GIS-based spatio-temporal analysis and modeling of COVID-19 incidence rates in Europe

Nima Kianfar [*], Mohammad Saadi Mesgari

Department of Geospatial Information Systems, Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, No. 1346, ValiAsr Street, Mirdamad Cross, Tehran, Iran 19967-15433

A B S T R A C T

The COVID-19 epidemic has emerged as one of the most severe public health crises worldwide, especially in Europe. Until early July 2021, reported infected cases exceeded 180 million, with almost 4 million associated deaths worldwide, almost a third of which are in continental Europe. We analyzed the spatio-temporal distribution of the disease incidence and mortality rates considering specific periods in this continent. Further, we applied Global Moran's I to examine the spatio-temporal distribution patterns of COVID-19 incidence rates and Getis-Ord Gi* hotspot analysis to represent high-risk areas of the disease. Additionally, we compiled a set of 40 demographic, socioeconomic, environmental, transportation, health, and behavioral indicators as potential explanatory variables to investigate the spatial variations of COVID-19 cumulative incidence rates (CIRs). Ordinary Least Squares (OLS), Spatial Lag model (SLM), Spatial Error Model (SLM), Geographically Weighted Regression (GWR), and Multiscale Geographically Weighted Regression (MGWR) regression models were implemented to examine the spatial dependence and non-stationary relationships. Based on our findings, the spatio-temporal distribution pattern of COVID-19 CIRs was highly clustered and the most high-risk clusters of the disease were situated in central and western Europe. Moreover, poverty and the elderly population were selected as the most influential variables due to their significant relationship with COVID-19 CIRs. Considering the non-stationary relationship between variables, MGWR could describe almost 69% of COVID-19 CIRs variations in Europe. Since this spatio-temporal research is conducted on a continental scale, spatial information obtained from the models could provide general insights to authorities for further targeted policies.

Introduction

Coronavirus disease (COVID-19), originating from Hubei province in China in early December 2019, became a global health crisis due to its rapid spread (World Health Organization (WHO) 2020a). As of June 30, 2021, more than 180 million infected cases and almost 4 million associated deaths have been reported globally (World Health Organization (WHO) 2021a). Europe accounts for almost a third of the total disease infections. On the mentioned date, this continent had about 56 million infected cases and almost 1,183,000 associated deaths caused by COVID-19. COVID-19 transmission is not confined to national borders or geographical territories since social, economic, and political communications and activities have risen dramatically worldwide. Accordingly, coronavirus spread is a complex issue, and it is better not to limit the study to a specific level (Kianfar et al., 2021). In Europe, the first COVID-19 infected case was found in France in December 2019

(Newyork Post, May 5, 2020). After that, other European countries became infected rapidly, and the number of cases and associated deaths increased dramatically and got out of control (World Socialist WebSite November 3, 2020).

In Europe, Sannigrahi et al. (Sannigrahi et al., 2020) analyzed the spatial relationship between the COVID-19 infected cases and associated deaths and socio-demographic determinants using spatial regression techniques. Their data investigation extended from December 31, 2019, to April 29, 2020, indicated that there was a significant statistical association between infected cases and two variables, namely income and poverty. They also demonstrated that the total population (of each country) had the most influence on COVID-19 associated deaths in European countries. Dye et al. (Dye et al., 2020) also implemented a flexible, empirical model (skew-logistic) to find the most influential variables on the geographical dynamics and variations of COVID-19 mortalities across the continental Europe. Considering population
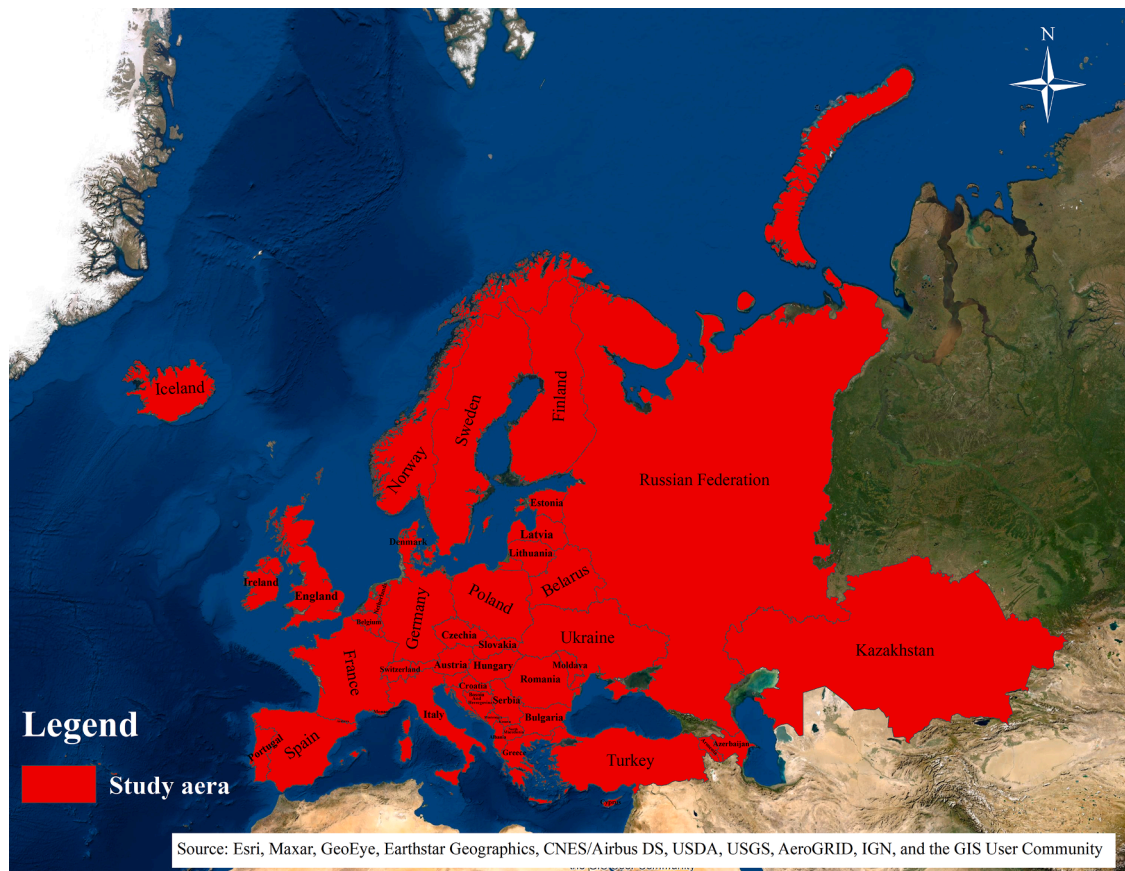

Fig. 1. Continental Europe (Location of the study area).

density, imposed contact rates (e.g., non-pharmaceutical interventions) among individuals, and the number of populations exposed to the infection as potential determinants, they indicated that countries with fewer COVID-19 death rates had smaller populations, had shorter epidemics which had peaked sooner, and also started their stay-at-home restrictions and quarantine policies earlier than other countries in Europe. Miller et al. (Miller et al., 2020) applied descriptive analysis to examine the spatial distribution of COVID-19 globally. The results proved that China, Italy, Iran, and Spain experienced the highest disease prevalence on March 17, 2020. As can be seen, much research has been done on different dimensions of COVID-19 to understand the spatial patterns of disease spread and identify the factors affecting its severe prevalence. The applicability of spatial analysis tools and techniques is remarkable in better understanding the behaviors of COVID-19 epidemiology (Kianfar et al., 2021). In addition, Geographic Information System (GIS) is a practical tool for analyzing various spatial distribution patterns of infectious diseases (Mollalo et al., 2018).

In this GIS-based study, we applied Global Moran's I and Getis-Ord Gi* analyses to examine the spatial distribution patterns of COVID-19 incidence rates and specify high and low-risk regions of the disease in continental Europe, which is the first objective of this study. The focus of this research was on the most specific dates of the disease in Europe. Considering the WHO table (World Health Organization (WHO) 2021a), which shows the situation of each continent, we extracted the seven most specific dates. Thus, the first COVID-19 peak in Europe, minimum incidence after the first peak, the second wave of the outbreak, second peak, and minimum infection after the second peak, start of vaccination, and the last time we used before vaccination proves its effectiveness (February 28, 2021), were the dates we considered in our study. Furthermore, we have included European Union (EU) countries and some other non-EU countries in the process of spatial analyses and

modeling for better comparing the situation of countries in terms of COVID-19 infections and enabling further comparative investigations. Countries such as Russia, Turkey, Belarus, Ukraine, Armenia, Moldova, Macedonia, and so on. Moreover, five global (OLS, SEM, SLM) and local (GWR, MGWR) regression models were implemented to identify how well these techniques can describe the distribution of COVID-19 incidence rates, based on several potential explanatory variables, which is the second objective of this study. Noteworthy, due to the unpredictable nature of COVID-19, considering a wide range of potential explanatory variables from different categories can help us discover the most influential variables on the disease prevalence and lead to more accurate and reliable results in modeling the disease.

Since this research is conducted on a continental scale, it can give us an overview of the COVID-19 situation across countries situated in one of the most affected continents. Furthermore, by identifying the factors that significantly impact the higher prevalence of the disease, it could be considered as a helpful resource for officials and authorities to make practical decisions (Mollalo et al., 2020).

2. Materials and Methods

2.1. Data collection and preparation

World Health Organization (WHO) is responsible for monitoring and collecting the daily data of COVID-19 worldwide. Considering the Europe continent as the study area of this research (Fig. 1), the country-level number of confirmed cases and associated deaths of COVID-19 for European countries were retrieved from WHO (World Health Organization (WHO) 2021b). Weekly and cumulative data of the disease for each country were retrieved for the seven most specific periods. These specific dates were March 30, 2020 (first peak in Europe), June 1, 2020

Table 1

Description of potential explanatory variables and data sources.

