



Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States

Arabinda Maiti^a, Qi Zhang^{b,c}, Srikanta Sannigrahi^{d,*}, Suvamoy Pramanik^e,
Suman Chakraborti^e, Artemi Cerda^f, Francesco Pilla^d

^a Geography and Environment Management, Vidyasagar University, West Bengal, India

^b Department of Earth and Environment, Boston University, Boston, MA, 02215, USA

^c Frederick S. Pardee Center for the Study of the Longer-Range Future, Frederick S. Pardee School of Global Studies, Boston University, Boston, MA, 02215, USA

^d School of Architecture, Planning and Environmental Policy, University College Dublin Belfield, Clonskeagh, Dublin, D14 E099, Ireland

^e Center for the Study of Regional Development, Jawaharlal Nehru University, New Delhi, Delhi, 110067, India

^f Soil Erosion and Degradation Research Group, Department of Geography, Valencia University, Blasco Ibàñez, 28, 46010, Valencia, Spain

ARTICLE INFO

Keywords:

COVID-19
Spatial regression
Temporal change
Confounding factors
Migration
Income

ABSTRACT

Since December 2019, the world has witnessed the stringent effect of an unprecedented global pandemic, coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). As of January 29, 2021, there have been 100,819,363 confirmed cases and 2,176,159 deaths reported. Among the countries affected severely by COVID-19, the United States tops the list. Research has been conducted to discuss the causal associations between explanatory factors and COVID-19 transmission in the contiguous United States. However, most of these studies focus more on spatial associations of the estimated parameters, yet exploring the time-varying dimension in spatial econometric modeling appears to be utmost essential. This research adopts various relevant approaches to explore the potential effects of driving factors on COVID-19 counts in the contiguous United States. A total of three global spatial regression models and two local spatial regression models, the latter including geographically weighted regression (GWR) and multiscale GWR (MGWR), are performed at the county scale to take into account the scale effects. For COVID-19 cases, ethnicity, crime, and income factors are found to be the strongest covariates and explain most of the variance of the modeling estimation. For COVID-19 deaths, migration (domestic and international) and income factors play a critical role in explaining spatial differences of COVID-19 deaths across counties. Such associations also exhibit temporal variations from March to July, as supported by better performance of MGWR than GWR. Both global and local associations among the parameters vary highly over space and change across time. Therefore, time dimension should be paid more attention to in the spatial epidemiological analysis. Among the two local spatial regression models, MGWR performs more accurately, as it has slightly higher Adj. R^2 values (for cases, $R^2 = 0.961$; for deaths, $R^2 = 0.962$), compared to GWR's Adj. R^2 values (for cases, $R^2 = 0.954$; for deaths, $R^2 = 0.954$). To inform policy-makers at the nation and state levels, understanding the place-based characteristics of the explanatory forces and related spatial patterns of the driving factors is of paramount importance. Since it is not the first time humans are facing public health emergency, the findings of the present research on COVID-19 therefore can be used as a reference for policy designing and effective decision making.

1. Introduction

The coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and first reported in December 2019 in Wuhan city of China, has soon become a

new public health concern across the world (Ge et al., 2020; Jin et al., 2020; Rumpler, Venkataraman, & Göransson, 2020; Sun & Zhai, 2020). The virus poses serious potential threats to the medical protection system all over the world (European Centre for Disease Prevention & Control, 2020; World Health Organization, 2020). As of January 29,

* Corresponding author at: School of Architecture, Planning, and Environmental Policy, University College Dublin, Belfield, Dublin 4, Ireland.

E-mail addresses: arabinda@mail.vidyasagar.ac.in (A. Maiti), qz@bu.edu (Q. Zhang), srikanta.sannigrahi@ucd.ie (S. Sannigrahi), suvamo60_ssf@jnu.ac.in (S. Pramanik), suman87_ssf@jnu.ac.in (S. Chakraborti), artemio.cerda@uv.es (A. Cerda), francesco.pilla@ucd.ie (F. Pilla).

<https://doi.org/10.1016/j.scs.2021.102784>

Received 18 November 2020; Received in revised form 13 February 2021; Accepted 15 February 2021

Available online 19 February 2021

2210-6707/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2021, there have been 100,819,363 confirmed cases and 2,176,159 deaths reported (World Health Organization, 2020). Geography that includes both spatial locations and characteristics of the spatial determinants has played a key role in the early outbreak and transmitting the virus across the scale (Andersen, Nielsen, Simone, Lewiss, & Jagsi, 2020; Sannigrahi, Pilla, Basu, & Basu, 2020, b). The spatial variability and clustered concentration of both COVID-19 mortality and morbidity in many countries have demonstrated a strong spatial dependency of the confounding factors (Desmet & Wacziarg, 2020; Ren et al., 2020; Zhang & Schwartz, 2020). Although several timely efforts (e.g., Luo, Yan, & McClure, 2020) have analyzed spatial heterogeneous patterns and uneven distributions of COVID-19 casualties, few studies have utilized the spatial time-varying dimension in spatial econometric modeling for analyzing geographic disparities in COVID-19 casualties in the United States (Sun, Matthews, Yang, & Hu, 2020). The present research, therefore, has made an effort to examine how spatial analysis can help with identifying the hotspots and vulnerable locations as well as exploring the spatial dependency of confounding factors that explain the overall casualties caused by COVID-19.

Spatial regression models can be useful for quantifying the risk of disease progression in the communities (Desmet & Wacziarg, 2020; Ehlert, 2020; Xiong, Wang, Chen, & Zhu, 2020). Previous spatial epidemiological research noted a strong spatial time-varying effect of the confounding factors on virus outbreaks (Auchincloss, Gebreab, Mair, & Diez Roux, 2012; Chakraborti et al., 2020; Fitzpatrick, Harris, & Drawve, 2020; Kirby, Delmelle, & Eberth, 2017; Sannigrahi, Pilla, Basu, Basu et al., 2020). Of them, a few studies have focused on the spatially heterogeneous characteristics of the COVID-19 transmission (Bashir et al., 2020; Conticini, Frediani, & Caro, 2020; Sarwar, Waheed, Sarwar, & Khan, 2020; Xiong et al., 2020; Yao et al., 2020). The disproportionate burden of COVID-19 could be due to place-based characteristics that include cluster concentration and spatial aggregation of infected population and the proximity of social interaction (Sannigrahi, Pilla, Basu, Basu et al., 2020; Sun et al., 2020). Therefore, both characteristics of the spatial confounding factors and spatial interconnection between the places should be carefully considered while inspecting the factors that exacerbate the spread of disease and identifying communities vulnerable to the infection (Mansour, Al Kindi, Al-Said, Al-Said, & Atkinson, 2021; Zhu et al., 2018). Hence, developing spatial models and understanding the confounding effects of the variables is critical to reveal the spatial variation of virus transmission at any spatial or administrative scale (Ren et al., 2020; Zhang & Schwartz, 2020).

Previous studies have utilized environmental, socio-economic and demographic factors to explain spatial variability of the COVID-19 incidents and discover the underlying risk of the outbreaks across multiple scales (Desmet & Wacziarg, 2020; Karaye & Horney, 2020; Qi et al., 2020; Ren et al., 2020; Sannigrahi, Pilla, Basu, Basu, & Molter, 2020). Among the explanatory factors, several have been found strongly linked to the early transmission of the virus and the overall casualties caused by COVID-19. These key factors include traveling distance (Fortaleza, Guimarães, de Almeida, Pronunciate, & Ferreira, 2020), concentration of particulate matter (Bolaño-Ortiz et al., 2020), ethnic composition (Oztig & Askin, 2020; Thakar, 2020), income and socio-demographic factors (Sannigrahi, Pilla, Basu, Basu et al., 2020, 2020b), migration (Chen et al., 2020; Xiong et al., 2020), and air transport (Christidis & Christodoulou, 2020). Considering the country-specific analysis, in Wuhan (China) for instance, population density, the proportion of construction land, aged population density, tertiary industrial output per unit land, are found to be strongly associated with the COVID-19 counts and the overall COVID-19 casualties (You, Wu, & Guo, 2020).

In the United States, from the thirty-five explanatory variables covering various types of characteristics, four variables (i.e., income inequality, median household income, the proportion of black females, and the proportion of nurse practitioners) are found the key determining factors in COVID-19 casualties (Mollalo, Vahedi, & Rivera, 2020). In another analysis covering 2,814 United States counties and using

COVID-19 data up to May 1, 2020, researchers found strong positive correlations between the socioeconomic factors such as proportions of elderly and COVID-19 incidence and mortality rate (Zhang & Schwartz, 2020). Considering the February 19 and June 14, 2020 COVID-19 data in Iran, several infrastructure and climate factors (distance from bus stations and the minimum temperature of the coldest month) were found strongly associated with COVID-19 incidences and exhibited high variable importance in the analysis (Pourghasemi et al., 2020). The cross-country comparison of virus spread and their interaction with demographic, economic, and environmental parameters are limited. Among them, Sannigrahi, Pilla, Basu, Basu, Molter et al. (2020) focused on the European region, and carried out the spatial models to understand the spatially heterogeneous properties among the factors in different European countries; this study found that income and socio-demographic variables have the highest impact on COVID-19 fatalities in Germany, Austria, Slovenia, etc. A similar association was found in Germany from another study (Ehlert, 2020). In cross-country analysis, several confounding factors, such as out-of-pocket expenditure, could significantly explain the global variation of COVID-19 casualties in 175 countries. Among these factors, the age composition and out-of-pocket expenditure were found to be positively related to COVID-19 counts (Iyanda et al., 2020). In another study with a world-level analysis, Chakraborti et al. (2020) had identified few key determinants including air pollution, migration, economy, and demographic factor, which had strong positive correlations with COVID-19.

Omitting the time variable in spatial models can lead to erroneous estimates and misleading conclusions. Moreover, assuming the time-independent and homogenous impact of the confounding factors on response variables (COVID-19 cases and deaths in the present study) may introduce ambiguity in parameter approximation and eventually produce unconvincing results. Therefore, the present research makes an effort to address the current research gap in spatial COVID-19 studies by conceptualizing time-dependent spatial regression models using open source data with information in the contiguous United States. The hypothesis of this study is framed as “*the spatial association between the confounding factors and COVID-19 counts strongly depend on time; thus, space entity alone cannot fully explain the associations and the spreading of diseases in the contiguous United States*”. The specific objectives of this study are to explore the overall associations between the explanatory factors and COVID-19 cases and deaths and examine local association between the explanatory drivers and COVID-19 incidences. The present study also develops dynamic spatial regression models for exploring the time-dependent local spatial association as well as measuring the relative importance of variables with parsimonious regression models.

2. Materials and methods

2.1. Data collection and pre-processing

This research utilized the most updated aggregated county-level datasets provided by Johns Hopkins University (Killeen et al., 2020). These datasets contain 348 relevant variables covering multiple domains, such as demography, education, economy, health care capacity, crime statistics, public transit, climate, and housing information (Table S1). Since the main aim of the present study is to establish a modeling framework to examine the space- and time-dependent associations between COVID-19 incidences and potential explanatory factors, all the relevant variables were pre-processed to connect the observations to their corresponding county units through the unique Federal Information Processing Standard (FIPS) code. Each FIPS code contains five digits, with the first two digits referring to state information and the last three digits describing county information. The Johns Hopkins team retrieves information from various governmental and institutional sources, including the United States Census Bureau, United States Department of Agriculture (USDA) Economic Research Service,

the National Oceanic and Atmosphere Administration (NOAA), the Association of American Medical Colleges (AAMC), Henry J. Kaiser Family Foundation (KFF), the Center for Neighborhood Technology (CNT), the Bureau of Justice Statistics, and Department of Justice (DOJ) (Killeen et al., 2020). The data also retrieved key information on the health care system at the county scale that indicates how a county's health care system performed in handling COVID-19 counts.

