Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Spatial and Spatio-temporal Epidemiology 43 (2022) 100534

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Spatiotemporal clusters and the socioeconomic determinants of COVID-19 in Toronto neighbourhoods, Canada

Nushrat Nazia [a,*], Jane Law [a,b], Zahid Ahmad Butt [a]

[a] School of Public Health Sciences, University of Waterloo, 200 University Ave W., Waterloo, ON N2L3G1, Canada
[b] School of Planning, University of Waterloo, 200 University Ave W., Waterloo, ON N2L3G1, Canada

ARTICLE INFO

Keywords:
Space-time clusters
Spatial regression
Multiscale geographically weighted regression (mgwr)
COVID-19
Clustering analysis

ABSTRACT

The aim of this study is to identify spatiotemporal clusters and the socioeconomic drivers of COVID-19 in Toronto. Geographical, epidemiological, and socioeconomic data from the 140 neighbourhoods in Toronto were used in this study. We used local and global Moran's *I*, and space-time scan statistic to identify spatial and spatiotemporal clusters of COVID-19. We also used global (spatial regression models), and local geographically weighted regression (GWR) and Multiscale Geographically weighted regression (MGWR) models to identify the globally and locally varying socioeconomic drivers of COVID-19. The global regression model identified a lower percentage of educated people and a higher percentage of immigrants in the neighbourhoods as significant predictors of COVID-19. MGWR shows the best fit model to explain the variables affecting COVID-19. The findings imply that a single intervention package for the entire area would not be an effective strategy for controlling COVID-19; a locally adaptable intervention package would be beneficial.

1. Introduction

In Canada, the coronavirus disease (COVID-19) burden is unevenly distributed spatially, while the highest number of cases are witnessed in Toronto (Detsky and Bogoch, 2020; COVID-19 Tracker Canada, 2021). COVID-19 is a part of a family of enveloped single-strained RNA viruses that can cause acute and chronic communicable respiratory diseases in humans (World Health Organization, 2020). Toronto, the most densely populated city in Canada, has been severely impacted by COVID-19 and has become an epicenter of COVID-19 outbreaks (City of Toronto, 2021). COVID-19 was first reported in Wuhan, China, in December 2019 (Shereen et al., 2020; Kang et al., 2020). As of March 7, 2022, the large-scale outbreaks of COVID-19 have contributed to over 440 million cases and 5.9 million deaths worldwide (WHO, 2021). The World Health Organization (WHO) declared the COVID-19 outbreak a global pandemic on March 11, 2020 (Cucinotta and Vanelli, 2020). The first two cases of COVID-19 in Canada were reported in Toronto on January 21, 2020, from a couple who had recently returned from Wuhan, China. The local government of Toronto has adopted several control strategies that include emergency lockdowns, stay-at-home orders, increased testing, contact tracing capacities, and closure of in-person schools and non-essential businesses. Despite these ongoing measures, Toronto continues to experience a rise in cases, creating an enormous challenge for public health as well as causing economic and social burdens.

Spatial studies in COVID-19 showed wide variances in the distribution of case and mortality rates among different communities across space. The lower socioeconomic groups have historically been shown to have a disadvantage in the diagnosis and mortality rates of infectious diseases. The city of Toronto has a diverse population of 2.99 million (2020) with varying socioeconomic statuses across the neighbourhoods (City of Toronto, 2021). People with low socioeconomic or marginalized status, such as minorities and low-income individuals, might be forced to leave their homes to maintain income or live in congregate settings, which places them and their neighbourhoods at a higher risk during this pandemic (Cordes and Castro, 2020a; Sun et al., 2020a; Vaz, 2021). Identifying disease clusters and understanding the driving factors for these clusters using spatial analytical approaches can provide us with a more realistic view of the issue in Toronto compared to the traditional simple maps (Cordes and Castro, 2020a).

Geographic Information Systems and spatial analysis have been established as important tools in infectious disease surveillance. Proximity is an important factor in the infectious disease distribution and diffusion processes. Spatial analysis is based on Tobler's first law of geography, stating that locations that are closer have more similar

* Corresponding author at: School of Public Health Sciences, University of Waterloo, 200 University Ave W., Waterloo, ON N2L3G1, Canada.
E-mail address: nnazia@uwaterloo.ca (N. Nazia).

https://doi.org/10.1016/j.sste.2022.100534
Received 30 May 2021; Received in revised form 19 July 2022; Accepted 24 August 2022
Available online 26 August 2022

attributes than locations that are further apart (Miller, 2004). Infectious diseases can show heterogeneity in cases and mortality rates, while the process of contagion can diffuse from the source to the neighbouring areas (Cordes and Castro, 2020a). Spatial analytical methods using community-based datasets are particularly important for new emerging diseases such as COVID-19 (Ali et al., 2009) to understand disease etiology and the transmission process in the communities (Chowell and Rothenberg, 2018). The results can help slow disease transmission rates in the study area.

In the case of COVID-19, previous spatial studies used space-scan statistic (Acharya et al., 2020; Alkhamis et al., 2020; Andersen et al., 2021a; Andrade et al., 2020; Arashi et al., 2020; Azmach et al., 2020; Ballesteros et al., 2020; Benita et al., 2020; Chakraborty, 2021; Chow et al., 2020; Cordes and Castro, 2020b; Da Silveira Moreira, 2020; Desjardins et al., 2020; Ferreira, 2020; Ferreira et al., 2020; Gomes et al., 2020; Greene et al., 2020; Han et al., 2021; Hohl et al., 2020; Islam et al., 2021; Kim and Castro, 2020; Leal-Neto et al., 2020; Masrur et al., 2020; Ladoy et al., 2021) and local Moran's *I* (Han et al., 2021; Bilal et al., 2020; Li et al., 2020a; Saffary et al., 2020; Xie et al., 2020; Xiong et al., 2020; Ye and Hu, 2020), MST-DBSCAN (Ridder et al., 2020; De Ridder et al., 2021), GeoMEDD (Curtis et al., 2020) to locate clusters of increased risk of COVID-19. A few studies have used global spatial regression models (Cao et al., 2020; Demenech et al., 2020; Feinhandler et al., 2020; Sannigrahi et al., 2020a; Sun et al., 2020b; You et al., 2020), Geographically Weighted Regression (GWR) (Han et al., 2021; Islam et al., 2021; Sannigrahi et al., 2020a; Das et al., 2020; Huang et al., 2020; Iyanda et al., 2020; Mollalo et al., 2020; Shariati et al., 2020a; Snyder and Parks, 2020) and MGWR (Multiscale Geographically Weighted Regression) (Mansour et al., 2021; Maiti et al., 2020a; Middya and Roy, 2021; Ma et al., 2022) to understand the contributing factors that may influence the risk of COVID-19. Some studies have linked socioeconomic influences to explain the variations in COVID-19 incidences or mortality rates. Some of the key factors to have an influence on the incidence or mortality rates of COVID-19 include income (Cordes and Castro, 2020a; Sannigrahi et al., 2020b; Chaudhry et al., 2020; Abedi et al., 2020; Maiti et al., 2020b), poverty rates (Sannigrahi et al., 2020b; Goutte et al., 2020; Chen and Jiao, 2020; Fielding-Miller et al., 2020; Richmond et al., 2020), education (Cordes and Castro, 2020a; Abedi et al., 2020; Goutte et al., 2020; Wu et al., 2020) ethnicity or minority status (Cordes and Castro, 2020a; Sun et al., 2020a; Maiti et al., 2020b; Chen and Jiao, 2020; Niedzwiedz et al., 2020; Sun et al., 2020c; Andersen et al., 2021b; Kathe and Wani, 2020), immigrant population (Fielding-Miller et al., 2020; Borjas, 2020), and unemployment rate (Sun et al., 2020a; Goutte et al., 2020).

Even though Toronto remains at heightened risk in this pandemic, spatial studies on the burden of COVID-19 in Toronto have so far been very limited. A detailed assessment is critical to identify the hotspots and the key spatial drivers for a more efficient intervention plan. This study is to fill the gaps of earlier studies by performing a comprehensive assessment of spatial dynamics of the COVID-19 outbreak in Toronto at the neighbourhood level. The main goal of this study is to improve our current understanding of the disease hotspots and provide information on the spatial association between the socioeconomic factors and COVID-19 outbreak in order for infection prevention and mitigation. The primary objective of this study is to identify the spatial and spatiotemporal clusters of COVID-19 incidences in Toronto, and the secondary objective is to identify globally and locally variable socioeconomic drivers of the COVID-19 incidences.