Category	Explanatory variable	Description	Source
Demographic	Population density	Counts of all residents per sq. km of land area	World Bank (The World Bank February 1, 2021)
	Population, male	Counts all male residents regardless of legal status or citizenship.	World Bank
	Population female	Counts all female residents regardless of legal status or citizenship.	World Bank
	Pop male rate	% of the population that is male	World Bank
	Pop female rate	% of the population that is female	World Bank
	Population in the largest city	% of a country's urban population living in that country's largest metropolitan area.	World Bank
	Urban population	Refers to people living in urban areas	World Bank
	Urban population growth	Annual growth of people living in urban areas	World Bank
	Rural population	Refers to people living in urban areas	World Bank
	Rural pop growth	Annual growth of people living in rural areas	World Bank
	pop ages 0-14	Population between the ages 0 to 14 as a percentage of the total population.	World Bank
	Population ages 15-64	Population between the ages 15 to 64 as a percentage of the total population.	World Bank
	*__Population ages 65 and above__	Population 65 years of age or older as a percentage of the total population	World Bank
	Hospital beds (per 1,000 people)	Including inpatient beds available in public and private rehabilitation centers.	World Bank
	Nurses and midwives (per 1,000 people)	Including professional, enrolled, and other associated personnel, such as primary care nurses.	World Bank
	Physicians (per 1,000 people)	Including generalist and specialist medical practitioners.	World Bank
Socioeconomic	Unemployment, total	The share of the labor force that is without work but available for employment.	World Bank
	Unemployment, male	The share of the male labor force that is without work but available for employment.	World Bank
	Unemployment, female	The share of the female labor force that is without work but available for employment.	World Bank
	Employment to population ratio, 15+	The proportion of a country's population that is employed.	World Bank
	Life expectancy at birth, total (years)	The number of years a newborn infant would live.	World Bank
	Out-of-pocket expenditure	% of current health expenditure spending on health directly out-of-pocket by households.	World Bank
	Inflation	The annual percentage change in the cost to the average consumer of acquiring a basket of goods	World Bank
	*__Poverty__	% of the population living below the national poverty line	World Bank
	GDP	Gross domestic product divided by midyear population.	World Bank
	GNI	Gross national income, converted to U.S. dollars	World Bank
Transportation	Air transport, passengers carried	Domestic and international aircraft passengers of air carriers registered in the country.	World Bank
	Railways, passengers carried	The number of passengers transported by rail times kilometers traveled.	World Bank
Health	Prevalence of HIV, total	% of people who are infected with HIV.	World Bank
	Diabetes prevalence (% of population ages 20 to 79)	% of people ages 20-79 who have type 1 or type 2 diabetes.	World Bank
	Incidence of tuberculosis	The number of new tuberculosis cases arising in a given year	World Bank
	Health-related mortality	Mortality from CVD, cancer, diabetes or CRD between ages 30 and 70	World Bank
Environmental	Altitude	Time Averaged Map of Tropopause Height (Daytime/Ascending, AIRS-only) daily 1 deg.	NASA, Giovanni (Giovanni March 1, 2021)
	Rain	Time Averaged Map of Total precipitation rate daily 0.25 deg.	NASA, Giovanni
	SO2	Time Averaged Map of SO2 Column Amount daily 0.25 deg.	NASA, Giovanni
	CO	Time Averaged Map of CO Emission (ENSEMBLE) monthly 0.5×0.625 deg.	NASA, Giovanni
	Temperature	Time Averaged Map of Air Temperature (Daytime/Ascending) daily 1 deg.	NASA, Giovanni
	NO2	Time Averaged Map of NO2 Total Column (30% Cloud Screened) daily 0.25 deg.	NASA, Giovanni
	PM2.5 air pollution, mean annual exposure	The average level of exposure to concentrations of particles measuring less than 2.5 microns in aerodynamic diameter	World Bank
Behavioral	Prevalence of current tobacco use (% of adults)	The percentage of the population ages 15 years and over who currently use any tobacco product	World Bank

(minimum incidence after the first peak), and September 24, 2020 (the second wave of the outbreak). November 2, 2020 (second peak in Europe), December 21, 2020 (minimum infection after the second peak), January 4, 2021 (start of vaccination), and February 28, 2021 (last time we used for modeling and the time before vaccination proves its effectiveness). We have extracted these times from WHO based on the weekly prevalence of COVID-19 in continental Europe (World Health Organization (WHO) 2021a). Based on these periods, we calculate the cumulative incidence rates (CIRs) (by dividing the number of infected cases by the population of each country) and cumulative mortality rates (CMRs) (by dividing the number of mortalities by the population of each country) of COVID-19 for each country, both from the beginning of the prevalence (cumulative) and weekly (seven days before each of the specific dates). Further, we joined both types of these CIRs and CMRs to the boundary shapefile of countries using ArcGIS Desktop 10.8. Noteworthy, retrospective data can be beneficial for related institutions and researchers to understand past behaviors and predict future trends afterward (Iyanda et al., 2020).

A set of 40 demographic, socioeconomic, environmental, transportation, health, and behavioral indicators was compiled at the country-level and considered as potential explanatory variables. Further, all variables were attached to the corresponding boundary shapefile of European countries in the ArcMap environment. Names, descriptions, and sources of all variables are provided in Table 1.

Global Moran's I and Getis Ord Gi*were applied as spatial analysis methods to identify the spatial distribution patterns of the disease incidence rates and specify high-risk countries across Europe, respectively. Moreover, five global (OLS, SEM, and SLM) and local (GWR and MGWR) models were implemented to investigate the relationship between COVID-19 CIRs (dependent variable) and selected explanatory variables.

2.2. Spatial pattern analysis

2.2.1. Spatial autocorrelation (Global Moran's I)

Global Moran's I was applied in the ArcMap environment to examine the spatial distribution of COVID-19 incidence rates (both weekly and cumulative) for all countries in the study area. This method performs by considering both locations and values of features simultaneously in the study area. Global Moran's Index ranges between -1 and +1, demonstrating that the spatial distribution of the disease is clustered (> 0), dispersed ($= 0$), or random (< 0). This analysis also measures a z-score and p-value. For a significant statistical value of z-score or p-value, when Moran's Index is positive, the distribution of the disease incidence rates has a propensity towards a clustered pattern. However, for a negative Moran's Index, the mentioned distribution tends towards a dispersed pattern (Moran, 1950).

2.2.2. Hotspot analysis

Getis-Ord Gi* approach was applied in the ArcMap environment to detect significant high and low-risk clusters of COVID-19 spread. The Gi* spatial statistics (Ord & Getis, 2001) identifies high and low-risk clusters of the disease based on distance. To be an intense hotspot or cold spot cluster, a country with a significant value of incidence rate should be surrounded by other countries with large incidence rates. A large positive value of z-score depicts a cluster of a hotspot. However, a small negative value of z-score demonstrates a cold spot cluster. The larger or smaller the z-score, the more intense pattern of clusters. When the z-score is close to zero, the spatial clusters are not obvious (Getis & Ord, 2010).

2.3. Spatial statistical models

2.3.1. Ordinary Least Squares (OLS)

The OLS regression is a global linear modeling technique that can understand the global relationships between the set of control and response variables. Considering the assumption of spatial stationary and homogeneity, this global regression method investigates the relationship between the set of explanatory and dependent variables (Oshan et al., 2019). The formula of OLS is characterized by: (Ward & Gleditsch, 2018)

$$y_i = \beta_0 + x_i\beta + \varepsilon_i \tag{1}$$

where y_i denotes the COVID-19 CIRs (dependent variable) at the ith location (country). β_0 is the estimated intercept, representing the value of y when x is equal to 0, x_i signifies the vector of selected explanatory variables, β indicates the vector of regression coefficients, and ε_i is a random error term.

One of the main functions of the OLS is to optimize the regression coefficients (β) by diminishing the sum of squared distances between the observed data and the values predicted by the model (Oshan et al., 2019). Moreover, Variance Inflation Factor (VIF) was used to quantify the intensity of any multicollinearity in the regression analysis. Multicollinearity occurs when there is a linear relationship between two or more explanatory variables. Large values of VIF represent redundancy among explanatory variables. If the VIF value for each explanatory variable is more than 7.5, it should be eliminated from the regression model. VIF is denoted by:

$$VIF_i = \frac{1}{1 - R_i^2} \tag{2}$$

where R_i^2 demonstrates the coefficient of determination for regressing the ith explanatory variable on other ones.

Regarding the case of COVID-19 spread, the observations are spatially correlated. However, OLS method considers no dependency or correlation among COVID-19 CIRs, leading to a bias in coefficient prediction (Goodchild et al., 1993). Although OLS might be an inefficient method in the case of COVID-19 (Ward & Gleditsch, 2018), it was applied in this study in order to help evaluate and compare the accuracy and robustness of the findings. Then, two variants of OLS, namely SEM and SLM, were implemented to consider spatial dependence and weights.

Spatial Error Model (SEM)

The OLS often refuses to consider the explanatory variables with spatial dependence (Wu et al., 2020). However, the SEM considers spatial dependency in the OLS error term (Mollalo et al., 2020) by decomposing the error term of OLS into two components, including the error term and the random error term. The formula of SEM is as follows: (Ward & Gleditsch, 2018)

$$y_i = \beta_0 + x_i\beta + \lambda W_i\xi_i + \varepsilon_i \tag{3}$$

where λ signifies the coefficient of spatial component errors, W_i represents the weight matrix (a vector of spatial weights) which determines the neighbors at country i and connects the independent variable to the explanatory variables at that country, and ξ_i describes the spatial error component.

2.3.3. Spatial Lag Model (SLM)

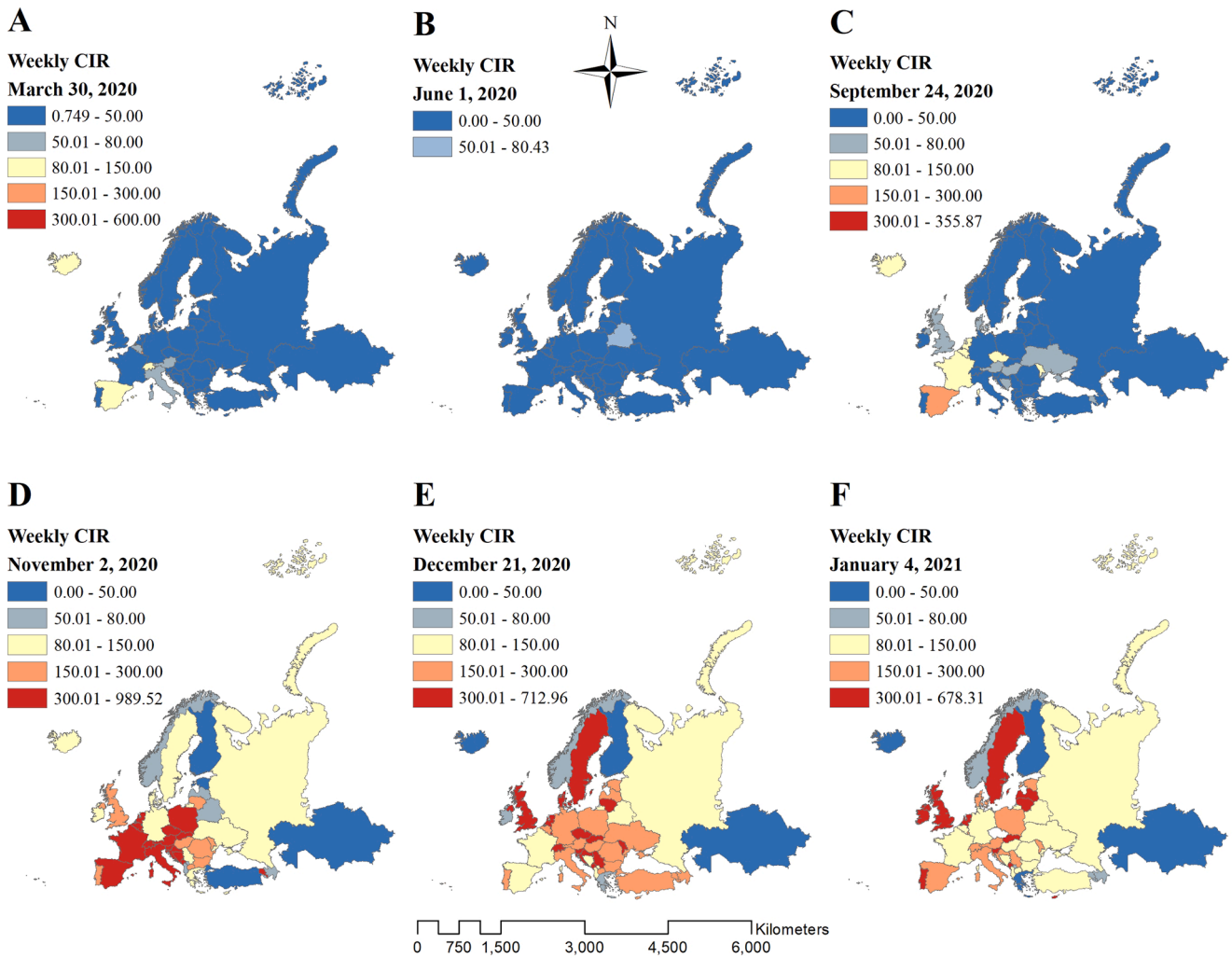Based on the "spatially-lagged dependent variable", the SLM model

Fig. 2. Weekly cumulative incidence rates (CIRs).

considers the spatial dependency between the response variable and explanatory variables. It delineates that in the neighborhood location, an independent variable could have impacts on the other independent variables (Wu et al., 2020). The SLM equation is as follows: (Ward & Gleditsch, 2018)

$$y_i = \beta_0 + x_i\beta + \rho W_i y_i + \varepsilon_i \qquad (4)$$

where ρ is the spatial autoregressive coefficient (spatial lag parameter).

