



**Geographia Polonica**  
2021, Volume 94, Issue 3, pp. 355-380  
<https://doi.org/10.7163/GPol.0209>



INSTITUTE OF GEOGRAPHY AND SPATIAL ORGANIZATION  
POLISH ACADEMY OF SCIENCES  
[www.igipz.pan.pl](http://www.igipz.pan.pl)

[www.geographiapolonica.pl](http://www.geographiapolonica.pl)

---

## TEMPORAL VARIATION OF SPATIAL AUTOCORRELATION OF COVID-19 CASES IDENTIFIED IN POLAND DURING THE YEAR FROM THE BEGINNING OF THE PANDEMIC

Alfred Stach 

Department of Geoinformation, Institute of Geoecology and Geoinformation  
Faculty of Geographical and Geological Sciences of the Adam Mickiewicz University  
Bogumiła Krygowskiego 10, 61-680 Poznań: Poland  
e-mail: [alfred.stach@amu.edu.pl](mailto:alfred.stach@amu.edu.pl)

### Abstract

The spread of the COVID-19 pandemic has a simultaneous temporal and spatial component. This pattern results from a complex combination of factors, including social ones, that lead to significant differences in the evolution of space-time distributions, both between and within countries. The aim of this study was to assess changes in the regularity of the spatial distribution of the number of diagnosed COVID-19 cases in Poland over more than a year of the pandemic. The analysis utilized daily and weekly data for 380 counties (poviats), using the local – Poisson risk semivariogram – measure of spatial autocorrelation. Despite the heterogeneity and errors in the source data, it was possible to identify clear patterns of temporal changes in the spatial distribution of COVID-19 cases, manifested by differences in the nature and extent of their autocorrelation.

### Key words

COVID-19 pandemic • COVID-19 cases • spatial autocorrelation • Poisson risk semivariogram • Poland • counties (poviats)

---

### Introduction

#### The current state of knowledge on the space-time dynamics of COVID-19 in the world

The rapid spread of the COVID-19 epidemic has a simultaneous temporal and spatial component. This pattern results from a complex combination of factors, including social ones, that lead to significant differences

in the evolution of space-time distributions both between and within countries.

Among the tens of thousands of scientific articles published, since the beginning of the pandemic over year ago, strictly medical issues, as well as those related to public health and socio-economic consequences, dominate. However, from the very beginning, there has been research thread that draws attention to the space-time relations taking

place during the development of the pandemic which uses for this purpose GIS and spatial econometrics instruments. A partial summary of the content of publications belonging to this trend is covered by the work of Fatima et al. (2021). The authors searched the PubMed database as of September 30, 2020 using the phrase: 'Spatial Analysis' OR 'Geographic Mapping' OR 'Spatial Regression' OR 'Space time Clustering' OR 'Spatio Temporal Analysis' AND 'COVID-19'. In this way, they identified 74 peer-reviewed papers published in English as of January 1, 2020. Fatima et al. (2021) identified, firstly, the dominance of studies from only three countries: China, Brazil and the USA. Secondly and thirdly, they diagnosed the main goals of the performed analyzes and the methodology used for this purpose. Thus, selected publications included: (a) cartographic visualization of the course of the epidemic, (b) development of SARS-CoV-2 virus infection risk maps, and / or (c) construction of time-space epidemiological models. In these papers, their authors used the methodology of clustering, hot / cold spot identification, spatial scanning statistics and spatial regression modeling. Factors limiting the use of these techniques were the deficiencies and shortcomings of data on the progress of the epidemic, which, together with the scarcity of small-scale demographic, environmental and socio-economic information, made it very difficult and challenging to study causal relationships with potential factors influencing the course of COVID-19.

Similar findings were made by the authors of a slightly earlier printed review, which included 63 items published between the end of January and the beginning of May 2020 (Franch-Pardo et al., 2020). They further emphasize that understanding the space-time dynamics of COVID-19 can help in making decisions, planning and implementing social activities. Interactions between public health officials, infected persons and health-care professionals to improve the quality of information on the disease proliferation and assess the likelihood of new outbreaks are important in this regard. Correlation

analyzes are used, i.e., to estimate the impact on the epidemic development of actions in the field of health policy (localization of health / sanitation services and control, mapping / tracking of human movement), the formulation of appropriate scientific and political concepts, and the projection of spatial diffusion and temporal trends.

Use of the same keywords and other criteria as Fatima et al. (2021) for the search of the PubMed database on April 15, 2021 gave a list of 596 items, 541 of which are publications available with full text. As many as 342 of them appeared in the PubMed database after September 30, 2020, i.e. they could not be included in the study by Fatima et al. (2021). This proves the high intensification of works in the field of spatial analysis (increase by over 700%) and probably the outdated of some of the conclusions provided by the above-mentioned authors (Fatima et al., 2021). Undoubtedly, the geographic scope of the analyzes performed has widened. The dominance of publications from China, the USA and Brazil is not so strong anymore. Most of the available studies are those that concern the development of the COVID-19 epidemic in individual countries (Danon et al., 2020; Gomes et al., 2020; Hernández-Flores et al., 2020; Hohl et al., 2020; Kim et al., 2020; Liu et al., 2020; Mollalo et al., 2020; Niu et al., 2020; Ramirez-Aldana, 2020; Castro et al., 2021; Gaudart et al., 2021; Gupta et al., 2021; Huang et al., 2021; Lipsitt et al., 2021; Mościcka et al., 2021; Vaz, 2021). Although so far rare, there are also available results of analyzes carried out on a supra-national scale, especially concerning the European Union (Mounir Amdaoud et al., 2020; Hass & Jekar Arsanjani, 2021; Sannigrahi et al., 2020), entire continents (Weiss et al., 2020 - Africa) and even the whole Earth (Shadi Nazari et al., 2020).

An undoubted drawback of most of the spatial analyzes of the development of the COVID-19 pandemic published so far is their time scope. Most authors rely on relatively short series of data ranging from a few weeks to several months. They are rarely selected

due to the need to take into account a specific phase of the epidemic. Therefore, the interpretation of these partial results must be carried out with great care. The calculations of the spatial autocorrelation and its changes over time during the epidemic development were usually a secondary element of the analyzes performed. They served mainly as a tool in the identification of local clusters (LISA – Local Indicators of Spatial Association, Gomes et al., 2020; Parvin et al., 2020; Ramírez-Aldana et al., 2020; Castro et al., 2021; Hass & Arsanjani 2021; Vaz, 2021) and their evolution over time.

### **The course of the COVID-19 epidemic in Poland and the current state of research**

The first case of COVID-19 in Poland was registered on March 4, 2020. The first victim of the epidemic was a 57-year-old woman who died on March 12, 2020. The state of epidemic threat in Poland was introduced on March 16, 2020 – 5 days after the World Health Organization announced a pandemic in the world. Further restrictions were added on March 25 and 31, and on April 16, the country introduced the obligation to cover the mouth and nose in all public places.

The number of identified cases of the disease has been growing rapidly since March 10, reaching the level of 300-400 per day at the turn of March and April. From the beginning of June to the beginning of July, the epidemic slowed down somewhat – the number of positive test results decreased from around 500 per day to around 250. By August 20, the increase continued from a maximum of around 900 new cases per day. After a short-term reduction, in mid-September, a rapid growth of new infections began – the Second Wave of the epidemic. It can be assumed that it was related to the beginning of the new school year on September 1. Wave Two peaked on November 7, when 27,875 new COVID-19 cases were registered. The situation began to improve significantly from November 20, 2020, when

more than 24,200 patients were registered. The next minimum of the incidence curve – around 5000 people a day – was around February 8, 2021. Since then, another increase has continued – the Third Wave of the Epidemic. In the analyzed period, its maximum occurred on March 10, 2021, when the daily number of identified cases of the disease exceeded 21,000. The viral genome research conducted in March 2021 showed that the so-called “British” mutation was responsible for the rapid increase in the incidence and soon also of deaths (its share reached 70%). Its spread in Poland is associated with the mass visits to the country before Christmas by thousands of Poles from Great Britain.

As of March 20 (Rynek Zdrowia, 2020) since the beginning of the pandemic, 2,036,700 infections have been confirmed in Poland, and 49,159 patients have died. According to conservative estimates, the real number of infected people may be several times greater and exceed 10 million. Random studies (Rosińska et al. 2020b) conducted in December 2020 showed that the percentage of infections that were detected varied depending on the place from 11% to 32%.

Research on the spatial aspects of the development of the COVID-19 epidemic in Poland is still scarce, and the state of knowledge in this area is still partial and very limited by the availability and quality of source data. Śleszyński (2020), basing on the data from the first 100 days of the pandemic, described the sequence of its phases, its spatial diffusion, and concentration. Other authors (Kowalski et al. 2021) pointed to the relationship between the incidence rates and the spatial distribution of air pollution. This corresponds to the results of the publication of Pozzer et al. (2020) estimating that as much as 28% of deaths from COVID-19 in Poland may be related to air pollution. In the world ranking presented in that article, Poland is placed second after the Czech Republic. Krzysztofiak et al. (2020) paid particular attention to the region that was most affected by the Pandemic in its first phase. It was Upper Silesia – a region

with a long mining and industrial tradition. The authors explain the mosaic distribution of cases and deaths from COVID-19 in the Śląskie Voivodeship referring to the processes of urban shrinkage, trans-industrialism, hard coal mining and polycentricity. Jarynowski et al. (2020a, 2020b) determined the relationship between the SARS-CoV-2 onset date and the number of confirmed COVID-19 cases in the first wave (March 4 to May 22, 2020), and socioeconomic variables at the poviats level (previously NUTS-4). They found that immigration and the logarithm of overall mobility are the best predictor of the time of arrival of SARS-CoV-2, while emigration, industrialization and air quality explain the intensity of the epidemic in poviats the most. Variables that were previously considered important, such as population and population density, income, and share of the elderly population, turned out to be less important. Jarynowski et al. (2020a, 2020b) for the entire analyzed period and the two indicators taken into account, calculated the Moran 'I' index, indicating a highly statistically significant ( $p < 0.001$ ) positive autocorrelation. An isolated, but very interesting study is the publication by Bochenek et al. (2021), who identified relationships between the dynamics of a pandemic, determined by the number of identified COVID-19 cases and the number of deaths caused by this disease, and meteorological conditions. However, when reading it, doubts arise as to whether the authors took into account a relationship between meteorological conditions and the efficiency of the pandemic data collection and reporting system.

### **Purpose and scope of the study**

The aim of this study was to identify, characterize and mathematically describe the spatial autocorrelation of daily and weekly data on the number of identified COVID-19 cases in Poland in poviats during the full year of the pandemic (from March 2020 to March 2021). The obtained spatial autocorrelation models can be used to determine

the regularity of the spatio-temporal development of the pandemic, as well as to assess the quality of official data. These results can also be used for planning and management within central and local government administration. At the moment, the author does not know of any similar type of study from any country in the world.

## **Analysis area and source data**

### **The administrative division of Poland and the system for collecting data on the COVID-19 epidemic**

The administrative division in Poland is three-tier: 2477 communes, 380 poviats and 16 voivodeships (as of January 1, 2020, Statistics Poland, 2020). Analyses of the spatial autocorrelation of COVID-19 cases identified in Poland were performed using data from poviats. This is because at this level there are the lowest-grade Sanitary and Epidemiological Stations (SANEPID) units belonging to the State Sanitary Inspection (<https://www.gov.pl/web/gis/o-gis>), responsible for collecting and publishing data on the COVID-19 pandemic.