The daily COVID-19 counts, including confirmed cases and deaths, were obtained for the period of January 22 to July 26, 2020 from USAFacts¹. The daily counts of COVID-19 cases and deaths were converted to cumulative sum for subsequent analysis and interpretation. The USAFacts team aggregates the most updated COVID-19 counts from various sources, including Centers for Disease Control and Prevention and state-level and local-level public health agencies. However, for most of the states, the USAFacts team gathers the daily county-level cumulative COVID-19 counts (positive cases and deaths) based on published tables, web dashboards, or PDF reports available on state public health websites through scraping or manual entry. The actual numbers (COVID-19 counts) reported in USAFacts sometimes may not exactly match with the statistics from the state public health organization reports. This can be due to the frequency in which the USAFacts are collecting and updating data is different from that of local governmental organizations. Additionally, there are a few states where up-to-date county-scale data is either not available on the public health website or data collection is not sufficiently frequent. For example, the updated COVID-19 counts in California and Texas are not available on the state public health websites. For these states, the USAFacts team extracted the latest available numbers from the county-specific public health websites.

Daily air pollution data were collected from the OpenAQ data repository system for extracting five key air pollutants, including two kinds of particulate matter (PM_{2.5}, PM₁₀), Sulfur Dioxide, Nitrogen Dioxide, and Carbon Monoxide. Daily concentrations of these atmospheric pollutants were converted to the monthly average unit for the examination of their associations with COVID-19 casualties. Currently, the OpenAQ platform consists of 686 million air quality measurements, 150 data sources, 13,000 locations, and 95 countries in their system, which is able to collect hourly air pollution concentration estimates from governmental and sensor sources. An R package, called "ropenaq: Accesses Air Quality Data from the Open Data Platform OpenAQ", was utilized to assess the large volume of data for the entire contiguous United States from January 22 to July 27 of 2020. The location wise air pollution data were further converted to raster surface using the "inverse distance weighting" interpolation method. Finally, the mean air pollution concentration of each county was calculated using zonal statistics as a table function in ArcGIS Pro v2.6.

2.2. Variable selection and dimensionality reduction

Dimensionality reduction and critical information extraction from datasets are crucial for regression modeling and effective decision analysis. This research employed a stepwise forward regression approach as a tool to separate the key variables from sets of unorganized variables. A total of nine groups (i.e., *crime*, *demography*, *education*, *employment*, *ethnicity*, *pollution*, *health*, *migration*, and *climate*), which were assumed to have both synergistic and trade-off associations with COVID-19 counts, were formed. Subsequently, key variables were extracted from each group based on Variable Inflation Factor (VIF) and model variability score, the latter of which is characterized by the coefficient of determination (R²) and adjusted coefficient of determination (Adj. R²). For the category of *crime*, totally 16 variables were incorporated into the modeling; for the other categories, a total of 14

(*demography*), 29 (*education*), 6 (*employment*), 72 (*ethnicity*), 63 (*health-care*), 5 (*pollution*), 7 (*migration*), and 4 (*climate*) variables were considered, respectively (see detail in Table S1). Multiple collinearity tests, including VIF, R² change, correlation coefficient, probability and t-statistics, were executed to detect the models' redundant variables. High collinearity would be evident in the model if the VIF value was greater than 10; therefore, all the filtered variables considered in the regression modeling were scrutinized to eliminate the redundancy in model parametrization. Followed by stepwise forward regression, the enter stepwise regression method was performed to measure the VIF value of each explanatory variable to ensure that the multicollinearity was entirely eliminated. The final parsimonious models that relied on fewer parameters and at the same time explained the maximum model variances with less uncertainty were parameterized for each category regarding both COVID-19 cases and deaths. These processes of variable selection and dimensionality reduction part were conducted in SPSS V26.

2.3. Spatial regression

2.3.1. Global spatial regression

Spatial regression models have been used extensively in the COVID-19 research across multiple spatial scales (Guliyev, 2020; You et al., 2020). Among all the available global spatial regression models, we used Ordinary Least Square (OLS), Spatial Error Model (SEM), and Spatial Lag Model (SLM) for measuring the global associations between the explanatory factors and COVID-19 counts at the county scale. The OLS model can be conceptualized as follows:

$$y_i = \beta_0 + \beta x_i + \varepsilon_i \quad (1)$$

Where y_i is the COVID-19 case or death counts at county i , β_0 is the model intercept, β is the slope parameter; x_i is the selected independent variable(s) at county i ; ε_i is the error term at model estimates. The global OLS assumes to have spatial stationarity across the scale, and therefore, also hypothesizes that a model conceptualized for a particular area can be applied effectively to other areas of interest (Fang, Liu, Li, Sun, & Miao, 2015). According to Anselin and Arribas-Bel (2013), the global OLS has fundamental assumptions: the observation in the feature space does not vary with space and therefore should be independent in nature, and the residual model errors should not be correlated (Oshan, Smith, & Fotheringham, 2020).

The Spatial Lag Model (SLM) has an assumption of spatial dependency between the explanatory and response variables in feature space and conceptualizes the global regression by incorporating spatial dependence attributes in the modeling process. The SLM also assumes to have spatially lagged dependent variable in the model estimation, which can be ensured by the spatial dependence test resulted from OLS. If the determinant factors, tested by Moran's I (error), Lagrange Multiplier (lag) and Robust LM (lag), exhibited statistically significant estimates at a defined probability level, then one should reconsider the model selection process and go for SLM as a replacement for OLS. The SLM can be formulated as:

$$y_i = \beta_0 + \beta x_i + \rho W_i y_i + \varepsilon_i \quad (2)$$

Where ρ is the spatial lag component; W_i contains spatial weights (spatial weights matrix in a row format). The spatial weight matrix was generated using multiple approaches, including the contiguity based methods (Queen contiguity and Rook contiguity) and the distance based methods (Euclidean distance, Arc distance and Manhattan distance). The contiguity-based weight was approximated using the first order of contiguity. The county unique identifier number was utilized as a base for weight calculation. Since the accuracy and performance of all the global regression models strongly rely on spatial weights, we adopted both contiguity and distance-based weights for comparing the results at various parameter setups. The reduced version of the SLM can be

¹ Link: <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>

expressed as:

$$Y = A^{-1}X\beta + A^{-1}\varepsilon \quad (3)$$

Where $A=I-\rho W$; I refers to the conformable identity matrix; A^{-1} is the spatial multiplier effect or Leontief inverse (Anselin, 2002; Lambert, Brown, & Florax, 2010). This inverted A matrix distinguishes this model from other spatial regression models as it gets feed-back/-forward effects of shocks between the defined spatial location and eventually makes the model sufficiently flexible to process spatial non-linearity (Lambert et al., 2010).

The Spatial Error Model (SEM) is an extension of global models that fundamentally stands on the assumption of spatial dependence in the residual error of OLS (Chi & Zhu, 2008; Fang et al., 2015; Guliyev, 2020; Song, Du, Feng, & Guo, 2014; Yang & Jin, 2010). The SEM posits that spatial autocorrelation among regression residuals is thus evident. Two standard spatial dependence tests, Lagrange Multiplier (error) and Robust LM (error), were executed to ensure statistical significance in spatial dependency in error terms, specified as follows.

$$y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it} \quad (4)$$

$$\varepsilon_{it} = \lambda W_{et} + \nu_{it} \quad (5)$$

Where λW_{et} is the spatial error term; λ denotes the autoregressive factor; ν_{it} refers to the random error term, which is normally conceptualized to be independent and ideally distributed in feature space; ε_{it} refers to the spatially uncorrelated error term (Guliyev, 2020). The SEM consists of two error terms, W_{et} and ε_{it} . The spatial dependence test derived from OLS suggested a statistically significant spatial dependency among the observations for SLM and SEM. To provide multiple perspective of model estimations, this study considered all the three standard global spatial regression models for modeling and subsequent interpretation. Meanwhile, the spatial dependence test showed that both LM (lag and error) and Robust LM (lag and error) exhibited the statistical significance estimates. Therefore, both SEM and SLM were utilized to assess the synergies and tradeoffs between COVID-19 counts and associated factors at the county scale. When estimating the global models, both dependent and independent variables were converted to cumulative sum units. Additionally, the global associations between the variables were assessed for all the seven sub-components for capturing the individual effect of each sub-component on COVID-19 counts over the feature space.

2.3.2. Local regression

In many real-life cases, the general global assumptions and spatial stationarity among the observations in feature space could be ineffective and thus produce inelastic and biased estimates at the local scale. Since the main objective of this research is to establish predictive spatial models at the local scale, two most used local spatial regression models, Geographically Weighted Regression (GWR) and Multiscale GWR (MGWR), were employed for local spatial regression modeling and result interpretation. The GWR model is developed following Toddler's first law of geography, "everything has some relationship with others, but near things are more related compared to distant things". In GWR, each observation in feature space can vary and hence be associated with locally varying coefficients of the regression parameters. This addition of local spatial context in GWR modeling favors exploring the spatial dependency among the parameters. GWR can be defined as:

$$y_i = \beta_{i0} + \sum_{j=1}^M \beta_{ij}X_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, N \quad (6)$$

Where y_i is the dependent variable (COVID-19 case or death counts) in county i ; β_{i0} refers to the regression intercept; β_{ij} refers to the independent regression parameter; X_{ij} is the value of the j th regression parameter; ε_i refers to the regression error.

Although GWR models have been embraced as a solution for global spatial stationarity in regression estimates, the same has been suffered in cases when a constant and straightforward bandwidth is not able to detect the spatial non-stationarity at varying bandwidths across the feature space. To address this problem, Fotheringham, Yang, and Kang (2017) and Oshan, Wolf et al. (2019) proposed a multiscale and multi bandwidth GWR, which allows exploring the local relationships among the varying factors across spatial scales by computing shifting bandwidth based on the distributions of observation. MGWR can be defined as:

$$y_i = \sum_{j=1}^M \beta_{bwj}X_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, N \quad (7)$$

Where β_{bwj} refers to the differential bandwidth at feature space. The rest is the same as discussed in GWR.

2.4. Variable importance

Machine Learning models have been used extensively in measuring feature importance in multi-parameter models. This research utilized a supervised machine-learning algorithm, Random Forest, for spotting the key explanatory factors in the models. Random Forest models (Breiman, 2001), fundamentally based on bootstrap aggregating of decision trees, can minimize the unexplained variance of models and thus improve prediction accuracy (Altmann, Tološi, Sander, & Lengauer, 2010). Random Forest models have been utilized for many domain-specific studies, such as gene expression-based cancer classification (Okun & Priisalu, 2007), biology of ageing (Fabris, Doherty, Palmer, De Magalhaes, & Freitas, 2018), remote sensing land cover mapping (Ma et al., 2017; Zhang, Yang, Wang, Zhan, & Bian, 2020; Zhang, Wang et al., 2020), screening underlying lead compounds (Cao et al., 2011), Structure damage detection (Zhou, Zhou, Zhou, Yang, & Luo, 2014). In this study, we measured the variable importance based on the overall capacity of the variables to explain the total model variances. Relative Importance and normalize importance scores were also computed for each variable to verify the predictive accuracy of the models and the individual contribution of each variable to the overall model performances.

