SPATIAL IMPACT OF COVID-19 IN TURKEY

Place-based factors affecting COVID-19 incidences in Turkey

Mehmet Ronael1· Tüzin Baycan1

Received: 21 March 2022 / Accepted: 1 September 2022 / Published online: 10 September 2022
© The Japan Section of the Regional Science Association International 2022

Abstract
In December 2019, COVID-19 infections first occurred in Wuhan City, China, after which it rapidly spread throughout the world. Today, COVID-19 has become a major disaster affecting countries physically, socially, and especially economically. However, reasons behind the spread of COVID-19 are still unclear. Therefore, many scholars from different disciplines try to understand the various leading indicators. Our study aimed to reveal place-based factors affecting COVID-19 incidences in Turkey while addressing and analyzing a set of indicators (physical, natural, economic, demographic, and mobility based) within the scope of the recent research findings in the literature on the COVID-19 Pandemic. Following this purpose, we addressed 81 provinces of Turkey using city-level data obtained from the Ministry of Health, and employed global and local regression methods through ArcGIS and GeoDa: Ordinary Least Square, Spatial Lag Model, Spatial Error Model, and Geographically Affected Weighted Regression to highlight place-based factors affecting the spread of the Pandemic. The results of our analyses demonstrated that three factors: (1) population density, (2) annual temperature, and (3) health capacity; are related to the COVID-19 incidences in Turkey. Our results also demonstrated that the impact of these factors causes varying spatial effects within the country, especially in the West–East direction. Although these results provide a base for future studies, COVID-19 is still spreading with several mutations. Therefore, the reliability of produced models and the effectiveness of factors should be retested using new and updated data for cities and at other geographical scales.

Keywords COVID-19 pandemic · Place-based factors affecting pandemic · Geographical dimensions · Ordinary least square regression · Geographically weighted regression · Turkey

JEL Classification C31 · C38 · C88 · R58

* Mehmet Ronael
mehmetronael@gmail.com

1 Department of Urban and Regional Planning, Faculty of Architecture, Istanbul Technical University, Istanbul, Turkey
1 Introduction

In December of 2019, the COVID-19 disease first occurred in Wuhan city, China, and rapidly spread across the globe. On March 11, 2020, WHO (World Health Organization) announced a pandemic, and as of March 11, 2021, more than 6,000,000 deaths and almost 450,000,000 incident cases have been officially recorded worldwide. Today, the COVID-19 Pandemic is one of the most critical disasters that humanity has faced and it has affected countries physically, socially, and especially economically (United Nations, 2020). According to data revealed by the Ministry of Health in Turkey, the COVID-19 Pandemic appeared in Turkey at the beginning of March 2020; it then rapidly spread across the country through metropolitan cities, such as Istanbul, Ankara, and Izmir. In particular, Istanbul, with its high total population and population density, has become the Wuhan of Turkey (Bahçetepe 2021). It is clear that COVID-19 can be transmitted easily across cities or regions, and researchers from different disciplines have investigated the reasons behind this dilemma (Franch-Pardo et al. 2020).

Although the reasons for variation in rate of spread of the COVID-19 pandemic remain somewhat unclear, many studies around the world have found that geo-environmental factors (Coccia 2020), demographic characteristics (Mollato et al. 2020; Arbel et al. 2021), and climatic conditions (Ma et al. 2020; Oto-Peralías 2020; Wu et al. 2020; Rios and Gianmoena 2021) can affect the severity and spread of COVID-19. Some researchers have also focused on the relationship between health and social geography (De Kadt et al. 2020; Gibson and Rush, 2020; Kuupiel et al. 2020; Hierro and Maza 2022). Additionally, some have made interdisciplinary correlations between COVID-19 and social behavior (Allcot et al. 2020; Kuchler et al. 2020; Martinho 2021; Boumahdi et al. 2021; Cutrini and Savati 2021; Florida and Mellander 2022).

While the COVID-19 Pandemic has become a hot topic in the academic community, GIS (Geographic Information System) has also become a vital method to observe the spatial distribution of infectious diseases. GIS is already a helpful resource for preventing an epidemic and enhancing the potential of care (Mollalo et al. 2018, 2019; Xiong et al. 2020). Guan et al. (2020) first used the spatial statistics of GIS to understand relations between patients’ demographic characteristics and the level of COVID-19. After these first geographic analyses provided a base and methodology for others, many researchers started to reproduce them for other countries on different scales, such as USA (Dong et al. 2020), South Korea (Rezaei et al. 2020), and Spain (Orea and Álvarez, 2020). Therefore, it is clear that GIS has become an essential tool for analyzing, capturing, and visualizing the COVID-19 epidemic. It profoundly affects decision-making in daily life and provides a predictive model for the onset of disease (Franch-Pardo et al. 2020).

The literature also contains some studies on the relationship between COVID-19 and spatial parameters in Turkey. Döker and Ocak (2020) discussed using GIS to monitor the geographic extent of the COVID-19 pandemic in Turkey. Zeren and Vildancı (2020) used hotspot analysis and the LISA model to understand the existing spatial pattern of the pandemic in Turkey. In addition, Zeren et al. (2021)
evaluated the transmission of virus and mobility factors between sub-regions using the Global and Local Moran’s I Indexes. Moreover, Istanbul Development Agency developed the vulnerability map of Istanbul using physical, economic, social, and environmental variables. However, no Turkish study has evaluated the impacts of spatial variables on the spread level of COVID-19. Therefore, this study aims to understand and analyze the association between determining physical, natural, economic, demographic, and mobility factors on Turkey’s pandemic within the scope of recent developments and information.

Turkey and its 81 provinces (both cities and rural areas) were chosen as a case area following this purpose. The dataset was also prepared at the province level and obtained from open-source websites. In the dataset, the dependent variable is the COVID-19 incidence, revealed by the Ministry of Health on 2 April 2021. The fundamental reason to set this date is that vaccination is not yet widely available in the community, and new variants have not begun to dominate the pandemic. On the other hand, explanatory variables consist of 20 different demographic, economic, physical, natural, and mobility indicators. After data wrangling, independent variables were examined in terms of their significance and multicollinearity in the correlation through explanatory regression analysis. Finally, the global and local regression models (OLS, SLM, SEM, and GWR) were employed to analyze their statistical and spatial relationship in ArcGIS and GeoDa. The results are intended to present the first attempt to use regional geographic modeling in COVID-19 studies throughout Turkey and may provide policy-makers with valuable information for targeted intervention.

The structure of the paper comprises five main sections. Section two summarizes the place-based factors affecting the spread of the COVID-19 Pandemic on different scales through the studies found in the literature. In the third section, the research methodology is explained step by step. This section also contains explanations and details about the process of acquiring data and for Moran’s I Index, Explanatory Regression Analysis, Ordinary Least Square, Spatial Lag Model, Spatial Error Model, and Geographically Weighted Regression. In section four, the results for each analysis type are explained separately. In the last section, variables associated with the spread of COVID-19 are also discussed in reference to the literature, and the paper is concluded by highlighting the study’s primary results, constraints, and potential for future studies.

2 Place-based factors affecting the propagation of COVID-19

As mentioned before, many researchers from different countries have concentrated on the influence of spatial variables on the pandemic level. The first paper, published on 28 February 2020, explained the distribution of patients by province in China and their characteristics such as age, gender, symptoms, etc. (Guan et al. 2020). Following that, Chen et al. (2020) investigated the effect of migration between provinces in China; one of this study’s notable results proving a significant correlation between migration and patient numbers. Gross et al. (2020) and Zhang et al. (2020) also worked on the ease of human mobility, short distance between provinces, and virus propagation in China.
Similarly, Hierro and Maza (2022) worked on Madrid’s inter-district mobility (and other socioeconomic factors) and its effect on disease propagation. Ignoring inter-city dynamism, Bogoch et al. (2020) underlined the global spread potential, examining the impact of air passenger transport on the global dispersion of the COVID-19 Pandemic. In short, all of these studies found the same result as the previous ones, the existence of a strong correlation between mobility and virus transmission.