Poviats are divided into urban (66) and non-urban (314). Their area ranges from 13.3 to 2976.4 km<sup>2</sup> (average 822.9 and standard deviation 520.0 km<sup>2</sup>), the number of inhabitants from approximately 19.8 to 1793.6 thousand (average 100.9, standard deviation 119.9 thousand). There is also a large variation in the population density in these administrative units, as it ranges from 18.9 to 3708.7 people / km<sup>2</sup> (the standard deviation is as much as 650.4 people per km<sup>2</sup>). Therefore, regardless of the high uncertainty as to the accuracy of SARS-CoV-2 virus infection registers, the problem of Small Numbers can be expected (Waller & Gotway, 2004). It occurs in areas where spatial units are characterized by a large diversification of the number of inhabitants. For some of them, having a very small population, small, random, differentiation in the number of cases of the analyzed feature, the calculation of standardized indicators for the population

(e.g. number of cases / 1000 inhabitants) is burdened with a large and unknown error.

Until the end of November last year, the only publicly available source of data on the development of the pandemic in poviats (morbidity and deaths in the 24-hour step) was a database collected by a group of enthusiasts led by Michał Rogalski (2020). However, when several times they discovered significant inaccuracies in the data published overnight on the website of the Ministry of Health, it was decided to fully centralize the flow of information – practically there was no possibility of their social control. Michał Rogalski's group was basing their pandemic figures on data provided directly by individual Poviats Sanitary and Epidemiological Stations, now such information is sent directly to the Ministry of Health in Warsaw and there is processed, compiled and made available to the public at: <https://wojewodztwa-rcb-gis.hub.arcgis.com/pages/dane-do-pobrania>.

The inaccuracies revealed by Michał Rogalski and his associates consisted of discrepancies between the sum of cases reported by individual poviat SANEPiDs and the figures published on the website of the Ministry of Health of the Republic of Poland. It is difficult to assess how accurate (reliable) the data collected at the poviat level are.

The assessment of the course of the COVID-19 pandemic in Poland is undoubtedly hindered by the relatively small number of tests performed, and according to WHO, effective testing of people suspected of being infected allows the epidemic to be controlled and the virus transmission to be limited (Medonet, 2020). From September 2020, people with symptoms indicating the possibility of COVID-19 infection are being tested in Poland as a priority. Preventive tests are also performed on people who report to stay in a sanatorium or are referred to hospices, health care facilities and nursing homes. Healthcare professionals are also tested on an ongoing basis. Poland, with the number of 270,835 tests per 1 million inhabitants (as of March 8, 2021), ranks very far in the world ranking. The maximum daily number of tests

performed in Poland exceeded 80,000, and usually oscillates between 20 and 40,000.

### Data used in the study

This analysis uses the available data on the daily number of new COVID-19 cases in poviats in the period from March 9, 2020 to March 14, 2021 (371 days). The first 5 days of the pandemic (from March 4, 2020) were omitted because they did not constitute a full week. The analysis was performed for 53 weekly totals – from Monday to Sunday. The inspiration for such a decision was weekly cyclicity, which was clearly visible in the data, discussed in more details below. The first Monday during the pandemic was March 9. Time and material scope of the analyzes made it necessary (1) to integrate the databases of the M. Rogalski's team (period from March 9, 2020 to November 23, 2020) and 'government' data (after November 24, 2020) and (2) to calculate weekly sums based on them. The latter dataset is 100% complete and does not show any obvious errors. In the first one, there were large gaps, and many errors consisting in a decrease in the cumulative number of cases in subsequent days. Such data was removed from the analysis. The lack of even one daily value was also the reason for the elimination of a given poviat from the analyzes of weekly totals. Overall, in the first period of 260 days (37 weeks), the mean completeness of daily data was 96.3%, ranging from 82.9 to 98.4%. The greatest shortcomings characterize the first month of the COVID-19 pandemic – until April 8, 2020. For this range, data are available for an average of 88.5% of poviats. Due to the above-mentioned rule, the losses in the weekly sums are greater. For the first 37 weeks, data are available for an average of 93.4% of poviats (from 80.8 to 98.4%), and for the first 5 – only 85.1%.

Unfortunately, neither the first nor the second database contained information on the gender and age structure of people identified with the infection in individual poviats. Aggregated information for the entire

country concerns mainly gender and age relationships among people who died as a result of COVID-19. Demographic characteristics of the disease in the period March-August 2020 included in the study by Rosińska et al. (2020a) is not relative to the general structure of the population. It shows that while the proportion among the sexes is similar, the age groups are dominated by people from 30 to 49 years old (35 to 40% of the total number of identified cases).

One of the most characteristic features visible on the chart of identified COVID-19 cases daily values of is their weekly cyclicity, mentioned above. Obviously, it is a function of many social and institutional factors, and directly depends on the number of tests performed. It should be noted that although this cyclicity is visible in the entire analyzed collection of data from Poland, it is particularly strong in the period after November 24, when data gathering and sharing was centralized at the Ministry of Health. Analyses made by Li (2020) on data from 149 countries of the world show significant linear correlations between this weekly periodicity and the total number of cases and deaths, ranging from 50% to 84% for the group of countries where the death rate is almost three orders of magnitude, from a few to about a thousand per million.

In addition to the data on the daily number of new COVID-19 cases, this study utilizes:

- data on the number of inhabitants in poviats as of December 31, 2019 obtained from the Local Data Bank of the Central Statistical Office (<https://bdl.stat.gov.pl/BDL/start>),
- population distribution in the 1x1 km grid from the National Census carried out in 2011, made available via the Geostatistics Portal of the Central Statistical Office at: Geostatistics Portal (2020),
- GIS layer of polygons of administrative borders of poviats as of August 8, 2019 from the "State register of borders and areas of territorial divisions of the country" (<http://www.gugik.gov.pl/pzgjik/dane-bez-oplat/dane-z-state-register-of-borders->

[and-area-units-territorial-subdivisions-country-prg](#)).

## Methodology of analysis

### Local autocorrelation: classical, population-weighted semivariogram and Poisson risk semivariogram

In this study, to analyze the spatial autocorrelation of data on COVID-19 cases identified during a day in poviats, instead of econometric methods commonly used in epidemiology (Waller & Gotway, 2004; Lawson, 2006; Lai et al., 2008; Pfeiffer et al., 2008; Getis, 2010) geostatistical methods were utilized (Goovaerts, 1997; Oliver, 2010). This decision was made for two reasons. First, this study is part of a wider project in which local spatial autocorrelation models (Poisson risk semivariograms) were used to correct (smooth) empirical data on the spatial distribution of identified COVID-19 cases using Poisson kriging (Goovaerts, 2005). 'Classic' measures of autocorrelation, such as the Moran, Geary or Getis-Orda correlograms, do not provide such possibilities. It was also decided that this part of the project should be included in a separate publication that would focus on the numerically defined regularities of the spatial distribution, and not to the distribution itself, presented cartographically. Secondly, the author of this study does not know of a publication that would indicate any shortcomings of the used autocorrelation measure in relation to the above-mentioned 'classical' measures.

In order for geostatistical methods to be widely used, it was first necessary for their processing to take into account their specificity in the calculation algorithms. The result of this consideration is development of Poisson kriging using as a model of spatial autocorrelation not the traditional Matheron semivariogram estimator, but the risk semivariogram (Oliver et al., 1992, 1998; Goovaerts, 2005; Monestiez et al., 2006; Krivoruchko et al., 2011; De Oliveira, 2014; Stach & Wysocka, 2014; Azevedo et al., 2020). The theoretical foundations of this methodology and an overview

of its applications were presented by Goovaerts (2017). The components of this methodology that were used in this study are briefly described below.

The infection/ disease incidence rate in the polygon  $\alpha$  (spatial unit  $\alpha$ )  $v_\alpha$  is its adjusted, if necessary - due to age and / or gender structure, number  $d(v_\alpha)$  in this polygon divided by the number of people inhabiting it  $n(v_\alpha)$  and multiplied by 100,000 or another constant (1000, 10,000, 1,000,000). Very often, when  $n(v_\alpha)$  is small, the index  $z(v_\alpha) = d(v_\alpha)/n(v_\alpha)$  gets an unrealistically small or large value. This feature is called the ‘Small Numbers’ problem mentioned earlier (Waller & Gotway, 2004).

Traditionally, in geostatistics, semivariance was adopted as a measure of spatial autocorrelation (Goovaerts, 1997; Oliver, 2010), defining the change in dissimilarity between data as a function of their distance and, possibly, direction (occurrence of anisotropy of the phenomenon) [1].

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha,\beta}^{N(\mathbf{h})} [z(v_\alpha) - z(v_\beta)]^2 \quad [1]$$

where:

$\mathbf{h}$  - the distance between any two locations (spatial units in this case) labeled  $\alpha$  and  $\beta$ .

In formula [1], only the “raw” values of the analyzed feature are taken into account. However, it may turn out that due to an important external factor, they cannot be treated as equivalent to each other. In the case of spatial analyzes of social data, such a factor is the number of inhabitants that varies greatly in individual spatial units. The solution to the problem may be the use of a semivariance estimator weighted by the population of compared units ( $n(v_\alpha)$  and  $n(v_\beta)$ ) [2].

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2 \sum_{\alpha,\beta}^{N(\mathbf{h})} n(v_\alpha) n(v_\beta)} \sum_{\alpha=1}^{N(\mathbf{h})} n(v_\alpha) n(v_\beta) [z(v_\alpha) - z(v_\beta)]^2 \quad [2]$$

However, this does not fully solve the problem, since the analyzed variable represents counts of the cases, and therefore a discrete random variable with a Poisson distribution. The classic formula for the measure of dissimilarity - semivariance [1] - must be modified in this type of analysis so that, taking into account the weight of the population, the significance of the ‘Small Numbers’ problem can be reduced. The risk semivariance  $\hat{\gamma}_{RV}(\mathbf{h})$  (Monestiez, 2006) for pairs of spatial units separated by the vector  $\mathbf{h}$  is as follows [3]:

$$\hat{\gamma}_{RV}(\mathbf{h}) = \frac{1}{2 \sum_{\alpha,\beta}^{N(\mathbf{h})} \frac{n(v_\alpha) n(v_\beta)}{n(v_\alpha) + n(v_\beta)}} \sum_{\alpha=1}^{N(\mathbf{h})} \left\{ \frac{n(v_\alpha) n(v_\beta)}{n(v_\alpha) + n(v_\beta)} [z(v_\alpha) - z(v_\beta)]^2 - m^* \right\} \quad [3]$$

where:

$N(\mathbf{h})$  - the number of pairs of spatial units ( $v_\alpha, v_\beta$ ) which (population-weighted) centroids are separated by the vector  $\mathbf{h}$ , and

$m^*$  - the average of the frequency of events for  $N$  areas weighted by the number of inhabitants.

The squared differences  $[z(v_\alpha) - z(v_\beta)]^2$  occurring in the traditional form of semivariance are modified (weighted) by the function of their number of inhabitants in the form of the expression  $n(v_\alpha) \cdot n(v_\beta) / [n(v_\alpha) + n(v_\beta)]$  which is inversely proportional to their standard errors. As a result, more trustworthy data pairs are given more importance. This importance depends on the number of inhabitants according to the rule: the greater it is, the lower the standard error. The modifications of the classical semivariance estimator introduced in the above formula are the more important the greater the diversity of the number of inhabitants in the set of analyzed spatial units.

### Parameterization of local autocorrelation calculations

Before starting the calculations of empirical semivariances, a decision should be made to determine their important parameters:

(1) taking into account or omitting any possible directional relations (anisotropy of the spatial distribution) and (2) the width and number of distance  $h$  intervals for which the calculations are performed. When analyzing the identified cases of COVID-19 in poviats, it was decided to ignore the potential anisotropy. First, there are no reasonable premises indicating the possibility of its existence. Besides, the tests performed on data from a few randomly selected dates did not show any directional relations. This does not mean, of course, that the problem is unequivocally resolved.