2. Methods

2.1. Study area, study population and data

Toronto is the capital city of Ontario and the fourth largest city in North America. The city, located on the Southwestern shores of Lake Ontario (Fig. 1), has a population density of 4692 people per square kilometer, making it the most densely populated city in Canada. The city of Toronto consists of four community council areas: Etobicoke York, North York, Toronto, and East York and Scarborough (Fig. 1). Toronto has a very diverse population, with a 51.2% of Toronto's population being a visible minority, and 51.2% of the population are immigrants (born outside of Canada) (City of Toronto, 2021). Toronto, based on the 2016 census, comprises immigrants from Asian (53.4%), European (23.6%), Americas (16.8%) and Africa (6.1%) origins. 49% of the Asian immigrants have immigrated from Chinese and the Philippines, and 11.7% of Asian immigrants are from India. 37.5% of the American immigrants are of Jamaican and Guyanese origin. 41.7% of the European immigrants are from Italy, Portuguese, and United Kingdom. 55% of the African immigrants are from Ethiopia, Egypt, Kenya, Nigeria, Somalia and South Africa.

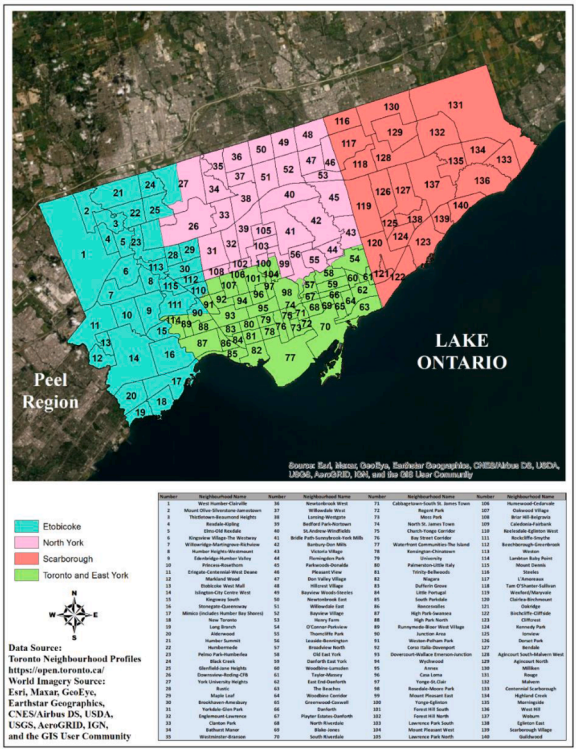

Fig. 1. The 140 neighbourhoods in the study area in Toronto, Canada.

The city has an unemployment rate of 6.4%, and 20.2% population lives in a low-income bracket. The direct and indirect impact of this pandemic's burdens can pose significant threats for a densely and diversely populated city like Toronto.

The study area is comprised of 140 geographically distinct neighbourhoods in Toronto (Fig. 1). The neighbourhoods were defined based on Statistics Canada census tracts for the purposes of statistical reporting (Statistics Canada, 2022). The140 neighbourhood profiles in Toronto contain social, economic, and demographic details using the census 2016 population dataset collected and released by Statistics Canada. Our study used a population size of 2.7 million (2016 census dataset) for analysis. Further details of these neighbourhood profiles datasets can be found on the city of Toronto website (Neighbourhood Profiles – City of Toronto) (City of Toronto, 2020a). The population dataset at the neighbourhood level was obtained from an open-source (https://open.toronto.ca/dataset/wellbeing-toronto-demographics).

The epidemiological data were collected by Toronto Public Health that contained the geographic and demographic details for all confirmed and probable cases of COVID-19. Confirmed cases are defined by a person with confirmation of SARS-CoV-2 infection documented by detection of at least one specific gene target by a validated laboratory-based nucleic acid amplification test (NAAT) assay (e.g. real-time PCR

or nucleic acid sequencing) performed at a community, hospital or reference laboratory (e.g. Public Health Ontario Laboratory or National Microbiology Laboratory) or a validated POC NAAT has been deemed acceptable by the Ontario Ministry of Health to provide a final result or demonstrated diagnostic rise with a 4-week interval in viral-specific antibody in serum or plasma using a validated laboratory-based serological assay for SARS-CoV-2. The probable cases are defined by a person with symptoms compatible with COVID-19 and had high-risk exposure with a confirmed case of COVID-19 or was exposed to a known cluster or outbreak and in whom a laboratory-based nucleic acid amplification test (NAAT-based assay (e.g. real-time PCR or nucleic acid sequencing) for SARS-CoV-2 are inconclusive or has not been completed or had SARS-CoV-2 antibody detected in a single serum, plasma, or whole blood sample using a validated laboratory-based serological assay for SARS-CoV-2 collected within four weeks of symptom onset or had a POC NAAT or POC antigen test for SARS-CoV-2 completed and the result is presumptive positive or had a validated POC antigen test for SARS-CoV-2 completed and the result is positive (Canada PHA of 2021).

The case datasets are extracted from the provincial Case & Contact Management System (CCM). The epidemiological datasets are updated weekly (City of Toronto 2020b). The case dataset includes the episode date and neighbourhood attributes of the infected individuals. For this study, the COVID-19 case data between January 1, 2020 – January 31, 2021, were extracted for analysis. The cases were plotted to the centroid of the infected individual's neighbourhood. During the study period, a total of 87,501 cases (71,940 sporadic cases and 15,561 outbreak-associated cases) of COVID-19 were diagnosed in Toronto, including the first reported case (January 21, 2020). Use of outbreak-associated cases, which are generally in healthcare (e.g., long-term care homes, hospitals) and residential or congregate settings (City of Toronto 2020b), can potentially create a bias towards clustering in outbreak areas (17.71% of total cases). Therefore, this study excluded the outbreak-associated cases to control for potential bias and included only the community-based sporadic cases. Among the total sporadic cases, 1185 (1.64%) cases were excluded due to missing neighbourhood information, leaving 70,755 sporadic cases for analysis.

We considered five socioeconomic covariates (2016 census) based on the literature review and available datasets. These covariates were previously identified as potential risk factors in some COVID-19 studies. The covariates included: (i) percentage of immigrants (individuals who were born outside of Canada) (Government of Canada SC, 2021), (ii) percentage of the population aged 25–64 years with a lower level of education (not having a university certificate, diploma or a bachelor degree) (A City of Toronto 2020), (iii) prevalence of low income (living in a low-income household based on the low-income cut-off (LICO) table representing the poverty line) (Government of Canada, 2021), and (iv) unemployment rate (population over 15 years and unemployed) (Government of Canada SC, 2021). While mortality and hospitalization rates may differ across different age groups, there is not enough evidence that the incidence rate of COVID-19 and transmission risks for sporadic cases vary significantly by sex or age (Li et al., 2020b). Since the entire population is at risk of contracting and transmitting the disease, we did not pursue any adjustment by age or sex.

2.2. Spatial and spatiotemporal cluster analyses

2.2.1. The global spatial autocorrelation

The global Moran's *I* by Anselin was first used in GeoDa, version 1.18.0, to assess whether the COVID-19 incidence rates (per 1000 population) in Toronto neighbourhoods display a tendency to cluster together and measure the extent of the correlation among neighbouring observations (Anselin, 2022). Global Moran's index is used to examine the absence or presence of spatial autocorrelation in disease diffusion processes by comparing location and attribute similarities in the area. The value of the global Moran's *I* must show a clustering distribution pattern to find high or low-risk clusters for further analysis (Shariati et al., 2020b). The formula for calculating global Moran's *I* index (Anselin, 2022) is shown in Eq. (1):

$$I = \frac{\Sigma_i \Sigma_j w_{ij} Z_i . Z_j / S_0}{\Sigma_i Z_i^2 / n} \quad (1)$$

where Z_i *and* Z_j represent the COVID-19 incidence rate variations in neighbourhood i and j, respectively, w_{ij} refers to the elements in the spatial weights matrix, neighbourhood i and j at study period, $S_0 = \Sigma_i \Sigma_j w_{ij}$, w_{ij}as the sum of all weights, and n represents the number of observations.

The value of global Moran's *I* can range between −1.0 and +1.0, where the positive value suggests the presence of a positive spatial correlation, while a negative value suggests a negative correlation. The higher the value of *I*, the stronger the spatial autocorrelation (Kim and Castro, 2020). Values close to 0 indicate no spatial autocorrelation and that the distribution of data is random (Li et al., 2020a).

2.2.2. Spatial clustering by local Moran's I statistics

The local Moran's *I*, a local indicator of spatial association (LISA), was used in GeoDa, version 1.18.0, to evaluate the local level of spatial autocorrelation or dependency of spatial data and to visualize the possible high-risk or low-risk clusters (Anselin et al., 2010) based on COVID-19 incidences in different neighbourhoods across Toronto.