2.5. Experimental design

In this study, we structured the entire analysis into a few sequential and logical steps (Fig. 1). The global and local spatial regression analysis has been carried out through four separate models:

2.5.1. Model 1: global regression model considering static dependent and independent variables

Model 1 was conceptualized for conducting global regression analysis between COVID-19 counts and the explanatory factors. The daily COVID-19 observations from January 22 to July 26 were converted to cumulative sum for changing the nature of the data from dynamic to static. Only the final filtered variables for cases and deaths were considered in Model 1. Group-wise assessment was not considered in Model 1. The final selected variables, 6 for cases and 6 for deaths, had exhibited acceptable VIF scores. This suggests that the multicollinearity problems in the model appeared not evident for all the multi-parameters regression models. All the global models, including OLS, SEM and SLM, were conducted using the GeoDa and GeoDa Space software. The first order Queen and Rook contiguity was applied for spatial weight estimation. The distance-based approach was utilized for generating the spatial weights of the observations. Specifically, the Euclidian distance method was adopted for distance-based spatial weight calculation.

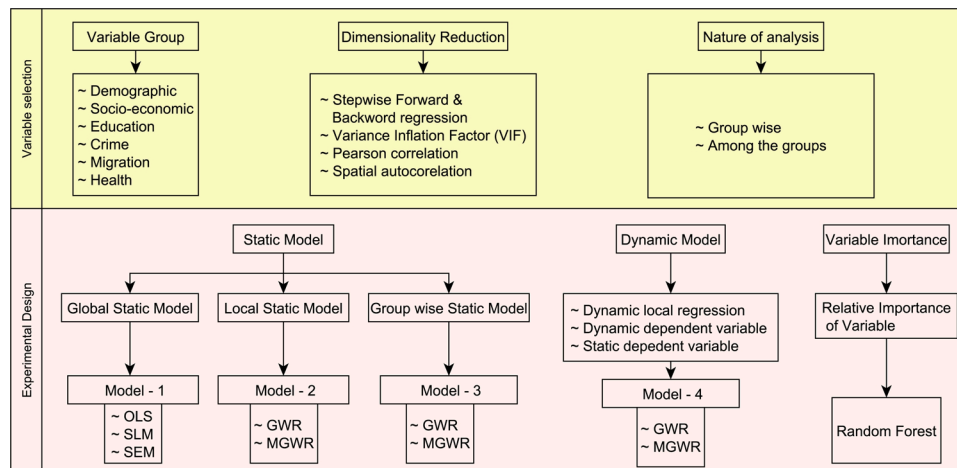


Fig. 1. Flowchart of the research methods and data analysis procedures in detail.

2.5.2. Model 2: local regression model using static dependent and independent variables

Model 2 was developed by incorporating both static independent and static dependent variables into the modeling process. Local GWR and MGWR modeling was undertaken to explore the local correlation and association between the explanatory and response variables. Both GWR and MGWR were performed with the MGWR software package (Oshan, Li, Kang, Wolf, & Fotheringham, 2019). For Model 2, only the final filtered variables (6 for cases and 6 for deaths) were taken as independent variables. Using these variables, seven parameters local regression models were developed for COVID-19 counts, with the cumulative sum values accounted.

2.5.3. Model 3: group-wise local regression model using static dependent and independent variables

Model 3 was conceptualized after incorporating group-wise (crime, demography, education, employment, ethnicity, health, and migration) variables into the modeling process. Using the stepwise forward and enter regression method, the filtered variables with VIF smaller than 4 for each group was identified. Among the seven major groups, a total of two variables (county population agency report crimes and ARSON), one variable (female age 85+), two variables (less than a high school diploma 2014–2018 and bachelor's degree or higher 2014–2018), three variables (unemployed 2018, median household income 2018, Median household income percent of state total 2018), two variables (HBAC_MALE and NH_FEMALE), two variables (Geriatric Medicine and Preventive Medicine), and three variables (Population estimate 2018, domestic migration 2018, and R international migration 2018), were selected for crime, demography, education, employment, ethnicity, health, and migration, respectively, for developing the local regression models regarding COVID-19 cases. Similarly, for COVID-19 deaths, totally two variables for crime (Robbery, Motor vehicle thefts), one variable for demography (female age85+), one variable for education (bachelor's degree or higher 2014–18), three variables for employment (unemployed 2018, median household income 2018, median household income percent of state total 2018), two variables for ethnicity (HBA Female, BA Female), one variable for health (endocrinology diabetes and metabolism specialists (2019)), and four variables for migration (Pop estimate 2018, domestic migration 2018, R international migration 2018, and R domestic migration 2018), were considered.

2.5.4. Model 4: dynamic local regression model using dynamic dependent and static independent variables

In Model 4, the monthly COVID-19 cases and deaths were chosen to be the dependent variables, while the annually averaged static group

variables were considered to be the independent variables. The monthly sum values of COVID-19 cases and deaths were derived for March, April, May, Jun, and July. A total of ten (five for cases and five for deaths) multi-parameter local spatial regression models were developed for exploring the dynamic associations between the response and the explanatory factors. The final filtered variables (six for cases, including ARSON, median household income 2018, median household income percent of the state total 2018, HBA male, domestic migration 2018, R international migration 2018; six for deaths, including median household income 2018, median household income percent of state total 2018, HBA Female, domestic migration 2018, R international migration 2018, and R domestic migration 2018) were incorporated for the dynamic local regression modeling.

The adaptive bi-square spatial kernel weighted method was employed for approximating the kernel bandwidth for GWR and MGWR models. The default golden bandwidth search approach was chosen for computing uniform (GWR) and locally varying (MGWR) bandwidths. Among the different optimization criteria, AICc, AIC, BIC, and CV, the AICc approach was considered for selecting the optimal bandwidth over feature space. Local correlation diagnostics, including condition number (CN), local spatial VIF, local variance decomposition proportions (VDP), were computed for evaluating the local collinearity among the observations and parameters. Bandwidth confidence intervals were also measured at different levels of probability to ensure reliable spatially varying bandwidths, derived from MGWR.

3. Results

3.1. Spatial patterns of COVID-19 cases and deaths in the contiguous United States

Spatial distributions and patterns of COVID-19 cases and deaths per 10,000 people in the contiguous United States is shown in Fig. 2. Multiple spatial clusters of simultaneously high numbers of cases and deaths are formed, which exhibit an unequal and heterogeneous distribution of COVID-19 counts across the counties. Among the clusters, four main clusters can be identified throughout the entire study period. The first cluster is formed over the North-Eastern coastal region, covering Massachusetts, Washington D.C., Maryland, Connecticut, Pennsylvania, and New Jersey, as well as part of New York (New York City in particular). The second cluster is observed in the South-Eastern region, which covers states of Mississippi, Alabama, Georgia, South Carolina, North Carolina, and Florida. The third cluster is detected in the Great Lakes region – Michigan, Wisconsin and Illinois, centered at Chicago of Illinois, one of the largest cities in the country. The last is located in the South-Western

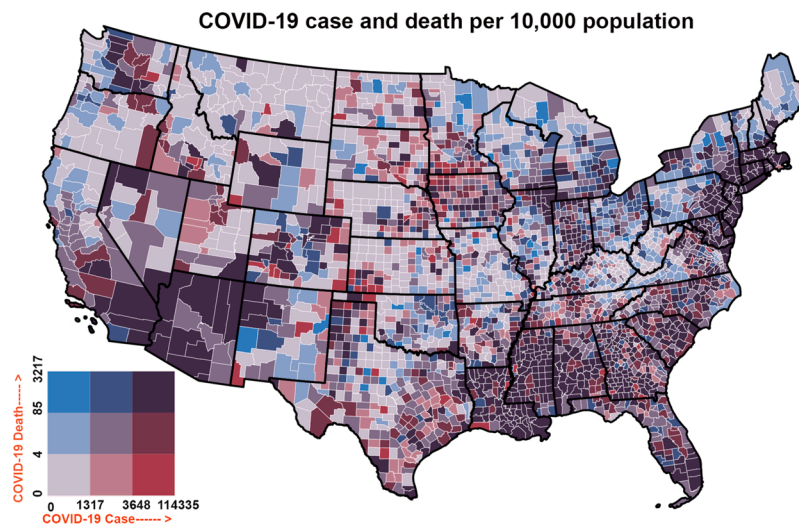


Fig. 2. Bivariate choropleth map demonstrates the county wise distribution (per 10,000 population) of COVID-19 cases and deaths from January 22 to July 26, 2020.

region including southern California, Arizona, New Mexico (north-western part), and Colorado is also notably among the areas with high numbers (Fig. 2).

3.2. Association between explanatory factors and COVID-19 cases and deaths

3.2.1. Model 1: static global regression analysis

Three global regression models, OLS, SLM and SEM, reveal the global and spatial non-stationary associations between the explanatory factors and the numbers of COVID-19 cases and deaths (Table 1).

For COVID-19 cases, the coefficient of determination (R^2) statistics, which denote the overall model strength and robustness, are measured as 0.78, 0.80, and 0.80 for OLS, SLM and SEM, respectively. The spatial dependence diagnostics criteria for the OLS model, namely LM Lag and LM error, are found statistically significant, thus indicating the requirement of more appropriate and relevant global models, such as SLM and SEM (Table S2). The AIC value, which denotes the overall model accuracy and parsimonious character of the models, is shown to be the lowest (most relevant) for SLM, followed by OLS and SEM. This suggests that the SLM model can be a more relevant global regression model with a better explanation of the model variability. Regarding the

Table 1
Global regression estimates derived from OLS, SLM, and SEM.

Variable	Cases								
	Ordinary Least Square			Spatial Lag			Spatial Error		
	Coefficient	t-Statistic	Probability	Coefficient	z-Statistic	Probability	Coefficient	z-Statistic	Probability
Case	—	—	—	0.34	23.34	0.00	—	—	—
CONSTANT	-120958	-8.49	0.00	-54546.7	-4.1	0.00	-103165	-6.02	0.00
ARSON	825.14	14.88	0.00	1073.06	21.2	0.00	1291.49	23.27	0.00
MHHInc	2.27559	5.36	0.00	-0.04	-0.1	0.92	2.32	3.40	0.00
MHHIncPer	224.50	0.78	0.43	633.62	2.43	0.02	20.76	0.05	0.96
HBACM	62.27	52.79	0.00	46.78	39.3	0.00	44.04	35.04	0.00
DomMig	-22.42	-20.13	0.00	-21.33	-20.94	0.00	-22.37	-20.73	0.00
RIntMig	274.51	0.17	0.86	-392.91	-0.27	0.78	-87.65	-0.06	0.95
Lambda	—	—	—	—	—	—	0.55	26.04	0.00
R ²	0.76	—	—	0.80	—	—	0.80	—	—
Adj. R ²	0.76	—	—	—	—	—	—	—	—
F	1611.37	—	—	—	—	—	—	—	—
P	0.00	—	—	—	—	—	—	—	—
AIC	83,825	—	—	83325.90	—	—	83458.30	—	—
SIC	83867.30	—	—	83374.30	—	—	83500.60	—	—
Deaths									
Death	—	—	—	0.73	56.08	0.00	—	—	—
CONSTANT	-9136.95	-5.65	0.00	332.33	0.29	0.76	1988.94	1.02	0.30
MHHInc	0.29	6.27	0.00	-0.05	-1.48	0.13	-0.13	-1.69	0.09
MHHIncPer	-48.05	-1.50	0.13	28.54	1.28	0.19	88.23	1.83	0.06
DomMig	-3.91	-37.32	0.00	-2.82	-37.14	0.00	-2.73	-36.2	0.00
RIntMig	1106.60	6.47	0.00	510.21	4.29	0.00	250.68	2.00	0.04
RDomMig	123.34	3.78	0.00	142.51	6.28	0.00	81.82	3.02	0.00
LAMBDA	—	—	—	—	—	—	0.81	63.64	0.00
R ²	0.36	—	—	0.69	—	—	0.69	—	—
Adj. R ²	0.36	—	—	—	—	—	—	—	—
F	349.83	—	—	—	—	—	—	—	—
P	0.00	—	—	—	—	—	—	—	—
AIC	70200.90	—	—	68324.50	—	—	68429.20	—	—
SIC	70237.10	—	—	68366.80	—	—	68465.50	—	—

Notes: MHHInc – Median household income, MHHIncPer – Median household income percent, DomMig – Domestic Migration, RIntMig – Rate of International Migration, HBACM – Not Hispanic, Black or African American alone or in combination male population, RDomMig – Rate of Domestic Migration, ARSON – Arson.

