A spatial autocorrelation for modelling the spread of coronavirus infections

Ilya Naumov[1], *Sergey* Krasnykh[1,*], and *Yuliya* Otmakhova[2]

[1]Institute of Economics of the Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia
[2]Central Economics and Mathematics Institute of the Russian Academy of Sciences, Moscow, Russia

Abstract. Spatial autocorrelation methods are used to study spatial disproportions in the socio-economic development of territories. The most common research methods are the analysis of Moran local indices, Moran global index, Getis-Ord hot spots. In this study, we used spatial autocorrelation methods to estimate COVID-19 distribution patterns. As a result of the study, we identified the formed growth poles, the epicenters of the spread of infection (St. Petersburg, Sverdlovsk and Nizhny Novgorod regions) and only emerging ones. The practical application of this methodological approach allowed us to predict further spatial directions of the spread of coronavirus infection (Vladimir, Kaluga, Smolensk, Tula, Tver, Yaroslavl, Ryazan and Leningrad regions).

1 Introduction

Spatial autocorrelation methods are used to study spatial disproportions in the socio-economic development of territories. The most common research methods are the analysis of Moran local indices, Moran global index, Getis-Ord hot spots. In this study, we used spatial autocorrelation methods to estimate COVID-19 distribution patterns. D. Kang, H. Choi, J.-H. Kim, and J. Choi in the study «Spatial epidemic dynamics of the COVID-19 outbreak in China» analyzed the presence of spatial autocorrelation in the spread of coronavirus infection between regions using the Moran local index. The authors found that COVID-19 is more likely to spread between geographically neighboring regions because higher population densities promote more inter-territorial interaction [1].

J. R. Cavalcante and A. J. L. Abreu in the study «COVID-19 in the city of Rio de Janeiro: a spatial analysis of the first confirmed cases and deaths» analyzed the dynamics of the spread and deaths of coronavirus infection. They identified clusters of high and low incidence rates of COVID-19 by calculating the local Moran's indices using the Anselin method. As a result of the study, the authors concluded that a high risk of COVID-19 infection and death was found in areas of the southern zone of the city of Rio de Janeiro. The neighborhoods in the northern zone of the city are also characterized by a high risk of death. A team of Chinese researchers analyzed the relationship between deaths from coronavirus infection and particulate matter pollution in 49 Chinese cities. The authors used the methods of the global and local Moran's index to determine the concentration level of mortality. As a result, they

---

[*] Corresponding author: krasnykh.ss@uiec.ru

identified areas affected by clusters with a high incidence of coronavirus infection. Also using multivariate linear regression, the authors found a positive relationship between air pollution and the spread of COVID-19 [2]. H. Li at al. analyzed the spread of the new coronavirus and characterized the spatial patterns of cumulative COVID-19 cases using ArcGIS based on spatial autocorrelation and cluster analysis using the global Moran's index, local Moran's indices, and Getis-Ord Gi hotspot analysis. Using this analysis, they identified areas of high and low concentration of COVID-19. This spatial information, according to the authors, can serve to develop strategies to combat this infection [3].

The study by W. Yang, M. Deng, C. Li, and J. Huang was aimed at identifying spatio-temporal patterns of the 2019-nCoV epidemic using spatio-temporal analysis methods at the county level in Hubei province. They used Mann-Kendall and Pettitt methods to determine time trends and abrupt changes in time series of daily new confirmed cases. They used the local Moran's index to identify spatial patterns of disease, including spatial clusters and outliers [4]. A theoretical review of works on this topic showed that these methods are actively used to determine clusters of high and low values, as well as inter-territorial relationships.

2 Research methodology

In this study, we used autocorrelation analysis according to the Moran P. method. The accuracy of the results obtained using this method directly depends on the type of spatial weight matrices that are used to take into account the distances between objects. In practice, researchers usually use matrices of linear distances, distances along roads and railways, matrices reflecting the difference in cultural values, market potential of regions, and others. Spatial autocorrelation can only be properly described by one set of spatial weights. Using other spatial weighting matrices can lead to false results. Therefore, we consider it important to use different matrix of spatial weights and select for further research those that allow us to calculate statistically significant coefficients. We propose the calculation of the coefficients of the global and local Moran's indices according to the traditional method of spatial autocorrelation, using the following formulas [5]:

$$I_G = \frac{N}{\sum_i \sum_j W_{ij}} * \frac{\sum_i \sum_j W_{ij} (x_i - \mu)(x_j - \mu)}{\sum_i (x_i - \mu)^2} \tag{1}$$

$$I_{Li} = N * \frac{(x_i - \mu) * \sum_i w_{ij}(x_j - \mu)}{\sum_i (x_i - \mu)^2} \tag{2}$$

I_G – global index of autocorrelation of regions;
I_{Li} – local index of autocorrelation of regions;
N – the number of regions;
W_{ij} – element of the matrix of spatial weights for regions i and j;
μ – the average value of the indicator;
x_i – analyzed indicator of one region;
x_j – analyzed indicator of another region.