The formula (Li et al., 2020a; Anselin, 1995) for calculating local Moran's index I_i is shown in Eq. (2:)

$$I_i = x_i \sum_j w_{ij} x_j \quad (2)$$

where x_i *and* x_j represents the COVID-19 incidence rates in neighbourhood i and j respectively, w_{ij} is the spatial weights matrix.

The global and local Moran's *I* tests were run using the first-order queen's contiguity spatial weights matrix that uses the values from all first-order neighbouring neighbourhoods in order to determine whether the area has a higher or lower mean assessing the degree of spatial autocorrelation. A permutation test was conducted using Monte Carlo simulations with 999 permutations to test the statistical significance of the clusters under the assumption that COVID-19 incidence rates are randomly distributed in the study area. The local Moran's *I* divide the neighbourhood polygons into four categories: high-high (hotspots), low-low (coldspots), high-low, and low-high, based on the type of spatial autocorrelations (Anselin, 1995). The high-high and low-low areas represent spatial clusters, and the high-low and low-high areas represent discordant patterns. The intensity value is calculated for each point, which shows the level of clustering of similar values around the point. The local Moran's *I* result showing local spatial autocorrelation of COVID-19 incidences in the Toronto neighbourhoods were presented in the form of cluster maps with a significance level of 0.1%, 1%, and 5%. Additionally, we have also performed the Bonferroni bound procedure to carry out an extensive sensitivity analysis to avoid the risk of obtaining false-positive results (Type I errors), and to check the robustness of the findings (not ultimately presented these results). We have presented the results that best represent the important High-High and Low-Low clusters.

2.2.3. The space-time scan statistic

Kulldorff's space-time scan statistic method was used in SaTScan™, version 9.7, to identify the space-time clusters of COVID-19 cases between January 2020 and January 2021 in Toronto neighbourhoods. SaTScan™ software is a widely used open-source spatial scan statistic software that utilizes Kulldorff's retrospective space-time permutation method to identify significant clusters in a study area (Kulldorff, 2009a). SaTScan™ uses a moving cylinder with circular or elliptical windows across a study area that locates the spatial clusters that are significant during a specific period. A discrete Poisson probability model was

chosen for the clustering analysis with the assumption that the disease cases have a Poisson distribution. The scan parameters with a time interval of one month ranged from January 1, 2020 – January 31, 2021. After a preliminary test, the spatial and temporal scanning windows were restricted to include 10% of the population at risk and 50% of the study period, respectively, to avoid a large cluster size. The clusters were tested for significance using 999 Monte Carlo simulations, and the clusters with a p-value<0.05 are considered to be the significant high-risk clusters. The neighbourhoods within the significant high-risk clusters are identified as high-risk neighbourhoods. The relative risk of COVID-19 for a cluster is calculated using the ratio of observed to expected cases, comparing the risk within a cluster to the areas outside the cluster. The relative risk (RR) (Kulldorff, 2009a; Rao et al., 2017) defined in Eq. (3):

$$RR = \frac{o/e}{(O-o)/(O-e)} \tag{3}$$

Where e is the expected number of cases in the cluster, o is the total number of observed cases within the cluster, and O is the total number of observed cases in the study area. The values of RR for a cluster greater than 1 indicate a high COVID-19 incidence rate. A spatiotemporal map with the clusters and the relative risks (RR) of the neighbourhoods was created in ArcGIS version 10.8.1 to show the spatial variations of COVID-19 risks in the Toronto neighbourhoods.

2.3. Regression analyses

We used five different global and local spatial regression models to understand the relationship between the socioeconomic variables and COVID-19 incidence rates. The models include three global regression models: ordinary least squares (OLS), spatial error model (SEM), spatial lag model (SLM), and two local regression models: geographically weighted regression (GWR) and multiscale GWR (MGWR). Before running these models, a bivariate regression analysis was conducted to select the explanatory variables. GeoDa version 1.18.0 was used for running the three global models. The local models were implemented in a stand-alone software (MGWR version 2.2: Spatial Analysis Research Center (SPARC), Tempe, USA), developed by Fotheringham et al. (2017). ArcGIS version 10.8.1 was used for mapping of all outputs.

2.3.1. Global regression models

The COVID-19 incidence rate per 1000 population was used as the dependent variable for the global models. A preliminary data analysis shows that the incidence rates were highly skewed, violating the normality assumptions of spatial regression models (Yu et al., 2010); thus, the log-transformed (based 10) incidence rates were used as the dependent variable in the models.

The ordinary least squares (OLS) is a regression method investigating the relationship between the dependent and explanatory variables (Li et al., 2020a). OLS makes two major assumptions: the observations are independent and constant across the study area, and there is no correlation between the error terms (Yandell and Anselin, 1990). A spatial error model (SEM) (Yandell and Anselin, 1990) is based on the assumption that there is a spatial dependence in the OLS model residuals generated from the OLS error term model. A spatial lag model (SLM) Spatial lag model (Yandell and Anselin, 1990) is based on a spatially-lagged dependent variable. The SLM model assumes dependency among the dependent and the independent variables. The SLM also assumes that an independent variable can depend on another independent variable in the neighbourhood region.

The OLS model is expressed in Eq. (4) as:

$$y_i = \beta_0 + x_i\beta + \varepsilon_i \tag{4}$$

The mathematical expression for the SEM is shown in Eq. (5) as:

$$y_i = \beta_0 + \beta x_i + \lambda W_i\mu_j + \varepsilon_i \tag{5}$$

The SLM model is expressed in Eq. (6) as:

$$y_i = \beta_0 + x_i\beta + \rho W_i y_i + \varepsilon_i \tag{6}$$

where yi is the COVID-19 incidence rate in neighbourhood i, x_i is the vector of the explanatory variable, ε_i is an error term, Wi is a vector of (nxn)spatial weights matrix, β is the vector of regression parameters, β_0 is the intercept, μ_i and μ_j are the error terms at neighbourhood i and j, respectively, λ is the coefficient of spatial correlated errors, and ρ is the spatial lag parameter.

However, in the case of COVID-19 in the Toronto neighbourhoods, as supported by SLM and SEM results shown later in the results, a spatial correlation exists between variables. Therefore, the interactions from the OLS are omitted from the results, and the spatial models (SEM and SLM) were considered better suited for this study. The AIC values were used in the final model selection process to evaluate overall model accuracy and how well the model fits the data, and a lower AIC value indicates an improvement in model performance.

2.3.2. Local regression models (GWR & MGWR)

Two local models, GWR and MGWR, were applied to the same set of predictors used in the global models to explore the local spatial variation in the relationships with the COVID-19 incidence rates.

The GWR model is a local spatial regression model that makes assumptions that spatial interactions are non-stationary and that parameter estimates may spatially vary that can not be explained by the global regression models (Lin and Wen, 2011). GWR takes spatial heterogeneity into consideration while calculating the spatial interaction among the dependent and explanatory variables and produces local regression parameter estimates at each observation location (Lin and Wen, 2011; Maiti et al., 2021). Geographically Weighted Regression (GWR) uses a local smoothing processing method to estimate the geographical functional form of regression coefficients non-parametrically (Nakaya, 2016). The GWR model is denoted in Eq. (8) as:

$$y_i = \sum_{j-0}^{m} \beta_j(\mu_j, v_j)Xi_j + \varepsilon_i \tag{8}$$

where at an area i, y_i is the dependent variable (log of COVID-19 incidence rate), $\beta_j(\mu_j, v_j)$ is the jth coefficient, (μ_j, v_j) is the vector form of x, y coordinates, X_{ij} is the value of the jth explanatory parameter, and $\varepsilon_i$is the random error term (Iyanda and Osayomi, 2021).

However, the GWR models produce a single optimal bandwidth for all variables, which assumes that all factors affect COVID-19 rates at the same spatial scale (Yu et al., 2020). This assumption is given that different processes may affect COVID-19 rates at different spatial scales. This can result in underestimation of the parameters, particularly in a large city such as Toronto with a high population density (Leong and Yue, 2017). Therefore, we have also applied the MGWR model, which is an extension of GWR that allows for studying the relationship between variables at different scales (Yu et al., 2020). MGWR obtains a set of optimal covariate-specific bandwidths in which each bandwidth indicates the spatial scale at which a factor impacts the outcome variable (Fotheringham et al., 2017). The MGWR model can be formulated in Eq. (9) as:

$$y_i = \sum_{j-0}^{m} \beta_{bwj}(\mu_j, v_j)Xi_j + \varepsilon_i \tag{9}$$

where $\beta_{bwj}$is the bandwidth used for calibration of the jth relationship (Iyanda and Osayomi, 2021), and the rest of the parameters are the same as Eq. (8).