2.3.4. Geographically Weighted Regression (GWR)

Traditional global regression techniques, namely OLS, SEM, and SLM, consider that the spatial relationship between explanatory variables and the dependent variable is stationary, meaning that the relationships are spatially constant across the study area (Brunsdon et al., 1996). Local GWR model was introduced by (Brunsdon et al., 1996) for relaxing this assumption and allowing the parameters to vary over space. Unlike global regression models, which produce a single regression equation to summarize global relationships (Kala et al., 2017), GWR detects spatial variation within relationships in a model and produces valuable information to explore and explain spatial non-stationarity

(Fotheringham et al., 2003). Thus, considering spatial context by GWR, this method predicts local regression parameters separately for all locations (Oshan et al., 2020). The GWR equation is as follows: (Fotheringham & Oshan, 2016)

$$y_i = \beta_{i0} + \sum_{j=1}^{m} \beta_{ij} X_{ij} + \varepsilon_i, \ i = 1, 2, 3, \ldots, n. \qquad (5)$$

where y_i represents the COVID-19 CIR value at the ith country, β_{i0} illustrates the local predicted intercept, β_{ij} denotes the jth regression parameter for the ith country, X_{ij} specifies the values of jth explanatory variables, and ε_i signifies a random error term. Parameter estimates at each country and for each independent variable in the form of a matrix is denoted by: (Fotheringham & Oshan, 2016)

$$\widehat{\beta}(i) = (X'W(i)X)^{-1}X'W(i)y \qquad (6)$$

where $\widehat{\beta}(i)$ represents the vector of parameter estimates $(v \times 1)$, X shows the selected independent variables in the form of matrix $(n \times v)$, $W(i)$ denotes the spatial weights in the form of matrix $(n \times n)$, and y demonstrates the vector observation of COVID-19 CIRs as the response
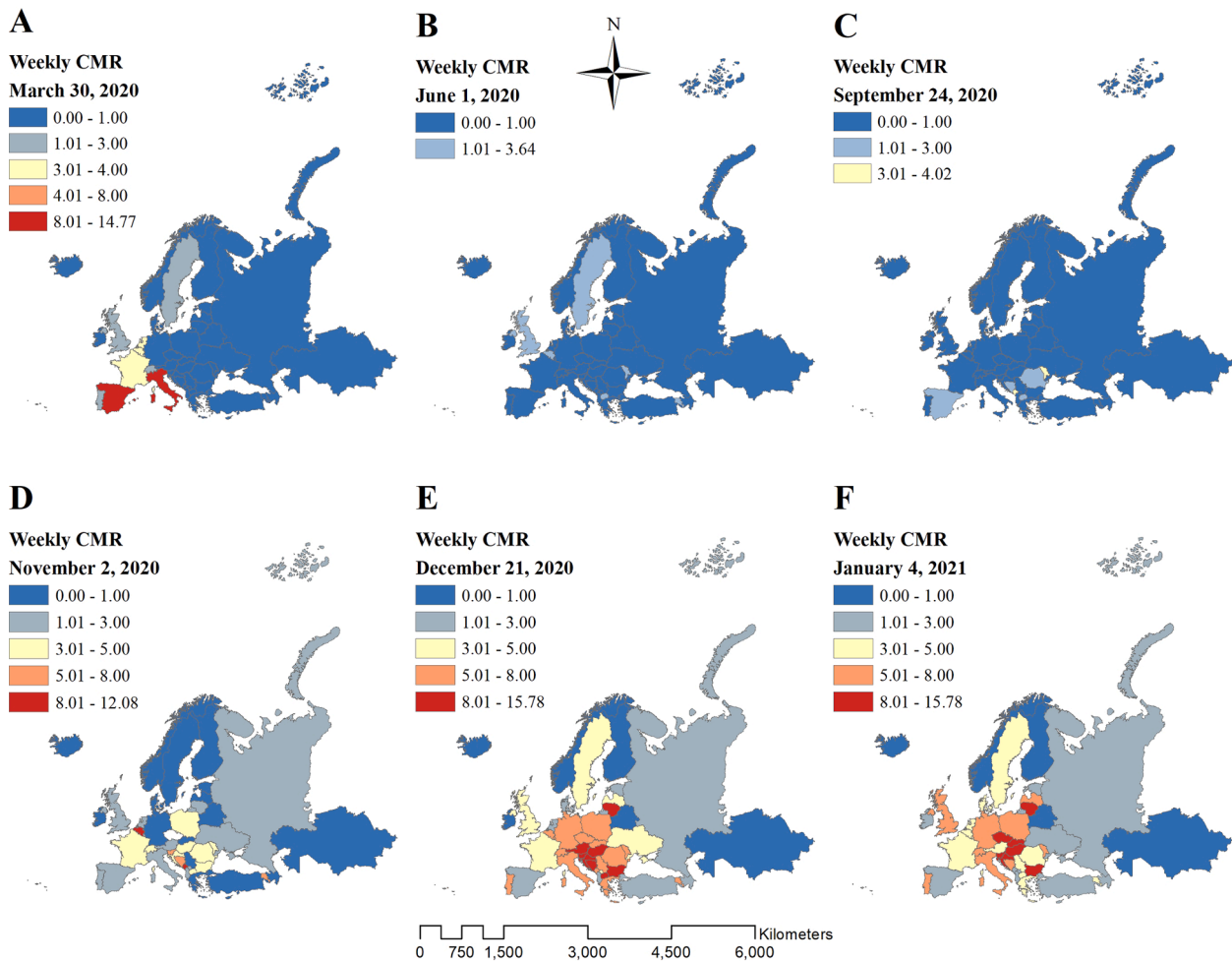
Fig. 3. Weekly cumulative mortality rates (CMRs).

variable ($v \times 1$) (Fotheringham & Oshan, 2016). The diagonal matrix of $W(i)$ is created by the weights of each spatial unit considering its distance from the ith location. The calibration of the matrix is based on a locally weighted regression (Brunsdon et al., 1998). A particular bandwidth and a kernel function need to be defined to compute $W(i)$ matrix. Gaussian and bisquare kernel functions are the most usual methods to perform the model calibration. Besides, the bandwidth is mostly examined based on the Euclidian distance and the number of nearest neighbor (Mollalo et al., 2020).

2.3.5. Multiscale geographically weighted regression (MGWR)

Compared to global methods, although GWR creates more benefits to the regression process in the context of geographic variations, it considers a fixed spatial scale for all the relationships in the modeling process. In the case of COVID-19, a constant spatial scale is not reliable due to the various spatial processes involved with several spatial scales (Fotheringham et al., 2017). While GWR restricts the relationships to vary at a constant spatial scale, MGWR allows the relationship between dependent and selected explanatory variables to vary at diverse spatial scales by applying different kernel bandwidths across the study area (Oshan et al., 2019). MGWR is denoted by: (Fotheringham et al., 2017)

$$y_i = \beta_{i0} + \sum_{j=0}^{m} \beta_{bwj} X_{ij} + \varepsilon_i, \ i = 1, 2, 3, \dots, n. \tag{7}$$

where β_{bwj} describes the bandwidth, which is used to calibrate the jth conditional relationship in the modeling process (Fotheringham et al., 2017).

Moreover, MGWR method can reduce collinearity and demonstrate the spatial heterogeneity more precisely (Wolf et al., 2018).

2.4. Regression modeling

To investigate which risk factors are associated with the COVID-19 CIRs, a variety of candidate explanatory variables are included to insert into the modeling process. Since there are a large set of candidate variables, a stepwise forward approach was carried out to remove the non-significant independent variables and obtain a model with the highest fitness. In doing so, Pearson's correlation analysis was implemented to identify the level of correlation between each pair of explanatory variables. In the final step, Variance Inflation Factor (VIF) was used to detect the most correlated variables with COVID-19 CIRs and measure the multicollinearity among the explanatory variables. Consequently, the most uncorrelated variables were identified as the
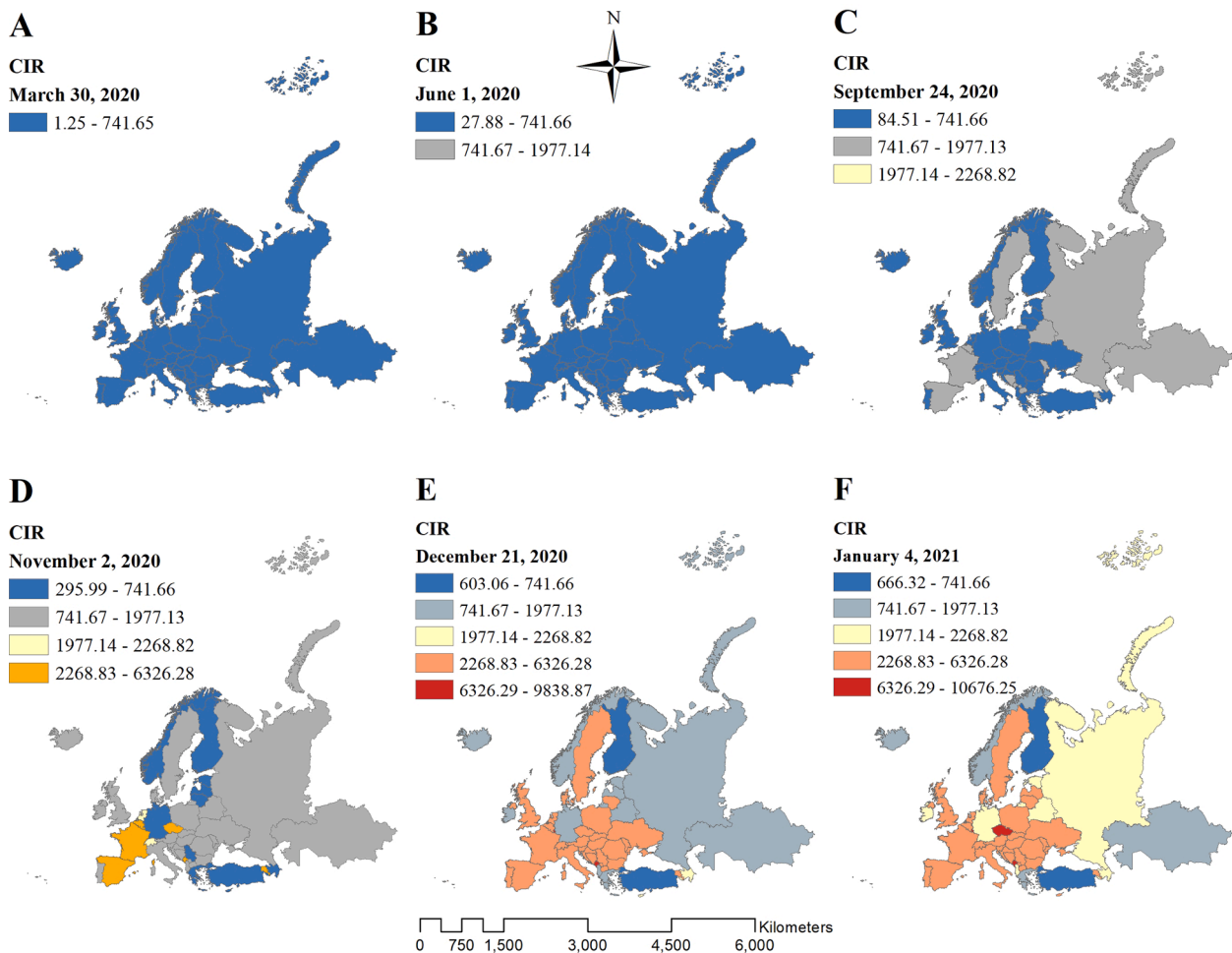
Fig. 4. Cumulative incidence rates (CIRs) from the beginning of the disease spread.

input of the regression analysis (final selected variables are starred in Table 1). A spatial weight matrix was implemented based on first-order Queens' contiguity, which determines whether countries are neighbors (by sharing a boundary) or not. This spatial weight matrix was applied to express the structure of spatial units and investigate the relationship between countries (Wang et al., 2020). Due to the existence of spatial autocorrelation within the local models, the spatial weight matrix is an essential part of these models (Brunsdon et al., 2002).