There are also some studies which have focused on health geography topics specifically. De Kadt et al. (2020) explored the difficulties of providing health control measures, and poverty in South Africa. They prepared two maps, one of which is the risk factor index based on six risk determinants that can be recognized as obstacles to realizing basic hygiene and social distancing: overcrowded living circumstances, distribution of hygiene and water usage, dependency on public health services, low accessibility to communication mechanisms, and reliance on public transportation. On the other hand, the second one represents the vulnerability of neighborhoods based on their risk level. Again, in South Africa, Gibson and Rush (2020) evaluated the poverty, population density, and lack of infrastructure related to the vulnerability of settlements to COVID-19. On an international level, Padula and Davidson (2020) investigated the relation of the number of nurses to the mortality ratio of COVID-19 disease across countries, and they emphasized the negative correlation between these factors. Similarly, Lakhani (2020) conducted spatial research for Queensland, Australia, pointing out the strong relationship between the access difficulties of people over 65 years to health services and the disease’s mortality rate. Finally, Florida and Mellander (2022) also investigated some place-based indicators and socioeconomic factors on the neighborhood scale in Sweden. They found that closeness to locations with higher risks of diseases, such as hospitals, was the most significant factor.

Other studies also built on interdisciplinary associations. Allcott et al. (2020) investigated the connection between cases of COVID-19, the community’s negative attitude toward quarantine, and the dominant political party in each US state. The results show an exact relationship between provinces with lower social distancing reactions and higher rates of Republican voters. Strong reactions against quarantine were also detected in areas with higher official numbers of COVID-19 patients. Moreover, Kuchler et al. (2020) analyzed the correlation between the social connectivity index based on Facebook friendship links and COVID-19 patients in two regions significantly affected by the disease. Their results underline that a social connectivity index could support epidemiologists in predicting the extent of infectious disease.

Expecting to find correlations between the physical environment, geography, socioeconomic factors, and COVID-19 is another motivation behind the pandemic studies. Coccia (2020) analyzed the relationship between the geo-ecological and demographic characteristics of 55 Italian cities and the prevalence of COVID-19. In this research, he categorized the variables in each city: distance from sea, latitude, population density, air pollution (PM10), and climate indicators (average temperature, relative humidity, and predominant wind speed), finding a link between stimulated COVID-19 and high levels of air pollution, weather conditions, and cities far from the sea. Sridhar (2021) also set out to understand the impact of urbanization
on COVID-19 prevalence in India. She determined socioeconomic factors of different levels for 600 districts of India. The results show that the high rate of urbanization, higher participation in the workforce, and higher income level contributed to increasing COVID-19 cases. Furthermore, while there is a strong positive correlation between urban poverty, slum areas, and the number of registered patients, there is a negative relationship between health services, open spaces, and COVID-19.

Similarly, Oto-Peralías (2020) worked in Spanish provinces, determining several geographical and socioeconomic indicators to describe the distinct differences in infected people between the regions. The results show significant correlation between temperature and distance from Madrid and COVID-19, suggesting that increased temperature and distance from Madrid adversely affect COVID-19 while reducing its contagion efficiency. Mollalo et al. (2020) also presented 35 explanatory determinants based on socioeconomic, behavioral, physical character, natural character, and demographic structure. They tested these factors through global and local spatial regression analysis. Results show that determinants related to the physical environment do not substantially impact the number of COVID-19 patients. However, four indicators (average household income, income imbalance, percentage of nurses, and percentage of black women) may define increased disease rate instability in the United States.

Many scholars have continued to use several global and local spatial regression techniques in their studies to test various place-based factors impacting the spread of COVID-19. Sannigrahi et al. (2020), Urban and Nakada (2021), Rahman et al. (2021), and Mansour et al. (2021) employed Ordinary Least Squares (OLS), Spatial Lag, and Spatial Error Regression models (SLM, SEM), and Multiscale Geographically Weighted Regression (MGWR) and Geographically Weighted Regression (GWR) as local models. Sannigrahi et al. (2020) evaluated the spatial relationship between 28 sociodemographic variables and COVID-19 deaths and cases in 31 European countries. They found total population, poverty, and income are key factors impacting the overall deaths caused by COVID-19. Urban and Nakada (2021) also focused on 18 demographic and socio-environmental variables, and they compared the efficiency of regression models to explain the spatial association between variables. Similarly, Rahman et al. (2021) analyzed 28 demographic, economic, built environment, health, and facilities-related factors for Bangladesh through OLS, SLM, SEM, and GWR. They also identified four risk factors affecting the COVID-19 incidence. Finally, Mansour et al. (2021) modeled the impact of 12 sociodemographic determinants on COVID-19 frequencies, and they underlined the effect of four variables and the success of MGWR in explaining spatial correlations.

In terms of economic factors, Curtini and Salvati (2021) investigated the spatial distribution of COVID-19 in Italy through selected global forces and economic drivers using regression models and local spatial analysis. The results show that urban agglomerations, including large-scale industries and companies, significantly affect COVID-19 incidence in different regions. Another study driven by Boumahdi et al. (2021) clarifies the role of industrial clusters in the propagation of viruses. They employed spatial statistical techniques, such as exploratory analysis and spatial econometrics, to prove the maximum impact of industrial activity; moreover,
they examined this issue over different periods: before and after the lockdown implementations.

In addition, some scholars focused on the relationship between demographic factors and the COVID-19 cases on the country level. Hamidi et al. (2020) analyzed the direct and indirect impacts of density on the COVID-19 incidences in dense metropolitan areas of U.S countries using structural equation modeling. They highlighted that one of the most critical indicators of infection rates is the urban population; more significant metropolitan regions have higher infection and mortality rates. Also, Arbel et al. (2021) evaluate the influence of population density and some socioeconomic measures (Gini Index) of Israel’s cities on COVID-19 infection. They highlight that those cities with high population density are more vulnerable to COVID-19 because of increased human interaction. They underline that urban policy-makers should enable and promote improved health infrastructure in denser cities. They also emphasized the moderate relationship between population density, size, income level, and COVID-19.

Finally, some researchers specifically concentrated on the influence of climate variables on the propagation of the disease. Sajadi et al. (2020) structured their study on temperature and humidity indicators, carrying out their investigations on a global perspective. On the other hand, Wang et al. (2020) evaluated the same topic on a national scale, observing that increased temperatures and humidity appeared to decrease the virus's transmission, compatible with similar behavior recorded for influenza transmission. Also, in China, Ma et al. (2020) investigated the correlation between COVID-19 mortality and Wuhan’s daily temperature range and associated humidity. This research was the first to associate reported deaths to climate indicators and was followed by various studies examining this correlation between climate and COVID-19 globally, nationally, and locally.

To sum up, although the COVID-19 Pandemic is still a riddle wrapped in an enigma, scientists from different disciplines have analyzed this phenomenon to understand which factors increase its transmission and mortality and to predict future scenarios (Table 1). All these researchers sometimes reached the same conclusion in these studies, which they prepared using different statistical methods; besides, their results, though at the elementary level now, will become even more comprehensive and produce much more data in the future. Studies focusing on COVID-19, which have increased day by day and been carried out in different parts of the world using different data sets, will undoubtedly continue to dominate the literature for a long time.