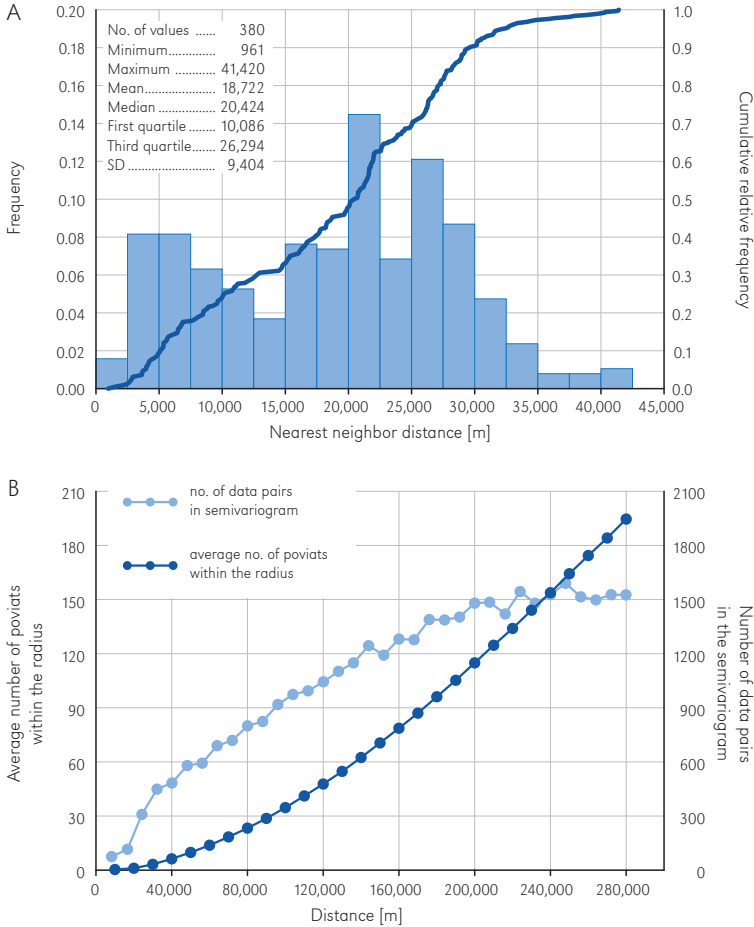
When determining the width and number of compartments, the intention was to achieve contradictory goals: on the one hand, to obtain the most accurate picture of spatial relations, especially over short distances, and on the other hand, to obtain statistically reliable results. The first postulate would require calculations for many narrow distance ranges, and therefore for relatively small subsets of pairs of poviats. With high uncertainty as to the quality of the source data, this would produce results with low statistical reliability. After a series of tests, it was finally decided to carry out calculations for 35 compartments (lags) 8 km wide. The total scope of the analysis was therefore 280 km, i.e. about half the diameter of the country's territory. This principle is quite commonly used in geostatistics. The number of data pairs for individual intervals with a complete data set (data from all 380 poviats) is presented in Fig. 1B. Only for the first interval it is smaller than 100 and equals 75. For intervals over 40 km this value is already over 500, and over 120 km – 1000 pairs. Such large samples certainly significantly reduced the negative impact of outliers.

In the analyzes of spatial autocorrelation performed in this study, standardized data of counts for individual spatial units had to be related to one point. Geometric centroids of these units' polygons are very often used for this purpose. However, in a situation where the distribution of the population in the polygons is highly uneven, it introduces

a considerable inaccuracy to the results obtained. A much better option is to use population distribution centroids, and this is also the chosen solution. Unfortunately, precise data on the distribution of the population of Poland are available only from the National Census carried out in 2011. However, it was recognized that in the period that has elapsed, there have been no significant changes in the spatial distribution of the population, and the benefits of this solution are much greater than possible inaccuracies resulting from the differences between the distributions in 2011 and 2020. The average distance between the poviat's geometric centroid and its center resulting from the population distribution is 2732 m (min. 123, max. 11,512, standard deviation 2129 m). Taking into account that the average distance of the nearest neighbor between poviat centroids weighted by the distribution of the population is 18,722 meters (Fig. 1A), the 'error' resulting from the use of the coordinates of the geometric center could reach even several dozen percent of this interval. In several situations, for example the city of Poznań and the Poznań poviat, when the latter surrounds the city on all sides, then the centroid designated for it is located inside the centrally located city. Similar situations apply to spatial units with a very irregular, crescent-like shape (e.g. Bielsko poviat). Unfortunately, the procedure in the QGIS program used to determine the center of the poviat does not allow 'forcing' such a point to be inside the range.

In order to fully interpret the results of the spatial autocorrelation analysis of daily and weekly data of the number of newly registered SARS-CoV-2 virus infections, a graphical and statistical summary of the distribution of distances between population-weighted centroids of poviats is required (Fig. 1A). As already mentioned, the average of these distances (of the nearest neighbor) is 18,722 meters. The median is nearly 2 km greater (20,424 m), the extreme values are between 961 and 41,420 meters. The distance range, defined by its standard deviation, amounts to 9404 meters.





**Figure 1.** Characteristics of the distance between weighted distribution of population by poviats. A. Nearest neighbor histogram, cumulative curve and basic distance statistics. B. Graph of the average number of poviats (centroids) within a specified radius from the selected one (in 10 km intervals), and the number of data pairs used to calculate the semivariance for individual 8 km wide intervals

The spatial ranges obtained as a result of the autocorrelation analysis should be compared with the average number of poviats (and more precisely their centroids), determined on the basis of the ‘peer-to-peer’ distance matrix, located within a given radius from the selected poviat (Fig. 1B). This is because it enables the assessment of the number of possible clusters grouping poviats with similar numbers of cases. From these calculations the conclusion is that there is on average 1 other poviat within a radius of 20 km (1.07). When the search radius

is increased to 50 km, there are on average about 10 poviats (9.82). For the radii of 100 and 200 km, the corresponding values are 34.7 and 114.9 poviats.

**Modeling of local autocorrelation**

The software used in this study (Goovaerts, 2005), apart from the calculation of the three types of empirical semivariograms discussed above, also performs the adjustment of parametric models to them. For this purpose, the VARFIT algorithm developed

by Pardo-Iguzquiza (1999) was used. In the implementation of the 'Poisson Kriging' software (Goovaerts, 2005), the procedure tests all possible combinations of one or two elementary functions selected from the following three available semivariogram models: spherical, exponential and cubic. Out of them a composite model is selected that is characterized by the smallest weighted sum of the squares of the differences between the curves: experimental  $\gamma(\mathbf{h}_i)$  and model,  $\hat{\gamma}_{RV}(\mathbf{h}_i)$  according to the formula [4]:

$$WSS = \sum_{i=1}^L w(\mathbf{h}_i) [\hat{\gamma}_{RV}(\mathbf{h}_i) - \gamma(\mathbf{h}_i)]^2 \quad [4]$$

where:

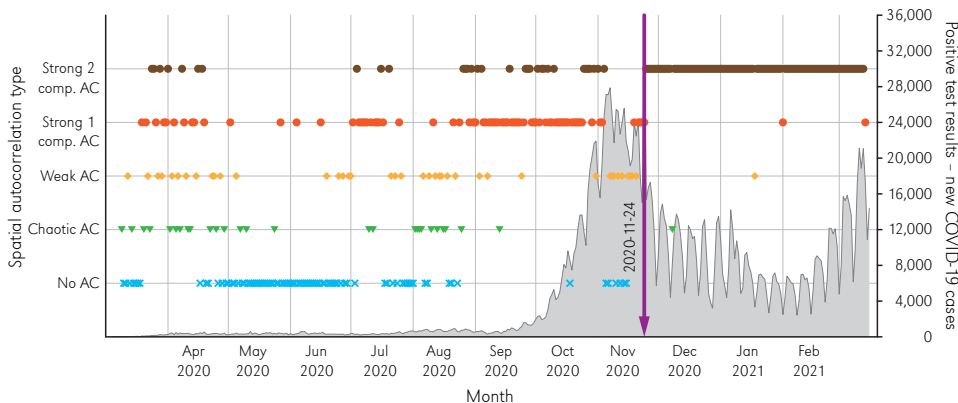
$L$  - the number of distance classes.

The software user can choose one of five weight variants of  $w(\mathbf{h}_i)$ . In the calculations, the results of which are presented in this study, a variant was selected in which the weight is directly proportional to the root of the number of pairs of points for a given interval and inversely proportional to the value of the empirical semivariogram for this interval. Such a scheme prefers semivariances computed from many pairs of data (and therefore more reliable), and at the same time their lower values, which usually occur at short distances, because a good fit of the semivariogram in its initial part has the greatest impact on the accuracy of the kriging estimation.

Due to the exponentiation procedure, semivariance as a derivative of variance is very sensitive to the occurrence of outliers. A single, extremely large or small value in close proximity to 'typical' values can completely distort the picture of actual spatial relations. Records of COVID-19 cases certainly contain many of such outliers. This is especially true in the first period of the pandemic, when the system of controlling the registered cases was not yet operational. At that time, there could have been many significant mistakes, mostly omissions. In analyzes performed on single data sets, various procedures can

be used that allow the identification of such situations and partial or complete elimination of their negative impact (Goovaerts, 1997). However, they are quite time-consuming and, in this study, in which the calculations and modeling of a total of 1272 empirical semivariograms (371 days + 53 weeks x 3 types of semivariograms) were performed, their use was not possible. To make so many calculations, a completely automatic solution was required, despite being aware of its potentially negative consequences. One of the problems was, for example, the alignment of functions by the algorithm even in a situation where the data clearly showed no spatial autocorrelation, or an unrealistic shape of the model. Unfortunately, the software used in this study did not generate any quantitative indicator of the quality of the model fit.

Therefore, it was decided that the summary of the autocorrelation modeling phase would be qualitative rather than quantitative. It consisted of a visual assessment, on the basis of the automatically generated graph, of the nature of the model and the quality of its fit, as well as subjective allocation to one of five classes: (1) no spatial autocorrelation, (2) chaotic autocorrelation, (3) weak autocorrelation, (4) strong autocorrelation with one spatial range, (5) strong autocorrelation with two spatial ranges. The determining factor in the division was the empirical semivariogram and the Poisson risk model. The decision to classify a specific case into class 4 or 5 was sometimes made after checking the numerical form of the model. In some cases, however, these results were ignored. This was the case, for example, when the algorithm fitted two functions to the empirical semivariogram, but the relative contribution of one of them to the total variance of the data was negligible - a situation often referred to as overfitting. Distribution to the distinguished classes could in some cases arouse controversy, because there were no strict, quantitative criteria for their distinction. However, the author estimates that there were few such cases - not more than 3 to 5% of the entire analyzed set of models.