The selected explanatory variables remained unchanged in all models. Global models, namely OLS, SEM, and SLM, were implemented in GeoDa 1.14 software package (geodacenter.github.io). Furthermore, local models such as GWR and MGWR were run in MGWR 2.2 software (https://sgsup.asu.edu/sparc/mgwr). An adaptive bisquare kernel function with the specified bandwidth size was used to eliminate the impact of spatial units outside the neighborhood. Corrected Akaike Information Criterion (AICc) was used to designate the optimal bandwidth (Oshan et al., 2019). Moreover, the adjusted R^2 and AICc were examined to measure the model fit/performance in expressing COVID-19 CIRs across continental Europe. The higher value of adjusted R^2 and the lower value of AICc demonstrates that the model has a better performance.

Results

We used specific periods to study the spatial-temporal distribution of the COVID-19 incidence rates across Europe. These periods were March 30 (first outbreak peak), June 1 (minimum incidence after the first peak), November 2 (second outbreak peak), and December 21 (minimum infection after the second peak), respectively. Other periods were September 24 (second wave of the outbreak), January 4 (start of vaccination), and February 28 (the last time we used). In this study, we first prepared weekly maps of COVID-19 CIRs and CMRs. As shown in Fig. 2, A shows the first peak, and D shows the second peak of the disease spread. From the onset of the second wave to the second peak of the disease, in about a month, more than ten countries in Central and Western Europe were classified as red and orange countries (dangerous regions) (Fig. 2 C, D). In Fig. 2, F indicates the rate of infection after vaccine delivery. B and E also illustrate the minimum prevalence after the first and second COVID-19 peaks, respectively.

In addition to the CIRs, we also calculated and mapped the CMRs for the same periods (Fig. 3). Mortality rates rose before the vaccination in Europe, especially after November 2, 2020. However, after COVID-19 vaccination, there has been a declining trend, proved by the results obtained in this study (Figs. 3, 5).

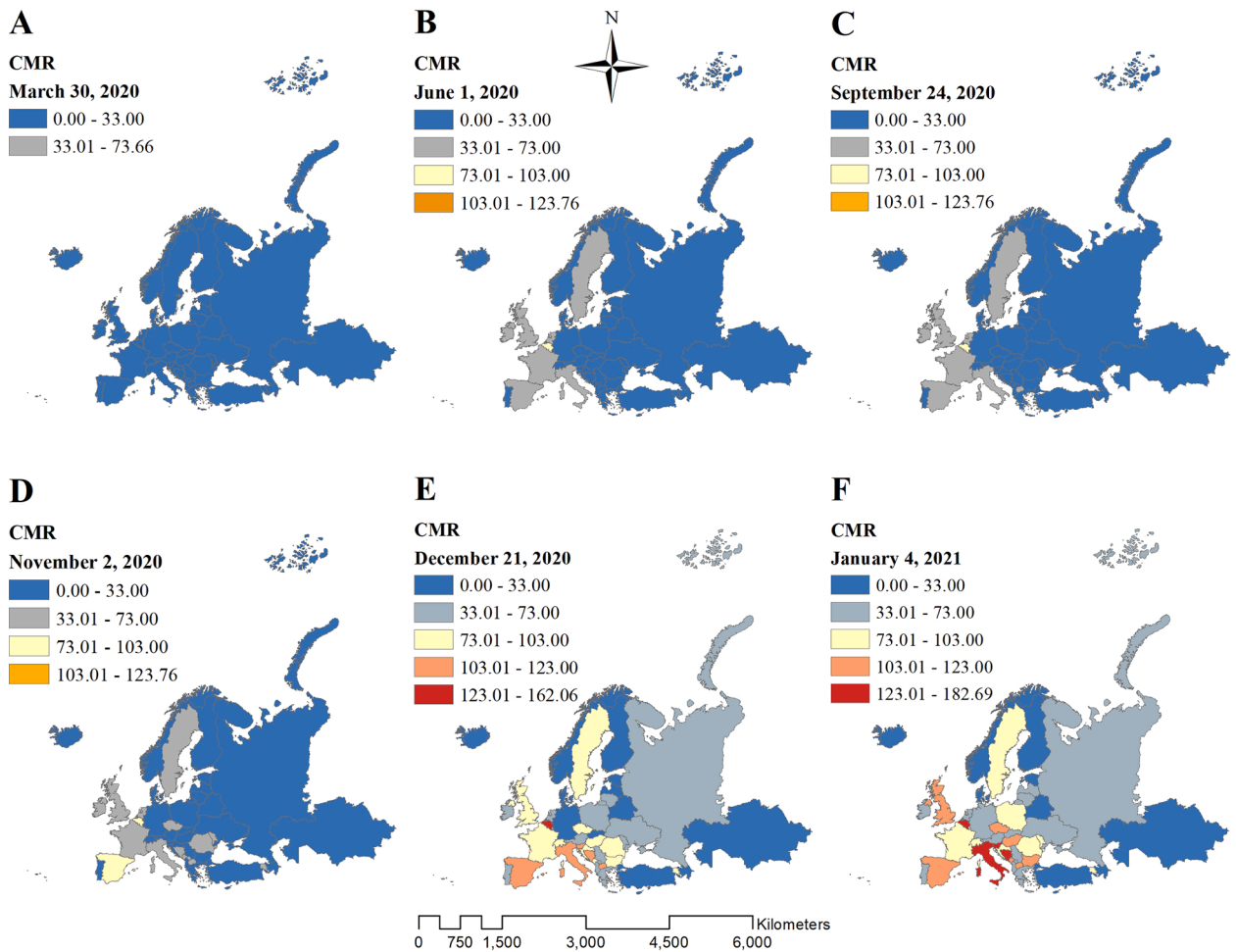In the cumulative spatial distribution map (since the beginning of the

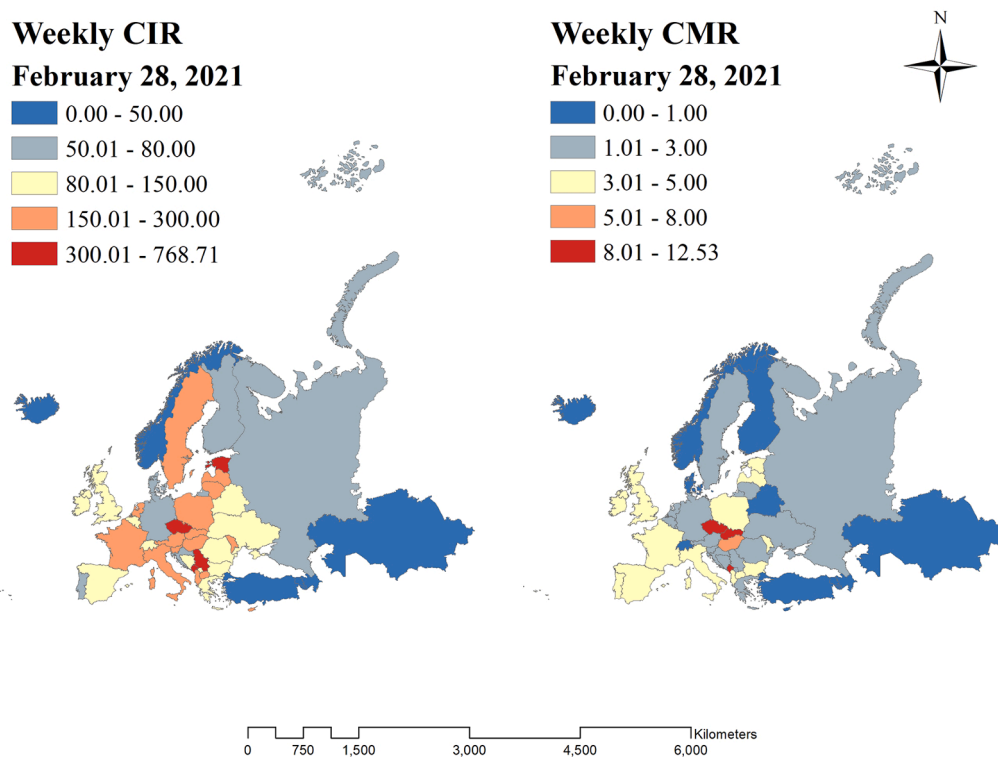**Fig. 5.** Cumulative mortality rates (CMRs) from the beginning of the disease spread.

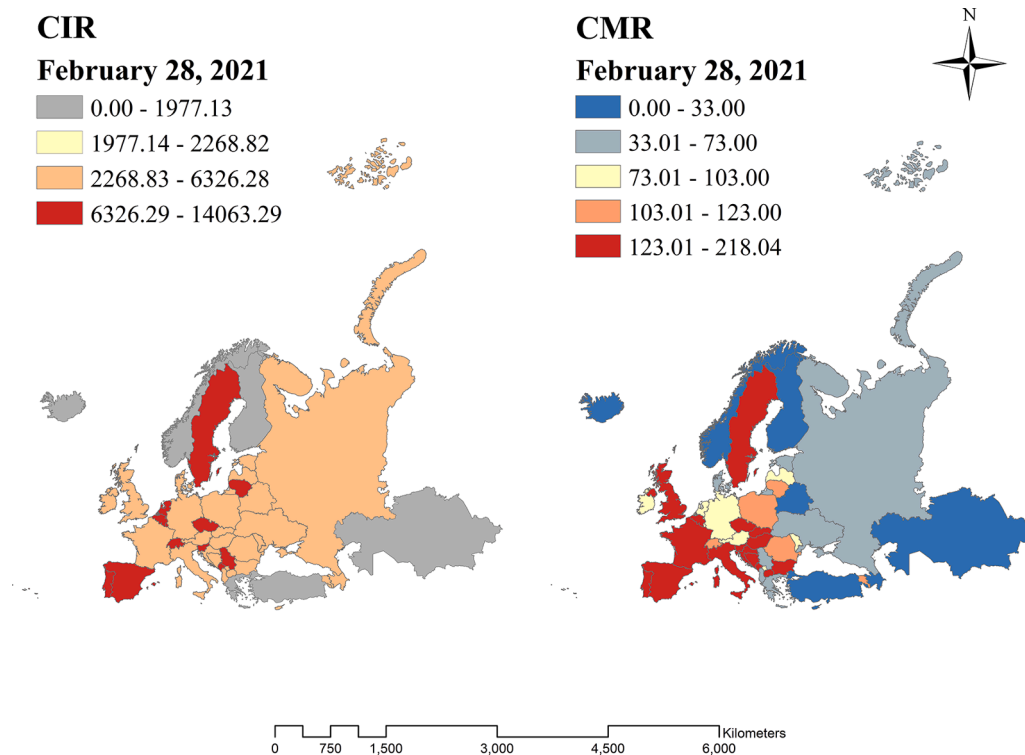**Fig. 6.** Weekly CIRs (left) and CMRs (right) of February 28, 2021.