3 Data and model development

The research methodology consists of five main steps (Fig. 1). The first stage is designing the dataset derived from studies in the literature review with a similar research scope. Physical, economic, natural, demographic, and mobility indicators mentioned in articles were combined and evaluated according to their relevance for the case area. Besides, the readily available data in online sources were considered more generally in the factor determination process. The data are provided for the
| Research area          | Authors                  | Location | Factors                                           | Methodology                      | General results                                                                 |
|-----------------------|--------------------------|----------|---------------------------------------------------|----------------------------------|---------------------------------------------------------------------------------|
| Dimensions of COVID-19| Guan et al. (2020)       | China    | The distribution of patients and their characteristics | Comparison                      | The degree of illness varied from the first occurrence according to the characteristics of patients |
|                       | Zhang et al. (2020)      | China    | The number of COVID-19 and SARS patients          | Spatial Statistics (Moran’s I and LISA) | Spatiotemporal clustering tendencies for both diseases were detected             |
| Mobility              | Bogoch et al. (2020)     | China    | Total number of international air passengers      | Comparison of Infectious Disease Vulnerability Index | The existence of a strong correlation between human mobility and virus was detected |
|                       | Chen et al. (2020)       | China    | Total number of immigrants/migration rate         | Correlation in ArcGIS            | There is a strong correlation between migration and the spread of disease (immigrants from Wuhan are the primary infection source) |
|                       | Gross et al. (2020)      | China    | Distance, population size, and human mobility     | Statistical Comparison          | The mobility of infected individuals affected the infection rate of cities         |
|                       | Hierro and Mazza (2022)  | Madrid   | Inter-district mobility and socioeconomic factors | Spatial Contagion Index          | It was found that the influential role of inter-municipal mobility and population density in the spatial distribution of COVID-19 |
| Research area          | Authors                        | Location         | Factors                                                                 | Methodology                           | General results                                                                                                                                 |
|------------------------|--------------------------------|------------------|-------------------------------------------------------------------------|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| Health Geography and   | De Kadt et al. (2020)          | South Africa     | Six risk factors about living conditions, public services, and hygiene  | Syndemic Theory                       | The production of a vulnerability map for neighborhoods based on the risk level                                                             |
| Vulnerability          |                                |                  |                                                                         |                                       |                                                                                                                                                  |
|                        | Gibson and Rush (2020)         | Cape Town        | Poverty, population density, the lack of infrastructure                 | Practicing Social Distance in ArcGIS  | Due to its high density in informal settlements, it can be challenging to implement social distances effectively                                  |
|                        | Padula and Davidson (2020)     | Global           | Total number of nurses, the mortality ratio                             | Regression                            | There is a negative correlation between the number of nurses and the mortality ratio of COVID-19 disease                                           |
|                        | Lakhani (2020)                 | Australia        | Accessibility and capacity of existing health services in rural areas  | Percent, Number, Availability, and Capacity Approach | There is a strong relationship between the access difficulties of people over 65 years to health services and the disease’s mortality rate |
| Research area                        | Authors  | Location | Factors                                    | Methodology                          | General results                                                                                                                                 |
|-------------------------------------|----------|----------|--------------------------------------------|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| Physical Environment and Urbanization | Coccia (2020) | Italy    | Geo-environmental and demographic features | Statistical analysis and log–log Model | There is a possible association between the stimulated spread of COVID-19 and high air pollution, special meteorological situations, and the cities distant from the sea |
|                                     | Sridhar (2021) | India    | Socioeconomic and physical factors (urbanization) | Regression Analysis                  | While there is a strong positive correlation between urban poverty, slum areas, and COVID-19, there is a negative link between health services, open spaces, and COVID-19 |
| Research area                        | Authors                  | Location       | Factors                                      | Methodology                        | General results                                                                                                                                 |
|-------------------------------------|--------------------------|----------------|----------------------------------------------|------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| Natural Environment                 | Sajadi et al. (2020)     | Global         | Temperature, humidity, latitude              | Statistical Analysis (Mann–Whitney Test) | The distribution of COVID-19 across limited latitude, temperature, and humidity measurements is consistent with seasonal respiratory virus behavior |
|                                     | Wang et al. (2020)       | China          | Temperature and humidity                     | Regression Analysis                | Higher temperatures and relative humidity in the summer may reduce virus propagation                                                                 |
|                                     | Ma et al. (2020)         | China          | Temperature, humidity, and diurnal temperature range | GAM Analysis                      | Increases in temperature and absolute humidity are associated with reduced mortality from COVID-19                                              |
|                                     | Rios and Gianmoena (2021)| Italy          | Temperature and climatic, demographic, social factors | Bayesian Model                    | Temperature is the most relevant factor in reducing the severity of the COVID-19                                                                 |
| Socio-economic and demographic and behavior | Mollalo et al. (2020)   | US Counties    | Thirty-five explanatory determinants are related to socioeconomic, behavioral, environmental, and demographic characteristics | GIS-based Spatial Modeling and Geographically Weighted Regression | Four variables (average household income, income inequality, percentage of nurse practitioners, and percentage of black women) could define a relatively high disease rate variability |
| Research area | Authors | Location | Factors | Methodology | General results |
|---------------|---------|----------|---------|-------------|-----------------|
|               | Allcott et al. (2020) | US Counties | Social behavior against quarantine and political stance | Survey and GPS Data | It has been determined that the reactions to quarantine differ according to the prevailing political attitude in the states |
|               | Kuchler et al. (2020) | US Counties | Population density and demographic features of regions | Hot Spot and Time Series Analysis | The social connectivity index could be beneficial for epidemiologists in indicating the extent of contagious disease |
|               | Oto-Peralías (2020) | Spain | Geographical and socioeconomic indicators | Correlation Analysis | While there is a negative correlation with annual temperature, there is a positive correlation between population density and COVID-19 |
|               | Hamidi et al. (2020) | US Counties | Population Density | Structural Equation Modeling | Population density is one of the most critical indicators for COVID-19 infection rates |
| Research area | Authors | Location | Factors | Methodology | General results |
|---------------|---------|----------|---------|-------------|-----------------|
| Sannigrahi et al. (2020) | 31 European Countries | 28 Sociodemographic variables | Global and Local Spatial Regression Models (OLS, SLM, SEM, GWR, and MGWR) | Total population, poverty, income, are the key factors in regulating overall casualties of COVID-19 in the European region |
| Urban and Nakada (2021) | Brazil | 18 Demographic and Socioeconomic Variables | The geographically weighted regression (GWR) model best explains the spatial distribution of COVID-19 |
| Rahman et al. (2021) | Bangladesh | 28 Demographic, economic, built environment, health, and facilities-related factors | Four factors (percentage of the urban population, monthly consumption, number of health workers, and distance from the capital city) as significant risk factors affecting the COVID-19 incidence rates |
| Mansour et al. (2021) | Oman | 10 Demographic and Socioeconomic Variables | Elderly population aged 65 and above, population density, hospital beds, and diabetes rates were found to be statistically significant determinants of COVID-19 incidence rates |
| Florida and Mellander (2022) | Sweden | Fourteen place-based socioeconomic factors | Proximity to places with higher risks of infections is the most significant factor |
| Research area                     | Authors                        | Location | Factors                                      | Methodology                        | General results                                                                 |
|----------------------------------|--------------------------------|----------|----------------------------------------------|------------------------------------|---------------------------------------------------------------------------------|
| Economic Drivers and Industry    | Arbel et al. (2021)            | Israel   | Population density and Gini Index            | Correlation Analysis               | Cities with high population density are more vulnerable to COVID-19 because of the increasing human interaction |
|                                  | Curtini and Salvati (2021)     | Italy    | Global forces and local economic factors (the power of industry) | Regression Analysis and Linear Model | The density of urban agglomerations have a significant impact on the COVID-19 incidence in different regions |
|                                  | Boumahdi et al. (2021)         | Morocco  | Demographic factors and Labor force employed in industry | Spatial Exploratory Analysis        | They found the impact of the industry sector on the spread of the virus, and this impact is changeable according to the time period, lockdown or not |
whole province, including urban and rural areas inside the administrative borders. Also, the data for certain factors, such as behavioral habits, humidity level, income level, percentage of people having chronic diseases, and the number of critical care units, do not exist at the city level, so they did not count in the model.

Consequently, 20 descriptive variables on the province scale were counted in the model (Table 2). The Turkish Statistical Institute’s database, the Ministry of Environment, Urbanization, and Climate Change statistics, and the Social Security Institution were the fundamental sources for designing the dataset. Since February of 2021, the Ministry of Health has shared the weekly COVID-19 incidents in the cities with the public. Therefore, the number of COVID-19 cases in provinces on 2 April 2021, the anniversary of sharing city-level data with the public, was determined as a dependent variable. Also, the dynamics of dataset including minimum, maximum, median and mean values, standard deviations and quartiles, is shown in Table 3 to give detailed information about variables.

In the second phase, GeoDa software was used to conduct Moran’s I Index and Hot Spot Analysis. The statistical distribution of the total COVID-19 patients on the city scale was tested using Moran’s I Index, which analyzes spatial autocorrelation of features according to their geographical position and properties. This method determines whether the presented samples are grouped, dispersed, or random; besides, it tests the validity of the hypothesis through the estimation of a z score and p value. A positive Moran’s I value means the existence of a tendency to cluster when the z score or p value shows statistical significance. In contrast, negative Moran’s I values indicate the direction of variance (Getis and Ord, 2010). For this analysis, a weight matrix of neighborhoods was provided through Queen contiguity. The order of contiguity was defined as one, meaning that a feature can be counted as a neighbor only if it has direct contact. Following the Moran’s I Index results, Univariate Local Moran’s I, which is a spatial analysis and mapping process to reveal the clustering tendency of the phenomena, was employed under the scope of COVID-19 incidence.