**Figure 2.** Occurrence of five types of distinguished spatial autocorrelation in the period from March 9, 2020 to March 14, 2021 in the daily data sets of the number of new COVID-19 cases in poviats. The vertical, purple arrow shows the division of the analyzed data set on COVID-19 cases in poviats in the country into two subsets: until November 24, 2020 and after that date (detailed explanations in the text)

## Results

### Autocorrelation of daily data: Poisson risk semivariogram models' types

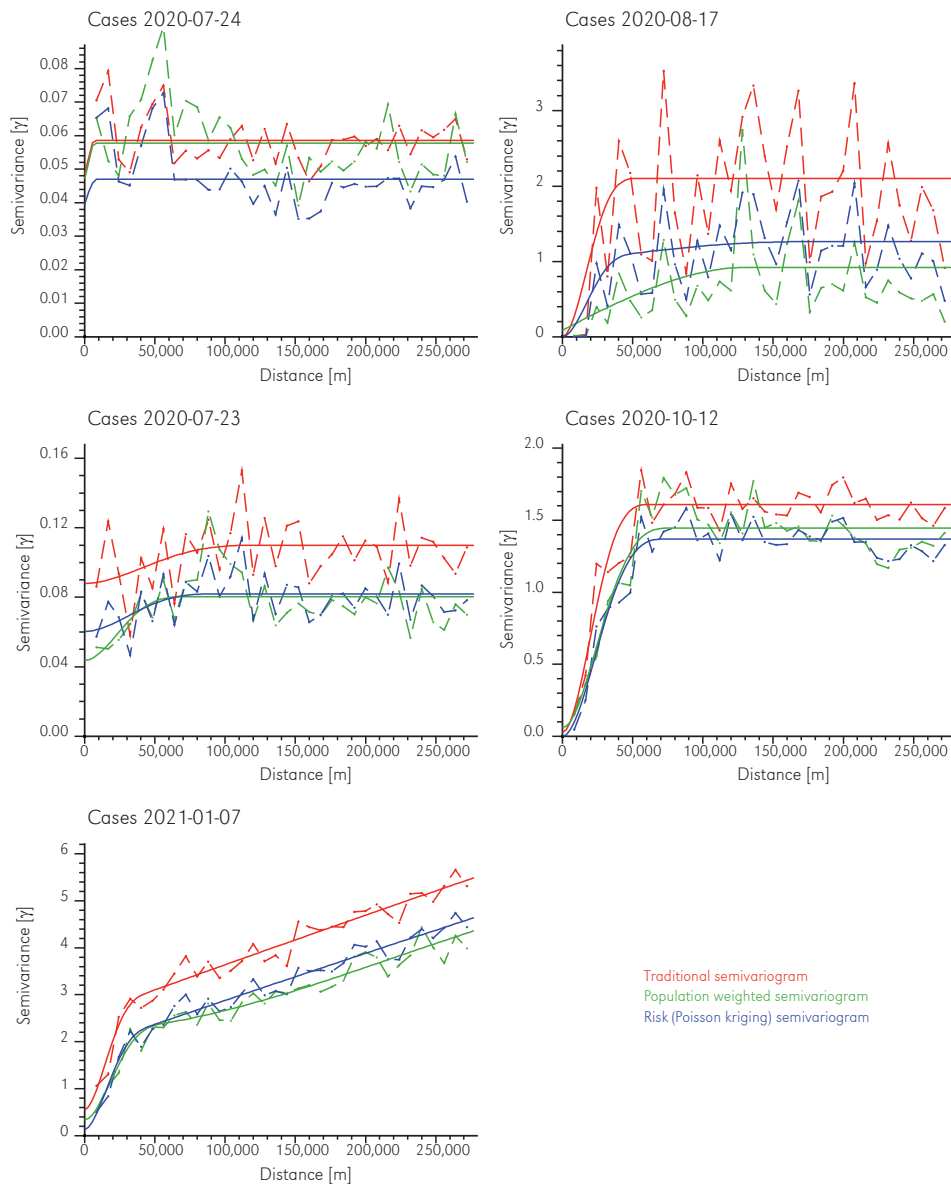
Examples of the 5 distinguished types (without autocorrelation, chaotic autocorrelation, weak autocorrelation, strong autocorrelation with one range, strong autocorrelation with two ranges) of Poisson risk semivariograms for daily data from specific days are presented in Figure 3. Figure 4 contains maps showing the number of identified cases of COVID-19 for the same selected dates. The individual types of autocorrelation of daily data during the year occurrence distribution is presented in Figure 2. To facilitate and extend the interpretation of the data contained therein, the curve of daily COVID-19 case sums for the whole Poland has also been added.

Out of 371 Poisson risk semivariograms for the daily number of new COVID-19 cases in poviats, as many as 85 (22.9%) were classified as not showing spatial autocorrelation. Their occurrence is limited to the period in which data were collected by a group of volunteers on the basis of information provided by individual poviat and voivodship SANEPIDs (before 24 November 2020, Fig. 2). These appearances were concentrated in four

sequences. The first (7 times) took place during the initial 10 days of the analyzed period (9-18 March 2020). The greatest number (as many as 55 cases of the lack of autocorrelation) occurred between April 20 and July 3, 2020. Another 15 cases, somewhat dispersed in time, were recorded between July 18 and August 23, 2020. The last sequence of 6 cases occurred during the Pandemic Wave Two peak, between November 5 and 15, 2020.

The overwhelming majority of dates with a lack of spatial autocorrelation expressed by the Poisson risk semivariogram are days with a low number of cases and a high proportion of poviats where no infected persons were identified. In those days, it was often observed that spatial units with extremely high and low values of the analyzed feature were directly adjacent to each other. The example of July 24 presented in Figure 4 is a situation with one cluster of high incidence rates in Upper Silesia and western Małopolska and a dozen or so randomly scattered individual poviats with a similar level of positive COVID-19 test results in relation to the entire population.

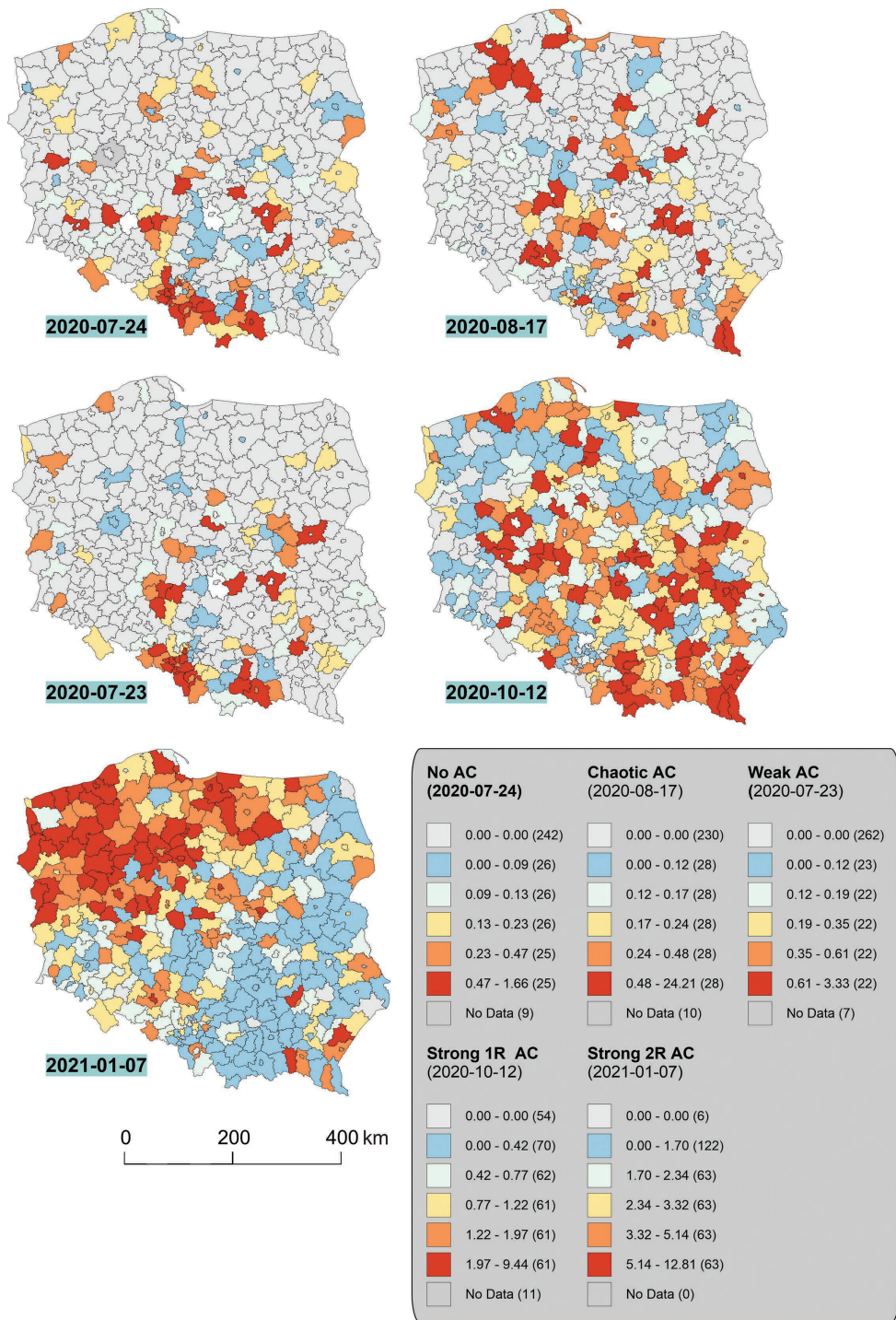
Situations with no 'local' autocorrelation occurred in different phases of the pandemic



**Figure 3.** Selected examples of empirical semivariograms (traditional, weighted by population and Poisson risk, dashed line) and their models (solid line) of COVID-19 cases for days without clear spatial autocorrelation (2020-07-24), with chaotic autocorrelation (2020-08-17), weak (2020-07-23), strong with one component (2020-10-12) and strong with two components (01-07-2021) recorded in poviats. The unit of the semivariance is the number of cases per 10,000 inhabitants squared

– the beginning and maximum of Phase One, the period between Phase One and Two, and at the peak of Phase Two. They intertwined during these periods with more regular

spatial patterns and ceased to appear after centralizing the procedures for collecting and processing information. All this shows quite clearly their genesis. Apart from



**Figure 4.** Examples of spatial distribution of COVID-19 cases per 10,000 inhabitants registered in poviats for selected days (see Figure 3) with different types of spatial autocorrelation. No color fill means no data in the spatial unit. The principle of division into classes for each map is identical, but the boundaries of the class intervals are different: class 1 with gray filling - no recorded COVID-19 cases, classes 2-6 - quintiles of the variable distribution

single cases of random appearance of large new outbreaks of the disease in its initial stage, it is a result of the shortcomings of the test execution system and the collection and processing of data on positive results. The main mechanism of virus spreading in the population is a diffusion process (Śleszyński, 2020), and as such it is characterized by spatial autocorrelation.

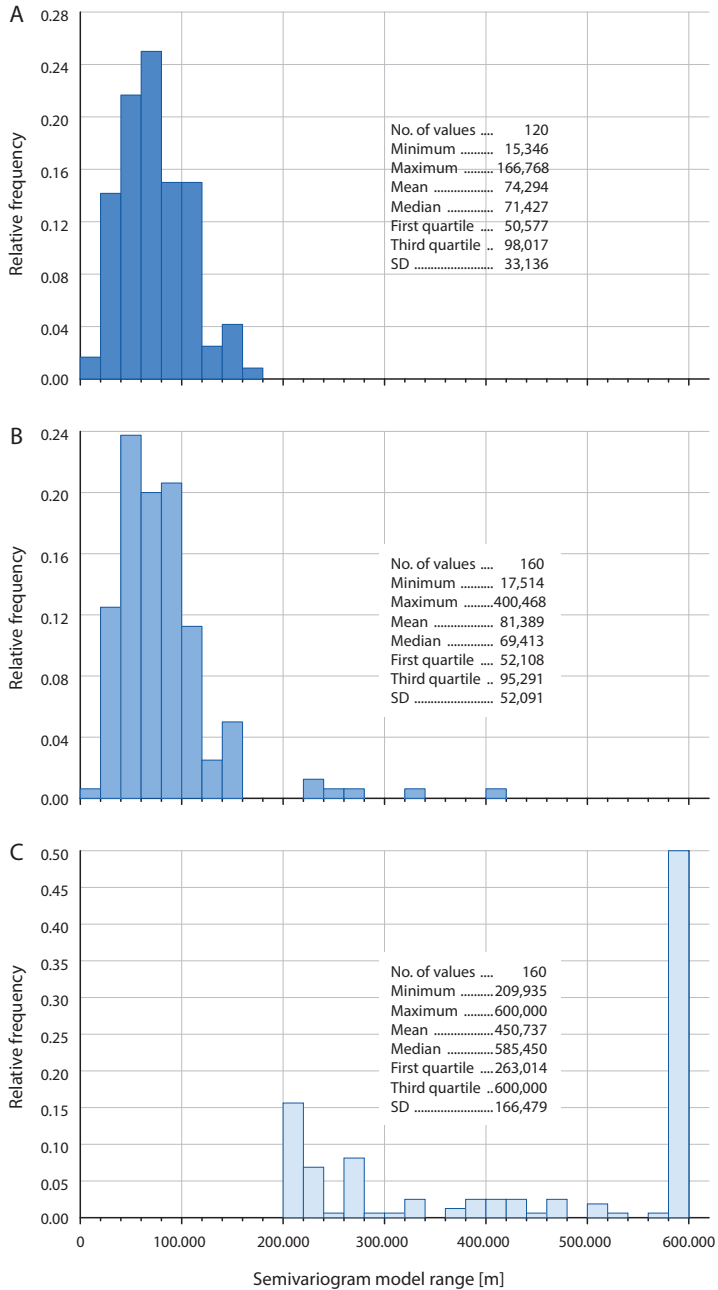
The chaotic type of autocorrelation and the corresponding example of the spatial distribution of daily COVID-19 cases are presented in Figures 3 and 4. The existence of autocorrelation is not in doubt here, but the empirical Poisson risk semivariograms are characterized by a large amplitude of differences between the values calculated for successive intervals. There were the fewest of such situations, because they were found only for 28 data sets (7.5%). Most often, they were scattered single days occurring in the period from March to September 2020. The last case was recorded on December 8, 2020. The chaotic semivariograms were noted more often for the reports of cases from Sunday and Monday (8 and 5 cases; from Tuesday to Saturday it was 4, 3, 2, 3 and 3 cases). The August 17, 2020 map, representing one of the terms with Poisson's chaotic risk semivariogram, shows a situation with a scattered few small clusters of high values (and sometimes even individual poviats) that are often adjacent to units with very low rates of SARS-CoV-2 infection.

