http://dx.doi.org/10.15576/GLL/2023.1.45

GLL

Geomatics, Landmanagement and Landscape No. 1 • 2023, 45–55

SIMPLIFIED DIFFUSION ANALYSIS – CARTOGRAPHY AS A TOOL FOR COMBATING PANDEMICS

Szczepan Budkowski

Summary

The transmission of the SARS-CoV-2 virus is a complex and intricate process, but it is possible to efficiently track and contain the spread of the pandemic in a given area by observing the regularities of the pathogen's diffusion. One of the basic measures to hamper the development of the disease was to reduce the intensity of social contact by banning free movement. An adequate response in selected regions, where the virus develops much more rapidly, is crucial and prevents serious economic damage to many industries. The modern perception of cartography as an interdisciplinary tool can contribute to limiting the diffusion of infection through spatial analyses. The aim of this paper is to present cartography as a tool to support the "management" of a pandemic. In terms of methodology, the well-known choropleth map method was employed along with spatial structure analyses. The basic category considered within the framework of statistics and econometrics is spatial relationships formulated for the purpose of achieving the set objective in the form of spatial weight matrices. In the analyses presented here, a modified Moran model was used, within which the Authors applied a row-standardised weight matrix using migration data of individual counties. The paper reviews what has been achieved so far, based mainly on European solutions. Insufficient availability of reliable data needed for advanced models (especially in the initial phase of virus spread) means that using migration data from the nearest neighbourhood can be a viable solution. This approach comes down to an analysis of migration and the population density in the county in question. A simplified analysis with a statistically significant probability allows the identification of counties that could potentially become sites of uncontrolled virus transmission in areas of high population density and high mobility. This is undoubtedly the main achievement of the publication.

The results obtained converge approximately with the actual development of a pandemic. The studies carried out indicate that the development of a pandemic is influenced not only by the number of infections, but above all by population density, as well as economic, social, educational and transport networks, as shown by the high Pearson coefficient correlation of 0.83. The analyses indicate the possibility of uncontrolled transmission of the virus in areas of high population density and high mobility.

Keywords

cartography • spatial weight matrix • SARS-CoV-2 • COVID-19 • choropleth map • pandemic combating model

1. Research issues and discussion with the literature on the subject

A milestone in the development of cartography came in the 20th century with the technological revolution. Since then cartography itself definitely ceased to be a science of how to draw maps, and until now map-making was dominated by aerial and satellite studies. The need to produce accurate cartographic studies required collecting huge amounts of spatial data provided by statistics or mathematics [Pluta 1986]. Cartography has definitely changed its nature and has become an interdisciplinary science whose task is to carry out data acquisition processes, i.e. to acquire raw data and then transform it into useful information.

Source: Wellcome Collection online archives

Fig. 1. Richard Grainger, Map of cholera in London 1849–1850

The concepts of data, information and knowledge as fundamental categories are difficult to define. In everyday language, they are often erroneously used interchangeably. Two fields are concerned with them in particular: knowledge management and information theory [Suchecka and Antczak 2010]. Taking the above into account, one can also often encounter the term DIKW, which is an acronym for: data, information, knowledge, wisdom (Fig. 2).

The spread of the SARS-CoV-2 virus in 2020 has created many challenges not only for healthcare systems worldwide, but also for scientists representing various, often divergent disciplines [Kraak and Ormelinf 2020]. Since the announcement of the pandemic by the World Health Organisation (WHO) on 11 March 2020, numerous publications have been made on the SARS-CoV-2 virus itself, as well as the COVID-19 disease it causes. In addition to papers presenting research in the broad field of medicine, those on cartography, spatial analysis, and geographic information systems (GIS) have also begun to appear with increasing frequency [Ma et al. 2020]. The ArcGIS programme and the ArcToolbox tool, which allows obtaining reports in graphic form, present great potential in this respect. ArcGIS was used in the analyses carried out for this paper.

Source: Grabowski and Zając [2009]

Fig. 2. Hierarchical organisation of DIKW

The predominant view in the studies of the latter group of researchers was that the spread of the SARS-CoV-2 virus was strongly correlated with mobility and, to some extent, also with weather conditions. The observations made by Wang [Wang and Liu 2020] had an important influence on the further development of the ongoing research. The study pointed to air travel as one of the factors indicating a strong positive correlation between the location of major airports in the United States of America and the provinces in China, where COVID-19 disease outbreaks were recorded. Human mobility data used from Google Community Mobility Reports generated from mobile phone data also proved helpful (https://www.google.com/covid19/mobility/)

The spread of COVID-19 disease caused by the SARS-CoV-2 coronavirus very quickly became a global phenomenon and triggered a pandemic, which at the same time became the catalyst for many scientists and research organisations to create thematic maps to visualise the spreading phenomenon.