Regression analysis was implemented to define what causes variation in the hotspot. Regression analysis provides a kind of model, and it examines and explores spatial relationships between dependent and explanatory variables. In other words, it is a statistical process for estimating the relationship between variables, and it helps to understand the factors behind observed spatial patterns

![Fig. 1 Research methodology](image-url)
| Theme                | Indicator                                                                 | Code | Year | Source                                      | References                                                                 |
|----------------------|---------------------------------------------------------------------------|------|------|---------------------------------------------|---------------------------------------------------------------------------|
| Dependent            | COVID-19 incidences—The number of COVID-19 cases for each province        | DE   | 2021 | Ministry of Health                         | In all studies                                                            |
| Demographic          | Young Population—The population under the age of 20                       | D1   | 2021 | Turkish Statistical Institution            | Coccia (2020); Hamidi et al. (2020); Oto-Peralías (2020); Florida and Mellander (2022); Arbel et al. (2021) |
|                      | Elderly population—The population over the age of 65                       | D2   | 2021 |                                             |                                                                           |
|                      | Active population—The population aged 15–64 who do not have a working disability | D3   | 2021 |                                             |                                                                           |
|                      | Population density—The ratio between area and total population (p/km2)     | D4   | 2021 |                                             |                                                                           |
|                      | Household size—The average number of people in the household              | D5   | 2021 |                                             |                                                                           |
|                      | Foreign immigrants—The number of foreign immigrants                        | D6   | 2020 |                                             |                                                                           |
|                      | Net migration rate—The difference between the number of immigrants and the number of emigrants (people leaving an area) throughout the year | D7   | 2021 |                                             |                                                                           |
| Economic             | GDP per capita—A country's economic output per person                     | E8   | 2019 | Curtini and Salvati (2021); Boumahdi et al. (2021) |                                                                           |
|                      | Active insured people—The number of registered employees                  | E9   | 2020 | Social Security Institution                |                                                                           |
|                      | Registered unemployment—The number of registered unemployed              | E10  | 2020 |                                             |                                                                           |
|                      | Industry sector—The number of registered employees in the industrial sector | E11  | 2020 | Turkish Union of Chambers and Commodity Exchanges |                                                                           |
| Theme                | Indicator                                                                 | Code | Year | Source                                               | References                                |
|---------------------|---------------------------------------------------------------------------|------|------|------------------------------------------------------|-------------------------------------------|
| Physical environment| Health capacity—*The number of hospitals per each province*              | PY12 | 2020 | Turkish Statistical Institution                      | Lakhani (2020); De Kadt et al. (2020)     |
|                     | Hospital capacity—*The number of hospital beds per each province*       | PY13 | 2020 |                                                      |                                           |
|                     | Built-up area—*The total area occupied by buildings and transportation networks (hectare)* | PY14 | 2018 | Ministry of Environment, Urbanization and Climate Change | Coccia (2020)                            |
|                     | Forest and semi-natural areas—*The total area of forest, scrubs, herbaceous plants and open areas with little or no vegetation (hectare)* | PY15 | 2018 |                                                      |                                           |
| Natural environment | Air quality—*The ratio of PM10*                                          | N16  | 2019 |                                                      | Rios and Gianmoena (2021)                 |
|                     | Annual temperature—*Dividing the sum of the air temperatures for everyday in a year by the number of days* | N17  | 2021 | General Directorate of Meteorology                   |                                           |
|                     | Sunshine duration—*A time indicating how many hours of sunshine per day on average a location falls on a yearly basis* | N18  | 2021 |                                                      |                                           |
| Mobility            | Car dependency—*The total number of automobile per each province*       | M19  | 2021 | Turkish Statistical Institution                      | Hierro and Mazza (2022)                  |
|                     | Overnight stays—*The number of domestic and foreign tourists staying in accommodation facilities* | M20  | 2021 |                                                      |                                           |
### Table 3  General dynamics of the dataset

| Code | Indicator                      | Minimum | Maximum   | Median   | Mean      | Std. Dev | Q1      | Q3      | IQR   |
|------|--------------------------------|---------|-----------|----------|-----------|----------|---------|---------|-------|
| DE   | COVID-19 incidences            | 102     | 93,691    | 1103     | 3150.84   | 10,527.82| 496     | 2268    | 1382  |
| D1   | Young population               | 16,051  | 4,407,103 | 146,929  | 311,181.13| 540,687.39| 82,206  | 299,906 | 191,056|
| D2   | Elderly population             | 9957    | 1,183,330 | 59,408   | 101,791.65| 153,747.34| 31,348  | 109,746 | 70,547|
| D3   | Active population              | 57,864  | 11,341,124| 343,680  | 709,372.66| 1,362,086.23| 196,917 | 727,932 | 462,424.5|
| D4   | Population density             | 11      | 3049      | 65       | 134.85    | 340.696  | 46      | 122     | 84    |
| D5   | Household size                 | 2.57    | 5.46      | 3.08     | 3.29      | 0.67     | 2.85    | 3.46    | 3.155 |
| D6   | Foreign immigrants             | 89      | 306,455   | 2094     | 8358.54   | 34,672.62| 918     | 4452    | 2685  |
| D7   | Net migration rate             | − 37.3  | 31.5      | 1.23     | 0.39      | 11.58    | − 6.84  | 6.92    | 0.04  |
| E8   | GDP per capita                 | 16,727  | 86,798    | 36,821   | 39,595.53 | 13,645.46| 30,267  | 46,529  | 38,398|
| E9   | Active insured people          | 19,291  | 5,649,571 | 114,215  | 288,204.28| 672,885.89| 68,428  | 291,319 | 179,873.5|
| E10  | Registered unemployment        | 2709    | 464,530   | 21,541   | 36,537.58 | 57,475.89| 11,401  | 37,441  | 24,421|
| E11  | Industry sector                | 396     | 756,527   | 15,732   | 44,143.88 | 97,697.83| 6166    | 31,467  | 18,816.5|
| PY12 | Health capacity                | 1       | 235       | 12       | 18.98     | 27.60    | 8       | 21      | 14.5  |
| PY13 | Hospital capacity              | 150     | 40,697    | 1460     | 2932.14   | 5138.80  | 743     | 3242    | 1992.5|
| PY14 | Built-up area                  | 141.2   | 79,800    | 800.4    | 1947.21   | 8781.87  | 562.8   | 1242.8  | 902.8 |
| PY15 | Forest and semi-natural areas  | 27,357.9| 1,695,468.16| 448,463.7| 500,545.85| 349,517.66| 237,532.57| 635,674.41| 436,603.49|
| N16  | Air quality                    | 18      | 99.34     | 47       | 51.11     | 17.89    | 40      | 60      | 50    |
| N17  | Annual temperature             | 3.6     | 18.8      | 12.8     | 12.78     | 3.20     | 10.9    | 14.7    | 12.8  |
| N18  | Sunshine duration              | 2.1     | 8.8       | 6.3      | 6.15      | 1.09     | 5.4     | 6.9     | 6.15  |
| M19  | Car dependency                 | 2204    | 3,162,884 | 64,025   | 169,210.67| 404,393.76| 30,662  | 153,316 | 91,989|
| M20  | Overnight stays                | 34,025  | 73,338,778| 432,259  | 2,174,018.50| 8,652,161.40| 254,419 | 965,019 | 609,719|
better (Chatterjee and Hadi, 2015). Due to making this understanding and prediction process more manageable, the explanatory regression analysis can be used in the third stage before the ordinary least square regression. This method tests all combinations of descriptive variables, and provides summaries about multicollinearity and significance of variables that are directive results to build a trustable model.

In the fourth stage, models were produced in ArcGIS to show each variable’s effect on the COVID-19 incidence. Ordinary Least Square (OLS) is a non-spatial regression model, but it is a suitable beginning for all spatial regression techniques. It offers a global measure including one equation for all indicators that are the main topic of the understanding or prediction process (Mitchell 2005).

OLS assumes that the statements on the city level are independent observations; moreover, it does not regard spatial dependency in the equation. It does not consider spatial dependence due to the assumption of homogeneity and spatial non-variability. However, on COVID-19 dissemination, previous studies show that defining factors are spatially connected with the dependent one. Therefore, Spatial Lag Model (SLM) and Spatial Error Model (SEM) were also used along with OLS. Due to the global models still have limitations as they cannot account for a spatial non-stationarity issue, Geographically Weighted Regression (GWR) was employed in the last step to specify the model according to each feature’s locations. GWR, introduced by Brunsdon et al. (1996), is a local regression model, and it claims that relationships between variables can vary across the study area. It provides one equation for each feature, so coefficients of variables can differ based on their locations (Wheeler and Paez 2010). GWR model was employed in ArcGIS software, and adaptive kernel type and AICc bandwidth were selected.