Table 2
Group wise GWR and MGWR estimates computed from COVID-19 cases and deaths.

Factors	Cases											
	R ²		Adj. R ²		Adj. alpha (95 %)	Adj. critical t value (95 %)	AIC		AICc		BIC	
	GWR	MGWR	GWR	MGWR	GWR	GWR	GWR	MGWR	GWR	MGWR	GWR	MGWR
Crime	0.96	0.953	0.954	0.95	0	3.577	-333.214	-272.611	-197.511	-241.079	2240.615	1014.878
Demography	0.935	0.93	0.927	0.925	0	3.612	989.264	973.424	1065.608	1002.018	2954.972	2201.44
Education	0.961	0.958	0.955	0.953	0	3.572	-395.974	-388.113	-265.737	-316.194	2129.215	1522.801
Employment	0.96	0.963	0.953	0.955	0	3.54	-220.113	-362.583	-33.805	-158.134	2758.259	2744.779
Ethnicity	0.953	0.952	0.946	0.946	0	3.572	136.66	81.321	266.864	171.804	2661.55	2211.106
Health	0.412	0.439	0.332	0.398	0	3.542	7915.288	7450.479	8017.685	7482.009	10172.44	8737.93
PopMig	0.964	0.962	0.957	0.957	0	3.552	-469.484	-574.647	-263.696	-462.762	2647.13	1778.057
All Variables	0.964	0.969	0.954	0.961	0.001	3.473	-133.397	-737.655	238.838	-434.883	3930.356	2971.075
Deaths												
Crime	0.936	0.941	0.927	0.934	0	3.566	1077.624	701.324	1202.174	782.08	3550.934	2719.931
Demography	0.892	0.887	0.879	0.879	0	3.612	2555.903	2460.028	2632.247	2490.847	4521.611	3733.372
Education	0.779	0.781	0.762	0.77	0	3.515	4577.562	4401.507	4612.928	4417.641	5938.388	5331.019
Employment	0.948	0.953	0.938	0.944	0	3.54	646.627	334.34	832.936	538.789	3624.999	3441.702
Ethnicity	0.925	0.926	0.914	0.918	0	3.546	1542.145	1303.461	1647.646	1361.208	3831.098	3025.062
Health	0.936	0.939	0.929	0.932	0	3.607	909.58	755.337	982.54	828.208	2833.558	2678.189
PopMig	0.98	0.98	0.975	0.977	0	3.549	-2090.17	-2484.22	-1759.22	-2371.52	1768.129	-123.531
All Variables	0.964	0.97	0.954	0.962	0.001	3.472	-138.548	-731.431	230.621	-358.146	3910.451	3337.357

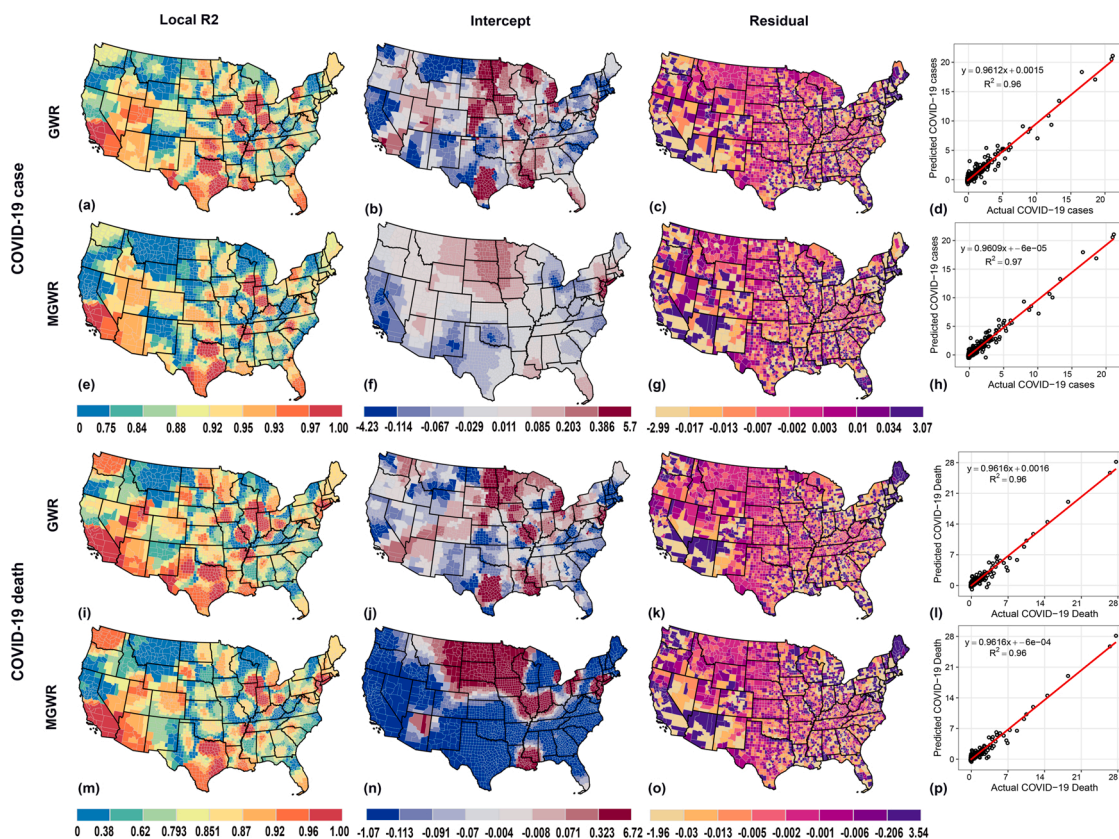


Fig. 3. Local associations between the confounding factors and COVID-19 incidences derived from GWR and MGWR. Model strength and spatial interactions of the parameters were demonstrated by local R², intercept, and residual.

correlations of the explanatory variables², ARSON, MHHInc, MHHIncPer, and HBACM have positive correlations with the number of COVID-19 cases. Among these four covariates, HBACM is found to have the most statistically significant relationship with the number of cases,

² The explanatory variables have different units with different value ranges, hence their coefficients are not comparable; the associated t-statistics (OLS) and z-statistics (SLM and SEM) instead can be compared in terms of the significance level of the associations.

given that its t-/z-statistic is the highest among the three models (52.79, 39.3, and 35.04 for OLS, SLM, and SEM, respectively). The next with substantial significant coefficients is ARSON, with its t-/z-statistics being 14.88, 21.2, and 23.27 in OLS, SLM, and SEM, respectively. MHHInc and MHHIncPer are found to have smaller significance values. The former is statistically significant (at the 5% level) only in SLM, while the latter is statistically significant (5% level) in OLS and SEM. Meanwhile, DomMig is found negatively (statistically significantly) correlated with COVID-19 cases in all three models. Last, RintMig shows statistically insignificant

associations with the cases, although the associations differ across the models.

Moving onto the number of COVID-19 deaths, the R^2 values are 0.36, 0.69, and 0.69 for OLS, SLM, and SEM, respectively. The AIC value is found to be the lowest in SLM, compared to those in OLS and SEM, indicating that the SLM model performs better under the given modeling framework. To interpret the explanatory variables, DomMig, RIntMig, and RDomMig are significantly associated with the number of deaths for all three models, and their associating directions are consistent. Specifically, RIntMig and RDomMig covariates are positively correlated with deaths, while DomMig (the one with the highest significance level measured by t-/z-statistics) negatively. As for MHHInc and MHHIncPer, however, the correlations between MHHInc and deaths are observed statistically significant in OLS and SEM, but the correlating directions are inconsistent between the two models of OLS and SEM; MHHIncPer is found to be significantly associated with deaths in only SEM and their relating direction is positive.

3.2.2. Model 2: static local regression analysis

The (M)GWR-derived local spatial heterogeneity of the determinant factors for COVID-19 cases and deaths are statistically and spatially displayed in Table 2 and Fig. 3, respectively. These numbers and figures collectively demonstrate the spatial variability of the local model at the county scale in the contiguous United States. Local R^2 estimates for both local regression models, MGWR and GWR, show high degrees of spatial agreement. The counties, for which the highest R^2 (i.e., $R^2 > 0.90$) values are derived, form spatially clustered patterns across the country. The high values of local R^2 are concentrated over the Wisconsin-Indiana-Michigan region, as well as several parts of states of Texas, California, Mississippi and Arkansas. The lowest R^2 scores are found in the Northern and North-Western states (Montana, Washington, Oregon, Wyoming), Southern states (New Mexico) and North-East coast region (North

Carolina and Georgia). For COVID-19 deaths, the spatial patterns of high, moderate and low R^2 values appear similar to those of the COVID-19 cases. Among the two local spatial regression models, MGWR performs more accurately, as it has slightly higher Adj. R^2 values (for cases, $R^2 = 0.961$; for deaths, $R^2 = 0.962$), compared to GWR's Adj. R^2 values (for cases, $R^2 = 0.954$; for deaths, $R^2 = 0.954$). Also, AICc values of the MGWR model (for cases, $AICc = -434.883$; for deaths, $AICc = -358.146$) are found much lower than those of GWR (for cases, $AICc = 238.888$; for deaths, $AICc = 230.621$), as shown in Table 2 and Fig. 3.

3.2.3. Model 3: group-wise static local regression analysis

The spatial associations between different groups (crime, demographic, education, employment, ethnicity, health and migration) and COVID-19 cases and deaths are depicted in Figs. 4 and 5. Among the seven groups, six groups viz. *demography, crime, education, ethnicity, employment, and population migration* show strong similarities in terms of their spatial patterns of local R^2 . The highest local R^2 values ($R^2 > 0.90$) are found in the Southern and South-Western states, mainly Texas, Arizona, California, Utah; in the Eastern United States, or the Wisconsin-Michigan-Indiana-Illinois region; in the tri-state area of Mississippi-Arkansas-Alabama. In contrary, the *health* factor exhibits a different association with the COVID-19 numbers. High local associations between the health factor and the COVID-19 cases are found in the Colorado-Utah and New Hampshire areas. For all groups, low spatial associations are found in states of Montana, North Dakota, Idaho, Oregon. Based on the R^2 and AICc values, the *population migration* factor is found to be the most critical component with the highest local estimates ($R^2 = 0.96$, $AICc = -462.76$), followed by *education* and *crime*. A similar spatial association is detected between the explanatory factors and COVID-19 deaths across the counties. High local associations are found over the South, South-Western United States (states of Texas, New