The Global Spatial Autocorrelation Index allows researchers to determine the type of spatial relationships between territories, the randomness or clustering of their distribution. Local Moran's indices are used by researchers to determine the spatial effects between study areas. To obtain more accurate results in the search for spatial clusters and spatial relationships between territories, we propose to use various matrix of spatial weights. We propose to carry out spatial clustering of territories using Patrick Moran's scatter diagram [6]. This diagram allows the researcher to correlate territories into four categories, depending on the incidence of COVID-19 and their spatial distribution. The novelty of our methodological

approach lies in the refinement of the Moran's scatter diagram. In this diagram, we highlight the territories, the value of the local autocorrelation index of which is above and below the average value. This approach makes it possible to select territories with the highest degrees of interconnection and to confirm the clustering results (Fig. 1).

This refinement of the traditional methodological approach to spatial autocorrelation analysis is due to the need to eliminate the ambiguity of the results.

For example, according to the Moran's scatter diagram, the HL category includes territorial systems with a high value of the analyzed indicator and surrounded by territories with relatively low values of this indicator, that is, territories that are «growth poles» for other territorial systems. However, this category often includes territories with weak interregional relationships with other territories that cannot be classified as «growth poles». Moreover, such regions may differ by not the highest values for the studied clustering feature.

LH		HH Spatial clusters with high values of the analyzed indicator	
High level of spatial interaction	Low level of spatial interference	High level of spatial interaction	Low level of spatial interference
The territories are affected by spatial clusters (HH) and growth poles (HL)	Territories are weakly influenced by spatial clusters (HH) and growth poles (HL)	These territories are not growth poles. These territories are influenced by growth poles (HL) and are located around them (formed clusters)	Territories are weakly affected by growth poles (HL). Territories are the periphery of the spatial cluster
LL Clustering territories with low indicator values		HL	
		High level of spatial interaction	Low level of spatial interference
Territories are not linked to other territorial systems Territories are not affected by spatial clusters (HH) and growth poles (HL)		Territories are growth poles (cores of a spatial cluster) Territories are zones with a high concentration of resources	Territories are outliers (extremes) Territories are not growth poles

Fig. 1. Patrick Moran's scatter diagram [7]

The territories belonging to the HH category have high values of the studied indicator and form a cluster. LH territories are influenced by spatial HH clusters and HL growth poles [8]. The LL territories have low values of the studied indicator and form a spatial cluster of low values among themselves [9]. To interconnected regional systems located in the HH and LH quadrants, we refer only to territories with a higher average value of the local autocorrelation index for all constituent entities of the Russian Federation [10].

This methodological approach to the study of spatial autocorrelation in the processes of the spread of COVID-19 allows not only to identify spatial effects between regions, but also to carry out spatial clustering of territories. It allows the researcher to establish growth poles in the area of COVID-19 spread, spatial clusters and zones of their influence, as well as close inter-territorial relationships in the implementation of these processes.

3 Results

Calculation of the global Moran's index using various matrices of spatial weights will allow the researcher to identify the features of spatial clustering of regions in the spread of coronavirus infection (Table 1). The positive value of the global Moran's autocorrelation index indicates the presence of spatial clustering, the unification of regions with a high incidence of coronavirus infection. This value of autocorrelation enables us to identify growth poles (the main epicenters of coronavirus infection), their zones of influence, as well as interregional relationships, that is, the direction of spread of coronavirus infection.

Table 1. The global Moran's spatial autocorrelation index and indicators of its statistical significance for various matrices of spatial weights as of 01.01.2021

	Distance matrix on roads (Normalized)	Linear Distance Matrix (Normalized)	Adjacent Boundary Matrix (Normalized)	Distance matrix on roads	Matrix on adjacent boundaries	Linear distance matrix
Global Moran's Index	-1,054	-1,085	13,3	0,003	0,064	0,003
Sd (Ii)	0,00	0,00	0,13	0,00	0,00	0,00
E(Ii)	0,000	0,000	0,002	0,000	0,000	0,000
Z-score	-238,4	-243,3	98,9	7,6	126,8	11,1
P-value	0,000	0,000	0,000	0,000	0,000	0,000

The calculation of z-scores and P-values confirmed the statistical significance and reliability of the global Moran's indices, calculated using simple and normalized matrices of linear distances, distances along roads and adjacent boundaries. The results of the study of the spatial clustering of Russian regions by the level of spread of coronavirus infection as of 01.01.2021 are presented in Figure 2. During the spatial autocorrelation analysis on January 1, 2021, we identified four main growth poles - the epicenter of the spread of COVID-19. These include the city of Moscow, St. Petersburg, Sverdlovsk and Nizhny Novgorod regions.