Additionally, geographically weighted regression models generally ignore the multiple testing issues that can lead to an excess of false

positives and therefore, the significance of the local parameter estimates may be questionable (da Silva and Fotheringham, 2016). Therefore, we have also used the newly developed correction method by da Silva and Fotehringham (da Silva and Fotheringham, 2016) for inference in the GWR/MGWR to solve the multiple testing issues to obtain reliable local parameter estimates. da Silva and Fotheringham proposed an effective correction to the significance level α to assess the significance of local parameter estimates and avoid the proportion of false positives exceeding α. The corrected significance level α value is calculated using Eq. (10):

$$\alpha = \frac{\xi}{\frac{ENP}{P}} \tag{10}$$

Where ξ is the expected type I error rate before correction, *ENP* is the effective number of parameters in the model, which is a function of the optimal bandwidth parameter, and *P* is the number of parameters in the model (da Silva and Fotheringham, 2016).

The adaptive bisquare spatial kernel function was applied to develop both local models, and the golden search was applied to select an optimal bandwidth (Yu et al., 2020). The adaptive bandwidth is defined as the proportion of data points involved in the calibration process of local estimates and eliminates the influence of outside the neighbourhood spatial units (Yu et al., 2020). AICc was used to evaluate and compare the global model fit and performance. The best model fit is indicated by a larger R-square and a smaller AICc value. The outputs from the best-fitted local model were used to map the local parameter estimates, their estimated standard errors, the significant parameter estimates (after adjusting for multiple tests), the local R-square and the local Condition numbers. A condition number greater than 20 can affect model accuracy and inferences, and a value less than 20 indicates no effect of multicollinearity (Oshan et al., 2019a).

3. Results

3.1. Descriptive statistics

Fig. 2 shows the distribution of the total cumulative sporadic cases of COVID-19 in Toronto by epidemiological week. The total monthly cases started to increase in March (week 12) and experienced a continuous decline from June 2020 (week 25) to August 2020 (week 34). However, the numbers continued to rise exponentially in the following months, with the maximum number of cases witnessed in January 2021 (20,506 total monthly cases).

3.2. Spatial and spatiotemporal cluster analyses

3.2.1. The global spatial autocorrelation

Before implementing the spatial and spatiotemporal cluster analyses, we examined the global Moran's *I* statistic results to evaluate the presence of spatial autocorrelation in the study area (Fig. 3). The outcome variable (incidence rate of COVID-19) demonstrated a positive spatial autocorrelation suggesting a strong clustering pattern, with a statistically significant Moran's *I* value of 0.62 (p-value=0.001, z-value =12.96). The results indicate that the distribution of the incidence rate of COVID-19 had a positively significant correlation with the incidence rate of the nearest neighbourhoods during the study period.

3.2.2. Local Moran's I

The results from the local Moran's *I* method show COVID-19 clusters with a significance level of 0.1%, 1% and 5% in the Toronto neighbourhoods. The impositions of the Bonferroni bounds result in only the low-low and high-high clusters being significant. Our goal is to understand the interesting locations rather than interpreting the most stringent tests (*p*-value=0.00012), and therefore we have interpreted the use of the traditional p-value of 0.05, 0.01 and 0.001. The results also show distinct clustering of statistically significant high-high clustering or 'hotspots' in the northwestern and southeastern parts of Toronto in 29 neighbourhoods (20.71% of all neighbourhoods) and Low-Low clustering or cold spots in 35 neighbourhoods (25% of all neighbourhoods) in central Toronto (Fig. 3).

3.2.3. The space-time scan statistic

We found eight statistically significant space-time clusters of COVID-19 in different parts of Toronto from January 2020 to January 2021 using the space-time statistic method in SaTScan™ (Fig. 4, Table 1). Table 1 provides the space-time cluster characteristics that include the *p*-values, the total population at risk, the observed and the expected number of cases, the total number of neighbourhoods and the relative risk rates for each cluster. The high-risk clusters varied in terms of size, the magnitude of relative risks, the total number of neighbourhoods, and the number of population at-risk. Forty-two neighbourhoods (30%) not within the significant clusters were considered low/no risk (of COVID-19) neighbourhoods. The highest relative risks (RR>3.0) were recorded in the western and the eastern part of Toronto. The temporal periods for the eight high-risk clusters fell between October 2020 -January 2021, with cluster 1 (located in the northwestern corner) experiencing a three-month-long clustering period.

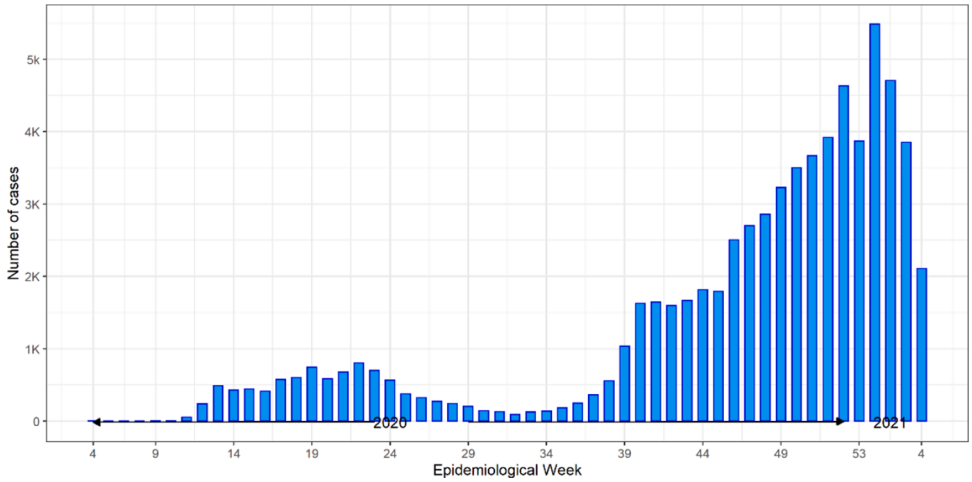

Fig. 2. Total cumulated sporadic COVID-19 cases by epidemiological week in Toronto between January 2020 and January 2021 (used for the statistical analysis).

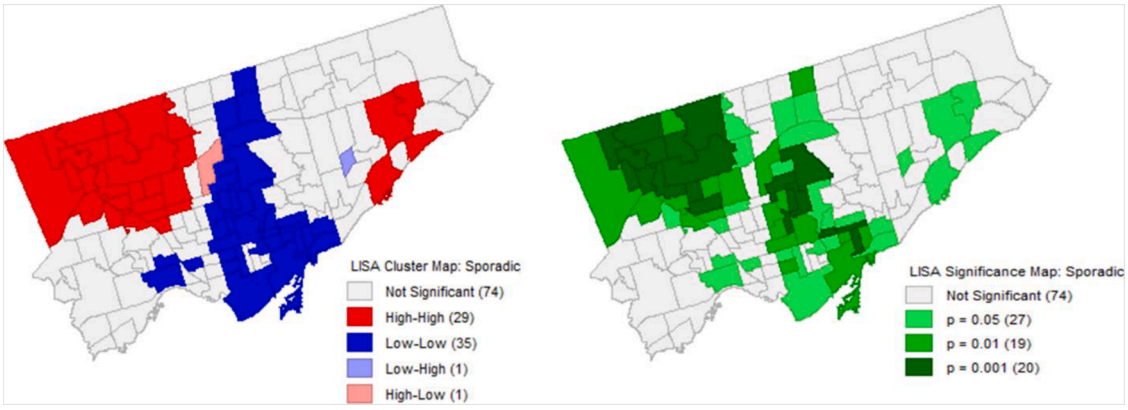

Fig. 3. Cluster and significance maps of COVID-19 incidence rates in Toronto using the local Moran's *I* approach.