Particularly important for controlling the course of the pandemic were not only *choropleth map*s [Budkowski and Litwin 2021], but also methods developed to predict the spread of the pandemic [Han et al. 2020]. Publications by scientists proved the correlation between weather conditions and the increase in COVID-19 incidence [Ma 2020, Sajadi 2020, Shi 2020]. It quickly became apparent that the role of weather alone as a factor contributing to the development of the pandemic was relatively small compared to factors such as human mobility and population density [Jamshidi 2020].

Methods that are used to monitor pandemic development require multiple inputs, such as infection location, human contact, environmental and climatic factors, and migration [Medyńska-Gulij 2011]. Examples of such advanced methods include the Spanish method [Perez-Fuentes et al. 2021] and the Czech method [Komenda et al. 2020], or the method developed by Polish scientists at a military technical academy [Araszkiewicz et al. 2021]. Each of these solutions is highly advanced and therefore generates significant problems at the development stage. In addition to the already problematic and time-consuming process of model generation, attention should be paid to the key problem of collecting data that will be characterised by high reliability. A central argument may also be that highly sophisticated models, which are characterised by a high degree of precision, require equally precise initial information in the form of reliable and accurate data for their analyses. It will be particularly difficult to obtain this kind of data in less developed countries and countries with low levels of health care. Another reason for the lack of such data may be the early phase of an epidemic, when such data are not collected, according to the author's analysis of a research paper [Chen et al. 2020]. With relatively little information available, it is therefore possible to develop a model quite quickly using the spatial structure of the area where viral infections occur [Ružičková et al. 2020].

The publication of the results of the study aims to present spatial relationships that allow an analysis of the spread of the virus using the available data.

2. Materials and methods

In the models developed so far, the research on SARS-CoV-2 was not based on theories applicable in econometrics. The study was conducted on the territory of Poland. The necessary population data were obtained from the Central Statistical Office, the geometric data were retrieved from the information provided free of charge by the Head Office of Geodesy and Cartography [GUGIK], "[…] and the data on the development of the epidemic, which included the increments of the incidence in individual counties, were taken from the website maintained by Michał Rogalski [https://docs.google.com/spreadsheets/d/1ierEhD6gcq51HAm433knjnVwey4ZE5DCnu1bW7PRG3E/edit#gid =1309014089]. The possibility of the use of data collected by an individual user is related to the fact that there is a noticeable gap in the Ministry of Health's data. Data on the pandemic was collected by the Ministry of Health from the beginning of the pandemic until the 10th of October, and then only from the 25th of November until now. This data was compared with Google's mobility data. The average Pearson correlation coefficient was 0.83, indicating that there was a strong correlation between infections and mobility. The analyses covered the period from the beginning of the pandemic to June 2022.

What is somewhat new about this approach is the use of basic information to support quick assessments of the spread of the pathogen. It is assumed that this will allow a course of action to be indicated before analyses requiring detailed data are retrieved, as these are often unavailable in the early stages of an outbreak. The shift from geographical distance to migration and the use of these data might create a new field of research, with a common ground for answering questions about the causes of a particular dependency structure. By applying this approach, the authors attempt to modify the existing Moran model [https://faculty.ksu.edu.sa/sites/default/files/moran.pdf].

In order to build such a model, spatial autocorrelation, one of the main categories considered within spatial statistics and econometrics, was adopted, which can be understood as a generalisation of the autoregressive process known from the time series analysis [Welfe 2009]. Consider the realisation of a spatial random function X at successive locations $s_i = 1, .., N$ forming a statistical sample $x(s_1), x(s_2), \ldots, x(s_n)$, where $x(s_i)$ is the measured value of x_i at sample point i. In addition, let the locations s_i belong to the domain of the function, which is the sample region D, contained in the n-dimensional Euclidean space $s \in D \subset R^n$). In the case of a two-dimensional space R^2, typical of spatial data, it can be said that the realisations of the random variable X at successive locations will be the realisations of the random field $\{x(s)\colon s \in D\}$ [Cressie 1993, p. 8–9]. For a random variable X defined in this way, the estimator of the covariance function (covariogram $C(.)$ of the random field takes the form:

$$\forall\, s_i, s_j \in D : C(s_i - s_j) = \mathrm{cov}[x(s_i), x(s_j)] = E\{[x(s_i) - \mu][x(s_j) - \mu]\} \tag{1}$$