The weak autocorrelation type was assigned to those Poisson risk semivariogram models, which were characterized by a very high share in the total variance of the sample of the random component represented by the nugget semivariance (model value for a 0-meter interval). An exemplary graph of such a semivariogram for the data on July 23, 2020 is shown in Figure 3, and the spatial distribution of the source incidence data for that day on the map in Figure 4. In this particular case, the nugget semivariance of 0.0603 accounted for over 73.6% of the entire model variance. Forty one similar models of semivariograms were identified, which

constitutes 11% of the entire analyzed period. Their occurrence, with the exception of one case in January 2021, is also concentrated in the period before November 24, 2020. Most often they were single days, showing no difference in the frequency of occurrence during the week. Only during the Pandemic Second Wave peak, between November 7 and 20, 2020, there was a sequence of 9 days with weak autocorrelation intertwining with those where autocorrelation was not found at all.

In 79 days (21.3%) a strong autocorrelation with one range was found. Only 3 of them took place after November 24, 2020 (Fig. 2). It differs from the weak autocorrelation type described above by the lack of a relatively low nugget semivariance. An example of such empirical semivariograms and their models from October 12, 2020 is shown in Figure 5. The range of the Poisson risk model was 77.5 km and the nugget was zero. In south-eastern and central Poland (Fig. 4), several clusters of poviats with very high and high incidence of people with identified SARS-CoV-2 virus infection (from 1.20 to 9.44 cases / 10,000 inhabitants) were formed that day. The distributions of the number of COVID-19 cases, characterized by this type of spatial autocorrelation, occurred quite frequently in the period of one month after March 20, 2020 (Fig. 2). Again, more of them began to appear in the first half of July 2020. The period of their greatest frequency, or rather domination, is two months starting from August 20, 2020, and therefore immediately before the beginning of the Second Pandemic Wave and during its growth phase. It should also be noted that considering the types with a chaotic and weak autocorrelation, the ones with one range occurred in 122 sets of daily data (32.9%).

A strong autocorrelation with two structures was recorded for 139 sets of daily data with the number of COVID-19 cases identified in poviats (37.5%, Fig. 2). As many as 106 of them occurred after November 24, 2020. This type of spatial autocorrelation distinguished on the basis of Poisson risk semivariograms was characterized by complex



**Figure 5.** Histograms and statistics of the ranges of Poisson risk semivariogram models for daily COVID-19 case data in poviats in the period from March 9, 2020 to March 14, 2021. The comparison also includes models with weak and chaotic autocorrelation. A. Single-range models. B. First structure in models with two ranges. C. Second structure in dual range models

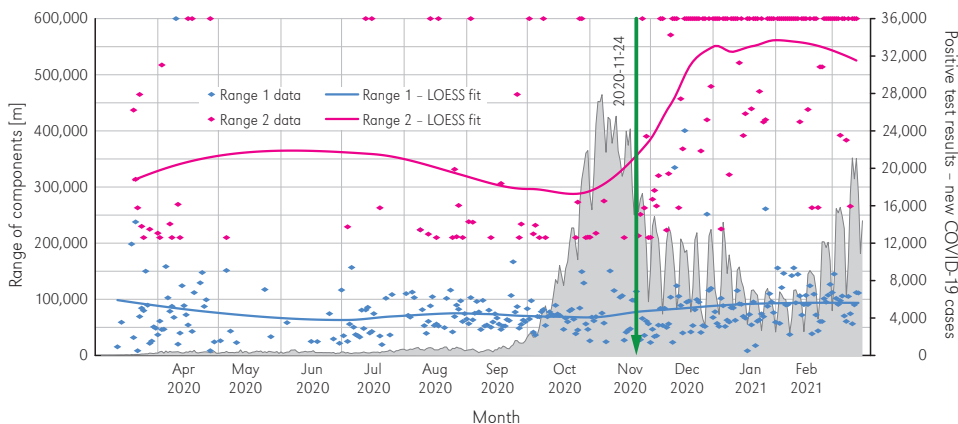
models being a combination of two acceptable mathematical functions. At the same time, as in the previous type, the nugget semivariance was relatively small or zero. The necessity to use more than one function to construct a semivariogram model is usually interpreted (Goovaerts, 1997) as a combination of the effects of various processes operating at different spatial scales. Besides, such a complex model of the semivariogram also makes it possible to evaluate the relative share of each component in the total variability of the analyzed feature (variable). Figure 3 shows an example of a strong spatial autocorrelation with two structures, dated January 7, 2021. The range of the first structure on that day was 51.5 km, and the second one was 1 259.5 km (extrapolated range; see explanations later in this section). This model also contained a nugget structure. Their relative ratio in the radius covered by the calculations (up to 280 km), related to the population-weighted variance of the data, amounted to 2.4, 34.9, and 55.6 (nugget, first and second structure, respectively). According to this the long-distance autocorrelation, perhaps covering the whole country, was of the greatest importance that day. This 'epidemiological background' should be 'superimposed' over clusters composed of several / a dozen poviats that created 'hot' and 'cold' spots of the epidemic. The actual picture of the spatial variability of the number of identified COVID-19 cases in poviats on January 7 is presented on the map (Fig. 4). At that time, Poland was divided into two parts: north-west with high infection rates ( $> 3.3$  people / 10,000 inhabitants) and south-east - with low ( $< 1.7$  people / 10,000). Both zones, however, were very differentiated, with clusters of poviats with higher or lower index values than the average. At the end of this thread of interpretation of the results of Poisson risk semivariogram modeling, it should be added that two spatial autocorrelation structures were also identified in 21 cases where it was chaotic or weak. Twenty of them took place before November 24, 2020.

### **Autocorrelation of daily data: parameters of Poisson risk semivariogram models**

Histograms and statistics of autocorrelation ranges determined by the parameters of Poisson risk semivariogram models are presented in Figure 5. Figure 6 presents their variability over time for individual sets of daily data. All types of models were included in both sets, including those with chaotic and weak autocorrelation. The median ranges of the first structure for the models with one and two ranges are very similar and equal 72.4 km and 69.4 km, respectively (Fig. 5A and 5B). However, the dispersion of the autocorrelation ranges is quite significant in both types of models, and in those with two ranges, there are also a lot of outliers. However, the interquartile range is similar and ranges from 50 to 95 km. The minimum values are also similar (15.5-17.5 km). Comparison with the data in Figure 3B allows to conclude that most often 10 to 30 poviats (radius 50-95 km) were within the range of strong autocorrelation, and the most typical clusters grouped them from 15 to 19 (radius approx. 70 km). It is also worth noting that the maximum range of the first structure, excluding outliers, is about 160-180 km. This is the typical diameter of a voivodeship, i.e. units of the highest level of the administrative division in Poland. The voivodships, grouping from 12 to 42 poviats, although they have only existed in their present form since 1999, often have a distinct natural, social and historical specificity.

The statistics of the range of the second data autocorrelation structure of the number of identified COVID-19 cases can only be estimated approximately (Fig. 5C). Only in 50 situations, out of 160 models with two structures (over 31%), it was shorter than 280 km, i.e. the maximum range of empirical semivariance calculations. For the remaining days, this range was estimated from the parameters of the fitted function. In 50% of cases (80 models), however, such a range was unrealistically large, and even





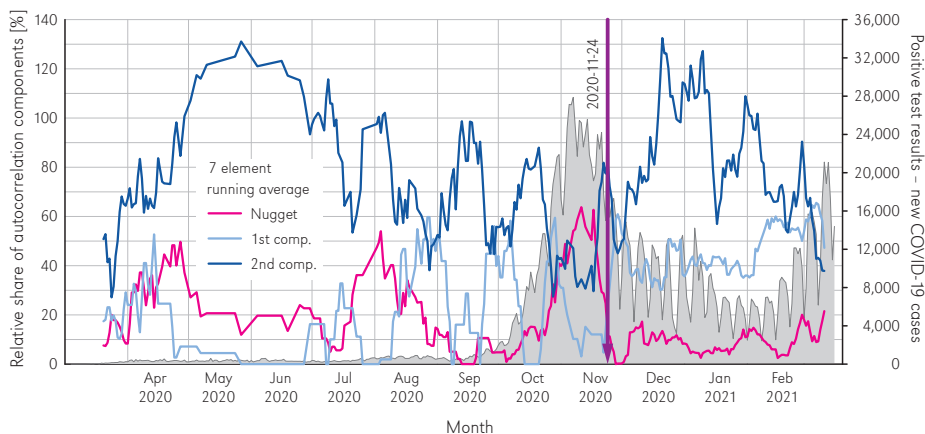
**Figure 6.** Temporal variability of the ranges of Poisson risk semivariogram models for daily COVID-19 case data in poviats in the period from March 9, 2020 to March 14, 2021. The comparison also includes models with weak and chaotic autocorrelation, and the range of the first structure also applies to models with one structure. For the extrapolated ranges of the second structure with a length of more than 600 km, such a value close to the diagonal average of the territory of Poland was arbitrarily assumed. Local polynomial smoothing (LOESS) was used to better illustrate long-term trends. The vertical, green arrow defines the division of the analyzed data set on COVID-19 cases in poviats in the country into two subsets: until November 24, 2020 and after that date (detailed explanations in the text)

amounted to over a thousand kilometers. For the construction of the histogram and the range statistics calculation for the second structure in these situations, a constant value of 600 km was adopted, representing approximately the average diameter of the territory of Poland. This decision was based on the assumption that, taking into account the strong limitations of people's mobility between individual countries, the maximum range of autocorrelation in such analyzes should be limited to the territory of each of them separately. As a result of such action, a bimodal histogram was obtained (Fig. 5C), with one lower frequency peak in the range between 200 and 280 km, and the other at distances of 600 km.