Fig. 7. CIRs (left) and CMRs (right) of February 28, 2021, from the beginning of the disease spread.

outbreak in Europe), the trends of CIRs and CMRs are logically increasing, as can be seen in Figs. 4 and 5. The last date used in this study to investigate the spatial distribution trend of the CIRs and CMRs can also be seen in Figs. 6 (Weekly) and 7 (from the beginning of the disease spread). At a glance at all spatial distribution maps, it can be concluded that the CIRs and CMRs distribution is often concentrated in central and western Europe, but to ensure this, we examined the spatial distribution patterns of the disease both globally and locally (Fig. 7).

For this purpose, we applied Moran's I spatial autocorrelation (global test) and Hotspot analysis (local test) tools in the spatial statistics toolbox in the ArcMap for all designated periods. We entered the CIRs into the Global Moran's I analysis for all the seven periods we used in this study. According to the results, except September 24 (the second wave), all other periods showed a cluster pattern (Fig. 8). In other words, the pattern of COVID-19 CIRs distribution in Europe has been clustered since the beginning of the outbreak (Fig. 8, left column). Then, where these clusters are concentrated should be sought in the hotspot analysis. After performing this analysis on periods with a cluster pattern, the maps in Fig. 8 (right column) were obtained. The figures in the left column depict Global Moran's I results. As illustrated in Fig. 8, March 30 and June 1 demonstrated relatively similar clusters. As can be seen, in both results, the Moran's index shows numbers greater than zero (0.1205, 0.1087) with high levels of significance (p-values: 0.0007, 0.0013) (z-scores: 3.3861, 3.2055) for the mentioned periods. The figures in the right column show the hotspot results. Hotspot analysis divides polygons into seven categories. Three cold spots (negative z-scores) with a 99%, 95%, and 90% confidence level. Three hotspots (positive z-scores) with a level of 99%, 95%, and 90% confidence, and finally, one non-significant category, meaning that no cluster was found due to a high p-value. All these seven categories are shown in various color schemes. On November 2 and December 21, the patterns of spatial

distribution of CIRs were highly clustered (Moran's index: 0.0956, 0.1019) with the high levels of significance (p-values: 0.0073, 0.0062) (z-scores: 2.6790, 2.7326). In addition, the high-risk coronavirus clusters are mostly located in Western Europe and then in the Central regions of this continent. Moreover, relatively similar results were found on January 4 and February 28, so that the distribution pattern of the disease CIRs were clustered, having the Moran's I of 0.0945, 0.0896, z-scores of 2.5549, 2.3838, and the p-values of 0.0106, 0.0171, respectively.

After extracting the high-risk COVID-19 clusters and determining their geographical locations in the study area, the question is what variables have made the geographical location of the high-risk clusters usually located in central and Western Europe. Therefore, at this stage, we were looking to find these influential variables that have led to the creation of these hotspot clusters of COVID-19. For this purpose, we used 40 potential explanatory variables in this research. After examining the relationship between explanatory variables and dependent variable (correlation analysis) for the CIRs of COVID-19 for all periods whose spatial pattern had a cluster distribution, only two variables had a significant relationship with COVID-19 CIRs. These two variables were poverty and population aged over 65 at the country level. The spatial distributions of these two influential variables are shown in Fig. 9.

Therefore, we inserted these two selected variables into various global and local regression models. To compare these regression models in terms of their applicability in explaining the variability of COVID-19 CIRs, we considered the corrected AIC and adjusted R^2 of each of these models. The model that gives us a smaller AIC and a larger R^2 would be more efficient than other models. The chosen variables in the OLS method exhibit minimum multi-collinearity since their VIF is lower than the threshold of 5 (VIF = 1.21) (Table 2) (O'brien, 2007). Also, due to their low P value (< 0.001), these selected variables were highly
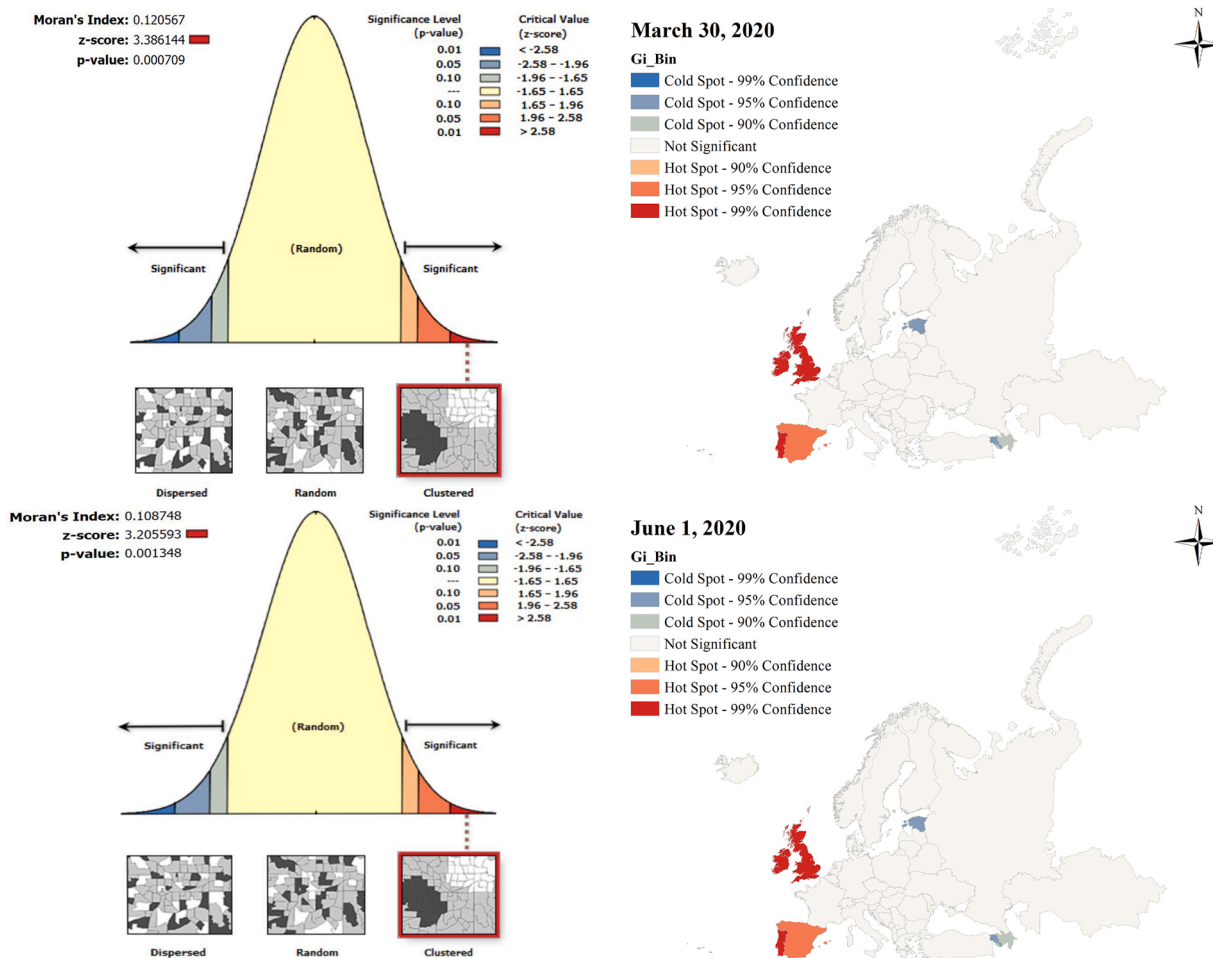
Fig. 8. Global Moran's I analysis (left column) and Hotspot analysis (right column).

associated with the disease CIRs. As illustrated in Table 2, the OLS method presented a poor performance due to the relatively low adjusted R^2 (52%) and a high AICc (208.48). Adjusted R^2 of 52% indicated that almost 0.48 of the disease incidence rates in the Europe continent are caused by unknown variables that might not be captured by the OLS model. To increase the efficiency of the OLS in the COVID-19 CIRs modeling, we applied two other global regression models, namely SLM and SEM to consider the spatial dependency.

As presented in Table 4, the adjusted R^2 values of SLM and SEM were 55% and 56%, respectively. Furthermore, the AICc decreased from 208.48 (obtained by the OLS) to 207.16 and 205.02 for SLM and SEM models, respectively. The higher adjusted R^2 and lower AICc indicate that the autoregressive models performed better in modeling COVID-19 CIRs. Moreover, the lag coefficients of both SLM and SEM, namely Lambda and Rho, were highly significant (P < 0.001) (Table 3). Noteworthy, the SEM outperformed SLM due to higher adjusted R^2 and lower AICc. However, performances of these autoregressive methods were relatively inefficient due to the non-consideration of the scale of geographical processes associated with COVID-19 CIRs modeling.

However, in the case of COVID-19, local regression techniques can produce better results by examining local spatial disparities in the relationships between explanatory variables and disease CIRs. Local

models, including GWR and MGWR, were applied for this purpose Table 4. shows that the value of adjusted R^2 boosted from 56% (SEM) to 62% in the GWR. Furthermore, the AICc considerably decreased from 205.02 to 97.329, demonstrating that the results derived from GWR model were strongly fitter than global models. Finally, by assuming different spatial scales, MGWR indicated the most efficient results due to the highest value of adjusted R^2 (69%) and lowest value of AICc (89.239) among all other regression models in this study.