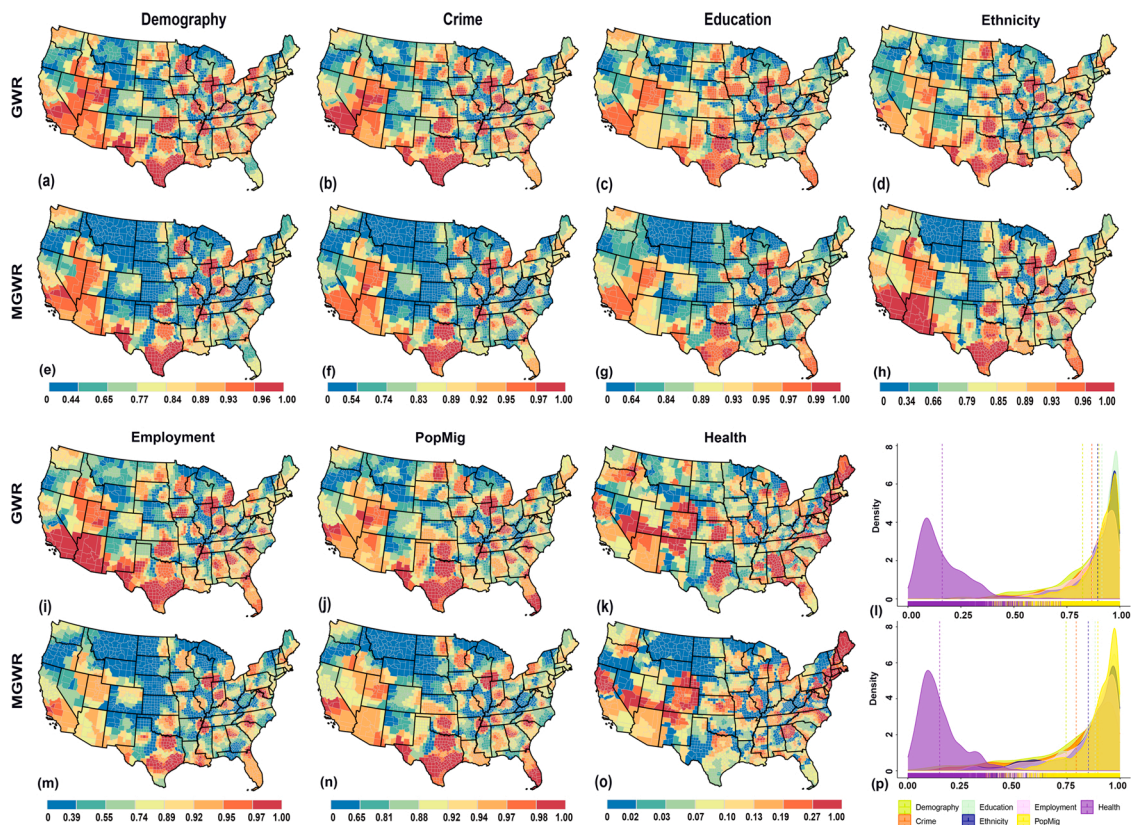


Fig. 4. Local effects of the driving factors (Demography, Crime, Education, Ethnicity, Employment, PopMig, and health) on COVID-19 cases at county scale derived from GWR and MGWR.

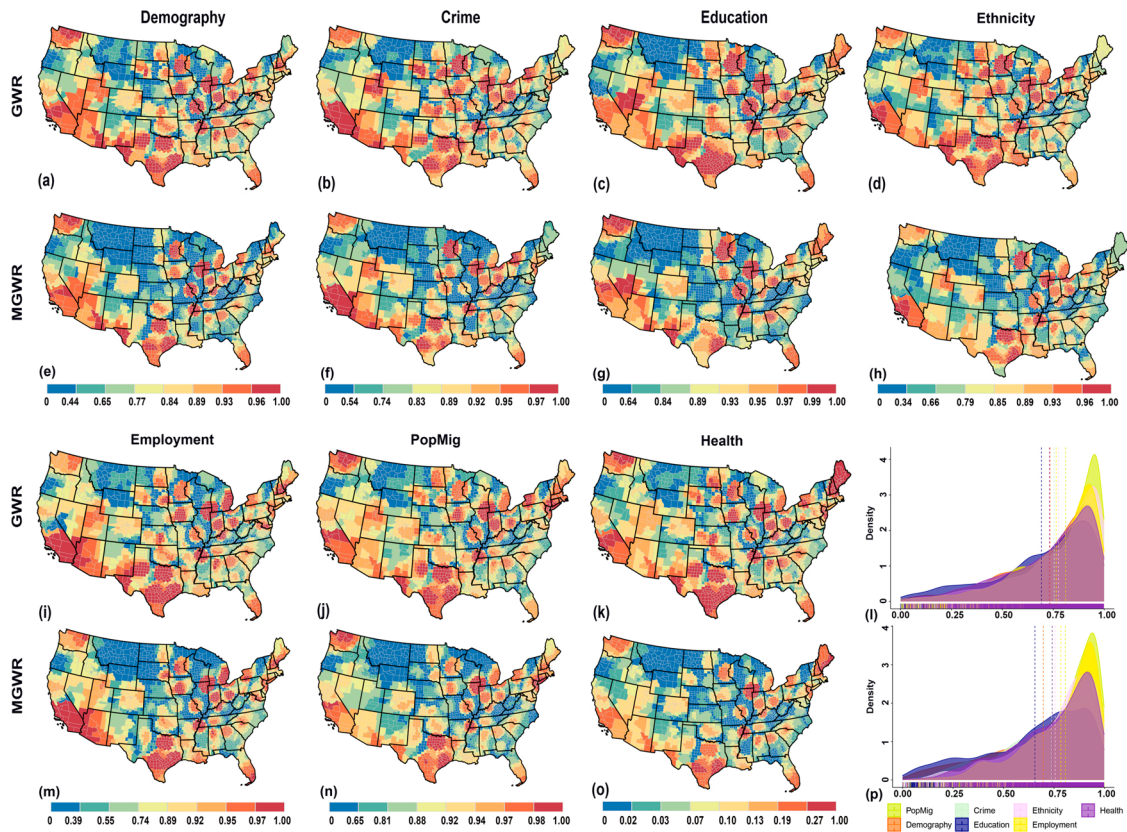


Fig. 5. Local effects of the driving factors (Demography, Crime, Education, Ethnicity, Employment, PopMig, and health) on COVID-19 deaths at county scale derived from GWR and MGWR.

Mexico, Arizona, and California) and the East Central states (Wisconsin, Michigan, Indiana, and Illinois). The *Population and migration* factors explains the maximum model variability ($R^2 = 0.98$ and $AICc = -2371.52$), followed by an order of *employment, crime, health, ethnicity, demography, and education* (Table 2).

3.2.4. Model 4: dynamic local regression analysis

Spatial and temporal associations between the final six selected factors and COVID-19 counts are presented in Figs. 6 and 7, and Table 3. Totally ten (five for cases and five for deaths) local regression models reveal local associations between the explanatory factors and COVID-19 counts in each of the five months, namely March, April, May, June, and July. High spatial associations between the explanatory variables and the response variables are found in states of Texas, New Mexico, Mississippi, Tennessee, Kentucky, Indiana, Illinois, Wisconsin and Michigan ($R^2 > 0.90$). In April and May, high spatial associations are found in Florida and California. In June and July, Arizona, Nevada, Oregon, Idaho states exhibit high spatial associations, characterised by large local R^2 values. On the contrary, low spatial associations are observed in Washington, Oregon, Idaho, Montana, North Dakota, and South Dakota. For COVID-19 deaths, the local association follows a similar pattern as observed for the cases. In March, a high spatial association is seen in the Wisconsin and Illinois states. In the later months, high spatial associations are shifted to multiple locations, such as Texas, California, Utah, Idaho, Wyoming region, Arkansas, Mississippi, Tennessee. On the contrary, low spatial associations are found in the northern (i.e., Montana and North Dakota) and eastern states (i.e., Florida, Georgia, and South

Carolina). All the dynamic models demonstrate the superiority of MGWR, as it is found to be a well-suited model for the local regression analysis throughout the study (Table 3, Figs. 6 and 7).

3.3. Variable importance

The levels of Relative Importance of the selected variables (final filtered variables, six for cases and six for deaths) measured by the Random Forest machine-learning model are presented in Fig. 8. For COVID-19 cases, among the variables, the highest level of Relative Importance is found for HBACM (44.31 %), followed by DomMig (15.56 %), ARSON (12.38 %), RIntMig (10.53 %), MHHIncPer (5.22 %), and MHHInc (3.7 %), respectively (Fig. 8a). For COVID-19 deaths, the HBAF explains the maximum variances, and therefore, the highest RI score appears in HBAF (26.56 %), followed by DomMig (13.23 %), RDomMig (8.07 %), MHHInc (6.84 %), RIntMig (5.88 %), and MHHIncPer (0.76 %), respectively (Fig. 8b).

4. Discussion

It has been nearly one year since the outbreak of COVID-19 started in Wuhan (China) and spread across the globe. The situation yet remains globally elusive as many countries have witnessed the re-emergence of COVID-19 incidents. Among all the countries, the United States is facing the most critical challenge in flattening the curve with urgent needs for more effective and appropriate control measurements. To inform the policy-makers at both national and state levels, understanding the

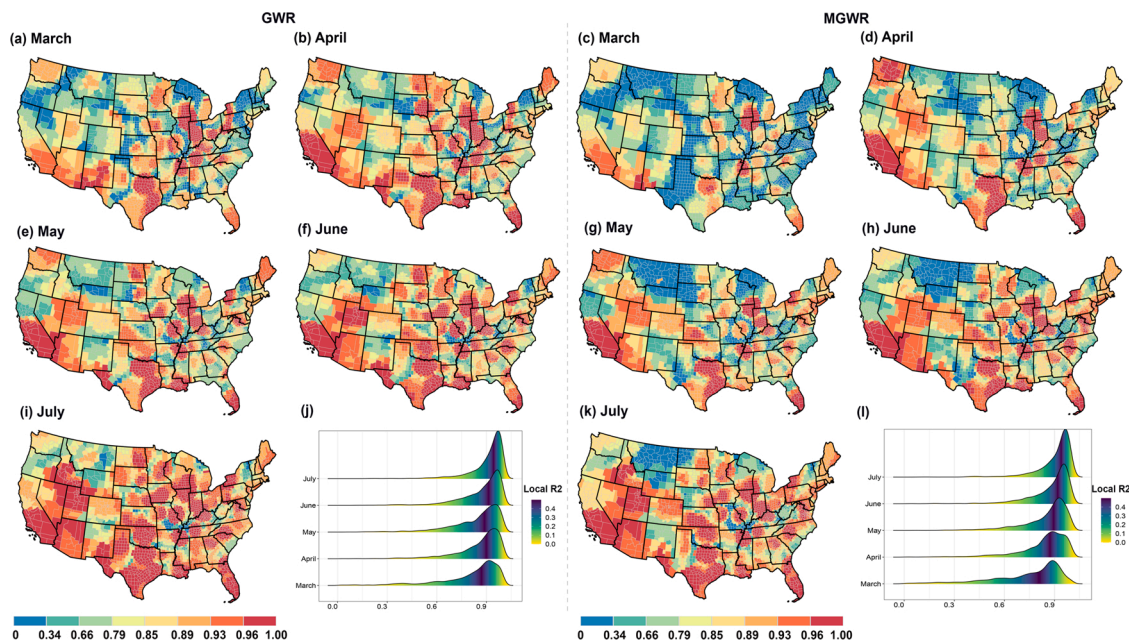


Fig. 6. Time-varying effects of the confounding factors on COVID-19 cases based on GWR and MGWR.

explanatory drivers and related confounding factors with spatial patterns and is of paramount importance. Timely studies have done much work of doing so (e.g., Beria & Lunkar, 2020; Hu, Roberts, Azevedo, & Milner, 2020; Rahman et al., 2020). However, this may not uncover the full picture since most of the factors change over time, namely being time-variant variables. The present study contributes to forwarding the knowledge of the outbreak by examining a set of factors over space and across time. Specifically, the most relevant variables are teased out from a large group of potential factors for explaining the COVID-19 cases and deaths at the county level, as well as for each month covering a five-month study period (Table 4).