Only two growth poles of them have a surrounding zone of influence, active spread of infection. The zone of influence of St. Petersburg in the spread of coronavirus infection is the Leningrad Region. The zone of influence of the epicenter of the spread of coronavirus infection in Moscow is the Vladimir, Kaluga, Ryazan, Smolensk, Tver, Tula and Yaroslavl regions. Other growth poles do not have a strong impact on the surrounding regions. The incidence rate between them is significantly different. A study of spatial autocorrelation using the modified Moran's method showed that several more potential poles of growth in the incidence of coronavirus infection are currently in the formation stage.

Such poles of growth are formed on the territory of the Arkhangelsk region and the Khanty-Mansiysk Autonomous Okrug, which are located near the Sverdlovsk region. This region is the current epicenter of the incidence of COVID-19. The growth pole is also being formed on the territory of the Rostov and Voronezh regions. The increase in the incidence of coronavirus infection in these regions will contribute to the formation of a spatial cluster with a core in these regions and a zone of influence on the surrounding regions in the future.
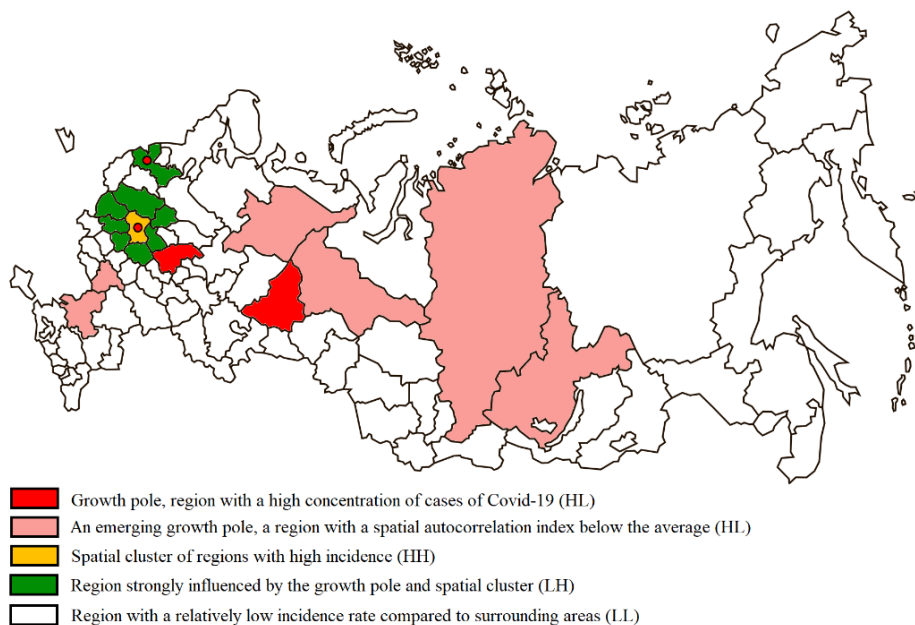
Fig. 2. Scatter diagram of Moran P. on the spread of coronavirus infection in the regions of Russia as of 01.01.2021

Active processes of the formation of a pole of growth in the incidence of coronavirus infection are currently observed in the Far Eastern Federal District, in particular, in the Krasnoyarsk Territory and the Irkutsk Region. During the period from March to December 2020, there have been major spatial changes in the dynamics of the spread of coronavirus infection. At the end of March, only one growth pole with a high incidence of COVID-19 (Moscow) was formed in Russia. Two growth poles (the city of St. Petersburg and the Sverdlovsk region) were in the stage of formation (Figure 3). By the end of 2020, these regions have become epicenters with a high incidence rate. During the noted period of time, the zone of distribution of the influence of the growth poles has also significantly expanded. The degree of their influence on the surrounding regions has increased. In early March 2020, Moscow and the Moscow region had little impact on the surrounding territories. However, by the end of 2020, these regions had a very strong influence on them. Spatial autocorrelation analysis using the Luc Anselin method confirmed the spatial directions of the spread of coronavirus infection, identified during the autocorrelation analysis. The generated matrix of local Anselin autocorrelation indices made it possible to establish direct relationships on the spread of coronavirus infection between the city of Moscow and St. Petersburg, Nizhny Novgorod region.