Where $i \neq j$, $\mu = E[x(s)]$, and μ is the expectation value of the random field, constant over the whole domain D when the spatial process is second-order stationary. Meeting the condition of second-order stationarity also means that the covariance function does not depend on the position but only on the shift vector $d = ||d||=||s_i - s_j||$, $d \in R^2$ (spatial process), and the overall variance of the process is constant and realised by the formula $C(0) = \mathrm{cov}\,[x(s_i), x(s_i)] = \varphi^2$ [Suchecka and Antczak 2010].

Determining the estimator of the spatial autocorrelation function:

$$\varphi(d) = \frac{C(d)}{C(0)} \tag{2}$$

it should be noted that the level of spatial correlation of the realised function X depending on the adopted distance d indicates the crucial importance of the distance between the aggregates (the basic element of the built model). This implies the necessity of examining each pair of aggregates realising the function of the variable X (for the i-th county) and then selecting according to the criterion:

a) included in the calculation on the basis of meeting the relationship $\varphi(d)$, if $d \geq ||s_i - s_j||$,
b) excluded from the calculations $\varphi(d)$, if $d < ||s_i - s_j||$.

The formal expression of the above relationships is provided by the spatial weight matrix [Legendre 1998]. It is a commonly used method in econometry to map the structure of connections. The scheme proposed by Moran [1950] presents the connections of the realisations of the function X in the form of a first-order neighbourhood for an arbitrarily defined structure of relations that can be expressed as follows:

$$I(d) = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} \left(x_i - \mu\right)\left(x_j - \mu\right)}{\sum_j \left(w_j - \mu\right)^2} = \frac{N}{\sum_i \sum_j w_{ij}} \frac{z^T W z}{z^T z} \tag{3}$$

where: z is a column vector whose elements can be expressed as $z_i = x_i - \mu$, while w_{ij} are the elements of an N × N matrix $W = [w_{ij}]$ that can be expressed as follows:

$$W = \begin{bmatrix} W_{11} & \cdots & W_{1N} \\ \vdots & \ddots & \vdots \\ W_{N1} & \cdots & W_{NN} \end{bmatrix} \tag{4}$$

3. Results

The relatively high statistical significance of the calculations performed, which is 0.67, and the results from the comparison of the actual value with the expected value are noteworthy. The spatial weight matrix was obtained on the basis of the neighbourhood matrix. The applied method of analysing the phenomenon allowed to eliminate the negative effect of spatial autocorrelation on the estimation of the model parameters and to identify spatial effects according to [Pietrzak 2010]. This possibility is provided by data on inter-county population migration retrieved from Google.

The choice of this criterion was motivated by the fact that in Poland counties have on average about five neighbours. The choice of the direct neighbourhood matrix would overestimate the results for the counties of central Poland, and would distort the results in the counties that are enclaves or lie on the country's borders. The neighbourhood matrix defined in this way, and consequently the consequence matrix and the spatial weight matrix, are asymmetric (5). The explanatory variable reflects the intensity of migration flows between n from-to parameters and therefore takes the form of a square matrix:

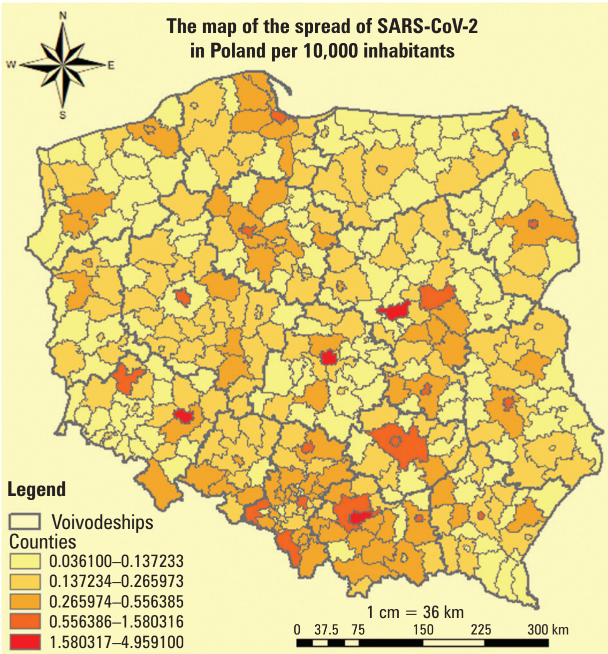
$$W_{5x5} = Y_{5x5} = \begin{pmatrix} 0 & 800 & 937 & 572 & 486 \\ 1250 & 0 & 1301 & 247 & 136 \\ 457 & 575 & 0 & 1079 & 1169 \\ 651 & 218 & 2383 & 0 & 984 \\ 564 & 365 & 1510 & 498 & 0 \end{pmatrix} \tag{5}$$