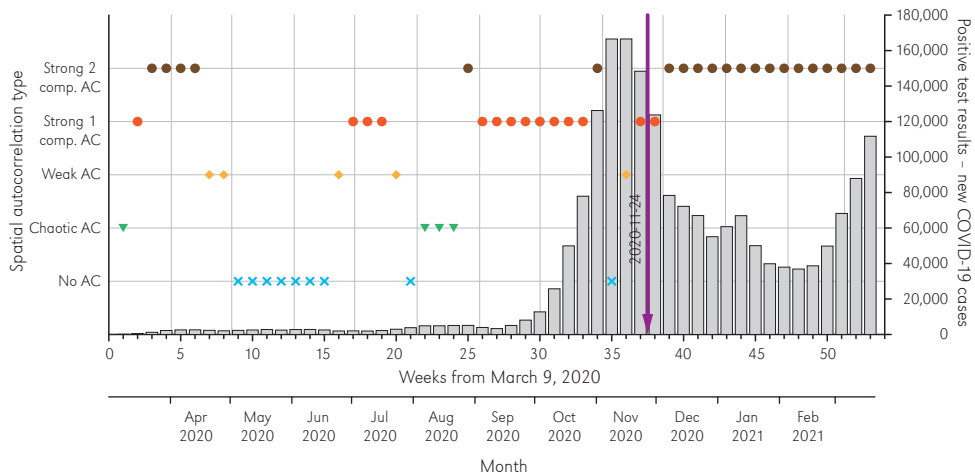
The values of the autocorrelation ranges shown in time, in Figure 6, determined on the basis of Poisson risk semivariogram models, at first glance give the impression of chaos and the absence of any trends. However, the application of smoothing by the local polynomial method (LOESS) reveals long-term tendencies masked by high variability of everyday values. In the first period of the

Pandemic, the range of both, the first and second autocorrelation structures was clearly shorter than in the second. The change took place in early November 2020, during the Second Wave maximum. Before that, the average range of the first structure was about 65 km, and of the second - 235 km. At the turn of December 2020 and January 2021, these values stabilized at around 90 and 540 km (an increase of 38 and 129%). Figure 6 also shows quite clear fluctuations in the LOESS curves occurring in the period from March to November 2020. It seems that taking into account the results obtained for the data aggregated into weekly periods (Fig. 10), only changes in the range of the first component can be considered reliable. They indicate that in the first period of development of the COVID-19 pandemic in Poland (March - April 2020), the range of autocorrelation was slightly larger (approx. 80-85 km), and it has stabilized at the above-mentioned level of 65 km since May.

It is quite difficult to interpret the graphs of changes in the relative shares of individual



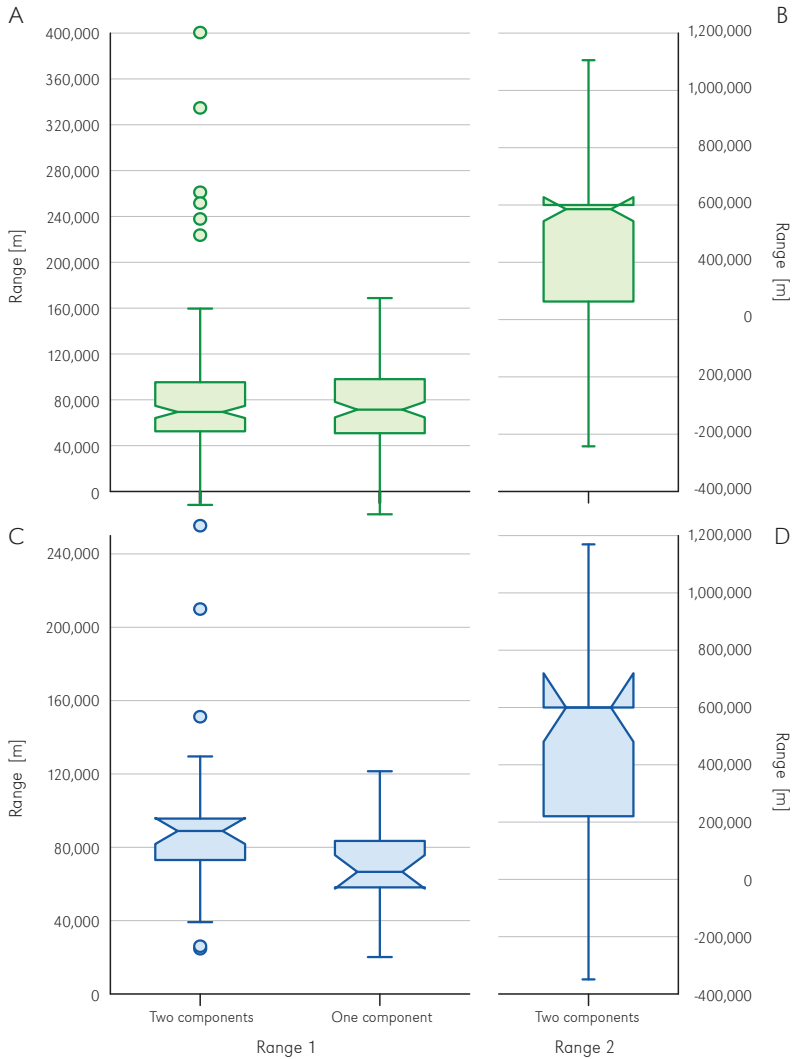
**Figure 7.** Temporal variability of the relative shares of the distinguished structures of Poisson risk semivariogram models for daily data of COVID-19 cases in poviats in the period from March 9, 2020 to March 14, 2021. The list also includes models with weak and chaotic autocorrelation, and the term “first structure” also applies to models with one structure. The shares were relativized in relation to the sample variance weighted by the number of inhabitants. To counteract the impact of single outliers, the data were smoothed with a 7-element moving average. The vertical, purple arrow defines the division of the analyzed data set on COVID-19 cases in poviats throughout the country into two subsets: until November 24, 2020 and after that date (detailed explanations in the text)



**Figure 8.** Occurrence in 53 weeks between March 9, 2020 and March 14, 2021 of 5 types of distinguished spatial autocorrelation in the data of weekly sums of new COVID-19 cases in poviats. The vertical, purple arrow defines the division of the analyzed data set on COVID-19 cases in poviats throughout the country into two subsets: until November 24, 2020 and after that date (detailed explanations in the text)

components of Poisson risk semivariogram models (Fig. 7). The values of partial variances of each structure were standardized

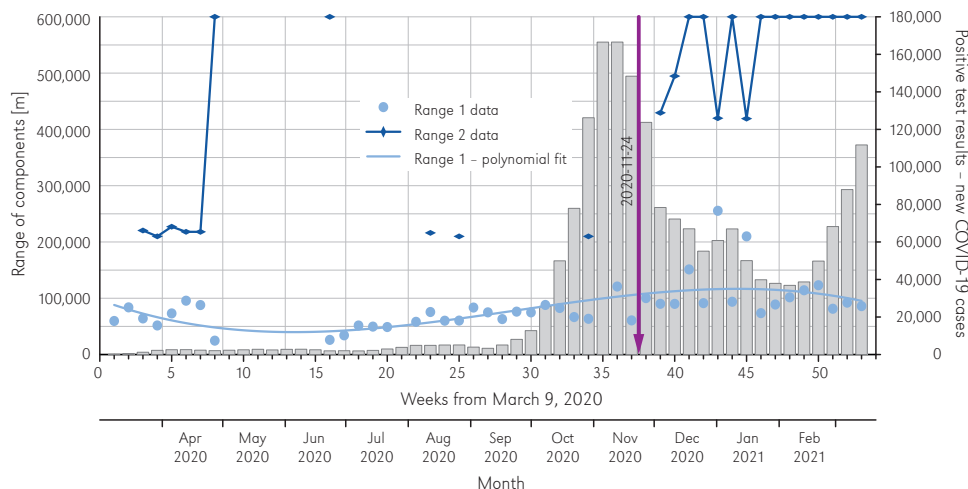
in relation to the variance weighted by populations in individual poviats of each set of daily data. Doubts relate mainly to the share of the



**Figure 9.** Box plots of the ranges of Poisson risk semivariogram models for daily (A and B) and weekly (C and D) data. Noteworthy are statistically significant differences in the range of the first structure of models with one and two structures for weekly data (part C of the figure)

long-distance structure, because its partial variance in most cases concerns the extrapolated range, exceeding many times, as mentioned earlier, the maximum radius of dissimilarity calculations. Therefore, it is limited to indicating very general tendencies of the other two components, because the short-term ones may be apparent. Increased shares of nugget semivariance are quite clearly

associated with the peaks of the infection curve in April, August and November 2020. After November 24, 2020, the share of the first – short-distance structure has clearly increased and stabilized. It also seems that the decline in the share of the second – long-distance structure from December 2020 to the end of the analyzed period also reflects the real changes.



**Figure 10.** Temporal variability of the ranges of Poisson risk semivariogram models for weekly COVID-19 case data in poviats in the period from March 9, 2020 to March 14, 2021. The comparison also includes models with weak and chaotic autocorrelation, and the range of the first structure also applies to models with one structure. For the extrapolated ranges of the second structure with a length of more than 600 km, such a value close to the diagonal average of the territory of Poland was arbitrarily assumed. A 3rd degree polynomial was used to better illustrate the long-term trend in the range of the first structure. The vertical, green arrow shows the division of the analyzed data set on COVID-19 cases in poviats in the country into two subsets: until November 24, 2020 and after that date (detailed explanations in the text)

### Autocorrelation of weekly data

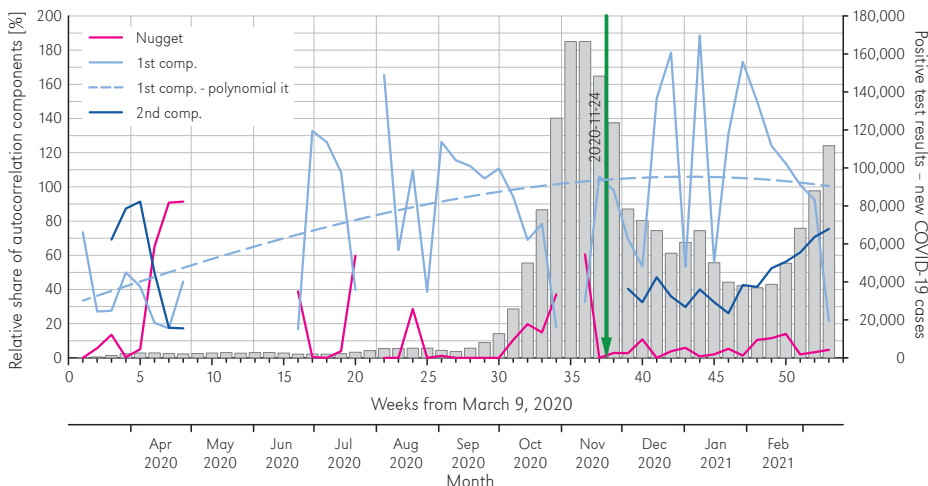
The results of the spatial autocorrelation analysis of the number of identified COVID-19 cases in poviats aggregated into weekly periods (53 data sets) are presented in Figures 8 to 11. They are almost 100% consistent with those presented above for the daily step of the analysis. Therefore, the following description will take into account only some of their more interesting aspects.

There is relatively less of Poisson risk semivariogram models for data of weekly non-autocorrelation cases in the set (17.0 to 22.9%), and more of those with strong autocorrelation with one range (26.4 to 21.2%) and two ranges (39.6 to 37.2%). However, these differences are not statistically significant. After November 24, 2020, all weekly Poisson risk semivariograms show a strong autocorrelation (Fig. 8). Noteworthy is the significant statistical difference ( $p < 0.05$ ) in the range of the first structure for models with one

and two components (Fig. 9C). The median of the former is 66.6 km, and the latter – 88.9 km. The comparison of the ranges of individual structures (spatial components) for weekly data (Fig. 10) suggests a slight decrease in the range of the first structure in the last three weeks of the analysis period, i.e. at the time of the build-up of the third wave of the pandemic. Such regularity in the daily data (Fig. 6) was not noticed. For weekly data, the opposite is the relation of the share of both structures in the total data variance (Fig. 11). The short-distance structure is more important. This is very clear for the period after November 24, 2020.