Figs. 10 and 11 map the coefficients of both GWR and MGWR models for two selected explanatory variables. As illustrated in Fig. 10, in both GWR and MGWR methods, population over 65 showed similar patterns in explaining the spatial variation of disease incidence rates across Europe. In describing the spatial distribution of COVID-19 CIRs, the elderly population was a significant variable in central, southern, western, southwestern, and northern European countries, including Germany, Italy, France, United Kingdom, Ireland, the Netherlands, Belgium, Czech, Sweden, Norway, Austria, Switzerland, Denmark, Iceland, and so on. This range of positive coefficient values represents that older people are at higher risks of COVID-19 in mentioned regions. However, considering this variable, both local methods showed relatively poor performances in describing CIRs at Eastern Europe regions, including Russia, Turkey, and Georgia (Fig. 10).
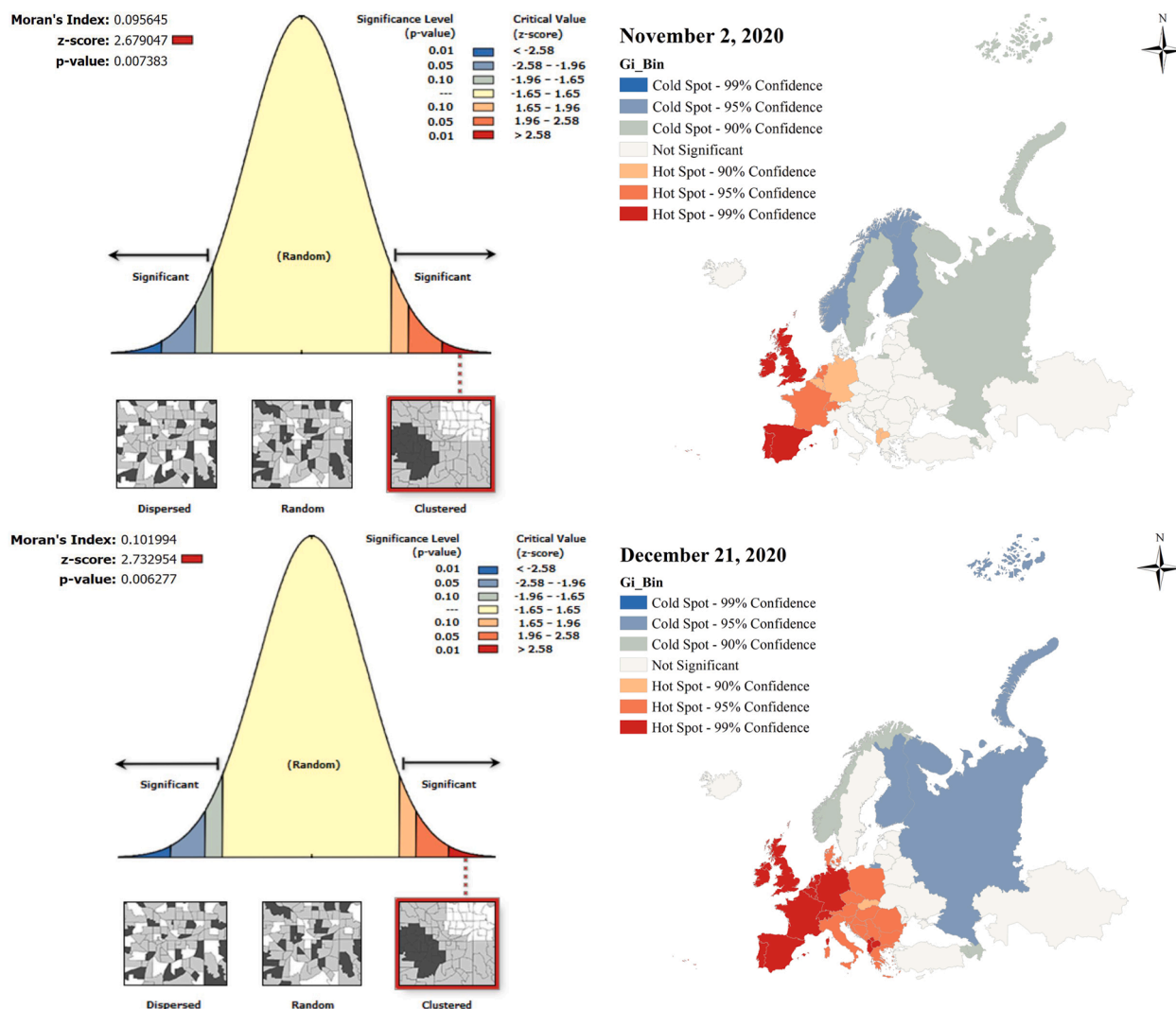
Fig. 8. (*continued*).

Fig. 11 indicates that poverty was another substantial variable in explaining the geographic distribution of disease incidence rates, specifically in western, central-western, northern, southern, and southwestern nations of continental Europe; Namely Spain, Portugal, France, Italy, Slovenia, Croatia, Bosnia and Herzegovina, Netherlands, Belgium, United Kingdom, Czech, Austria, Ireland, Iceland, and Norway. Like the previous variable, both local models represented weak performances in eastern areas of the study area. It is worth mentioning that the influence of both selected explanatory variables (poverty and elderly population) on disease incidence rates was often consistent between GWR and MGWR approaches (Figs. 10,11).

Fig. 12 demonstrates the geographic variation of local R^2 values, delineating the performances of GWR and MGWR models in terms of predicting the values of COVID-19 CIRs across Europe. Higher local R^2 values depict better performance of estimation. As observed, the values obtained from these local models are significantly inconsistent in the study area. MGWR offered highly better performance in all regions. As is illustrated by Fig. 12, considering MGWR, countries located in central,

western, southwestern, and northwestern of the Europe continent showed higher values of local R^2 compared to the GWR model. Iceland (0.7352), Ireland (0.7295), Portugal (0.7271), United Kingdom (0.7266), Spain (0.7251), France (0.7225), Andorra (0.7221), Belgium (0.7219), Netherlands (0.7208), Luxembourg (0.7199), were the top ten countries having the highest Local R^2 obtained from MGWR model. The highest Local R^2 obtained from GWR model were for Iceland (0.6593), Ireland (0.6499), Portugal (0.6455), United Kingdom (0.6454), Spain (0.6423), France (0.6384), Belgium (0.6378), Andorra (0.6377), Netherlands (0.6362), and Luxembourg (0.6346), respectively. It illustrates that the predictions obtained by MGWR model in the study area were relatively better.

Discussion

The COVID-19 pandemic has become one of the most dangerous and damaging crises in recent years, especially in Europe. As of June 24, 2021, the disease has claimed the lives of almost 4 million people
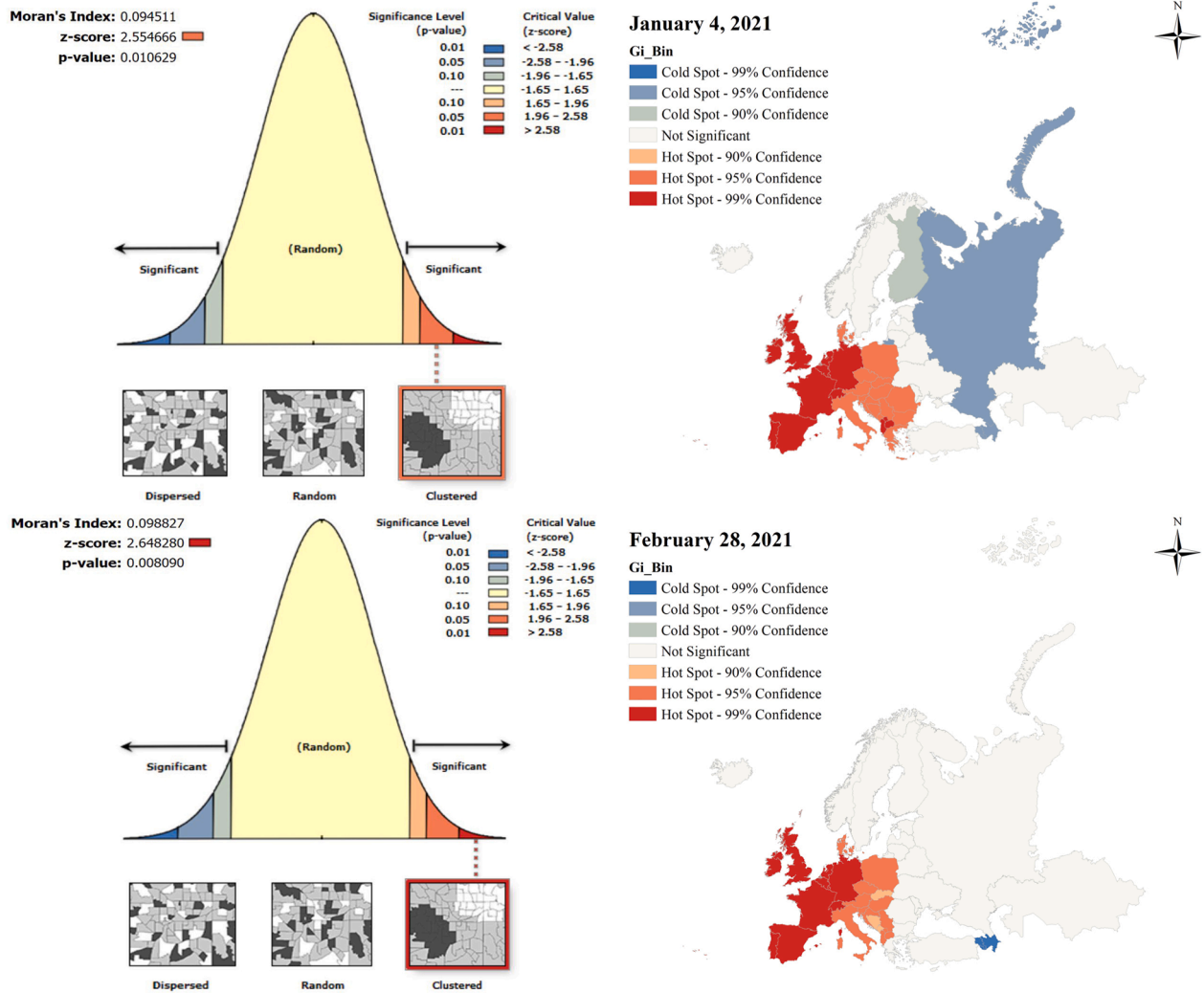
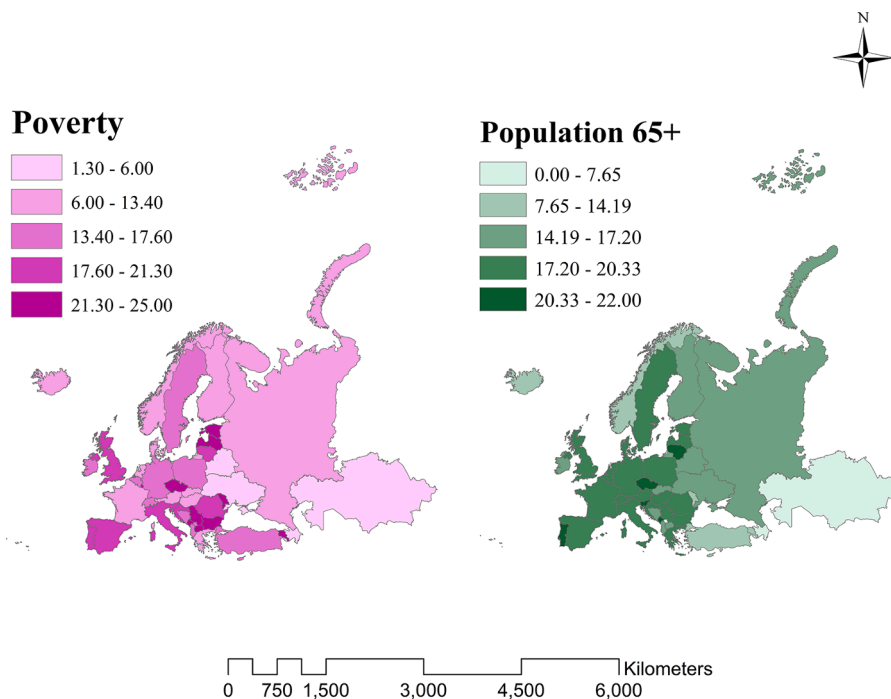**Fig. 8.** (*continued*).