A row standardised spatial weight matrix considering the five nearest neighbours was used for local data analyses.

$$W_{5x5} = Y_{5x5} = \begin{pmatrix} 0.00 & 0.29 & 0.34 & 0.20 & 0.17 \\ 0.43 & 0.00 & 0.44 & 0.08 & 0.50 \\ 0.14 & 0.18 & 0.00 & 0.33 & 0.36 \\ 0.15 & 0.05 & 0.56 & 0.00 & 0.23 \\ 0.19 & 0.12 & 0.51 & 0.17 & 0.00 \end{pmatrix} \tag{6}$$

The calculation of spatial autocorrelation used a weight matrix derived from a row standardisation of the matrix (which reduces the interpretation to shares in total migration flows). To simplify the analyses, the so-called five nearest neighbours rule was adopted.

The results obtained from the Moran spatial correlation suggest that units between which the distance is small should interact in a stronger way than those farther from each other. This principle forms the basis for the choice of the distance weighting function, which is assumed to be a decreasing function. In practice, thus, potentiometric and exponential functions are used, with varying rates of decrease in impact, depending on the nature of the process and data being analysed.



Source: Author's own study

Fig. 3. Analysis showing the spread of the SARS-CoV-2 virus in Poland as of April 2022

4. Discussion of the results

Analysing and interpreting the spatial correlation results using the Moran statistics method, it should be noted that the positive spatial autocorrelation indicates an increase in infections caused by decreasing distances between individuals and larger human clusters.

The advantage of such a solution over advanced models is that it can be generated and implemented relatively quickly as additional data become available. However, it should be borne in mind that the shift from geographical distance to social flows and links raises questions concerning: the scale of migration, the intensity of contacts or trade flows, and thus the occurrence of various aberrations that radically affect these behaviours (pandemic, war, etc.). The matrices of weights thus obtained do not assume exogeneity anymore, although potential exogeneity can provide a field of research whose common denominator is to answer a question concerning a particular dependency structure. While in case of social connections we mostly do not know the actual structure of interactions and we can only approximate, in the case of flow matrices it is possible to quantify the social connections, and the lack of data does not pose a significant problem.

The spatial analyses presented by the diffusion map (Fig. 3) are heterogeneous, and the interrelationship of migration between neighbouring administrative units plays a key role in the spread of infection.

The analyses presented here suggest that geography, understood as "a description of the spatial differentiation of a phenomenon and as a way of interpreting and explaining it" [Śleszyński 2020], is particularly important for studying the spread of the SARS-CoV-2 virus epidemic. Emerging analyses from the beginning of the pandemic suggested a fairly obvious link with mobility [Jarynowski et al. 2020] and large human concentrations [Krzysztofik et al. 2020], which research has confirmed. The necessity to reformulate previous views of the virus in relation to many aspects of socio-economic life, from goods and services to their distribution. to tourism and recreation, confirms the need to use spatial autocorrelation to "manage" the pandemic. International discussions along these lines [Chowaniak 2020, Florida 2020, Haywood 2020, Łon 2020, Napierała et al. 2020] legitimise the manuscript as a valuable study.

5. Summary and conclusions

Cartography as a branch of map science, map theory, map-making methods and map use has accompanied mankind since ancient times. Spatial visualisations have also been used to deal with current problems, as evidenced by an 1854 map showing the development of a cholera epidemic in London (Fig. 1), through which John Snow correctly deduced the origin of the disease from water, by superimposing the location of water pumps in the city on the location of deaths.