Briefly, this study has accepted a holistic and top–down approach as a research method. While the effectiveness of factors on the general distribution of COVID-19 was measured with OLS, the effectiveness of these factors was discussed over space with SLM, SEM, and GWR. Both modeling techniques are designed to work in coordination with one another. However, the R$^2$ and AICc values of the four developed models are compared at the end through their performances, explaining COVID-19 incidence rates across Turkey.

4 Results

Before evaluating models, the distribution of COVID-19 patients in Turkey is mapped based on the data revealed on 2 April 2021 (Fig. 2). According to the data, Istanbul, Ankara, Izmir, Konya, Samsun, and Bursa were the most affected by the pandemic; on the other hand, Hakkari, Sirnak, Tunceli, Bitlis, Siirt, and Ardahan had the fewest COVID-19 incidents over the same time period.
4.1 Spatial autocorrelation and Moran’s I

The virus distribution is tested statistically using Moran’s I index. Based on the first-order Queen’s contiguity, a weight matrix was created. Also, the descriptive statistics for the weight matrix are provided in Table 4. Accordingly, the Univariate Moran’s I was employed through the produced matrix.

The results show that the z score is 0.556, which means that the pattern appears to be significantly clustering in 0.01 confidence interval (Fig. 3). Given this score, it can be said that there is less than a 1% likelihood that this clustered pattern could be the result of random chance. When we examine the statistically significant clusters in detail through Univariate Moran’s I, it can be said that there are two hot spots and two cold spots. In terms of hot spots, the first region is located in the north-western part of Turkey, and it includes eight cities: Istanbul, Tekirdag, Kirklareli, Edirne, Canakkale, Bursa, and Kocaeli. This result can be explained by the existence of Istanbul, which dominates the Marmara Region for many reasons. The second hot spot is formed around six provinces in the Black Sea Region: Samsun, Corum, Sinop, Amasya, Tokat, and Ordu. On the other hand, while first cluster consists of eleven cities from the eastern part of Turkey which are Sanliurfa, Diyarbakir,
Mardin, Batman, Mus, Siirt, Bitlis, Sirnak, Hakkari, Van, and Agri, other cold spot includes Denizli and Burdur from the western part of Turkey. Although the first hot spot can be conceptualized as partly due to the existence of Istanbul, explaining all these clusters is one of the main aims of this study. Such an explanation is attempted in the analyses that follow.

### 4.2 Explanatory regression

As mentioned before, the explanatory regression method examines all sets of descriptive indicators. While the explanatory regression was employed; the minimum acceptable $R^2$ was determined as 0.5, the maximum VIF value was selected as 7.5, and the minimum acceptable Jarque Bera value and spatial autocorrelation value was arranged as 0.05. This regression analysis also produces results about multicollinearity and the importance of variables to create a trustable equation. When the summary of multicollinearity between variables is examined (Table 5), it can be said that 11 factors explain the same part of the equation, which negatively affects the model’s reliability:

- Young population (D1), elderly population (D2), active population (D3), population density (D4), and foreign immigrants (D6) in the demographic variables,
- Active insured people (E9), registered unemployment (E10), and industry sector (E11) in the economic variables,
- Health capacity (PY12) and the hospital capacity (PY13) in the physical factors, and
- Car dependency (M19) in the mobility.

When choosing among the 11 indicators mentioned, it is essential to know at what level and how the factors explain the model, and Table 6 explains the significance of variables in the equation. According to the results, population density (D4)
| Code | Variable                  | VIF  | Violations                                                                 | Covariates                                                                 |
|------|---------------------------|------|-----------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| D1   | Young population         | 191.46 | 4567                                                                       | PY12 (99.80), PY13 (99.80), E10 (99.80), M19 (99.80), E9 (98.18), D2 (98.08), D3 (98.08), D6 (75.23), E11 (73.00), D4 (54.60) |
| D2   | Elderly population       | 131.66 | 4368                                                                       | PY12 (99.80), PY13 (99.80), E10 (99.70), M19 (99.70), E9 (98.08), D1 (98.08), D3 (97.98), E11 (74.72), D6 (64.31), D4 (47.62) |
| D3   | Active population        | 965.96 | 4726                                                                       | PY12 (99.80), PY13 (99.80), E11 (99.70), E10 (99.70), M19 (99.70), D6 (99.70), E9 (98.08), D1 (98.08), D2 (97.98), D4 (87.06) |
| D4   | Population density       | 46.87  | 2947                                                                       | D6 (99.90), E9 (99.80), D3 (87.06), PY12 (63.90), D1 (54.60), E11 (53.89), E10 (49.34), M19 (48.84), PY13 (48.63), D2 (47.62) |
| D5   | Household size           | 3.64   | 0                                                                          | –                                                                           |
| D6   | Foreign immigrants       | 52.17  | 4006                                                                       | PY12 (99.90), M19 (99.90), D4 (99.90), E9 (99.80), D3 (99.70), D1 (75.23), E10 (72.50), PY13 (65.52), D2 (64.31), E11 (47.02) |
| D7   | Net migration rate       | 2.58   | 0                                                                          | –                                                                           |
| E8   | GDP per capita           | 3.92   | 0                                                                          | –                                                                           |
| E9   | Active insured people    | 763.23 | 4762                                                                       | E11 (99.80), PY12 (99.80), PY13 (99.80), D6 (99.80), D4 (99.80), E10 (99.70), M19 (99.70), D1 (98.18), D2 (98.08), D3 (98.08) |
| E10  | Registered unemployment  | 180.29 | 4489                                                                       | PY12 (99.90), PY13 (99.90), M19 (99.90), D1 (99.80), E9 (99.70), D2 (99.70), D3 (99.70), D6 (72.50), E11 (57.33), D4 (49.34) |
| E11  | Industry sector          | 26.68  | 3307                                                                       | E9 (99.80), D3 (99.70), D2 (74.72), D1 (73.00), PY12 (66.94), PY13 (59.76), E10 (57.33), M19 (54.50), D4 (53.89), D6 (47.02) |
| PY12 | Hospital capacity        | 103.21 | 4625                                                                       | E10 (99.90), M19 (99.90), PY13 (99.90), D6 (99.90), E9 (99.80), D2 (99.80), D3 (99.80), D1 (99.80), E11 (66.94), D4 (63.90) |
| PY13 | Hospital capacity        | 146.69 | 4413                                                                       | PY12 (99.90), E10 (99.90), M19 (99.90), E9 (99.80), D2 (99.80), D3 (99.80), D1 (99.80), D6 (65.52), E11 (59.76), D4 (48.63) |
Table 5 (continued)

| Code | Variable                              | VIF  | Violations | Covariates        |
|------|---------------------------------------|------|------------|--------------------|
| PY14 | Built-up area                         | 1.20 | 0          | –                  |
| PY15 | Forest and semi-natural areas         | 2.38 | 0          | –                  |
| N16  | Air quality                           | 1.27 | 0          | –                  |
| N17  | Annual temperature                    | 1.76 | 0          | –                  |
| N18  | Sunshine duration                     | 1.54 | 0          | –                  |
| M19  | Car dependency                        | 135.09 | 4526   | PY12 (99.90), PY13 (99.90), D6 (99.90), E10 (99.80), D1 (99.80), E9 (99.70), D2 (99.70), D3 (99.70), E11 (54.50), D4 (48.84) |
| M20  | Overnight stays                       | 5.99 | 0          | –                  |
and hospital capacity (PY12) have the highest score among other variables. While population density explains the model at a 99.97% significance level and 100% positive, hospital capacity also describes the equation at a 99.58% significance level and 100% positive. Therefore, selecting these two factors and continuing the process with them are the most optimal option to ensure the accuracy of the possible model. As a result of selecting two indicators among 11 variables showing multicollinearity, the Ordinary Least Square (OLS) regression was employed with the rest of the 11 variables in total.