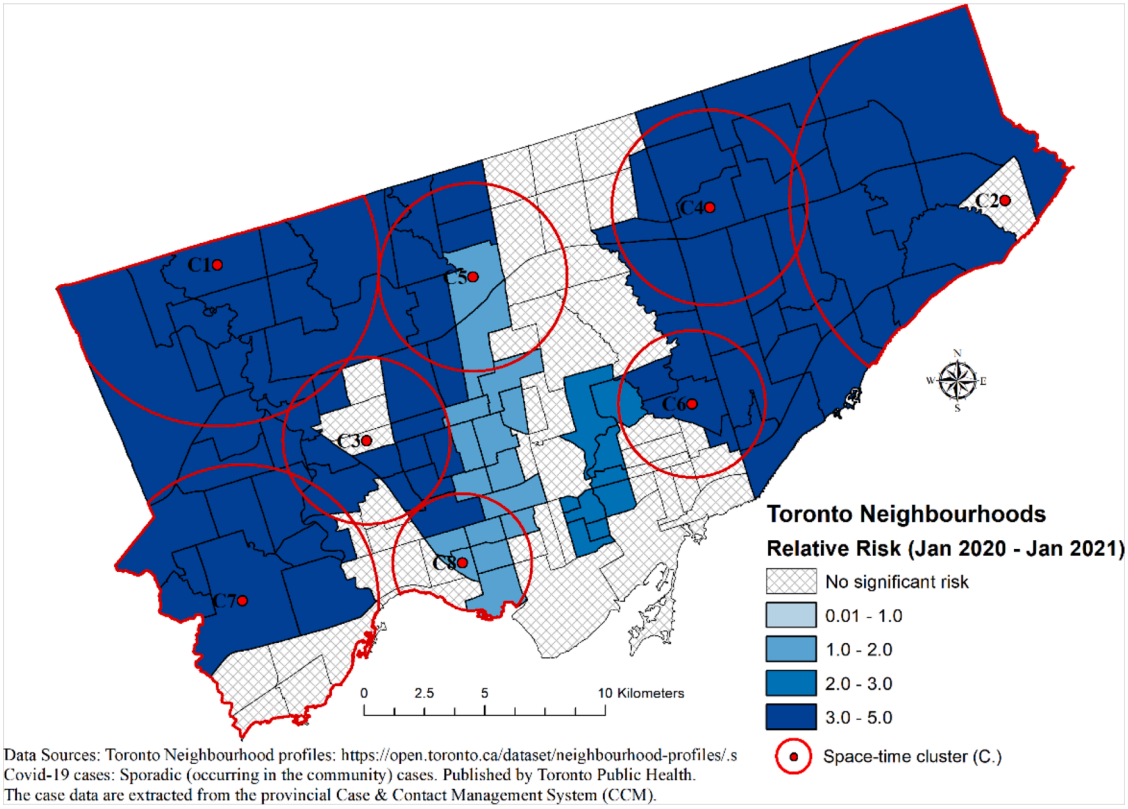

Fig. 4. The space-time clusters in Toronto between January 2020 and January 2021.

Table 1
COVID-19 space-time clusters characteristics (January 2020-January 2021, in Toronto).

CLUSTER	P-VALUE	Radius	START DATE	END DATE	# of Neighbours	OBSERVED	EXPECTED	RR	POP
1	<0.001	6691	2020/10/1	2021/1/31	13	11,690	2184.21	6.21	272,167
2	<0.001	8912	2020/11/1	2021/1/31	10	7137	1584.07	4.90	263,895
3	<0.001	3463	2020/10/1	2021/1/31	16	5967	1765.93	3.60	220,046
4	<0.001	4049	2020/11/1	2021/1/31	10	5298	1609.89	3.48	268,198
5	<0.001	3905	2020/12/1	2021/1/31	11	3447	987.09	3.62	244,011
6	<0.001	3045	2020/11/1	2021/1/31	12	3674	1206.91	3.16	201,063
7	<0.001	5643	2020/12/1	2021/1/31	15	2665	1026.03	2.66	253,639
8	<0.001	2880	2020/12/1	2021/1/31	10	1608	760.13	2.14	187,906

RR = Relative risk, Pop = Poulation at risk.

3.3. Regression analyses

The results of the bivariate regression analysis show a significant positive correlation of COVID-19 incidences with the four independent variables: percentage of immigrants (p-value<0.0001), the prevalence of low income (p-value<0.0001), unemployment rate (*p*-value<0.0001) in the neighbourhoods, whereas a significant negative correlation was observed with the percentage of the population with a lower level of education (*p*-value<0.0001) in the neighbourhood (Table 2). The spatial regression models were run with the four independent variables, and the multicollinearity among predictor covariates was tested using the multicollinearity condition number method. The model yielded a multicollinearity condition number of 18.9. Since the multicollinearity condition number was under 30, it indicates that the predictor variables in the regression model are not highly correlated (Yandell and Anselin, 1990; Anselin et al., 2006). The presence of autocorrelation among the predictor variables was further investigated using a correlation matrix. The results indicated no strong correlation between the four independent variables and were included in the local and global regression models.

3.3.1. Global regression models

The results from the OLS diagnostics show that the LM lag and LM error were both statistically significant at the 0.05 significance level, indicating that spatial models such as SLM and SEM are more appropriate and relevant global models. The Akaike information criterion (AIC), R-square and the log-likelihood values were compared to find the most appropriate (Yandell and Anselin, 1990) spatial regression model (SLM or SEM) that explains the global spatial relationship between the socioeconomic variables and COVID-19 incidence rates. Table 3 shows that the SEM model had the lowest AIC value and highest R^2value. Therefore, the SEM model is chosen as the most appropriate model that successfully incorporates spatial effects in the model and can better explain the model variability. The results from the SEM model showed a one percent increase of the immigrants in a neighbourhood was associated with a 1% (exponentiation of the estimate of 0.0110) increase in the COVID-19 incidence rate in the neighbourhood. Moreover, a one-unit increase in lower educated individuals in a neighbourhood was associated with a 2% (exponentiation of the estimate of 0.0179) increase in the COVID-19 incidence rate in the neighbourhood. The spatial autoregressive coefficient (LAMDA) of the SEM model has a positive sign and is highly significant (p-value<0.05), indicating a positive spatial dependence between the neighbourhoods (Table 4).

3.3.2. Local regression models (GWR and MGWR)

Table 5 shows the covariate-specific bandwidths, the effective number of parameter estimates, critical t values, adjusted alpha, R square and AICc values for both GWR and MGWR models. The diagnostics of the local models indicated that the MGWR model presented the largest R-square and lowest AICc among all models and, therefore, a better-fitted model (Table 5). The R-square value indicates that the MGWR model explains 88.4% of the variations in the COVID-19 rates in Toronto. The summary results of the MGWR are listed in Table 6.

The spatial associations between the socioeconomic factors and COVID-19 rates using the outputs from the MGWR model are shown in Figs. 5–8, where the local parameter estimates and associated standard errors are mapped in (a) and (b), respectively, while (c) indicates only the significant local parameter estimates defined based on the significance (α) value given in Table 4 which has been adjusted for multiple tests.

Significant positive associations between COVID-19 incidence rates and the percentage of immigrants were observed in all neighbourhoods in Toronto (Fig. 5c). The higher local parameter estimates were observed in the northwestern and southern parts of Toronto, whereas lower local parameter estimates were observed in the northeastern parts of Toronto. Similar to the percentage of immigrants, a higher percentage of the population with a lower level of education was positively associated with COVID-19 incidence rates throughout all neighbourhoods of Toronto (Fig. 6c). The higher local parameter estimates were observed in northeastern Toronto, whereas relatively lower local parameter estimates were observed in southern Toronto.

The association between COVID-19 incidence rate and the two factors: prevalence of low income and unemployment rate, were not found to be significant in the majority of the neighbourhoods in Toronto (Figs. 7c and 8c). The prevalence of low-income local parameter estimates varied from positive to negative values (Fig. 7a). However, only six neighbourhoods in northwestern Toronto had a significant negative association with COVID-19 incidence rates (Fig. 7c). The unemployment rate was found to be significantly associated with COVID-19 incidence rates in only five neighbourhoods in the northeastern part of Toronto (Fig. 8c)

Fig. 9 presents spatial variations in local R-square values in the study area and the local condition numbers (CN) from the MGWR model for each neighbourhood. The highest local R-square values ($R^2 \geq 0.86$) were observed in the western and central parts of Toronto, namely in the Etobicoke and North York regions (Fig. 9a). The local condition number (CN) was observed higher in the northern and central parts of Toronto. However, the overall the CN value was <20 with maximum value being 9.5, suggests that there were no presence of multicollinearity in the model (Fig. 9b).

Table 2
Results of the bivariate regression analysis (Outcome variable: log of COVID-19 incidence rate).

Variables	Estimate (95% CI)	*p*-value
Percentage of immigrants	0.69 (0.60, 0.77)	<0.0001
Percentage of population with lower level of education	0.82 (0.75, 0.86)	<0.0001
Prevalence of low income	0.30 (0.15, 0.45)	<0.0001
Unemployment rate	0.64 (0.53, 0.72)	<0.0001

Table 3
Summary of output from the global spatial regression models.