Fig. 9. Spatial distribution of Poverty (left) and Population 65+ (right) in continental Europe (as the most influential variables on COVID-19 prevalence).

Table 2

Summary statistics of global OLS regression model.

Variable	Coefficient	St. Error	T-statistic	Probability	VIF
Intercept	-2.8556	1.2005	-2.3786	0.0331	—
Poverty	0. 2537	0.0537	4.7226	0.0000	1.2172
Population 65+	0.2307	0.0708	3.2583	0.0008	1.2172

worldwide and has left irreversible social, economic, and environmental impacts. Discovering the pattern of COVID-19 distribution and identifying the variables affecting the prevalence of the disease can help authorities create more targeted policies to reduce the speed of the outbreak. Using GIS-based techniques, we analyzed the spatial distribution pattern of the disease incidence and mortality in seven specific periods in continental Europe. Further, we specified hotspot regions of the disease in different periods. Additionally, by obtaining various potential variables, we applied five regression methods to identify the most influential variables on the disease spread and model the disease incidence rates across the continent.

Based on our study, by examining the weekly spatial-temporal distribution of COVID-19 incidence and mortality rates, the trend is increasing, and there are two peaks that are seven months apart (from March 30, 2020 (first peak) to November 2, 2020 (second peak)). Most of the high-risk regions and countries affected by COVID-19 were located in central and western Europe. Besides, poverty and the elderly population (the two most influential explanatory variables) had a significant relationship with the disease incidence rates in Europe. These two variables can well describe the spread of COVID-19 in Europe. Noteworthy, this result was consistent with hotspot clusters in central and western Europe, meaning that most countries with higher COVID-19 incidence rates were either relatively poorer or had older populations, namely Luxemburg, Switzerland, Slovenia, France, Italy, Liechtenstein, and Slovenia (Fig. 9). Monitoring these two influential variables can be helpful in improving the condition and reducing the disease incidence. In addition, among the regression methods applied to the variables, local methods provided more accurate results by inserting spatial scales in the modeling process. We found that MGWR was the most parsimonious model, which could better describe the geographical context of the disease incidence rates (Mansour et al., 2021).

Many articles considered the elderly population as one of the most influential variables (Mansour et al., 2021, Sun et al., 2020, Tian et al., 2020, Wang et al., 2020). For instance, by implementing various global and local regression models, (Mansour et al., 2021) showed that population over 65 was among the other selected variables that were found to be statistically significant regarding COVID-19 prevalence in Oman. Moreover, (Sannigrahi et al., 2020) presented that for both infected cases and associated deaths of COVID-19, higher records were identified in western Europe (Italy, Spain, France, Germany, United Kingdom, Belgium, and the Netherlands). They concluded that this result might be attributed to the sociodemographic variables of these regions. For

instance, Italy has the oldest population (aged 65 and above) among all European countries. Other countries, namely Spain, France, and the Netherlands, had been greatly influenced by this factor (Sannigrahi et al., 2020). Mentioned research also presented that socioeconomic variables, specifically poverty could be a significant variable in increasing the prevalence of coronavirus disease in Europe (Sannigrahi et al., 2020). Since poor people have less financial ability to access health centers than others, they may be present in the community without being treated or hospitalized and communicate with others, increasing the risk of disease transmission. Another hypothesis is that poor people, who are generally illiterate, are less likely to be vaccinated and underestimate the positive effects or overestimate the risks of vaccination, leading to a more severe outbreak of the disease in the community (Mollalo & Tatar, 2021). This variable also was among the potential determinants affecting higher prevalence and mortality of COVID-19 worldwide (Abedi et al., 2021, Bhayani et al., 2020, Ramírez & Lee, 2020, Cordes & Castro, 2020).

In this study, even though environmental variables were extracted and considered for continental European countries, we did not find any substantial relationship between such variables and COVID-19 spread. This result was also obtained by other studies (Briz-Redón & Serrano-Aroca, 2020, Baker et al., 2020).

However, there were a considerable number of papers reported that environmental factors, including temperature (Guo et al., 2011, Bashir et al., 2020), air quality (Zhang et al., 2020, Yao et al., 2020), and humidity (Matthew et al., 2021, Qi et al., 2020) could be among the factors influencing the more severe prevalence of COVID-19.

Following the vaccination process in Europe by February 15, the weekly incidence of COVID-19 dropped significantly. The weekly number of cases in this continent decreased from 1,900,152 on January 4 to 973,796 on February 15 (World Health Organization (WHO) 2021a). This weekly decrease occurred mainly in the United Kingdom, Ireland, Ukraine, Greece, and Romania. Besides, the weekly CMRs have risen significantly since the second outbreak wave (September 24, 2020). This trend continued until early February 2020, when the vaccination process received a great deal of attention in Europe. After vaccination, we saw a declining trend in both CIRs and CMRs in Europe, based on WHO situation by region table (World Health Organization (WHO) 2021a). Considering the CIRs from the beginning of the epidemic, it can be seen that the number of orange-colored countries has increased significantly since the second wave (September 24, 2020). These countries (high-risk category), which included only Andorra on

Table 4

Comparison of goodness of fit for OLS, SEM, SLM, GWR and MGWR regression models.

Criterion	OLS	SLM	SEM	GWR	MGWR
Adj. R^2	0.52	0.55	0.56	0.62	0.69
AICc	208.48	207.16	205.02	97.329	89.239

Table 3

Summary statistics of SEM and SLM regression models.

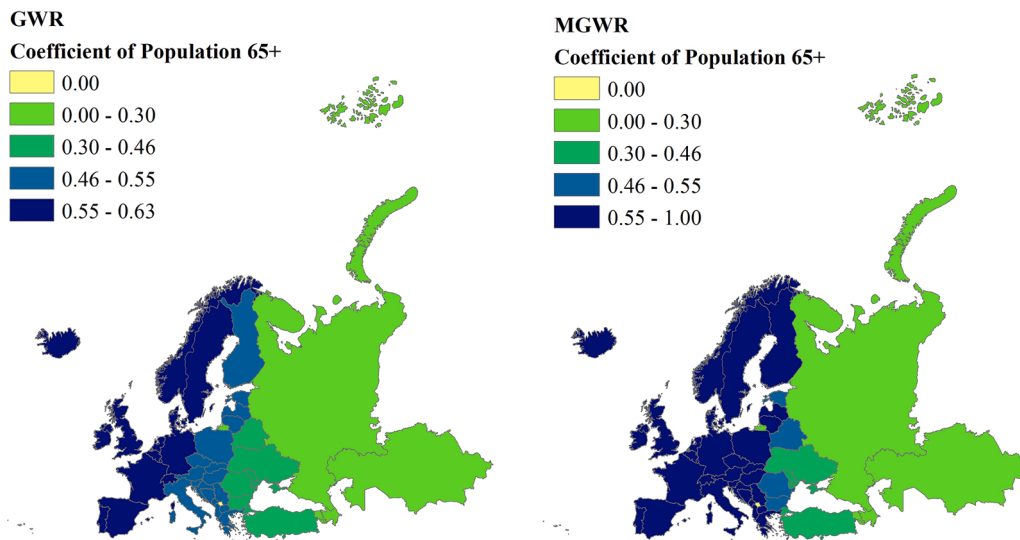| Variable	Coefficient		St. error		z-score		p-value	
	SLM	SEM	SLM	SEM	SLM	SEM	SLM	SEM
Intercept	-3.0482	-2.9315	1.1788	1.1870	-2.585	-2.4696	0.0097	0.0135
Poverty	0.2473	0.2484	0.0523	0.0520	4.7294	4.7744	0.0000	0.0000
Population 65+	0.2265	0.2420	0.0696	0.0709	3.2506	3.4118	0.0011	0.0006
Lag coefficient (Rho)	0.0790	—	0.1189	—	0.6647	—	0.0001	—
Lag coefficient (Lambda)	—	0.1462	—	0.1739	—	0.8407	—	0.0004

Fig. 10. The effects of Population aged 65+ in describing COVID-19 incidence rates using GWR (left) and MGWR (right) across continental Europe.
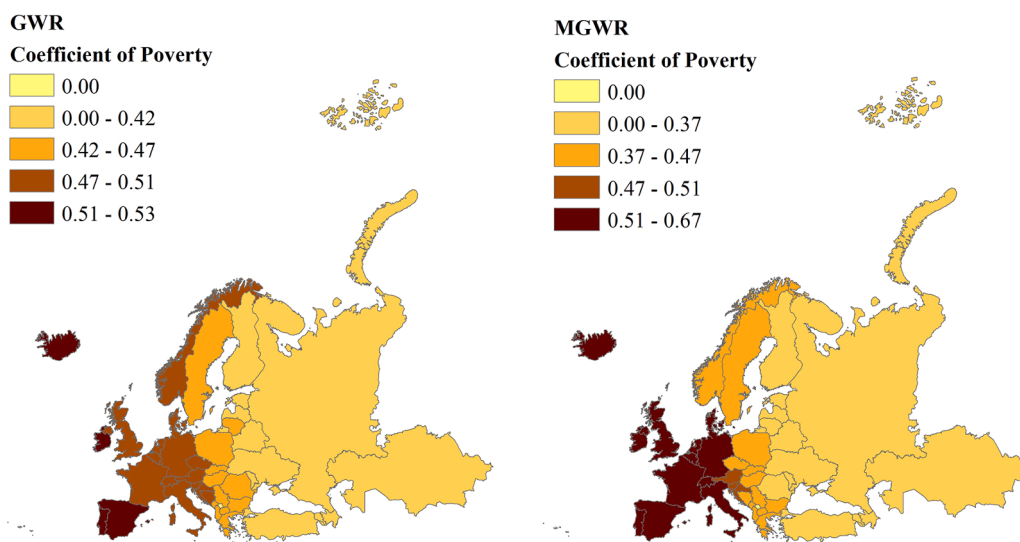

Fig. 11. The effects of Poverty in describing COVID-19 incidence rates using GWR (left) and MGWR (right) across continental Europe.

September 24, rose to seven countries in the second peak (November 2, 2020), including Andorra, Belgium, Czech Republic, Luxemburg, San Marino, Spain, and France. The number of these high-risk countries dramatically increased in the following periods. In all periods, high-risk clusters were located in central and western Europe. The spatial distribution patterns of COVID-19 incidence rates at March 30, and June 1 were highly clustered. In these two times, which show relatively similar patterns, the United Kingdom, Ireland, Spain, and Portugal were the hotspots of COVID-19. Over time, the number of high-risk countries increased, as shown in the maps. The distribution of COVID-19 clusters, as in previous times, was located mainly in western and central Europe. Northern European countries, including Finland, Norway, and Russia, were among the cold spot areas of the disease on November 2, 2020, December 21, 2020, and January 4, 2021. In general, the Nordic countries were in the low-risk clusters at the time of the study, and the countries in western and central Europe were hotspot regions with the highest levels of confidence (Sannigrahi et al., 2020).