Choosing the best models when taking into account spatial and temporal features have always been a crucial point in spatial epidemiological research. Previously, several methodological approaches have evolved to capture the influence of explanatory variables on the response variables in the epidemiological study (Bashir et al., 2020). Among these are Spearman's, Pearson's and Kendall's Correlation Coefficient, Ordinary Least Square regression (Méndez-Arriaga, 2020), Poisson regression, Distributed Lag Nonlinear Model (Runkle et al., 2020), cluster-based analysis (Andersen, Harden, Sugg, Runkle, & Lundquist, 2021), spatial lag model, spatial error model (Sun et al., 2020). These models are mainly global models in nature and therefore have proven ineffective to capture the local or spatial patterns between explanatory and response variables.

Based on the present research, notably, the overall regression models reveal that population migration, as indicated by domestic migration and the rate of international migration, is highly correlated with the numbers of COVID-19 cases and deaths. The move of people across continents internationally is accompanied with high risk of virus spread, as the air traveling means by its nature increases the likelihood of person-to-person COVID-19 transmissions (Zhang, Yang et al., 2020; Zhang, Wang et al., 2020). Given this evidence, air flight restrictions could be effective in undermining the virus spread, which is in line with the conclusion of positive associations between travel restrictions and COVID-19 spread from previous findings (Christidis & Christodoulou, 2020), although this involves trade-offs between air-transporting public health and social-economics risks (Cotfas, Delcea, Milne, & Salari, 2020). The other population moving variable, domestic migration, is found to be negatively related to numbers of both cases and deaths, which may be because that the redistribution of population from high

density areas (e.g., megacities) to low population density areas (e.g. mountainous suburban regions) can diffuse the infected people while decreasing the frequency of person-to-person contact. A study suggests that residents from New York City, especially those in high wealth status, tend to flee the city to lower physical exposure to COVID-19 (Coven & Gupta, 2020). Apart from domestic migration and population flows that have been recorded during the outbreak, the intra/inter city and county transport connectivity plays a crucial role in spreading the disease especially at the early transmitting phase. Although this study includes both domestic and international migration into the assessment, the explicit role of transport network in transmitting the virus spatially is not focused. It should be noted that this relationship is based on the overall regression model, lacking heterogeneity over time and space. Socioeconomically, median household income at the county level is shown to be positively related to COVID-19 spread, as it indicates that the larger cities and higher population densities with more burden of virus transmissions.

Interestingly, when viewing different time periods (monthly from March to July) as revealed from the dynamic local regression analysis, there exists high spatial heterogeneity in how the explanatory variables are associated with COVID-19 cases and deaths. Such heterogeneity is dynamic over time, which is also supported by the better performance of MGWR than GWR (Figs. 6 and 7). In the early phase of the COVID-19 outbreak (mainly in March), associations between the potential factors and the infected numbers in most regions have not been well manifested except for the Chicago-centred Great Lake region and the Tennessee-Arkansas-Mississippi region (Fig. 6c). However, since April, several prominent hotspots of such correlations have been discovered including the states of California and Florida as well as many regions in the middle east part of the country (Fig. 6d, g, h, k). These regions identified as hotspots have characteristics of high population densities and hence the outbreak outcomes are more likely to be explained by the selected factors, particularly the migration-related variables of domestic migration behaviours. This implication again demonstrates the importance of controlling people mobility as effective measures to combat the virus spread by the government in high populated states (Badr et al., 2020), as those actions taken in other countries including China (Kraemer et al., 2020). In terms of COVID-19 deaths, the spatial patterns of the modeling outcomes also begin to exhibit high explanatory powers over large scales after April and remain stable during April-July, covering most of the

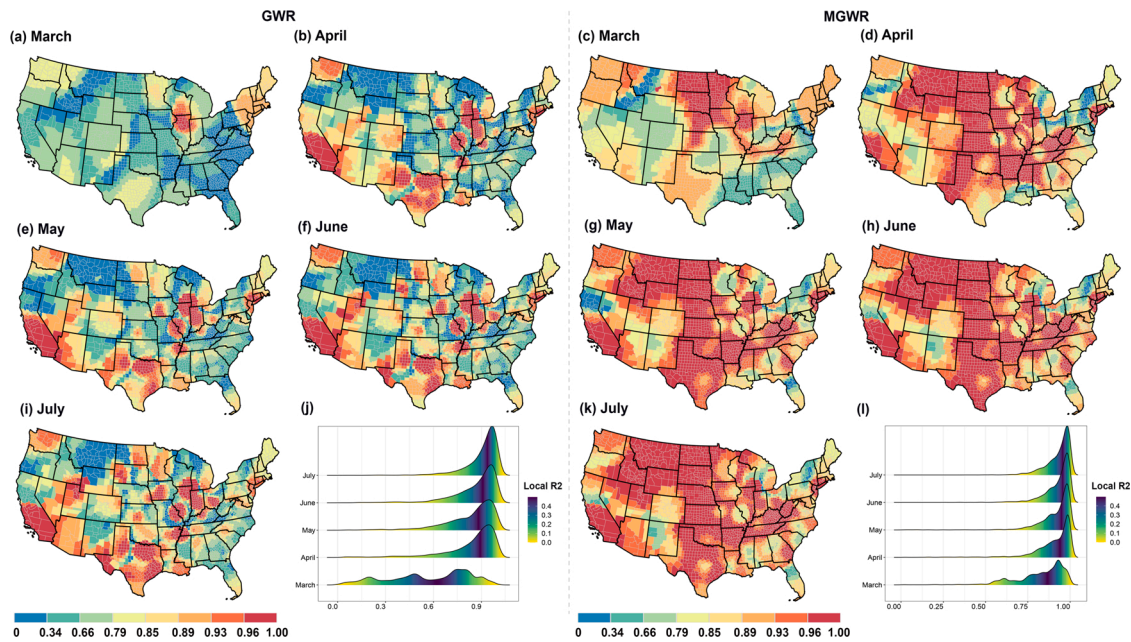


Fig. 7. Time-varying effects of the confounding factors on COVID-19 deaths based on GWR and MGWR.

Table 3

Month wise GWR and MGWR estimates for cases and deaths.

Months	Cases													
	R ²		Adj. R ²		Adj. alpha (95 %)		Adj. critical t value (95 %)		AIC		AICc		BIC	
	GWR	MGWR	GWR	MGWR	GWR	MGWR	GWR	MGWR	GWR	MGWR	GWR	MGWR	GWR	MGWR
March	0.886	0.887	0.858	0.87	0.001	3.447	3290.311	2854.937	3588.543	2974.945	6971.588	5284.201		
April	0.931	0.944	0.914	0.932	0.001	3.447	1719.785	943.653	2018.018	1169.966	5401.063	4195.497		
May	0.953	0.962	0.941	0.953	0.001	3.447	541.379	-141.82	839.612	158.405	4222.657	3550.403		
Jun	0.966	0.971	0.956	0.964	0.001	3.473	-332.63	-939.287	40.276	-631.086	3731.492	2796.307		
July	0.974	0.976	0.966	0.97	0	3.49	-1077.61	-1533.93	-647.896	-1217.58	3246.505	2245.248		
	Death													
March	0.855	0.912	0.844	0.897	0.002	3.161	3262.754	2180.349	3297.051	2343.071	4602.697	4977.473		
April	0.957	0.965	0.945	0.957	0.001	3.472	371.277	-470.549	741.026	-224.624	4420.234	2905.75		
May	0.959	0.969	0.95	0.961	0.001	3.42	-11.612	-677.245	229.337	-360.743	3333.698	3102.754		
Jun	0.963	0.969	0.953	0.96	0.001	3.472	-120.738	-586.839	249.011	-212.922	3928.219	3482.117		
July	0.962	0.966	0.951	0.957	0.001	3.472	37.51	-372.678	407.259	-6.848	4086.467	3657.341		

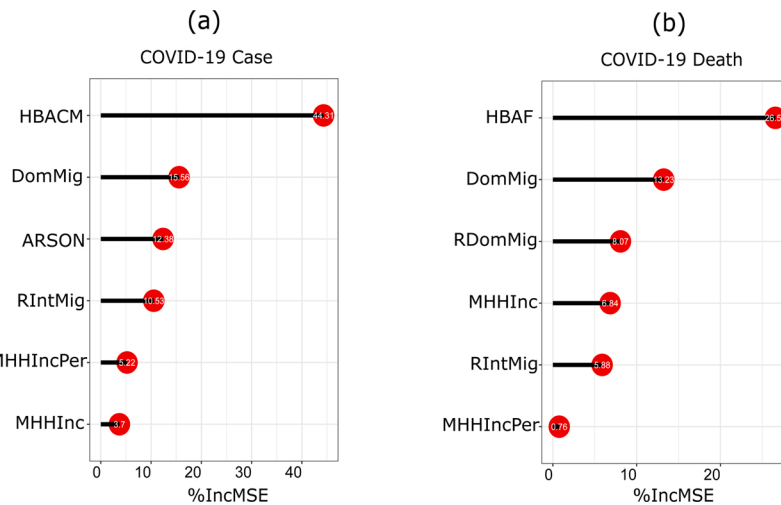


Fig. 8. Relative influence of the variables utilized for developing parsimonious regression models.

Table 4
Changes in Local R^2 values in different months.

R^2 range	Case									
	March		April		May		June		July	
	GWR	MGWR	GWR	MGWR	GWR	MGWR	GWR	MGWR	GWR	MGWR
0 - 0.34	379	851	167	362	104	311	76	214	41	145
0.34 - 0.66	384	553	300	413	366	403	219	321	133	185
0.66 - 0.79	391	449	378	462	541	410	420	384	224	304
0.79 - 0.85	347	277	356	409	306	320	352	313	243	281
0.85 - 0.89	349	341	337	397	333	321	303	308	263	299
0.89 - 0.93	465	358	492	419	452	511	428	401	447	394
0.93 - 0.96	347	161	427	253	387	428	438	468	519	509
0.96–1.00	447	119	652	394	620	405	873	700	1239	992
Death										
0 - 0.34	878	63	538	80	414	45	358	17	264	11
0.34 - 0.66	892	414	642	102	680	62	575	60	512	45
0.66 - 0.79	643	501	575	178	559	177	547	130	511	122
0.79 - 0.85	367	420	351	325	302	270	348	201	418	227
0.85 - 0.89	57	316	243	319	304	321	253	287	262	260
0.89 - 0.93	186	449	192	421	281	391	330	376	325	411
0.93 - 0.96	31	504	193	336	233	379	280	454	328	475
0.96–1.00	55	442	375	1348	336	1464	418	1584	489	1558

contiguous United States (except for a few regions such as northern California and northern New York). These results confirm that the selected factors of migration and household economic status can be useful for understanding the deaths caused by COVID-19 across counties and states during the study period.

Although the United States is equipped with best healthcare facilities in the world, the high-level response to the pandemic has been argued as inadequate and leading to “surprisingly” resurgence of COVID-19 cases in for example California³. Currently, despite the authorization of vaccines, the most effective measures to protect people from virus spread and minimize exposure risk are keeping social distances, wearing masks, and high frequency of washing hands (Badr et al., 2020). At the state level, local governments have been sufficiently vigilant to anticipate the situations and have taken preventive and protective measures (e.g. implementing anti-contagion policies) beyond federal guidance to minimize the potential damage. These government-imposed containment policies include, for instance, large event bans, school closures, and mandating social distances, which could reduce the growth of new cases (Courtemanche, Garuccio, Le, Pinkston, & Yelowitz, 2020). State travel restrictions as well as quarantine rules for out-of-state visitors have been put into practices by many states such as Vermont⁴. Educational institutions transferred from in-person classes to online meetings, or otherwise designed protocols specifying different categories of students/staff/faculty members, regular testing, restricted public room usages, etc.