Spatial visualisations can now be used to analyse the spread of the SARS-Cov-2 virus. Studies have shown that the development of a pandemic is influenced not only by the number of infections, but above all by population density, as well as economic, social, educational and transport links, as indicated by the high Pearson correlation coefficient of 0.83. The solutions implemented made it possible to map the regions with high infection rates. It is also noteworthy that the spread of the virus from these regions causes an increase in incidence in neighbouring regions. Figure 3 shows a simulation of the development of a pandemic in counties characterised by high population migration and high population numbers. The diffusion map (virus transmission) is presented as a *choropleth* map highlighting the abatements of the epidemic in counties with low population migration and small overall population. The analyses conducted indicate the possibility of uncontrolled transmission of the virus in areas of high population density coinciding with their mobility.

In its current form, the proposed method does not take into account many factors that are important in controlling a pandemic, such as the number of hospital beds, the location of temporary hospitals or the availability of medical staff. Instead, the analysis presented here helps to select areas with the greatest potential for pathogen spread. A method that allows for quick identification of areas at risk facilitates an efficient response by indicating the hotspots, where action should be taken to protect human health and life. Knowledge of the nature of the pandemic's development could lead to targeted measures (at county, municipal or city level) without the need for nationwide lockdowns. A rapid response based on regional increases in infection will also be important for business owners and entrepreneurs. Pre-emptive action will allow them to manage their businesses in a controlled way, limiting financial losses.

Funded by a subsidy from the Ministry of Education and Science for the University of Agriculture in Krakow for 2023.

References

Araszkiewicz A., Mościcka A., Wabiński J., Kuźma M., Kiliszek D. 2021 Modeling of Various Spatial Patterns of SARS CoV-2. The Case of Germany, Journal of Clinical Medicine.

Budkowski S., Litwin U. 2021. Jak kłamać mapą? Kartogram, czyli metody podziału klasowego. Przegląd Geodezyjny.

Chen Z.L., Zhang Q., Lu Y., Guo Z.M., Zhang X., Zhang W.J., Guo C., Liao C.H., Li Q.L., Creise N.A.C. 1993. The exclusive point of minimum travel. Annals of the Association of American Geographers, 54.

Chowaniak K. 2020. Biznes w czasie epidemii: wpływ koronawirusa na polskie przedsiębiorstwa. Personel i Zarządzanie, 5, 10–15.

Florida R. 2020. The geography of coronavirus. Lity Lab. https://www. bloomberg.com/news/articles/2020-04-03/what-we-know-about-density-and-covid-19-s-spread [accessed: 23.12.2022].

Grabowski M., Zając A. 2009. Dane, informacja, wiedza – próba definicji. Zeszyty Naukowe, Uniwersytet Ekonomiczny w Krakowie, 798, 99–116.

Han X.H. et al. 2020. Distribution of the Covid-19 epidemic and correlation with population emigration from Wuhan, China. Chin. Med., 133, 1044–1050.

Haywood K.M. 2020. A post Covid-19 future: Tourism re-imagined and re-enabled. Tourism Geographies, 22, 3, 599–609.

Jamshidi S., Baniasad M., Niyogi D. 2020. Global to USA county scale analysis of weather, urban density, mobility, homestay, and mask use on Covid-19. Int. J. Environ. Res. Public Health, 17, 7847.

Jarynowski A., Wójta-Kempa M., Płatek D., Krzowski Ł., Bielik V. 2020. Spatial diversity of Covid-19 cases in Poland explained by mobility patterns: Preliminary results, SSRN, Elsevier.

Komenda M., Bulhart V., Karolyi M., Jarkovský J., Mužík J., Májek O., Šnajdrová L., Łon E. 2020. Polityka pieniężna banków centralnych w dobie koronawirusa. Pieniądze i Więź, 1(86), 69–76.

Kraak M.J., Ormeling F.J. 2020. Cartography, 4th end. CRC Press, Boca Raton.

Krzysztofik R., Kantor-Pietraga I., Spórna T. 2020. Spatial and functional dimensions of the Covid-19 epidemic in Poland. Eurasian Geography and Economics, 99.

Legendre P., Legendre L. 1998. Numerical ecology. 2nd edition. Elsevier Science BV, Amsterdam.

Ma Y., Zhao Y., Liu J., He X., Wang B., Fu S., Yan J., Niu J., Zhou J., Luo B. 2020. Effects of temperature variation and humidity on the death of Covid-19 in Wuhan, China. Sci. Total Environ., 724, 138226.

Medyńska-Gulij B. 2011. Kartografia i geowizualizacja. Wydawnictwo Naukowe PWN, Warszawa.