4.3 Global models (OLS, SEM, and SLM)

The Ordinary Least Square regression was the first step to analyze the effects of selected indicators behind the clustering tendency of COVID-19 incidence before using Geographically Weighted Regression. According to the first model’s results (Table 7), only 5 of 11 variables are statistically significant based on their probability values to explain the relationship with the distribution of COVID-19 patients. These are population density (D4), net migration rate (D7), GDP per capita (E8), health capacity (PY12), and annual temperature (N17). However, before proceeding to the last stage of OLS, the five indicators that were found to be statistically

| Code | Variable                              | % Significant | % Negative | % Positive |
|------|---------------------------------------|---------------|------------|------------|
| D4   | Population density<sup>a</sup>        | 99.97         | 0          | 100        |
| PY12 | Health capacity<sup>a</sup>           | 98.95         | 0          | 100        |
| D6   | Foreign immigrants<sup>a</sup>        | 81.93         | 0.28       | 99.72      |
| E9   | Active insured people<sup>a</sup>      | 79.75         | 2.85       | 97.15      |
| N18  | Sunshine duration                     | 69.16         | 96.03      | 3.97       |
| D3   | Active population<sup>a</sup>         | 67.55         | 20.77      | 79.23      |
| N17  | Annual temperature                    | 65.12         | 95.29      | 4.71       |
| E10  | Registered unemployment<sup>a</sup>    | 51.35         | 63.81      | 36.19      |
| PY15 | Forest and semi-natural areas         | 50.53         | 79.76      | 20.24      |
| D2   | Elderly population<sup>a</sup>        | 49.77         | 55.79      | 44.21      |
| M19  | Car dependency<sup>a</sup>            | 45.01         | 46.58      | 53.42      |
| PY13 | Hospital capacity<sup>a</sup>         | 43.75         | 48.28      | 51.72      |
| D1   | Young population<sup>a</sup>          | 39.48         | 43.74      | 56.26      |
| PY14 | Built-up area                          | 39.32         | 41.53      | 58.47      |
| D5   | Household size                         | 35.51         | 71.50      | 28.50      |
| E11  | Industry sector<sup>a</sup>           | 33.41         | 30.49      | 69.51      |
| N16  | Air quality                            | 32.02         | 91.66      | 8.34       |
| M20  | Overnight stays                        | 26.45         | 19.64      | 80.36      |
| E8   | GDP per capita                         | 25.74         | 75.37      | 24.63      |
| D7   | Net migration rate                     | 17.50         | 38.21      | 61.79      |

<sup>a</sup>Factors having high multicollinearity
| Variable                      | Coefficient \([a]\) | StdError      | \(t\) statistic | Probability \([b]\) | Robust_SE | Robust_t | Robust_Pr \([b]\) | VIF \([c]\) |
|-------------------------------|---------------------|---------------|-------------------|---------------------|-----------|----------|-------------------|----------|
| Intercept                     | 5055.224704         | 1664.251406   | 3.037537          | 0.003367*           | 1478.505779 | 3.419144 | 0.001062*         | –        |
| Population density            | 16.748035           | 1.902271      | 8.804232          | 0.000000*           | 1.943669   | 8.616709 | 0.000000*         | 15.138060 |
| Household size                | −255.351339         | 414.764394    | −0.615654         | 0.540149            | 318.710196 | −0.801202 | 0.425761          | 2.817576 |
| Net migration rate            | 46.819095           | 23.353721     | 2.004781          | 0.048912*           | 18.738076  | 2.498607 | 0.014851*         | 2.637123 |
| GDP per capita                | −0.047952           | 0.020118      | −2.383357         | 0.019900*           | 0.022438   | −2.137067 | 0.36138*          | 2.716018 |
| Health capacity               | 175.777588          | 22.541176     | 7.798067          | 0.000000*           | 24.350325  | 7.218696 | 0.000000*         | 13.953717|
| Hospital capacity             | −0.035947           | 0.020222      | −1.777625         | 0.079875            | 0.009489   | −3.788251 | 0.000324*         | 1.136598 |
| Forests and semi-natural areas| −0.000353           | 0.000703      | −0.502095         | 0.617202            | 0.000548   | −0.643963 | 0.521734          | 2.176826 |
| Air quality                   | −0.093373           | 10.209985     | −0.091945         | 0.992729            | 8.481715   | −0.011099 | 0.991247          | 1.203196 |
| Annual temperature            | −289.483502         | 59.237136     | −4.886858         | 0.000007*           | 52.742190  | −5.488651 | 0.000001*         | 1.295994 |
| Sunshine duration             | −159.826768         | 188.030241    | −0.850006         | 0.398257            | 157.975614 | −1.011718 | 0.315203          | 1.536294 |
| Overnight stays               | 0.000690            | 0.000337      | 2.045815          | 0.044588*           | 0.000397   | 1.739928 | 0.086334          | 3.902299 |
significant were re-evaluated according to their significance level in the equation. According to Table 6, while GDP per capita (E8) explains the model with a 25.74% significance level and 75.37% negative, net migration rate (D7) also describes the equation at a 17.50% significance level and 61.79% positive. Therefore, these factors were not included in the final stage of the OLS, because their low significance level would be misleading in terms of the model’s accuracy. With this in mind, OLS regression was run with three factors: population density (D4), the number of hospitals per 100,000 people (PY12), and the annual temperature (N17) (Table 8). According to probability values, each variable is statistically significant in the equation. When the VIF (Variance Inflation Factor) is examined, all values are less than 7.5, meaning that each variable defines a different part of the story. Also, according to Jarque–Bera Statistics, the value is 0.059301, higher than 0.01, meaning that model predictions are not biased. Finally, the Adjusted R-Squared value is 0.975890, so it can be said that the model performance is 98%. After verifying the model, it can be said that population density (D4) affected the propagation of COVID-19 most (99.97%), and health capacity (PY12) comes in second place at 98.95%. On the other hand, the annual temperature (N17) negatively correlates with the dependent variable. It explains the model at a 63.46% significance level, which is 96.3% negative. Briefly, this result emphasizes that about 98% of the COVID-19 incidence rates across Turkey are associated with the model’s three factors.

When the distribution of error terms is examined (Fig. 4), the value was found to be more than zero (0.146), which means that the error terms tend to cluster and are not normally distributed. This indicates the importance of geography; therefore, SLM and SEM models were produced by integrating spatial dependence among the variables to enhance the outcomes of the overall OLS model (Table 9).

According to the results, three factors (population density, health capacity, and annual temperature) sustain their significance in explaining the COVID-19 incidence rates across Turkey. On the other hand, the models’ autoregressive lag coefficients (Rho and Lamda) were statistically significant between the 90% and 95% confidence levels.

When models are compared through their $R^2$ and AICc values, Table 10 shows that both SLM and SEM have higher $R^2$ values and lower AICc values than the OLS. SEM model provides the highest $R^2$ value among other global models. This result indicates that the SEM model is the best option compared with OLS and SLM. However, the modeling performance of the COVID-19 incidence rates in Turkey might be enhanced if the model could be produced locally instead of globally. Therefore, GWR was employed to model the COVID-19 incidence on a local scale.

### 4.4 Geographically weighted regression (GWR)

After establishing a trustable model through OLS, SLM, and SEM, GWR was used to analyze whether the variables differ depending on their locations, and each factor was evaluated separately. As mentioned before, the highest $R^2$ value comes from the SEM model; however, this value increased from 0.979 in the SEM to 0.988 in the GWR model (Table 10). Also, the AICc value decreased to the lowest ratio, 1393.33
### Table 8: Summary of OLS results, final model

| Variable            | Coefficient \([a]\) | StdError  | \(t\) Statistic | Probability \([b]\) | Robust_SE | Robust_t | Robust_Pr \([b]\) | VIF \([c]\) |
|---------------------|---------------------|-----------|------------------|---------------------|------------|----------|-------------------|-----------|
| Intercept           | 1198.829            | 745.0299  | 1.609102         | 0.111693            | 609.3793   | 1.967295 | 0.052751          | –         |
| Population density  | 18.07338            | 1.292372  | 13.98466         | 0.000000*           | 1.733652   | 10.42503 | 0.000000*         | 6.055026  |
| Health capacity     | 167.5658            | 15.99216  | 10.478           | 0.000000*           | 19.93559   | 8.405359 | 0.000000*         | 6.086485  |
| Annual temperature  | −286.779            | 57.06227  | −5.02573         | 0.000004*           | 52.13934   | −5.50025 | 0.000001*         | 1.042146  |
Fig. 4  The distribution of error terms

Table 9  Results of the SLM and SEM models

| Variable              | Coefficient | Std. error | p value |
|-----------------------|-------------|------------|---------|
|                       | SLM         | SEM        | SLM     | SEM     | SLM     | SEM     |
| Intercept             | 1185.35     | 876.12     | 727.453 | 841.773 | 0.103   | 0.297   |
| Population density    | 18.018      | 19.413     | 1.298   | 1.296   | 0.000*  | 0.000*  |
| Health capacity       | 168.134     | 151.601    | 15.953  | 15.571  | 0.000*  | 0.000*  |
| Annual temperature    | −288.219    | −247.688   | 56.144  | 62.828  | 0.000*  | 0.008*  |
| Rho                   | 0.009       | 0.036      | 0.080***|         |         |         |
| Lambda                | 0.319       | 0.142      |         |         | 0.024** |

*Significant at 99% confidence level  
**Significant at 95% confidence level  
***Significant at 90% confidence level

Table 10  Comparison of OLS, SLM, SEM, and GWR models

| Criterion | OLS | SLM | SEM | GWR |
|-----------|-----|-----|-----|-----|
| $R^2$     | 0.975 | 0.977 | 0.979 | 0.988 |
| AICc      | 1431.80 | 1430.94 | 1425.06 | 1393.33 |
in the GWR, compared to other global models. This result shows that the GWR model explains 98.8% of the variations of COVID-19 incidence across Turkey, and it is the most effective option in this context.