During the formation of this matrix, we found that an increase in the incidence of COVID-19 in one of the regions contributes to an increase in the number of cases in the associated territories. We have established inverse relationships between the city of Moscow and the Vladimir, Kaluga, Smolensk, Tula, Tver, Yaroslavl, Ryazan regions, as well as between the city of St. Petersburg and the Leningrad region.
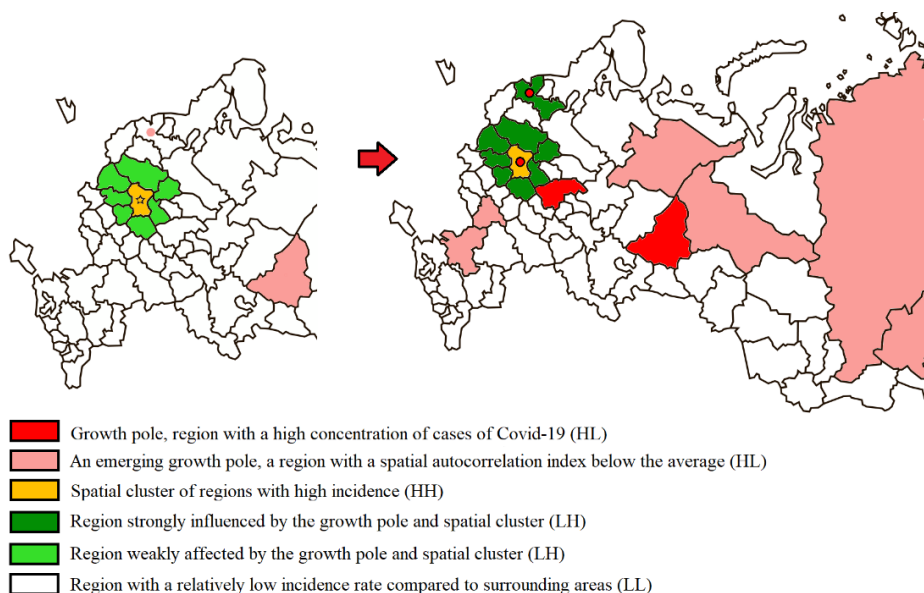
Fig. 3. Transformation of the Moran P. scatter diagram for the spread of coronavirus infection in the regions of Russia for the period from March to December 2020

The revealed close inverse relationships between the epicenters of the coronavirus infection and the surrounding regions characterize the main directions of the further spread of the infection. A decrease in the incidence of coronavirus infection in the city of Moscow and St. Petersburg contributes to an increase in the incidence rate in the regions interconnected with them.

4 Conclusion

As a result of the study, an approach was presented that makes it possible to assess the dynamics of the transformation of the spatial clustering of the processes of the spread of coronavirus infection at the macroeconomic level. This approach allows the researcher to identify the formed growth poles with a high concentration of infected people, spatial clusters of interconnected territories and zones of their influence, close interregional relationships.

The novelty of this approach is the use of various matrices of spatial weights in spatial autocorrelation analysis to obtain more substantiated conclusions, as well as the modification of the P. Moran's scattering diagram presented in the work. The allocation of territories with different strengths of spatial mutual influence in each quadrant of this matrix makes it possible to establish already formed and just emerging growth poles, spatial clusters and their periphery, zones of strong and weak influence. The use of this differentiation in the matrix of local Ancelin autocorrelation indices makes it possible to identify stable, close direct and inverse relationships between regions in the spread of coronavirus infection.

The approach presented in the work made it possible to establish a spatial cluster that unites the city of Moscow and the Moscow region. This cluster is distinguished by a high level of COVID-19 infection, a strong zone of influence and stable interregional relationships with surrounding regions.

As a result of the study, we identified the formed growth poles, the epicenters of the spread of infection (St. Petersburg, Sverdlovsk and Nizhny Novgorod regions) and only emerging ones. The practical application of this methodological approach allowed us to

predict further spatial directions of the spread of coronavirus infection (Vladimir, Kaluga, Smolensk, Tula, Tver, Yaroslavl, Ryazan and Leningrad regions).

References

1. D. Kang, H. Choi, J.-H. Kim, J. Choi, IJID **94**, 96 (2020)
2. J. R. Cavalcante, A. J. L. Abreu, Epidemiol Serv Saude **29(3)**, e2020204 (2020)
3. Y. Yao, J. Pan, W. Wang, Z. Liu, H. Kan, Y. Qiu, X. Meng, W. Wang, Sci. Total Environ, 140396 (2020)
4. H. Li, H. Li, Z. Ding, Z. Hu, F. Chen, K. Wang, Z. Peng, H. Shen. Geospat Health **15** (2020)
5. W. Yang, M. Deng, C. Li, J. Huang, Int. J. Environ. Res. Public Health **17(7)**, 2563 (2020)
6. P. Moran, JRSSB **10**, 243 (1948)
7. L. Anselin, Geogr. Anal. **2(27)**, 93 (1995)
8. I. Boal-San Miguel, LC. Herrero-Prieto, Sustainability **12(16)**, 6376 (2020)
9. S. S. Krasnykh, The Bulletin of the Academy of Knowledge **4(39)**, 228 (2020)
10. I. V. Naumov, A. Z. Barybina, R-economy **6(1)**, 14 (2020)