Moran P.A.P 1950. Notes on continuous stochastic phenomena. Biometrika, 37, 1/2.

Napierała T., Leśniewska-Napierała K., Burski R. 2020. Impact of Geographic Distribution of Covid-19 Cases on Hotels' Performances: Case of Polish Cities, Sustainability, 12, 4697, doi.org/10.3390/su12114697.

Neft D. 1966. Statistical analysis for areal distribution. Regional Science Research Institute, Filadelfia.

Pérez-Fuentes M.D.C., Herrera-Peco I., Jurado M.D.M.M., Oropesa N.F., Gázquez Linares J.J. 2021. Predictors of Threat from Covid-19: A Cross-Sectional Study in the Spanish Population. J. Clin. Med., 10, 692.

Pietrzak M.B. 2010. Dwuetapowa procedura budowy przestrzennej macierzy wag z uwzględnieniem odległości ekonomicznej. Oeconomia Copernicana, 1.

Pluta W. 1986. Wielowymiarowa analiza porównawcza w modelowaniu ekonometrycznym. PWN, Warszawa.

Ružičková P., Rážová J., Prymula R. et al. 2020. Complex Reporting of the Covid-19 Epidemic in the Czech Republic: Use of an Interactive Web-Based App in Practice. J. Med. Internet Res., 22.

Sajadi M.M., Habibzadeh P., Vintzileos A., Shokouhi S., Miralles-Wilhelm F., Amoroso A. 2020 Temperature, humidity, and latitude analysis to predict potential spread and seasonality for Covid-19. SSRN Electron. J.

Shi P., Dong Y., Yan H., Li X., Zhao C., Liu W., He M., Tang S., Xi S. 2020. The impact of temperature and absolute humidity on the coronavirus disease 2019 (Covid-19) outbreak – evidence from China. Med. Rxiv.

Suchecka J., Antczak E. 2010. Elementy geostatystyki i metody analiz przestrzennych danych punktowych, C.H. Beck, Warszawa.

Śleszyński P. 2020. Prawidłowości przebiegu dyfuzji przestrzennej rejestrowanych zakażeń koronawirusem SARS-Cov-2 w Polsce w pierwszych 100 dniach epidemii. Instytut Geografii i Przestrzennego Zagospodarowania PAN. Czasopismo Geograficzne, 91 (1–2).

Tang K., Feng K., Lv W. 2020. High Temperature and High Humidity Reduce the Transmission of Covid-19. SSRN Electron. J.

Wang Y., Liu Y., Struthers J., Lian M. 2020. Spatiotemporal Characteristics of the Covid-19 Epidemic in the United States. Clin. Infect. Dis., 72.

Welfe A. 2009. Ekonometria. Metody i ich zastosowanie, PWE, Warszawa.

https://wellcomecollection.org/works?query=%22ARCHIVE%22. Wellcome Collection online archives [accessed: 16 February 2022].

https://covid19.who.int/.iAccessedi6iJulyi2020. World Health Organisation (WHO). https://www.who.int/director-general/speeches/detail/who-directorgeneral-s-opening-remarks-at-the-media-briefing-on-covid-19 [accessed: 16 February 2022].

https://www.google.com/covid19/mobility/, data on mobility provided by Google [accessed: 16 February 2022].

https://geoforum.pl/news/28786/mapy-pandemii-koronawirusa-przeglad-opracowan. Geoforum.pl [accessed: 16 February 2022].
https://commons.wikimedia.org/wiki/File:Tissot_mercator.png. Mercator mapping [accessed: 16 February 2022].
https://stat.gov.pl/. Central Statistical Office [GUS] [accessed: 16 February 2022].
https://www.gov.pl/web/gugik. Head Office of Geodesy and Cartography [GUGIK] [accessed: 16 February 2022].
https://docs.google.com/spreadsheets/d/1ierEhD6gcq51HAm433knjnVwey4ZE5DCnu1bW7PRG3E/edit#gid=1309014089. Rogalski Michał [accessed: 16 February 2022].
https://faculty.ksu.edu.sa/sites/default/files/moran.pdf. Moran model [accessed: 23 December 2022].

Dr inż. Szczepan Budkowski
University of Agriculture in Krakow
Department of Geodesy
30-198 Kraków, ul. Balicka 253a
e-mail: szbudkowski@o2.pl
ORCID: 000-0002-1806-1173