### Summary and conclusions

The results of the analysis of daily and weekly indicators of infection numbers in poviats during the year from the beginning of the pandemic are heavily burdened with shortcomings of the data collection, verification



**Figure 11.** Temporal variability of the relative shares of the distinguished structures of Poisson risk semivariogram models for weekly sums of COVID-19 cases in poviats in the period from March 9, 2020 to March 14, 2021. The list also includes models with weak and chaotic autocorrelation, and the term “first structure” also applies to models with one structure. The shares were relativized in relation to the sample variance weighted by the number of inhabitants. A 3rd degree polynomial was used to better illustrate the long-term trend of the share of the first structure. The vertical, purple arrow defines the division of the analyzed data set on COVID-19 cases in poviats throughout the country into two subsets: until November 24, 2020 and after that date (detailed explanations in the text)

and processing system. The lack of a well-organized flow of information during the first eight and a half months of its duration, currently limits the possibilities of correctly identifying the spatial regularities of disease progression in Poland, and in real time hindered making the right decisions about socio-economic activities. It should be noted that the aggregation of daily data into weekly periods did not significantly improve their quality – it did not limit the share of periods with no spatial autocorrelation. This means that the problem was not related to delays in collecting information from individual poviats (“shifting” data from day to day), but to the non-systematic appearance of large errors in time and space. This may also be proven by the presence, in days and weeks with no local autocorrelation, of highly statistically significant values of the global Moran autocorrelation index (data not presented in this article). Certainly, the relatively small number of tests performed and the rules, according

to which people were qualified to have these tests done, did not promote the precise assessment of the actual spatial dynamics of COVID-19 in Poland.

Assuming, however, that the distribution of data errors was random both in time and space, one can try to “filter them out”, at least partially. The basic tool for this is a Poisson risk semivariogram computed from the data and parametrically smoothed. The model of such a semivariogram can be used to identify outliers and eliminate their impact on the analyzed spatial distribution. On the other hand, the time series of such models enables the quantification of the changing regularities of the spatial process, and thus gives grounds for making genetic hypotheses. This was the main goal of this study. It used data from poviats on the number of identified COVID-19 cases for the period of 371 days, starting from March 9, 2020 and ending on March 14, 2021. Weekly totals (from Monday to Sunday) for full 53 weeks were also analyzed.

The discussion on the obtained results must, unfortunately, be limited, because in the literature available to the author, there is nothing so far that would allow them to be confronted with the characteristics of the dynamics of local autocorrelation of the COVID-19 epidemic progress for other areas and / or other periods.

The lack of spatial autocorrelation was demonstrated in 23% of daily and 17% – weekly data. They occurred during the first period of the pandemic from March to August and at the Second Wave maximum. After changing the system for collecting and publishing data, they did not appear even once. Before this date, there were also frequent cases of chaotic and weak autocorrelation.

Most, i.e. 58.4% of daily data sets and 66.0% of weekly data in total, showed a very strong autocorrelation with a small share of the random component. Autocorrelation models with two structures: short-distance and long-distance, dominated (respectively 37.2 and 39.6%). This indicates the presence of something that, by analogy with geochemical mapping, can be called an 'epidemiological background', in which there are clusters of poviats with high (so-called hot-spots) and low (so-called cold-spots) numbers of diagnosed COVID-19 cases. This, by analogy to terrain representation, creates a landscape with local hills and depressions superimposed on gradual changes in the height of the area on a regional scale. Such situations occurred throughout the analyzed year, but their absolute dominance (96.3%) characterizes the period after November 24, 2020, when the data collection and processing system was reorganized. Situations with strong autocorrelation and one structure represent a picture with isolated clusters of poviats with high and low epidemic intensity rates, showing no regional connections. Such cases were frequent in the first period of the development of the Pandemic, especially between the First and Second Waves of the disease and during the build-up of the Second Wave. It should also be noted that the systems with one and two spatial structures also

occurred on days and weeks with autocorrelation classified as weak or chaotic. It can be assumed that in most of such cases, the number and magnitude of errors in the data made it difficult to correctly identify the real spatial regularities.

The autocorrelation of disease data shows two ranges: an average of about 70 km and over 280 km (possibly also the scale of the entire country). Within the radius determined by the first range, there are from a few to a dozen poviats. The maximum range of the first structure is about 160-180 km, which corresponds quite well with the size of the voivodships. The first range decreased, and then increased (85-65-90 km). This layout applies to both daily and weekly data. The second range increased significantly at the start of the Second Wave of the Pandemic. Noteworthy is the significant statistical difference in the range of the first structure for models with one and two components for data aggregated up to weeks. It shows that when there was no regional 'epidemiological background', the clusters of poviats with similar COVID-19 incidence rates were slightly smaller. The systematic appearance of poviat clusters composed of a few or a dozen units (autocorrelation radius approx. 70 km) indicates the need for coordinated administrative activities on such a scale, and not, as was the case most often, on the scale of entire voivodeships. On the other hand, focusing on individual poviats with extremely high numbers of cases will also not be effective, due to the strong links resulting from social and economic activity.

The share of the random component (nugget semivariance) decreased significantly after November 24, 2020. It reached its maximum at the height of the Second Wave of pandemic (first half of November 2020). The daily data is dominated by the second component (long-distance), and the weekly data – by the first (short-distance). The importance of the short-distance component in the weekly data clearly increased during the analyzed period. According to this the differences between the values typical for clusters

of poviats, both of the 'hot-spot' and 'cold-spot' type, and the regional background increased in the course of the epidemic progression. This tendency has been reversed since the beginning of the Third Wave of the pandemic. The analysis of data aggregated into weekly periods also shows, which seems logical, that changes in the 'level' of epidemics in clusters occur faster and on a larger scale (shorter distances) than changes in the regional 'epidemiological background'. It is not visible in the comparison for daily data, but it may be the result of high uncertainty in the assessment of the scope of the long-distance structure.

Undoubtedly, the most important conclusion from the analysis carried out in this study

is the need to reorganize the system of collecting, verifying and reporting epidemic data in Poland. This must be done without waiting for the next COVID-19 phase, or for the next pandemic, which will come sooner or later. Many publications in the media indicated both financial and personal limitations of the State Sanitary Inspection and the lack of administrative actions that would improve this sad state of affairs immediately after the pandemic began.

Editors' note:

Unless otherwise stated, the sources of tables and figures are the authors', on the basis of their own research.

## References

- Azevedo, L., Pereira, M.J., Ribeiro, M.C., Soares, A. (2020). Geostatistical COVID-19 infection risk maps for Portugal. *International Journal of Health Geographics*, 19(25). <https://doi.org/10.1186/s12942-020-00221-5>
- Bhunja, G.S., Roy, S., Shit, P.K. (2021). Spatio-temporal analysis of COVID-19 in India – a geostatistical approach. *Spatial Information Research*, 1-12. <https://doi.org/10.1007/s41324-020-00376-0>
- Bochenek, B., Jankowski, M., Gruszczynska, M., Nykiel, G., Gruszczynski, M., Jaczewski, A., Ziemianski, M., Pyrc, R., Figurski, M., Pinkas, J. (2021). Impact of meteorological conditions on the dynamics of the COVID-19 pandemic in Poland. *International Journal of Environmental Research and Public Health*, 18, 3951. <https://doi.org/10.3390/ijerph18083951>
- Castro, R.R., Santos, R.S.C., Sousa, G.J.B., Pinheiro, Y.T., Martins, R.R.I.M., Pereira, M.L.D., Silva, R.A.R. (2021). Spatial dynamics of the COVID-19 pandemic in Brazil. *Epidemiology and Infection*, 149, e60, 1-9. <https://doi.org/10.1017/S0950268821000479>
- Danon, L., Brooks-Pollock, E., Bailey, M., Keeling, M. (2020). A spatial model of COVID-19 transmission in England and Wales: early spread and peak timing. *MedRxiv*, 2020.02.12.20022566. <https://doi.org/10.1101/2020.02.12.20022566>
- De Oliveira, V. (2014). Poisson kriging: A closer investigation. *Spatial Statistics*, 7, 1-20. <https://doi.org/10.1016/j.spasta.2013.11.001>
- Fatima, M., O'Keefe, K.J., Wei, W., Arshad, S., Gruebner, O. (2021). Geospatial analysis of COVID-19: A scoping review. *International Journal of Environmental Research and Public Health*, 18(5), 2336. <https://doi.org/10.3390/ijerph18052336>
- Feng Y, Li Q, Tong X, Wang R, Zhai S, Gao C, et al. (2020). Spatiotemporal spread pattern of the COVID-19 cases in China. *PLoS ONE*, 15(12), e0244351. <https://doi.org/10.1371/journal.pone.0244351>
- Franch-Pardo, I., Napoletano, B.M., Rosete-Verges, F., Billa, L. (2020). Spatial analysis and GIS in the study of COVID-19. A review. *Science of The Total Environment*, 739. <https://doi.org/10.1016/j.scitotenv.2020.140033>

- Gaudart, J., Landier, J., Huiart, L., Legendre, E., Lehot, L., Bendiane, M.K., Chiche, L., Petitjean, A., Mosnier, E., Kirakoya-Samadoulougou, F., Demongeot, J., Piarroux, R., Rebaudet, S. (2021). Factors associated with the spatial heterogeneity of COVID-19 in France: a nationwide geo-epidemiological study. *Lancet Public Health*, 6(4), E222-E231. [https://doi.org/10.1016/S2468-2667\(21\)00006-2](https://doi.org/10.1016/S2468-2667(21)00006-2)
- Geostatistics Poland. (2020). [https://geo.stat.gov.pl/start/-/asset\\_publisher/jNfJiljcyRp/content/id/36734](https://geo.stat.gov.pl/start/-/asset_publisher/jNfJiljcyRp/content/id/36734)
- Getis, A. (2010). B.3 Spatial Autocorrelation. In M.M. Fischer, A. Getis (Eds.), *Handbook of applied spatial analysis: Software tools, methods and applications* (pp. 255-278). Berlin-Heidelberg: Springer-Verlag. [https://doi.org/10.1007/978-3-642-03647-7\\_14](https://doi.org/10.1007/978-3-642-03647-7_14)
- Gomes, D.S., Andrade, L.A., Ribeiro, C.J.N., Peixoto, M.V.S., Lima, S.V.M.A., Duque, A.M., Cirilo, T.M., Góes, M.A.O., Lima, A.G.C.F., Santos, A.D. (2020). Risk clusters of COVID-19 transmission in northeastern Brazil: Prospective space-time modelling. *Epidemiology and Infection*, 148, e188, 1-8. <https://doi.org/10.1017/S0950268820001843>
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York: Oxford University Press.
- Goovaerts, P. (2005). Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. *International Journal of Health Geographics*, 4(31), 1-33. <https://doi.org/10.1186/1476-072X-4-31>
- Goovaerts, P. (2017). Geostatistical interpolation of rate data using Poisson kriging. In S. Shekhar, H. Xiong, X. Zhou (Eds.), *Encyclopedia of GIS: Second Edition* (pp. 767-774). Springer International Publishing AG. <https://doi.org/10.1007/978-3-319-17885-1>
- Gupta, D., Biswas, D., Kabiraj, P. (2021). COVID-19 outbreak and Urban dynamics: Regional variations in India. *GeoJournal*. <https://doi.org/10.1007/s10708-021-10394-6>
- Hass, F.S., Arsanjani, J. (2021). The geography of the COVID-19 pandemic: A data-driven approach to exploring geographical driving forces. *International Journal of Environmental Research and Public Health*, 18, 2803. <https://doi.org/10.3390/ijerph18062803>
- Hernández-Flores, M. de la L., Escobar-Sánchez, J., Paredes-Zarco, J.E., Franyuti Kelly, G.A., Carranza-Ramírez, L. (2020). Prediction and potential spatially explicit spread of COVID-19 in Mexico's megacity North Periphery. *Healthcare*, 8(4), 453. <https://doi.org/10.3390/healthcare8040453>
- Hohl, A., Delmelle, E.M., Desjardins, M.R., Lan, Y. (2020). Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spatial and Spatio-temporal Epidemiology*, 34, 100354. <https://doi.org/10.1016/j.sste.2020.100354>
- Huang, X., Zhou, H., Yang, X., Zhou, W., Huang, J., Yuan, Y. (2021). Spatial characteristics of Coronavirus disease 2019 and their possible relationship with environmental and meteorological factors in Hubei Province, China. *GeoHealth*, 5, e2020GH000358. <https://doi.org/10.1029/2020GH000358>
- Jarynowski, A., Wójta-Kempa, M., Platek, D., Krzowski, Ł., Belik, V. (2020a). Spatial diversity of COVID-19 cases in Poland explained by mobility patterns – Preliminary results (June 6, 2020). <https://doi.org/10.2139/ssrn.3621152>
- Jarynowski, A., Wójta-Kempa, M., Krzowski, Ł. (2020b). An attempt to optimize human resources allocation based on spatial diversity of COVID-19 cases in Poland. *medRxiv*, 2020.10.14.20090985. <https://doi.org/10.1101/2020.10.14.20090985>
- Kim, S., Marcia, C. Castro, M.C. (2020). Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). *International Journal of Infectious Diseases*, 98, 328-333. <https://doi.org/10.1016/j.ijid.2020.07.004>
- Kowalski, P.A., Szwagrzyk, M., Kiełpińska, J., Konior, A., Kusy, M. (2021). Numerical analysis of factors, pace and intensity of the corona virus (COVID-19) epidemic in Poland. *Ecological Informatics*, 63, 101284. <https://doi.org/10.1016/j.ecoinf.2021.101284>
- Krivoruchko, K., Gribov, A., Krause, E. (2011). Multivariate areal interpolation for continuous and count data. *Procedia Environmental Sciences*, 3, 14-19. <https://doi.org/10.1016/j.proenv.2011.02.004>