Index	Spatial Lag Model (SLM)	Spatial Error Model (SEM)
R^2	00.79	0.81
Log-likelihood	102.67	107.60
Akaike information criterion	−193.35	−205.21

Table 4
Regression outputs from the Spatial Error Model (SEM) model ($n = 140$).

Variable	Coefficient	Std. Error	z-value	Probability
Constant	1.4605	0.1562	9.3487	<0.0001
Percentage of Immigrants	0.0110	0.0035	3.0842	0.0020
Percentage of population with lower level of education	0.0179	0.0026	6.8249	<0.0001
Prevalence of low income	−0.0024	0.0061	−0.3988	0.6899
Unemployment rate	0.0402	0.0255	1.5738	0.1155
LAMDA	0.4455	0.1077	4.1337	<0.0001

Dependent variable: base-10 logarithmically transformed rates of COVID-19.
Akaike Information Criterion (AIC): −63.46.

4. Discussion

The results of our analysis of the COVID-19 incidences displayed non-random spatial distribution patterns in our study area in Toronto. Both Anselin's local Moran's *I* and Kulldorff's space-time statistic produced similar patterns of hotspots/clusters in the Northwestern part of

Table 5
GWR and MGWR Summary Statistics for the COVID-19 Data.

Diagnostic	GWR Entire Model	MGWR Entire Model	Intercept	Immigrants	Lower level of education		Low Income	Unemployment Rate
Bandwidth	71	n/a	43	65	99	43		139
Effective No. of Parameters	20.23	21.09	6.79	3.83	1.98	7.20		1.28
Adjusted α	0.0125	0.0018	0.007	0.013	0.025	0.006		0.039
Critical t (95%)	2.53		2.72	2.51	2.26	2.74		2.08
AICc	158.105	149.45						
R2	0.874	0.884						

n=140.

Table 6
Summary of coefficients results from the local MGWR Model.

Variables	Mean	STD	Min	Median	Max
Intercept	0.019	0.195	−0.248	−0.022	0.526
Percentage of Immigrants	0.307	0.095	0.077	0.310	0.489
Percentage of population with lower level of education	0.505	0.048	0.437	0.492	0.609
Prevalence of low income	−0.019	0.128	−0.368	0.027	0.231
Unemployment rate	0.119	0.008	0.110	0.116	0.136

Toronto in the Etobicoke region. An assessment of the MGWR maps suggests that the higher percentage of immigrants in the neighbourhoods may have contributed to the higher COVID-19 incidence in that part of Toronto. Additionally, in the eastern part of Toronto, the lower level of education of the people in the neighbourhood may have contributed to higher COVID-19 incidence.

Although there were some similarities, we also observed a few dissimilarities in the clusters detected by the two clustering approaches. It is important to note that the local Moran's *I* do not take the temporal factors into account, and the magnitude of the risk is not provided, which is covered in the scan statistic. While the scan statistic identified clusters in the southwestern parts of Toronto, the local Moran's *I* found no significant clusters in those areas. Overall, the clusters identified by the scan statistic were more localized and covered a broader area than that identified by the local Moran's *I*. In local Moran's *I* method, the clusters are identified in a strictly bounded area, where the correlation is assessed between the disease rate of a certain neighbourhood and the average disease rate of its surrounding (first-order contiguity) neighbourhoods (Laohasiriwong et al., 2018). Scan statistic might be a more sensitive method (Laohasiriwong et al., 2018) and provides more elaborate details of the cluster characteristics such as the total population at risk, radius, and relative risks compared to local Moran's *I*. Since each method has its own set of strengths and limitations, we believe a combined approach to identify the disease clusters would provide an in-depth understanding of the disease clusters and ensure logical consistency of the results of the analyses.

The results of the global regression yielded that the percentage of immigrants and a lower level of educated people in the neighbourhood influenced to vary COVID-19 incidence rates in Toronto, consistent with the previous studies (Goutte et al., 2020; Fielding-Miller et al., 2020; Borjas, 2020). We found that a higher concentration of immigrants in the neighbourhood was associated with the increased COVID-19 incidence in the neighbourhood, suggesting that COVID-19 might have

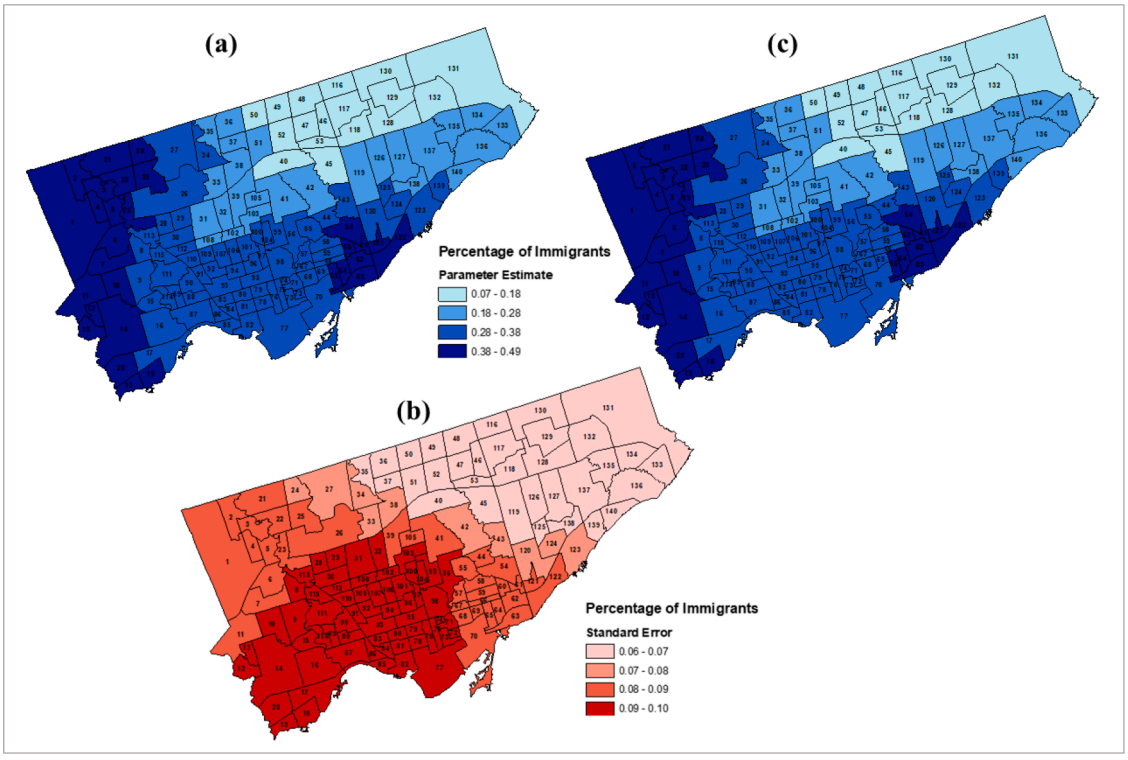

Fig. 5. (a) Local parameter estimates (b) Standard errors (c) Significant local parameter estimates (after adjusting for multiple tests) for the percentage of immigrants factor from the MGWR model.