However, effort has been regarded as seemingly being put in vein based on the possible rebounding trend of newly found cases⁵. Given the critics based on the fact that the contiguous United States has the size of confirmed cases far more than any other places, policy-makers have been placed on a verge of taking critically adaptive and learning actions by referring to successful examples. China, the world’s second largest economy (after the United States), has put tremendous resources for controlling virus spread (primarily through city lockdown), which was reported as effective as potentially prevented hundreds of thousands of cases outside Hubei province (World Health Organization, 2020). Challenges such as those rooted in difference in political systems are

admittedly persistent when learning from the way in which China respond to the virus crisis, yet quick actions as the Chinese government has taken should be undoubtedly encouraged as the priority by other countries (Kupferschmidt & Cohen, 2020). With more evidence accumulated for testing the underlying forces of COVID-19 spread, it is urgent to call for taking serious and sophisticated consideration by the federal government of socioeconomics and demographics especially population migration at the county or state level in addition to physical protection at the individual level. Without taking these temporally and spatially dynamic factors into account, the COVID-19 mitigation outcomes and the future of public health of the country in response to the pandemic would remain uncertain and risky.

The findings in the present studies are generally in agreement with previous investigations, meanwhile not only adding values to the existing knowledge of COVID-19 spread in the United States but also possessing international relevance for combating the crisis worldwide. Consistent with what have been previously found, several (e.g., demographic, economic) factors have played key role in determining the casualties incurred by COVID-19 across countries. Bashir et al. (2020) showed that minimum temperature and average temperature are greatly related to the spread of COVID-19 spreading in New York city. Apart from that, specific humidity are found positively related with COVID-19 in four cities – New Orleans, LA; Albany, GA; Chicago, IL; Seattle, WA (Runkle et al., 2020). Different socio-economic factors such as median household income equality are also found to be determining drivers of COVID-19 related casualties (Mollalo et al., 2020). In addition, demographic profile of the health care professional (over 55 years old population) is found substantially correlated with the disease (Dowd et al., 2020). Economic profile of the communities including unemployed population and existence of socio-economic disparities, s also found to be one of the key regulating factors of COVID-19 casualties in the United States. The present study, however, did not find any significant relationships between climate, air pollution and COVID-19 cases or deaths (Fig. S2). This finding is in line with the observation of Mollalo et al. (2020).

This research has explored local and global spatial associations between the explanatory factors and COVID-19 casualties at the county scale in the contiguous United States. This study adopts many relevant approaches and methods to allow multiple-perspective model estimates, which can further be used as a reference for similar research interest and policy design. Still, there exist unavoidable uncertainties and biases both in parameter approximation and model design. Cumulated COVID-19 deaths and cases were used as a dependent variable in the spatial models. Though, we consider the latest COVID-19 counts (COVID-19

³ Website: <https://www.latimes.com/opinion/story/2020-07-02/u-s-was-prefectly-equipped-to-beat-coronavirus-federal-government-failed>

⁴ Website: <https://accd.vermont.gov/covid-19/restart/cross-state-travel>

⁵ Websites: 1) <https://www.cnn.com/videos/politics/2020/04/12/anthony-fauci-polls-november-rebound-jake-tapper-sotu-vpx.cnn>; 2) <https://coronaviru.s.jhu.edu/testing/individual-states>

datasets from January 22 to July 26, 2020, was collected) for the modeling, there is high chance to have different estimates if the proposed models are performed considering different time frame datasets. To clearly understand this uncertainty, we compare our modeled estimates with Mollalo et al. (2020) observations; this study has conducted the analysis considering 90 days of aggregated COVID-19 data. While, in the present research, we consider 348 variables and sort out few final uncorrelated variables for the explanation of COVID-19 cases and deaths, respectively, after processing nearly 184 days of data (both aggregated and daily COVID-19 counts were considered). The final filtered variables identified in our study has not matched perfectly with others' estimation. This can be due to the difference in time frame taken between Mollalo et al. (2020) (90 days of COVID-19 data) and our study (184 days of COVID-19 data). Moreover, in our study, we consider seven groups of factors (crime, demography, education, ethnicity, employment, health, and population & migration) for the modeling and subsequent interpretation. The causal effects of the other factors, such as the lockdown date, the strictness of lockdown (partial or complete), restrictions on social gathering and human mobility, have not been explored in the present research, which can be an issue for future research.

5. Conclusion

The present research aims to explore the local and global associations between explanatory factors and COVID-19 counts in the contiguous United States with local and global spatial regression and machine-learning models. To capture the time varying effects of the potential factors on COVID-19 counts, several dynamic local parsimonious models have been conceptualized. Among the confounding factors, crime, income, and migration are found to be strongly associated with COVID-19 casualties, and hence explain the maximum model variances. Interestingly, when viewing different time periods (monthly from March to July) as revealed from the dynamic local regression analysis, there exists high spatial heterogeneity in how the explanatory variables are associated with COVID-19 cases and deaths. Additionally, both global and local associations among the parameters vary highly over space and change across time. This spatial variability of the model estimates exhibit the varying behavior of the explanatory factors and COVID-19 incidences at the county scale. Thus, the application of various models can be effective to uncover the global and local spatial associations from multiple perspectives. The findings in the present studies are generally in agreement with previous investigations, meanwhile not only adding values to the existing knowledge of COVID-19 spread in the United States but also possessing international relevance for combating the crisis worldwide. To inform policy-makers at the nation and state levels, understanding the explanatory forces and related confounding factors with spatial patterns is of paramount importance. The present study can be a reference for future spatial epidemiological research and informing decision making in the case of crisis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors are grateful to three anonymous reviewers and handling editor for making constructive comments that helps to improve the quality of the manuscript. The authors also acknowledged Dr. Prasenjit Acharya for the continuous help and support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.scs.2021.102784>.

References

- Altmann, A., Tološi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, 26(10), 1340–1347. <https://doi.org/10.1093/bioinformatics/btq134>
- Andersen, L. M., Harden, S. R., Sugg, M. M., Runkle, J. D., & Lundquist, T. E. (2021). Analyzing the spatial determinants of local Covid-19 transmission in the United States. *Science of the Total Environment*, 754, Article 142396. <https://doi.org/10.1016/j.scitotenv.2020.142396>
- Andersen, J. P., Nielsen, M. W., Simone, N. L., Lewiss, R. E., & Jagsi, R. (2020). Meta-Research: COVID-19 medical papers have fewer women first authors than expected. *Elife*, 9, e58807. <https://doi.org/10.7554/eLife.58807.sa2>
- Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3), 247–267. <https://doi.org/10.1111/j.1574-0862.2002.tb00120.x>
- Anselin, L., & Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in a single cross-section. *Papers in Regional Science*, 92(1), 3–17. <https://doi.org/10.1111/j.1435-5957.2012.00480.x>
- Auchincloss, A. H., Gebreab, S. Y., Mair, C., & Diez Roux, A. V. (2012). A review of spatial methods in epidemiology, 2000–2010. *Annual Review of Public Health*, 33, 107–122. <https://doi.org/10.1146/annurev-publhealth-031811-124655>
- Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., & Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases*, 20(11), 1247–1254. [https://doi.org/10.1016/S1473-3099\(20\)30553-3](https://doi.org/10.1016/S1473-3099(20)30553-3)
- Bashir, M. F., Ma, B., Komal, B., Bashir, M. A., Tan, D., & Bashir, M. (2020). Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Science of the Total Environment*, 728, Article 138835. <https://doi.org/10.1016/j.scitotenv.2020.138835>
- Beria, P., & Lunkar, V. (2020). Presence and mobility of the population during the first wave of Covid-19 outbreak and lockdown in Italy. *Sustainable Cities and Society*, Article 102616. <https://doi.org/10.1016/j.scs.2020.102616>
- Bolaño-Ortiz, T. R., Camargo-Cacedo, Y., Pullafito, S. E., Ruggeri, M. F., Bolaño-Díaz, S., Pascual-Flores, R., et al. (2020). Spread of SARS-CoV-2 through Latin America and the Caribbean region: A look from its economic conditions, climate and air pollution indicators. *Environmental Research*, 191, Article 109938. <https://doi.org/10.1016/j.envres.2020.109938>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cao, D. S., Liang, Y. Z., Xu, Q. S., Zhang, L. X., Hu, Q. N., & Li, H. D. (2011). Feature importance sampling-based adaptive random forest as a useful tool to screen underlying lead compounds. *Journal of Chemometrics*, 25(4), 201–207. <https://doi.org/10.1002/cem.1375>
- Chakraborti, S., Maiti, A., Pramanik, S., Sannigrahi, S., Pilla, F., Banerjee, A., et al. (2020). Evaluating the plausible application of advanced machine learnings in exploring determinant factors of present pandemic: A case for continent specific COVID-19 analysis. *Science of the Total Environment*, 765, 142723. <https://doi.org/10.1016/j.scitotenv.2020.142723>
- Chen, Z. L., Zhang, Q., Lu, Y., Guo, Z. M., Zhang, X., Zhang, W. J., et al. (2020). Distribution of the COVID-19 epidemic and correlation with population emigration from Wuhan, China. *Chinese Medical Journal (English)*, 133, 1044–1050. <https://doi.org/10.1097/CM9.0000000000000782>
- Chi, G., & Zhu, J. (2008). Spatial regression models for demographic analysis. *Population Research and Policy Review*, 27(1), 17–42. <https://doi.org/10.1007/s11113-007-9051-8>
- Christidis, P., & Christodoulou, A. (2020). The predictive capacity of air travel patterns during the global spread of the COVID-19 pandemic: Risk, uncertainty and randomness. *International Journal of Environmental Research and Public Health*, 17(10), 3356. <https://doi.org/10.3390/ijerph17103356>
- Coticini, E., Frediani, B., & Caro, D. (2020). Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environmental Pollution*, 261, Article 114465. <https://doi.org/10.1016/j.envpol.2020.114465>
- Cotfas, L. A., Delcea, C., Milne, R. J., & Salari, M. (2020). Evaluating classical airplane boarding methods considering COVID-19 flying restrictions. *Symmetry*, 12(7), 1087. <https://doi.org/10.3390/sym12071087>
- Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., & Yelowitz, A. (2020). Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States. *Health Affairs*, 39(7), 1237–1246. <https://doi.org/10.1377/hlthaff.2020.00608>
- Coven, J., & Gupta, A. (2020). Disparities in mobility responses to covid-19. *NYU stern working paper*. Available at: <https://static1.squarespace.com/static/56086d00e4b0fb7874bc2d42/t/5ebf201183c6f016ca3abd91/1589583893816/DemographicCovid.pdf>
- Desmet, K., & Wacziarg, R. (2020). *Understanding spatial variation in COVID-19 across the United States (No. w27329)*. National Bureau of Economic Research. Available at: <https://www.nber.org/papers/w27329>.