- Krzysztofik, R., Kantor-Pietraga, I., Spórna, T. (2020). Spatial and functional dimensions of the COVID-19 epidemic in Poland. *Eurasian Geography and Economics*, 619(4-5), 573-586, <https://doi.org/10.1080/15387216.2020.1783337>
- Lai, P.-C., So, F.-M., Chan, K.-W. (2008). *Spatial epidemiological approaches in disease mapping and analysis*. CRC Press. <https://doi.org/10.1201/9781420045536>
- Lawson, A.B. (2006). *Statistical methods in spatial epidemiology, Second Edition*. John Wiley & Sons. <https://doi.org/10.1002/9780470035771>
- Li, S. (2020). The relationship between weekly periodicity and COVID-19 progression. *medRxiv*, preprint 2020.11.24. <https://doi.org/10.1101/2020.11.24.20238295>
- Lipsitt, J., Chan-Golston, A.M., Liu, J., Su, J., Zhu, Y., Jerrett, M. (2021). Spatial analysis of COVID-19 and traffic-related air pollution in Los Angeles. *Environment International*, 153. <https://doi.org/10.1016/j.envint.2021.106531>
- Liu, Y., He, Z., Zhou, X. (2020). Space-time variation and spatial differentiation of COVID-19 confirmed cases in Hubei Province based on extended GWR. *ISPRS International Journal of Geo-Information*, 9(9), 536. <https://doi.org/10.3390/ijgi9090536>
- Medonet. (2020). <https://www.medonet.pl/koronawirus/koronawirus-w-polsce,testy-na-covid-19-w-polsce---aktualizacja,artykul,58274591.html>
- Mollalo, A., Vahedi, B., Rivera, K.M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of The Total Environment*, 728, <https://doi.org/10.1016/j.scitotenv.2020.138884>
- Monestiez, P., Dubroca, L., Bonnin, E., Durbec, J.-P., Guinet, C. (2006). Geostatistical modelling of spatial distribution of *Balaenoptera physalus* in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. *Ecological Modelling*, 193(3-4), 615-628. <https://doi.org/10.1016/j.ecolmodel.2005.08.042>
- Mościcka, A., Araszkiwicz, 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*, 10(7), 1409. <https://doi.org/10.3390/jcm10071409>
- Mounir Amdaoud, M., Arcuri, G., Levratto, N., Succurro, M., Costanzo, D. (2020). Geography of COVID-19 outbreak and first policy answers in European regions and cities. <https://halshs.archives-ouvertes.fr/halshs-03046489>
- Niu B, Liang R, Zhang S, Zhang, H., Qu, X., Su, Q., Zheng, L., Chen, Q. (2020). Epidemic analysis of COVID-19 in Italy based on spatiotemporal geographic information and Google Trends. *Transboundary and Emerging Diseases*. <https://doi.org/10.1111/tbed.13902>
- Oliver, M.A. (2010). B.6 The Variogram and Kriging. In M.M. Fischer, A. Getis (Eds.), *Handbook of applied spatial analysis: Software tools, methods and applications* (pp. 319-352). Berlin Heidelberg: Springer-Verlag. [https://doi.org/10.1007/978-3-642-03647-7\\_17](https://doi.org/10.1007/978-3-642-03647-7_17)
- Oliver, M.A., Muir, K.R., Webster, R., Parkes, S.E., Cameron, A.H., Stevens, M.C., Mann, J.R. (1992). A geostatistical approach to the analysis of pattern in rare disease. *Journal of Public Health*, 14(3), 280-289. <https://doi.org/10.1093/oxfordjournals.pubmed.a042744>
- Oliver, M.A., Webster, R., Lajaunie, C., Muir, K.R., Parkes, S.E., Cameron, A.H., Stevens, M.C.G., Mann, J.R. (1998). Binomial cokriging for estimating and mapping the risk of childhood cancer. *Mathematical Medicine and Biology: A Journal of the IMA*, 15(3), 279-297, <https://doi.org/10.1093/imammb/15.3.279>
- Pardo-Iguzquiza, E. (1999). VARFIT: A Fortran-77 program for fitting variogram models by weighted least squares. *Computers and Geosciences*, 25, 251-261. [https://doi.org/10.1016/S0098-3004\(98\)00128-9](https://doi.org/10.1016/S0098-3004(98)00128-9)
- Parvin, F., Ali, S.A., Hashmi, S.N.I. Ateeque, A. (2021). Spatial prediction and mapping of the COVID-19 hotspot in India using geostatistical technique. *Spatial Information Research*. <https://doi.org/10.1007/s41324-020-00375-1>

- Pfeiffer, D., Robinson, T., Stevenson, M., Stevens, K., Rogers, D., Clements, A. (2008). *Spatial analysis in epidemiology*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198509882.001.0001>
- Pozzer, A., Dominici, F., Haines, A., Witt, C., Münzel, T., Lelieveld, J. (2020). Regional and global contributions of air pollution to risk of death from COVID-19. *Cardiovascular Research*, 116(14), 2247-2253. <https://doi.org/10.1093/cvr/cvaa288>
- Ramírez-Aldana, R., Gomez-Verjan, J.C., Bello-Chavolla, O.Y. (2020). Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level. *PLoS Neglected Tropical Diseases*, 14(11), e0008875. <https://doi.org/10.1371/journal.pntd.0008875>
- Rogalski, M. (2020). Internetowa baza danych o zakażeniach COVID według województw i powiatów, aktualizowana codziennie. <https://docs.google.com/spreadsheets/d/1ierEhD6gcq51HAM433knjVwey4ZE5DCnu1bW7PRG3E/edit?usp=sharing>
- Rosińska, M., Sadkowska-Todys, M., Stępień, M., Kitowska, W., Milczarek, M., Juszczyk, G. (2020a). COVID-19 epidemic in Poland in spring and summer 2020. In B. Wojtyński, P. Goryński (Eds.), *Health status of Polish population and its determinants 2020* (pp. 333-350). Warsaw: National Institute of Public Health, National Institute of Hygiene. <https://www.pzh.gov.pl/download/21915/>
- Rosińska, M., Sadkowska-Todys, M., Stępień, M., Kitowska, W., Milczarek, M., Juszczyk, G. (2020b). Badanie seroprewalencji w populacji ogólnej i w grupie pracowników medycznych. Supplement do Rozdziału 7. Epidemia COVID-19 w Polsce na wiosnę i w lecie 2020. In B. Wojtyński, P. Goryński (Eds.), *Health status of Polish population and its determinants 2020* (pp. 1-9). Warsaw: National Institute of Public Health, National Institute of Hygiene. <https://www.pzh.gov.pl/wp-content/uploads/2021/02/Suplement-do-Rozdzialu-7-seroprewalencja.pdf>
- Rynek Zdrowia. 2020. <https://www.rynekzdrowia.pl/Polityka-zdrowotna/Koronawirus-w-Polsce-2-036-700-potwierdzonych-zakazen-zmarlo-49-159-osob,204119,14.html>
- Sannigrahi, S., Pilla, F., Basu, B., Basu, A.S., Molter, A. (2020). Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustainable Cities and Society*, 62, 102418. <https://doi.org/10.1016/j.scs.2020.102418>
- Shadi Nazari, S., Norouzi, S., Asghari Jafar-abadi, M. (2020). How is Coronavirus distributed in the world? A Spatial-Temporal Assessment Using Geographic Information System Approach. *Jorjani Biomedicine Journal*, 8(1): P 24-33. <https://doi.org/10.29252/jorjanibiomedj.8.1.24>
- Stach, A., Wysocka, P. (2014). Zastosowanie metody krzygu Poissona w badaniach rozkładu przestrzennego problemów społecznych na przykładzie Poznania. *Acta Universitatis Lodzensis, Folia Geographica Socio-Oeconomica*, 16, 169-188. [http://dspace.uni.lodz.pl:8080/xmlui/bitstream/handle/11089/10674/Strony%20od%20FOLIA\\_16-10-STACH\\_WYSOCKA.pdf?sequence=1&isAllowed=y](http://dspace.uni.lodz.pl:8080/xmlui/bitstream/handle/11089/10674/Strony%20od%20FOLIA_16-10-STACH_WYSOCKA.pdf?sequence=1&isAllowed=y)
- Statistics Poland. 2020. <https://stat.gov.pl/en/regional-statistics/classification-of-territorial-units/administrative-division-of-poland/>
- Śleszyński, P. (2020). Prawidłowości przebiegu dyfuzji przestrzennej rejestrowanych zakażeń koronawirusem SARS-CoV-2 w Polsce w pierwszych 100 dniach epidemii. *Czasopismo Geograficzne*, 91(1-2), 5-18. <http://czasgeo.ptgeo.org.pl/ojs31/index.php/geo/issue/view/25/0045-9453%202020%20%281-2%29>
- Vaz, E. (2021). COVID-19 in Toronto: A Spatial exploratory analysis. *Sustainability*, 13, 498. <https://doi.org/10.3390/su13020498>
- Waller, L.A., Gotway, C.A. (2004). *Applied spatial statistics for public health data*. New Jersey: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471662682>
- Weiss, D.J., Bertozzi-Villa, A., Rumisha, S.F., et al. (2020). Indirect effects of the COVID-19 pandemic on malaria intervention coverage, morbidity, and mortality in Africa: A geospatial modelling analysis. *The Lancet Infectious Diseases*, 21(1), 59-69. [https://doi.org/10.1016/S1473-3099\(20\)30700-3](https://doi.org/10.1016/S1473-3099(20)30700-3)