The spatial distribution of the coefficient values of the GWR model is examined in Fig. 5. The results show that population density (D4), annual temperature (N17), and health capacity (PY12) have almost similar patterns. According to Fig. 5, population density is essential in explaining the COVID-19 incidence in the western part of Turkey. This situation can be related to highly populated metropolitan cities in the west, such as Istanbul, Kocaeli, Bursa, Izmir, Ankara, and Konya, which also resulted in provinces with a higher number of cases. In terms of the health capacity factor, the situation is similar. In the northwest part of Turkey, health capacity influences COVID-19 incidence positively. When the annual temperature is examined, OLS regression already underlined the existence of a negative correlation with COVID-19 incidence. Similarly, the map shows that while the annual temperature is an influential variable in the western part of Turkey, it is less influential in the country’s north-eastern region.

Figure 6 also shows the local $R^2$ values of the GWR model. According to the results, it is clear that provinces located in the western part of Turkey have higher values than the east. In addition, it was determined that this value increased again toward the east after a specific region. GWR estimated the local $R^2$ value of the province based on its surroundings, and provinces with the lowest $R^2$ values are central provinces with neither dependent nor independent variables. Therefore, this situation can be explained through the hot and cold spots.

5 Discussion

The COVID-19 Pandemic is the most dangerous disaster of the twenty-first century, and it has affected countries physically and socially, but especially economically. Although their causes and effects are still unclear, many disciplines try to shed light on the matter by understanding the leading indicators behind the spread of the disease. Similarly, this study aims to examine the correlation between determining physical, natural, economic, demographic, and mobility factors on Turkey’s pandemic within the scope of these recent developments and information. Following this purpose, Turkey and its 81 provinces were chosen as a case area, and cities were investigated using spatial/non-spatial and global/local regression methods: Explanatory Regression, Ordinary Least Square, Spatial Lag Model, Spatial Error Model, and Geographically Weighted Regression.

Three of the twenty factors, namely population density, health capacity, and annual temperature, are highly correlated with the spread of COVID-19 in Turkey. Arbel et al. (2021) emphasized that high population density increases the vulnerability of cities to COVID-19; moreover, Florida and Mellander (2022) underlined that the existence of hospitals increases the risk of infection due to the interaction between people. Also, many studies found a negative correlation between temperature and the spread of COVID-19 (Wang et al. 2020; Ma et al. 2020; Rios and Gianmoena 2021). Considering the findings from the literature, it can be said that the
Fig. 5  Spatial distribution of coefficient values
statistical significance of these factors in explaining the model is a sensible conclusion to draw.

On the other hand, many factors posited as influential in the propagation of virus by the literature were actually ineffective according to the OLS results. Hierro and Mazza (2022) and Chen et al. (2020) had already underlined the influential role of migration, which is related to human mobility, in the spatial distribution of COVID-19 disease. In addition, most of the studies focused on international mobility between countries, and underlined the positive correlation between air traffic and COVID-19. In terms of the socioeconomic perspective, while GDP per capita as a global variable impacts the dissemination of the virus in terms of income level (Curtini and Salvati 2021), Gibson and Rush (2020) evaluated the household size under the scope of social distance issues. They claim that the high density in informal settlements could increase vulnerability to the virus. Sridhar (2021) examined the relationship between the urbanization level of Indian cities and the spread of the virus and highlighted some positive and negative correlations in terms of selected physical factors such as open spaces. Coccia (2020) investigated the impact of high air pollution and special meteorological conditions on COVID-19 for different zones of Italy. Despite all these validating statements in the literature, the absence of a statistical relationship with COVID-19 within the scope of the current study can only be explained through Turkey’s internal dynamics.

According to GWR results, it was also found that the impact of these factors differs spatially, especially in the West–East direction. In terms of population density, it is clear that this variable is an influential factor in explaining COVID-19 incidences in the provinces in the west, especially Istanbul, Kocaeli, Bursa, Tekirdag, Kirklareli, and Edirne. The situation shows a similar pattern when the health capacity factor is examined. This situation can be described through the density of metropolitan cities in the west compared to the east. They have the highest values regarding COVID-19 incidence, population density, and health capacity;
therefore, they dominate the equation and determine the direction of correlation. The last factor, annual temperature, reflects the same result with other variables in a different direction. It is an influential variable in the western part of Turkey and less influential in the country’s north-eastern region. Briefly, only 3 of the 20 factors are statistically associated with the spread of COVID-19 in Turkey. While these three factors differ spatially, metropolitan cities, especially Istanbul, Kocaeli, Izmir, and Ankara, have most affected the equation. Since each variable has a different dynamic based on the location, the direction of the correlation varies. However, it can be said that this correlation is generally formed on the East–West axis of Turkey.

This study also offers a chance to investigate different models and their performance in evaluating the role of place-based factor in the COVID-19 incidence. When three global (OLS, SLM, and SEM) and one local model (GWR) are compared, it is clear that while SEM is the best option between global model because of the higher $R^2$ (0.979) and lower AICc values (1425.06), GWR performs the highest $R^2$ (0.988) and lowest AICc values (1393.339). On the other hand, the high $R^2$ values provided by all models stand out. This situation can be explained through the distribution of variables in the data set and their impact rates. In future studies, analyses can be repeated by normalizing the data, and comparisons can be made over $R^2$ values.

Although OLS, SLM, SEM, and GWR can explain the correlation between dependent and explanatory variables on the global and local scale, this study’s fundamental limitation was data availability. Although the Ministry of Health started to share the weekly COVID-19 cases on the city scale after February 2021, data about the mortality and ventilation cases which are used in correlation studies in the literature are still unclear. This situation affected the reliability index of the study. Besides, similar to COVID-19 mortality and ventilation rates, many explanatory factors underlined in the literature, such as behavioral habits, humidity level, and income level, do not exist at the city level, so they did not count in the model.

6 Conclusion

In conclusion, despite difficulties in finding data, the study did offer many remarkable results. Three factors are statistically associated with the spread of COVID-19 in Turkey. Three factors, population density, health capacity, and annual temperature, are statistically significant in explaining the spread of COVID-19 in Turkey’s provinces. While these factors differ spatially, the correlation is generally constructed on the East–West axis of Turkey. Moreover, while all global and local methods explain the correlation between dependent and explanatory variables, it is clear that GWR is the best option among them because of its highest $R^2$ and lowest AICc values. On the other hand, it should be considered that the findings are shaped in line with Turkey’s internal dynamics and may vary according to the space-based conditions. These results provide a base for future studies, and they fill an important gap in the literature. Similar studies can be developed using new datasets and different statistical methodologies, spatial or non-spatial. In this way, urban policy-makers can
develop appropriate approaches toward expected future pandemics, including that of COVID-19.

Author contributions All authors contributed to the study’s conception and design. Material preparation, data collection, and analysis were performed by MR and TB. The first draft of the manuscript was written by MR, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding No funding was received for the preparation of this manuscript.

Declarations

Conflict of interest The authors have no competing interests to declare relevant to this article’s content.