- Dowd, J. B., Andriano, L., Brazel, D. M., Rotondi, V., Block, P., Ding, X., et al. (2020). Reply to Nepomuceno et al.: A renewed call for detailed social and demographic COVID-19 data from all countries. *Proceedings of the National Academy of Sciences*, 117(25), 13884–13885. <https://doi.org/10.1073/pnas.2009408117>
- Ehlert, A. (2020). The socioeconomic determinants of COVID-19: A spatial analysis of German county level data. *MedRxiv*. <https://doi.org/10.1101/2020.06.25.20140459>, 2020.06.25.20140459.
- European Centre for Disease Prevention and Control. (2020). *COVID-19 situation update worldwide*. Available at: <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>.
- Fabris, F., Doherty, A., Palmer, D., De Magalhães, J. P., & Freitas, A. A. (2018). A new approach for interpreting random forest models and its application to the biology of ageing. *Bioinformatics*, 34(14), 2449–2456. <https://doi.org/10.1093/bioinformatics/bty087>
- Fang, C., Liu, H., Li, G., Sun, D., & Miao, Z. (2015). Estimating the impact of urbanization on air quality in China using spatial regression models. *Sustainability*, 7(11), 15570–15592. <https://doi.org/10.3390/su71115570>
- Fitzpatrick, K. M., Harris, C., & Drawve, G. (2020). Fear of COVID-19 and the mental health consequences in America. *Psychological Trauma: Theory, Research, Practice, and Policy*, 12(S1), S17–S21. <https://doi.org/10.1037/tra0000924>, 2020.
- Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, 107(6), 1247–1265. <https://doi.org/10.1080/24694452.2017.1352480>
- Fortaleza, C. M. C. B., Guimarães, R. B., de Almeida, G. B., Pronunciato, M., & Ferreira, C. P. (2020). Taking the inner route: Spatial and demographic factors affecting vulnerability to COVID-19 among 604 cities from inner São Paulo State, Brazil. *Epidemiology & Infection*, 148. <https://doi.org/10.1017/S095026882000134X>
- Ge, X. Y., Pu, Y., Liao, C. H., Huang, W. F., Zeng, Q., Zhou, H., et al. (2020). Evaluation of the exposure risk of SARS-CoV-2 in different hospital environment. *Sustainable Cities and Society*, 61, Article 102413. <https://doi.org/10.1016/j.scs.2020.102413>
- Guliyev, H. (2020). Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial Statistics*, 38, Article 100443. <https://doi.org/10.1016/j.spaSta.2020.100443>
- Hu, M., Roberts, J. D., Azevedo, G. P., & Milner, D. (2020). The role of built and social environmental factors in Covid-19 transmission: A look at America's capital city. *Sustainable Cities and Society*, 65, Article 102580. <https://doi.org/10.1016/j.scs.2020.102580>
- Iyanda, A. E., Adeleke, R., Lu, Y., Osayomi, T., Adaralege, A., Lasode, M., 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. <https://doi.org/10.1016/j.jiph.2020.07.006>
- Jin, T., Li, J., Yang, J., Li, J., Hong, F., Long, H., et al. (2020). SARS-CoV-2 presented in the air of an intensive care unit (ICU). *Sustainable Cities and Society*, Article 102446. <https://doi.org/10.1016/j.scs.2020.102446>
- Karaye, I. M., & Horney, J. A. (2020). The impact of social vulnerability on COVID-19 in the US: An analysis of spatially varying relationships. *American Journal of Preventive Medicine*, 59(3), 317–325. <https://doi.org/10.1016/j.amepre.2020.06.006>
- Killeen, B. D., Wu, J. Y., Shah, K., Zapaischkyova, A., Nikutta, P., Tamhane, A., et al. (2020). A county-level dataset for informing the United States' response to COVID-19. *arXiv preprint arXiv:2004.00756*.
- Kirby, R. S., Delmelle, E., & Eberth, J. M. (2017). Advances in spatial epidemiology and geographic information systems. *Annals of Epidemiology*, 27(1), 1–9. <https://doi.org/10.1016/j.annepidem.2016.12.001>
- Kraemer, M. U., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490), 493–497. <https://doi.org/10.1126/science.abb4218>
- Kupferschmidt, K., & Cohen, J. (2020). Can China's COVID-19 strategy work elsewhere? *Science*, 367(6482), 1061–1062. <https://doi.org/10.1126/science.367.6482.1061>
- Lambert, D. M., Brown, J. P., & Florax, R. J. (2010). A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application. *Regional Science and Urban Economics*, 40(4), 241–252. <https://doi.org/10.1016/j.regsciurbeco.2010.04.001>
- Luo, Y., Yan, J., & McClure, S. (2020). Distribution of the environmental and socioeconomic risk factors on COVID-19 death rate across continental USA: A spatial nonlinear analysis. *Environmental Science and Pollution Research*, 1–13. <https://doi.org/10.1007/s11356-020-10962-2>
- Ma, L., Fu, T., Blaschke, T., Li, M., Tiede, D., Zhou, Z., et al. (2017). Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. *ISPRS International Journal of Geo-Information*, 6(2), 51. <https://doi.org/10.3390/ijgi6020051>
- 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). *Sustainable Cities and Society*, 65, Article 102627. <https://doi.org/10.1016/j.scs.2020.102627>
- Méndez-Arriaga, F. (2020). The temperature and regional climate effects on communitarian COVID-19 contagion in Mexico throughout phase 1. *Science of the Total Environment*, 735, 139560. <https://doi.org/10.1016/j.scitotenv.2020.139560>
- Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of the Total Environment*, 728, Article 138884. <https://doi.org/10.1016/j.scitotenv.2020.138884>
- Okun, O., & Priisalu, H. (2007). Random forest for gene expression based cancer classification: Overlooked issues. *Iberian conference on pattern recognition and image analysis* (pp. 483–490). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-72849-8_61
- Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (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. <https://doi.org/10.3390/ijgi8060269>
- 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. <https://doi.org/10.1186/s12942-020-00204-6>
- Oshan, T., Wolf, L. J., Fotheringham, A. S., Kang, W., Li, Z., & Yu, H. (2019). A comment on geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, 33(7), 1289–1299. <https://doi.org/10.1080/13658816.2019.1572895>
- Oztig, L. I., & Askin, O. E. (2020). Human mobility and coronavirus disease 2019 (COVID-19): A negative binomial regression analysis. *Public Health*, 185, 364–367. <https://doi.org/10.1016/j.puhe.2020.07.002>
- Pourghasemi, H. R., Pouyan, S., Heidari, B., Farajzadeh, Z., Fallah Shamsi, S. R., Babaei, S., et al. (2020). Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). *International Journal of Infectious Diseases*, 98, 90–108. <https://doi.org/10.1016/j.ijid.2020.06.058>
- Qi, H., Xiao, S., Shi, R., Ward, M. P., Chen, Y., Tu, W., et al. (2020). COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis. *Science of the Total Environment*, 728, Article 138778. <https://doi.org/10.1016/j.scitotenv.2020.138778>
- Rahman, M. A., Zaman, N., Asyhari, A. T., Al-Turjman, F., Bhuiyan, M. Z. A., & Zolkplli, M. F. (2020). Data-driven dynamic clustering framework for mitigating the adverse economic impact of Covid-19 lockdown practices. *Sustainable Cities and Society*, 62, Article 102372. <https://doi.org/10.1016/j.scs.2020.102372>
- Ren, H., Zhao, L., Zhang, A., Song, L., Liao, Y., Lu, W., et al. (2020). Early forecasting of the potential risk zones of COVID-19 in China's megacities. *Science of the Total Environment*, 729, Article 138995. <https://doi.org/10.1016/j.scitotenv.2020.138995>
- Rumpler, R., Venkataraman, S., & Göransson, P. (2020). An observation of the impact of CoVid-19 recommendation measures monitored through urban noise levels in central Stockholm, Sweden. *Sustainable Cities and Society*, 63, Article 102469. <https://doi.org/10.1016/j.scs.2020.102469>
- Runkle, J. D., Sugg, M. M., Leeper, R. D., Rao, Y., Matthews, J. L., & Rennie, J. J. (2020). Short-term effects of specific humidity and temperature on COVID-19 morbidity in select US cities. *Science of the Total Environment*, 740, 140093. <https://doi.org/10.1016/j.scitotenv.2020.140093>
- Sannigrahi, S., Pilla, F., Basu, B., & Basu, A. S. (2020). The overall mortality caused by COVID-19 in the European region is highly associated with demographic composition: A spatial regression-based approach. *Working Paper* (pp. 1–43). Available at: <https://arxiv.org/abs/2005.04029>.
- Sannigrahi, S., Pilla, F., Basu, B., & Molter, A. (2020). Examining the association between sociodemographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustainable Cities and Society*, 62, Article 102418. <https://doi.org/10.1016/j.scs.2020.102418>
- Sarwar, S., Waheed, R., Sarwar, S., & Khan, A. (2020). COVID-19 challenges to Pakistan: Is GIS analysis useful to draw solutions? *Science of the Total Environment*, 730, Article 139089. <https://doi.org/10.1016/j.scitotenv.2020.139089>
- Song, J., Du, S., Feng, X., & Guo, L. (2014). The relationships between landscape compositions and land surface temperature: Quantifying their resolution sensitivity with spatial regression models. *Landscape and Urban Planning*, 123, 145–157. <https://doi.org/10.1016/j.landurbplan.2013.11.014>
- Sun, F., Matthews, S. A., Yang, T. C., & Hu, M. H. (2020). A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: Where geography matters? *Annals of Epidemiology*, 52, 54–59. <https://doi.org/10.1016/j.annepidem.2020.07.014>
- Sun, C., & Zhai, Z. (2020). The efficacy of social distance and ventilation effectiveness in preventing COVID-19 transmission. *Science of the Total Environment*, 62, Article 102390. <https://doi.org/10.1016/j.scs.2020.102390>
- Thakar, V. (2020). Unfolding events in space and time: Geospatial insights into covid-19 diffusion in Washington state during the initial stage of the outbreak. *ISPRS International Journal of Geo-Information*, 9(6), 382. <https://doi.org/10.3390/ijgi9060382>
- World Health Organization. (2020). *Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19)*. Available at: [https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)).
- Xiong, Y., Wang, Y., Chen, F., & Zhu, M. (2020). Spatial statistics and influencing factors of the COVID-19 epidemic at both prefecture and county levels in Hubei Province, China. *International Journal of Environmental Research and Public Health*, 17(11), 3903. <https://doi.org/10.3390/ijerph17113903>
- Yang, X., & Jin, W. (2010). GIS-based spatial regression and prediction of water quality in river networks: A case study in Iowa. *Journal of Environmental Management*, 91, 1943–1951. <https://doi.org/10.1016/j.jenvman.2010.04.011>
- Yao, Y., Pan, J., Wang, W., Liu, Z., Kan, H., Qiu, Y., et al. (2020). Association of particulate matter pollution and case fatality rate of COVID-19 in 49 Chinese cities. *Science of the Total Environment*, 741, Article 140396. <https://doi.org/10.1016/j.scitotenv.2020.140396>
- You, H., Wu, X., & Guo, X. (2020). Distribution of COVID-19 morbidity rate in association with social and economic factors in Wuhan, China: Implications for urban development. *International Journal of Environmental Research and Public Health*, 17(10), 3417. <https://doi.org/10.3390/ijerph17103417>

- Zhang, C. H., & Schwartz, G. G. (2020). Spatial disparities in coronavirus incidence and mortality in the United States: An ecological analysis as of May 2020. *The Journal of Rural Health*, 36(3), 433–445. <https://doi.org/10.1111/jrh.12476>
- Zhang, Q., Wang, Y., Tao, S., Bilsborrow, R. E., Qiu, T., Liu, C., et al. (2020). Divergent socioeconomic-ecological outcomes of China's conversion of cropland to forest program in the subtropical mountainous area and the semi-arid Loess Plateau. *Ecosystem Services*, 45, Article 101167. <https://doi.org/10.1016/j.ecoser.2020.101167>
- Zhang, L., Yang, H., Wang, K., Zhan, Y., & Bian, L. (2020). Measuring imported case risk of COVID-19 from inbound international flights—a case study on China. *Journal of Air Transport Management*, 89, Article 101918. <https://doi.org/10.1016/j.jairtraman.2020.101918>
- Zhou, Q., Zhou, H., Zhou, Q., Yang, F., & Luo, L. (2014). Structure damage detection based on random forest recursive feature elimination. *Mechanical Systems and Signal Processing*, 46(1), 82–90. <https://doi.org/10.1016/j.ymssp.2013.12.013>
- Zhu, G., Xiao, J., Zhang, B., Liu, T., Lin, H., Li, X., et al. (2018). The spatiotemporal transmission of dengue and its driving mechanism: A case study on the 2014 dengue outbreak in Guangdong, China. *Science of the Total Environment*, 622, 252–259. <https://doi.org/10.1016/j.scitotenv.2017.11.314>