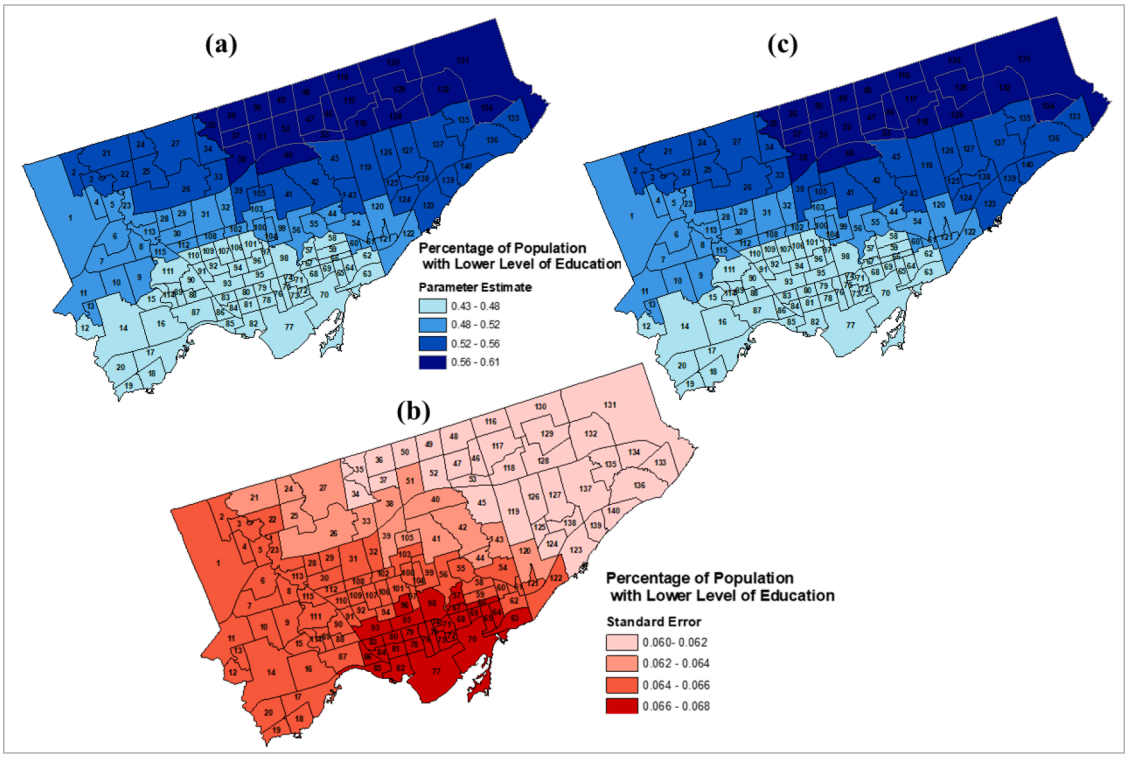

Fig. 6. (a) Local parameter estimates (b) Standard errors (c) Significant local parameter estimates (after adjusting for multiple tests) for the lower level of education factor from the MGWR model.

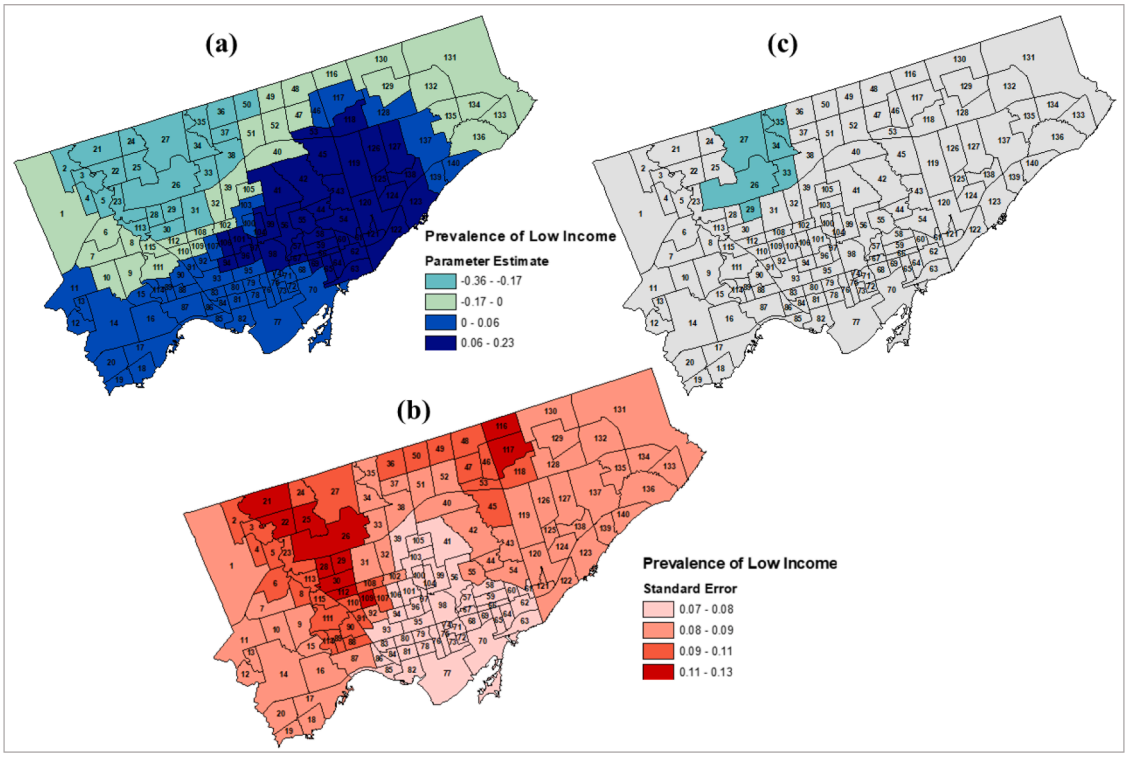

Fig. 7. (a) Local parameter estimates (b) Standard errors (c) Significant local parameter estimates (after adjusting for multiple tests) for the prevalence of low income factor from the MGWR model.

disproportionately impacted the immigrant population in Toronto. This could be attributed to immigrants living in overcrowded housing conditions, lower-income neighbourhoods and/or working in an environment where physical distancing is often challenging (COVID-19 Disproportionately Impacted Immigrants and Refugees in Ontario, 2021; Choi et al., 2020). It is worth noting that Asian countries account

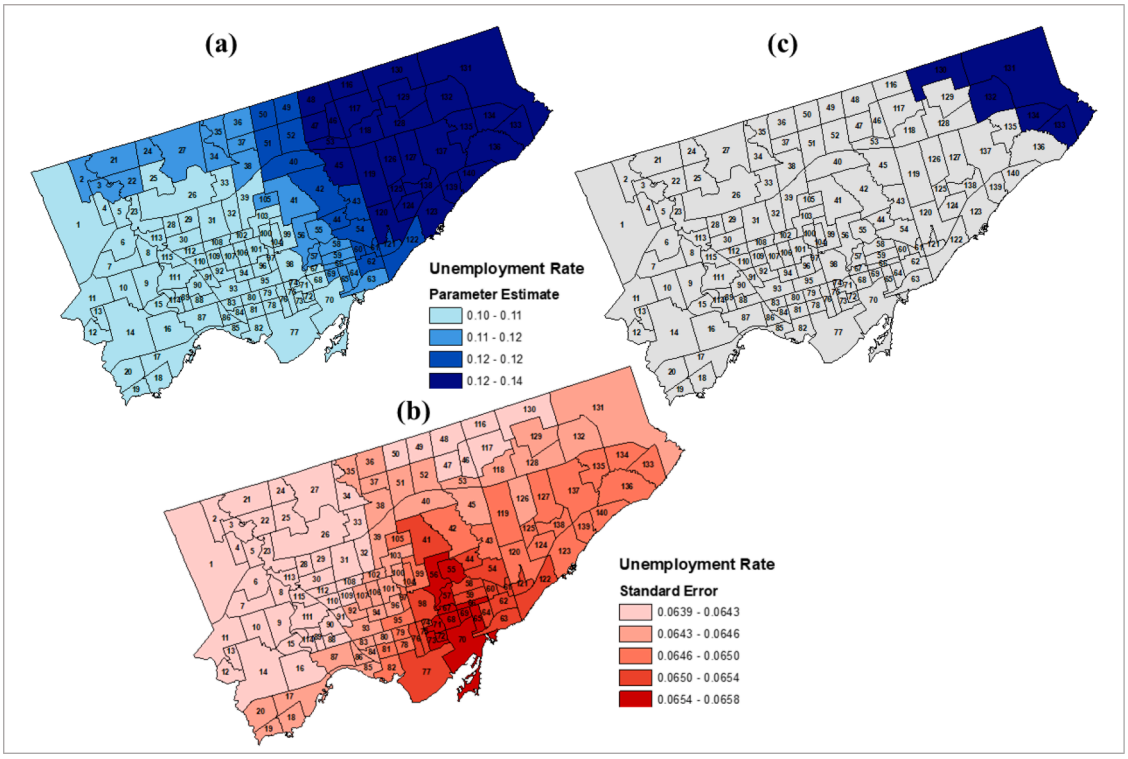

Fig. 8. (a) Local parameter estimates (b) Standard errors (c) Significant local parameter estimates (after adjusting for multiple tests) for the unemployment rate factor from the MGWR model.

for the majority (53%) of immigrants in Toronto, with China alone accounting for 20%. Heterogeneity in risk factors across Asian-origin subgroups may have led to distinct patterns of geographic disparities of risk in COVID-19.

We also found that neighbourhoods with lower levels of education had higher outbreaks of COVID-19 in Toronto. This suggests that the least educated people may have a lower understanding and awareness of COVID-19, highlighting the importance of knowledge to fight this pandemic. It is also possible that the less educated people are employed in professions where they are at higher risk of contracting COVID-19, and it is not possible for them to self-isolate at home. However, note that the global models have some limitations since these models assume that the interactions between the variables are stationary over space which is often not realistic (Thayn and Simanis, 2013). A variable that is not found significant at the global scale may not be true at a local spatial scale. The results of our study also suggest that although the prevalence of low income and the unemployment rate were not found to be significant predictors in the global model, the individual influences of these two predictors exhibited a significant association (negative and positive) in the local MGWR model in a few neighbourhoods in the city. These suggest that ignoring the locally variable indicators may not help to control the disease transmission.