Due to the large scale considered in this study, data availability was one of the most significant limitations. Moreover, due to the particular circumstances of some countries and the lack of accurate information sharing, access to accurate disease data and other information about explanatory variables faced some difficulties, which may bias the results. Although we tried to display and model the COVID-19 situation for all countries simultaneously, a study on a higher spatial resolution (sub-country level) can provide more accurate results. Another limitation of this study was the lack of considering each country's lockdown policies. Some countries started their quarantine policies faster and more strictly after the epidemic announcement than other countries that did not consider any specific restrictions. Furthermore, though we did not consider any dynamic data as potential variables in our modeling, they need to be included in future research. For instance, (Kraemer et al., 2020) showed that real-time human mobility data could be an
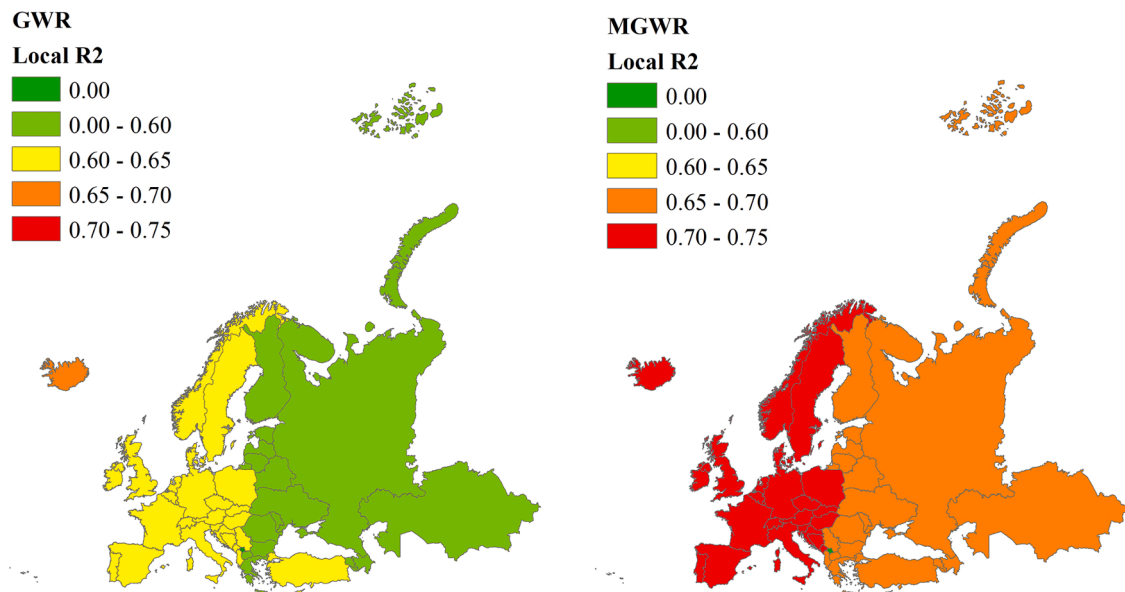
Fig. 12. Spatial distribution of Local R2 of GWR and MGWR models for COVID-19 incidence rate associated with Population aged 65+ and Poverty.

influential factor in spreading COVID-19. Besides, including individual characteristics and considering a smaller group of infected people in the modeling process could improve the results (Kwan, 2012).

Conclusion

Currently, despite quarantine-related restrictions and policies imposed by various countries in Europe, the COVID-19 is still widespread. Thus, spatial techniques can be useful in identifying various spatial distribution patterns and hotspots of the disease and detecting the most significant risk factors. By applying spatial analysis methods, we found that the most high-risk areas of the disease are located in central and western Europe. Besides, poverty and the elderly population were the most influential factors related to higher COVID-19 prevalence. MGWR could explain the highest goodness-of-fit among all the applied regression models, indicating the most parsimonious model. Given that the countries of continental Europe were among the most COVID-19 infected areas throughout the world, trying to control the factors influencing the spread of the disease can lead to a considerable reduction in near future. Moreover, the outcomes of this study can become helpful for policymakers to take necessary actions in combating the COVID-19 pandemic.

Declaration of Competing Interest

None

References

World Health Organization (WHO), 2020a. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) aAvailable from. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.

World Health Organization (WHO), 2021a. WHO Coronavirus (COVID-19) Dashboard aAvailable from. https://covid19.who.int/.

Kianfar, N., et al., 2021. Spatio-temporal modeling of COVID-19 prevalence and mortality using artificial neural network algorithms. Spatial and Spatio-temporal Epidemiology, 100471.

Newyork Post, L.B., 2020. Europe's first coronavirus case was in December: scientists. May 5Available from. https://nypost.com/2020/05/05/europes-first-coronavirus-case-was-in-december-scientists/.

World Socialist WebSite, 2020. A.L. *Europe's COVID-19 pandemic spirals out of control.* November 3Available from. https://www.wsws.org/en/articles/2020/11/03/pers-n03.html.

Sannigrahi, S., et al., 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 p.

Dye, C., et al., 2020. The scale and dynamics of COVID-19 epidemics across Europe. R. Soc. Open Sci. 7 (11), 201726.

Miller, L.E., Bhattacharyya, R., Miller, A.L., 2020. Spatial analysis of global variability in Covid-19 burden. Risk Management and Healthcare Policy 13, 519.

Mollalo, A., et al., 2018. Machine learning approaches in GIS-based ecological modeling of the sand fly Phlebotomus papatasi, a vector of zoonotic cutaneous leishmaniasis in Golestan province, Iran. Acta Trop. 188, 187–194.

Mollalo, A., Vahedi, B., Rivera, K.M., 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci. Total Environ. 728, 138884.

World Health Organization (WHO), 2021b. Coronavirus Disease (COVID-19) Dashboard bAvailable from. https://covid19.who.int/. </Dataset>.

Iyanda, A.E., et al., 2020. A retrospective cross-national examination of COVID-19 outbreak in 175 countries: a multiscale geographically weighted regression analysis (January 11-June 28, 2020). Journal of infection and public health 13 (10), 1438–1445.

Moran, P.A., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23 (1/2).

Ord, J.K., Getis, A., 2001. Testing for local spatial autocorrelation in the presence of global autocorrelation. Journal of regional science 41 (3), 411–432.

Getis, A., Ord, J.K., 2010. *The analysis of spatial association by use of distance statistics*, in *Perspectives on spatial data analysis*. Springer, pp. 127–145.

Oshan, T.M., et al., 2019. mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. ISPRS International Journal of Geo-Information 8 (6), 269.

Ward, M.D., Gleditsch, K.S., 2018. Spatial regression models, Vol. 155. Sage Publications.

Goodchild, M.F., B.O. Parks, and L.T. Steyaert, *Environmental modeling with GIS*.1993.

Wu, Z., et al., 2020. Identifying the influencing factors controlling the spatial variation of heavy metals in suburban soil using spatial regression models. Sci. Total Environ. 717, 137212.

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical analysis 28 (4), 281–298.

Kala, A.K., et al., 2017. A comparison of least squares regression and geographically weighted regression modeling of West Nile virus risk based on environmental parameters. PeerJ 5, e3070.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons.

Oshan, T.M., Smith, J.P., Fotheringham, A.S., 2020. Targeting the spatial context of obesity determinants via multiscale geographically weighted regression. International journal of health geographics 19, 1–17.

Fotheringham, A.S., Oshan, T.M., 2016. Geographically weighted regression and multicollinearity: dispelling the myth. Journal of Geographical Systems 18 (4), 303–329.

Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically weighted regression. Journal of the Royal Statistical Society: Series D (The Statistician) 47 (3), 431–443.

Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale geographically weighted regression (MGWR). Annals of the American Association of Geographers 107 (6), 1247–1265.

Oshan, T., et al., 2019. A comment on geographically weighted regression with parameter-specific distance metrics. International Journal of Geographical Information Science 33 (7), 1289–1299.

Wolf, L.J., Oshan, T.M., Fotheringham, A.S., 2018. Single and multiscale models of process spatial heterogeneity. Geographical Analysis 50 (3), 223–246.

Wang, A.-L., Zhong, X., Hurd, Y., 2020. Comorbidity and sociodemographic determinants in COVID-19 mortality in an US urban healthcare system. medRxiv.

Brunsdon, C., Fotheringham, A., Charlton, M., 2002. *Geographically weighted summary statistics—a framework for localised exploratory data analysis.* Computers. Environment and Urban Systems 26 (6), 501–524.

O'brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41 (5), 673–690.

Mansour, S., et al., 2021. Sociodemographic determinants of COVID-19 incidence rates in Oman: Geospatial modelling using multiscale geographically weighted regression (MGWR). Sustainable cities and society 65, 102627.

Sun, P., et al., 2020. Understanding of COVID-19 based on current evidence. J. Med. Virol. 92 (6), 548–551.

Tian, S., et al., 2020. Characteristics of COVID-19 infection in Beijing. J. Infect. 80 (4), 401–406.

Wang, C., et al., 2020. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int. J. Environ. Res. Public Health 17 (5), 1729.

Mollalo, A., Tatar, M., 2021. Spatial Modeling of COVID-19 Vaccine Hesitancy in the United States. Int. J. Environ. Res. Public Health 18 (18), 9488.

Abedi, V., et al., 2021. Racial, economic, and health inequality and COVID-19 infection in the United States. Journal of racial and ethnic health disparities 8 (3), 732–742.

Bhayani, S., et al., 2020. Dialysis, COVID-19, poverty, and race in greater Chicago: an ecological analysis. Kidney medicine 2 (5), 552–558 e1.

Ramírez, I.J., Lee, J., 2020. COVID-19 emergence and social and health determinants in Colorado: a rapid spatial analysis. Int. J. Environ. Res. Public Health 17 (11), 3856.

Cordes, J., Castro, M.C., 2020. Spatial analysis of COVID-19 clusters and contextual factors in New York City. Spatial and Spatio-temporal Epidemiology 34, 100355.

Briz-Redón, Á., Serrano-Aroca, Á., 2020. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. Sci. Total Environ. 728, 138811.

Baker, R.E., et al., 2020. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. Science 369 (6501), 315–319.

Guo, Y., et al., 2011. The impact of temperature on mortality in Tianjin, China: a case-crossover design with a distributed lag nonlinear model, 119. Environmental health perspectives, pp. 1719–1725.

Bashir, M.F., et al., 2020. Correlation between climate indicators and COVID-19 pandemic in New York, USA. Sci. Total Environ. 728, 138835.

Zhang, Z., Xue, T., Jin, X., 2020. Effects of meteorological conditions and air pollution on COVID-19 transmission: Evidence from 219 Chinese cities. Sci. Total Environ. 741, 140244.

Yao, Y., et al., 2020. Association of particulate matter pollution and case fatality rate of COVID-19 in 49 Chinese cities. Sci. Total Environ. 741, 140396.

Matthew, O.J., Eludoyin, A.O., Oluwadiya, K.S., 2021. Spatio-temporal variations in COVID-19 in relation to the global climate distribution and fluctuations. Spatial and Spatio-temporal Epidemiology 37, 100417.

Qi, H., et al., 2020. COVID-19 transmission in Mainland China is associated with temperature and humidity: a time-series analysis. Sci. Total Environ. 728, 138778.

Kraemer, M.U., et al., 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 368 (6490), 493–497.

Kwan, M.-P., 2012. The uncertain geographic context problem. Annals of the Association of American Geographers 102 (5), 958–968.

The World Bank. *World Bank Open Data.* February 1, 2021; Available from: https://data.worldbank.org/. </Dataset>.

Giovanni. *NASA/GSFC, Greenbelt, MD, USA, NASA Goddard Earth Sciences Data and Information Services Center (GES DISC).*March 1, 2021; Available from: https://giovanni.gsfc.nasa.gov/. </Dataset>.