References

Allcott H, Boxell L, Conway J, Gentzkow M, Thaler M, Yang D (2020) Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. J Public Econ. https://doi.org/10.1016/j.jpubeco.2020.104254

Arbel Y, Fialkoff C, Kerner A, Kerner M (2021). Do population density, socioeconomic ranking, and Gini index of cities influence infection rates from Coronavirus? Israel as a case study. Ann Region Sci:1–26

Bahçetepe S (2021) Türkiye’nin yeni Wuhan’ı İstanbul. Cumhuriyet. https://www.cumhuriyet.com.tr/haber/turkiyeinin-yeni-wuhani-istanbul-1822736

Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MU, Khan K (2020) Potential for global spread of a novel coronavirus from China. J Trav Med. https://doi.org/10.1093/jtm/taaa011

Boumahdi I, Zaoujal N, Fadlallah A (2021) Is there a relationship between industrial clusters and the prevalence of COVID-19 in the provinces of Morocco? Reg Sci Policy Pract 13:138–157. https://doi.org/10.1111/rsp3.12407

Brusdon C, Fotheringham AS, Charlton ME (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr Anal 28(4):281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Chatterjee S, Hadi AS (2015) Regression analysis by example. Wiley, Oxford

Chen ZL, Zhang Q, Lu Y, Guo ZM, Zhang X, Zhang WJ, Lu JH (2020) Distribution of the COVID-19 epidemic and correlation with population emigration from Wuhan, China. Chin Med J. https://doi.org/10.1097/CM9.0000000000000782

Coccia M (2020) Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2020.138474

Cutrini E, Salvati L (2021) Unraveling spatial patterns of COVID-19 in Italy: global forces and local economic drivers. Reg Sci Policy Pract 13:73–108. https://doi.org/10.1111/rsp3.12465

Döker MF, Ocak F (2020) COVID-19 salgınının Türkiye’deki coğrafi dağılışının izlenmesinde Web CBS kullanımı. Türk Coğrafya Dergisi. https://doi.org/10.17211/tdc.778712

Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 20(5):533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

De Kadt J, Gotz G, Hamann C, Maree G, Parker A (2020) Mapping vulnerability to COVID-19 in Gauteng. https://gcro.ac.za/outputs/map-of-the-month/detail/mapping-vulnerability-to-covid-19/. Accessed 13 Mar 2022

Florida R, Mellander C (2022) The geography of COVID-19 in Sweden. Ann Reg Sci 68(1):125–150. https://doi.org/10.1007/s00168-021-01071-0

Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L (2020) Spatial analysis and GIS in the study of COVID-19. A review. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2020.140033
Getis A, Ord JK (2010) The analysis of spatial association by use of distance statistics. Perspectives on spatial data analysis. Springer, Berlin, pp 127–145

Gibson L, Rush D (2020) Novel Coronavirus in Cape Town informal settlements: feasibility of using informal dwelling outlines to identify high risk areas for COVID-19 transmission from a social distancing perspective. JMIR Public Health Surveill. https://doi.org/10.2196/18844

Gross B, Zheng Z, Liu S, Chen X, Sela A, Li J, Havlin S (2020) Spatio-temporal propagation of COVID-19 pandemics. EPL (Europhysics Letters). https://doi.org/10.1209/0295-5075/131/58003

Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Zhong NS (2020) Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 382(18):1708–1720. https://doi.org/10.1056/NEJMoa2002032

Hamidi S, Sabouri S, Ewing R (2020) Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners. J Am Plann Assoc 86(4):495–509. https://doi.org/10.1080/01944363.2020.1777891

Hierro M, Maza A (2022) Spatial contagion during the first wave of the COVID-19 pandemic: some lessons from the case of Madrid, Spain. Region Sci Policy Pract. https://doi.org/10.1111/rsp3.12522

Kuchler T, Russel D, Stroebel J (2020) The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook. J Urban Econ. https://doi.org/10.2139/ssrn.3587255

Kuupiel D, Adu KM, Bawontuo V, Adogboba DA, Drain PK, Moshabela M, Mashamba-Thompson TP (2020) Geographical accessibility to glucose-6-phosphate dioxynegase deficiency point-of-care testing for antenatal care in Ghana. Diagnostics. https://doi.org/10.3390/diagnostics10040229

Lakhani A (2020) Introducing the percent, number, availability, and capacity [PNAC] spatial approach to identify priority rural areas requiring targeted health support in light of COVID-19: a commentary and application. J Rural Health. https://doi.org/10.1111/jrh.12436

Ma Y, Zhao Y, Liu J, He X, Wang B, Fu S, Luo B (2020) Effects of temperature variation and humidity on the death of COVID-19 in Wuhan. China Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2020.138226

Mansour S, Al Kindi A, Al-Said A, Al-Said A, Atkinson P (2021) Sociodemographic determinants of COVID-19 incidence rates in Oman: Geospatial modelling using multiscale geographically weighted regression (MGWR). Sustain Cities Soc 65:102627. https://doi.org/10.1016/j.scs.2020.102627

Martinho VJP (2021) Impact of Covid-19 on the convergence of GDP per capita in OECD countries. Reg Sci Policy Pract 13:55–72. https://doi.org/10.1111/rsp3.12435

Mitchell A (2005) The ESRI guide to GIS analysis: Volume 2: Spatial measurements and statistics. ESRI Press, Redlands

Mollalo A, Sadeghian A, Israel GD, Rashidi P, Sofizadeh A, Glass GE (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 Tropica 188:187–194. https://doi.org/10.1016/j.actatropica.2018.09.004

Mollalo A, Mao L, Rashidi P, Glass GE (2019) A GIS-based artificial neural network model for spatial distribution of tuberculosis across the continental United States. Int J Environ Res Public Health 16(1):157

Mollalo A, Vaheedi B, Rivera KM (2020) GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2020.138884

Orea L, Álvarez IC (2020) How effective has the Spanish lockdown been to battle COVID-19? A spatial analysis of the coronavirus propagation across provinces. Documento De Trabajo 3:1–33. https://doi.org/10.1002/hec.4437

Oto-Peralías D (2020) Regional correlations of COVID-19 in Spain. OSF Preprints. https://doi.org/10.31219/osf.io/tjdgw

Padula WV, Davidson PM (2020) Countries with high registered nurse (RN) concentrations observe reduced mortality rates of coronavirus disease 2019 (COVID-19). Soc Sci Res Netw. https://doi.org/10.2139/ssrn.3566190

Rahman MH, Zafri NM, Ashik FR, Waliullah M, Khan A (2021) Identification of risk factors contributing to COVID-19 incidence rates in Bangladesh: a GIS-based spatial modeling approach. Heliyon 7(2):e06260. https://doi.org/10.1016/j.heliyon.2021.e06260

Rezaei M, Nouri AA, Park GS, Kim DH (2020) Application of geographic information system in monitoring and detecting the COVID-19 outbreak. Iran J Public Health. https://doi.org/10.18502/ijph.v49iS1.3679
Rios V, Gianmoena L (2021) On the link between temperature and regional COVID-19 severity: Evidence from Italy. Reg Sci Policy Pract 13:109–137. https://doi.org/10.1111/rsp3.12472
Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A (2020) Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (COVID-19). JAMA Netw Open. https://doi.org/10.1001/jamanetworkopen.2020.11834
Sannigrahi S, Pilla F, Basu B, Basu AS, Molter A (2020) Examining the association between sociodemographic composition and COVID-19 fatalities in the European region using spatial regression approach. Sustain Cities Soc 62:102418. https://doi.org/10.1016/j.scs.2020.102418
Sridhar KS (2021) Urbanization and COVID-19 Prevalence in India. Reg Sci Policy Pract. https://doi.org/10.1111/rsp3.12503
Urban RC, Nakada LYK (2021) GIS-based spatial modelling of COVID-19 death incidence in São Paulo, Brazil. Environ Urban 33(1):229–238. https://doi.org/10.1177/0956247820963962
Wang J, Tang K, Feng K, Lv W (2020) High temperature and high humidity reduce the transmission of COVID-19. SSRN Electron J. https://doi.org/10.2139/ssrn.3551767
Wheeler DC, Páez A (2010) Geographically weighted regression. Handbook of applied spatial analysis. Springer, Berlin, pp 461–486
Wu X, Nethery RC, Sabath BM, Braun D, Dominici F (2020) Exposure to air pollution and COVID-19 mortality in the United States. MedRxiv. https://doi.org/10.1289/fseec.2020.virtual.O-OS-638
Xiong Y, Wang Y, Chen F, Zhu M (2020) Spatial statistics and influencing factors of the novel coronavirus pneumonia 2019 epidemic in Hubei Province, China. https://doi.org/10.21203/rs.3.rs-16858/v2
Zeren F, Yılançı V (2020) Analysing Spatial Patterns of the COVID-19 Outbreak in Turkey. Bingöl Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi 4(2):27–40
Zeren F, Yılançı V, İşlek H (2021) İtalya’da COVID-19’un Bölgesel Arası Yayılımı: Keşfedici Mekansal Veri Analizi. Elektronik Sosyal Bilimler Dergisi 20(79):1432–1442
Zhang X, Rao H, Wu Y, Huang Y, Dai H (2020) Comparison of spatiotemporal characteristics of the COVID-19 and SARS outbreaks in mainland China. BMC Infect Dis 20(1):1–7. https://doi.org/10.1186/s12879-020-05537-y

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.