rare approach, it is famous and relatively popular among authors conducting studies in the human geography domain. However, most of authors are aimed onto enhancement of the approach. In some cases, the efforts are directed onto enhancement of map visualization capabilities, including three dimensional representation development (Bashirov, 2017). Another one research question attracting investigators is the transportation network accounting (Kushnyr, 2015), that is discovered in our paper.

We have not observed any known study attracted road graph routing to estimate distances when modeling settlement field potential. In this meaning, our study can be denoted as new approach study. However, we cannot pose that at the current stage the development and implementation of road-network-based settlement field potential modeling is finalized. We designed and tested methodology, and it shows applicability. Nevertheless, a set of other different questions was introduced during the study conduction.

The first and probably most significant among other is a question on representativeness of interpolated settlement field potential. As in general case when interpolation techniques are attracted, and we have to gain interpolation errors due to initial data gaps, in the case of spatial interpolation between settlements we produce errored settlement field potential values. Some authors propose to apply a regular grid computations when operating in extensive areas (Kolosov, 2014), and then to reduce interpolation

spots and interpolation errors. This approach appears to be more natural in the meaning of gravity model idea, while it assumes estimation of a potential at a point (grid node) directly instead of interpolation of the variable for lowly settled area basing on surrounding settlements having high population density.

To visualize and estimate a problem, we produced a settlement field potential map using pixel-by-pixel computations (Fig. 3). To visualize a map, we took raster grid nodes produced at initial stages, when we interpolated settlement field potential gridded maps. These nodes were used as the locations for estimation of settlement field potential related to settlement locations. In whole, we used 80,586 nodes and provided 34,007,292 direct distance computations. This technique allowed us in fact to exclude the need of interpolation, while other authors use sparse grids (Kolosov, 2014) and provide two-step computations (estimation of settlement field potential at the sparse grid nodes with consequent spatial interpolation of the estimated variable between nodes). Produced map demonstrates more smooth picture of settlement field potential (in comparison to the map in Fig. 2) that is more natural in the meaning of settlements distribution in the studied area. Similarly we can forecast that road-network-based settlement field potential map have to be more natural if/when the computations will be provided on the basis of regular grid.

To involve the regular-grid-based routing computations at the time, we have to redevelop elaborated previously computational algorithms and (probably) explore the computations' optimization issue. In the case of 422 by 422 distance matrix, we built 177,662 routes and spent 15 minutes for matrix filling by

distance values using ordinary office desktop computer. However, for the regular-grid-based computations forecasted processing time for one map is ~48 hours. Additionally, we detected a road graph discontinuity problem when using *Valhalla* to compute distance to the settlement not connected to the road network (ordinary case is the settlement location on the island in some lake or gulf). In such a case *Valhalla* routing engine producer computation error. The problem can be resolved possibly using grid node mapping to the nearest road segment. However, the algorithm of such a mapping have to be developed and tested. Moreover, the needed algorithm appears to be complex as road closest to selected location in the meaning of direct distance can be not closest in the meaning of accessibility. Gained in this case result, drives us to enhance proposed methodology and involve the regular-grid-based routing computations, and we discover this option for a future work.

5. CONCLUSIONS

Implementation of road network distance accounting in computational studies of settlement field involves another one metric that can be used commonly with other previously developed to form more complex description of settlement field in any studied region.

Posed study aim was achieved, we implemented and applied in a test mode the instruments for road network accounting when mapping settlement field potential in transborder regions. Generally, we have no known examples of road-distance-based settlement field potential mapping proposed by other Authors. Current results of our study incorporates the data conversion and processing methodology designed to estimate and map settlement field potential for the studied region, a set of algorithms (program code) that implements massive distance computations basing on OSM road graph, and a set of draft maps (Fig. 1, 2) produced for the Russia-Kazakhstan transborder region that illustrates performance of elaborated methodology.

We can conclude however, that future work have to be comprehensive, as a higher automation level is needed to ensure big area mapping. Particularly the geocoding of population amount data (that was conducted manually) can be (Obuhov, Panidi, 2021) and have to be automated as it will give a possibility to account in the model a higher number of settlements of different rang.

The analysis of road-distance-based settlement field potential maps lead to another one conclusion on settlement field potential spatial interpolation. In fact, we compute settlement field potential at the point settlement location, where the settlement population value appears to be a super scaler of settlement field potential value, while population of other settlements impact the settlement field potential dramatically lower. Gravity model on the other hand, makes it possible to exclude the need of spatial interpolation, and to estimate the settlement field potential value in any location out of settlements using all available or distance-filtered settlement population values. Consequently, it will lead to the excluding of interpolation aberrations. However, such an approach demands a redesign of routing and road distance computation algorithms we implemented previously. In this situation, almost all estimated locations are allocated out of road network, and it leads to different routing collisions. In simplest case, the routing algorithm appears unable to build the rout.

Bearing in mind abovementioned conclusions, we expect to dive deeper into the study of settlement field potential /gravity model modification with the aim of open source software package development to ensure mapping of regional settlement field potential basing on different distance metrics. Here it have to be pointed that real-world transportation networks are multimodal, while currently we discuss the car transportation capabilities.

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

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

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

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

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

Obuhov, L., Panidi, E., 2021. Toward correctness control of postal addresses geocoding. *InterCarto, InterGIS*, 27, 114-127. (in Russian) doi:10.35595/2414-9179-2021-2-27-114-127

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)

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