A major advantage of using the local regression model in this study is that it allows us to visually demonstrate the magnitude of risk deriving from a factor at a spatial scale, which would allow us to know where a particular type of intervention will be required based on the different associations between characteristics of given populations located at different places. The intervention may also include informing on COVID protection measures using which languages and where to ask people to stay at home and work remotely, etc. The parameter estimates and significant maps allowed a comparison with the cluster maps to identify the socioeconomic drivers of the clustering of the disease in neighbourhoods.

While the global models performed well, the local models provided a more parsimonious quantitative output of the socioeconomic determinants that may influence COVID-19 rates. The MGWR allowed the relationship between COVID-19 and explanatory variables to vary spatially and at different scales. The MGWR has the advantage of more accurately depicting spatial heterogeneity, diminish multicollinearity and lessening the bias in the parameter estimates (Yu et al., 2020; Oshan et al., 2019b). Our findings indicated that the MGWR does not suffer from the effect of multicollinearity and is robust as it allowed each of the parameters to be processed at flexible and varying scales. The use of the correction method proposed by da Silva and Fotheringham (da Silva and Fotheringham, 2016) allowed us to identify the significant association with each predictor variable at a local scale for a more validated output. The MGWR model explained 88.4% of the model variances, which is higher than the global four-parameter SEM model (81%), SLM (79%), and the local GWR model (87.4%), suggesting that the results of the MGWR model provided a more enhanced model accuracy. Therefore, we believe using the MGWR approach to demonstrate the influence of socioeconomic components on COVID-19 incidence patterns in Toronto benefitted our study.

There are several limitations to our study. First, we used a limited number of variables based on the publicly available data at the neighbourhood level. Further study by including additional explanatory variables may benefit our understanding of the spatial variations of the disease incidences. Second, this study used confirmed and probable COVID-19 cases collected from Toronto Public Health. However, COVID-19 is often known to be asymptomatic. Thus, we could have missed the infected individuals who had mild symptoms and not visited any hospital or testing centers and were unreported. Furthermore, there is a possibility of misclassification and variation in the propensity to test. However, we believe that these unreported cases could be randomly distributed and may not have a significant impact on our analysis. Third, the scan statistic uses a circular window to detect disease clusters and is unable to detect clusters irregular in shape (Kulldorff, 2009b; Hughes and Gorton, 2013).

In Toronto, COVID-19 has rapidly evolved, creating a dire public health crisis. Several ongoing preventative measures were adopted in

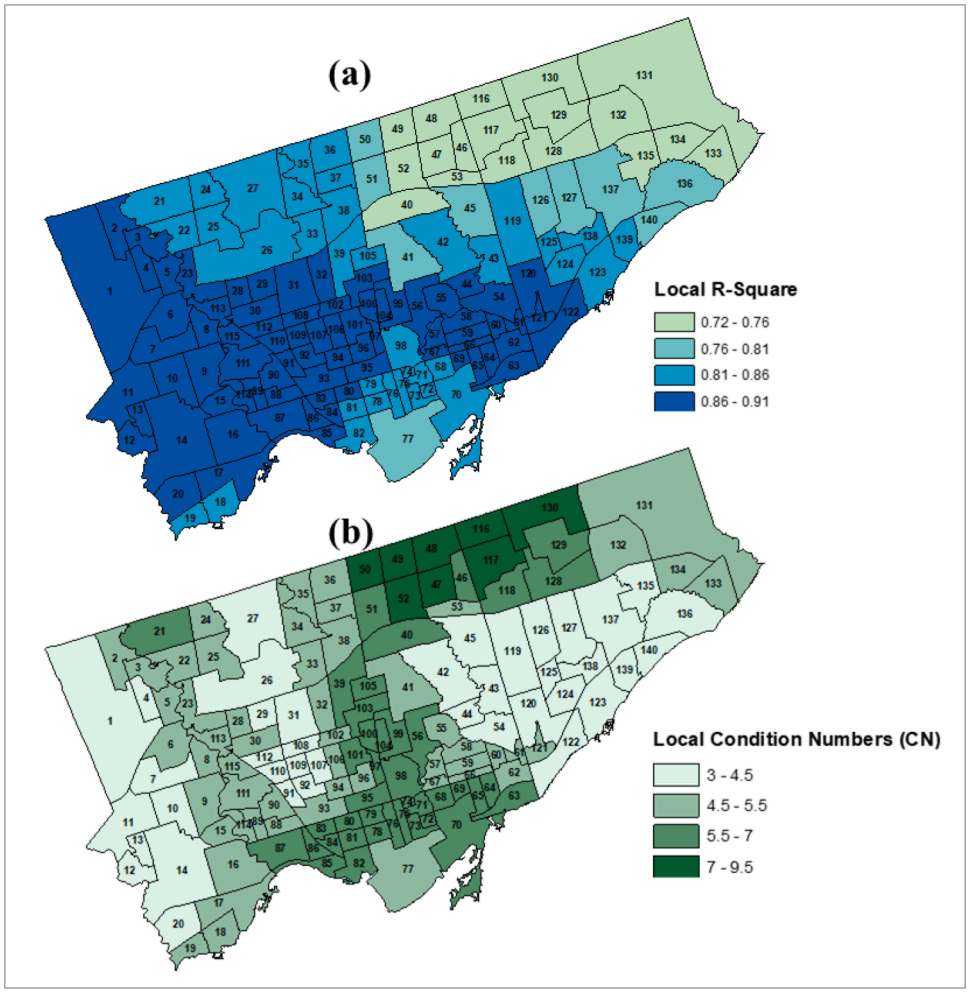

Fig. 9. Mapping of the (a) locally weighted coefficient of determination (R^2) between the observed and fitted values and (b) the local condition numbers (CN) of the MGWR model.

Toronto by the government to control the spread of COVID-19 due to the higher rates of infections among the residents. However, the number of cases continued to rise. Spatial modeling of disease is important to assess a newly emerging infectious disease such as COVID-19 to understand the magnitude of the risk in a densely populated city like Toronto. Some studies have emphasized the importance of early implementation of intervention strategies to mitigate the disease risk from COVID-19,[111,112]. Our study has several policy implications for mitigating the risk from COVID-19: (i) it may serve as a spatial guideline for the decision-makers to formulate mitigation strategies focusing on the hot-spots across the neighbourhoods with effective and clear guidelines, (ii) the socioeconomic determinants of COVID-19 may provide spatially explicit information about the spatial drivers of COVID-19 to identify the localities, and the policymakers can establish disease surveillance based on the socioeconomic drivers that were influencing disease risk in the neighbourhoods, (iii) it may help in developing plans for decreasing the socioeconomic inequalities in the high-risk neighbourhoods to mitigate the disease risk, and (iv) it may help to identify areas where interventions will be required to improve public knowledge and awareness of COVID-19.

Future research may focus on how the epidemics are disseminated by using diffusion modeling techniques to display the diffusion direction, magnitude, and dynamics by taking into account of the knowledge related to the virus, such as the incubation time and injected into modeling to better represent the reality.

5. Conclusion

The use of spatial methods to model COVID-19 incidence in Toronto is warranted to improve the current control and vaccination strategies at the postal code level. Our study adopts multiple spatial and spatiotemporal models to provide a deeper insight into the magnitude of the risk of COVID-19 in Toronto. We detected several high-risk clusters in different parts of Toronto, and the socioeconomic conditions in the neighbourhoods could be the underlying factors for clustering the cases. Addressing the socioeconimic determinants of health such as education, diversity, income, and unemployment status is important in infectious disease surveillance. People living in a lower socioeconomic condition may struggle to adhere to proper social distancing measures. The findings of our study could allow a closer focus on the COVID-19 incidence and the socioeconomic predictors to mitigate the disease risk and control it. The policymakers could benefit from the findings of this study.

Funding

No funding was received for this study.

CRediT authorship contribution statement

Nushrat Nazia: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Jane Law:** Conceptualization, Writing – review & editing. **Zahid Ahmad Butt:**

Writing – review & editing.

Data Availability

Data will be made available on request.

Acknowledgement

The authors of this study wish to acknowledge the city of Toronto, Toronto Public Health, Ontario Government and Statistics Canada for the free open access to the data.

References

Detsky, A.S., Bogoch, I.I., 2020. COVID-19 in Canada: experience and response. JAMA J. Am. Med. Assoc. https://doi.org/10.1001/jama.2020.14033. Published online.
COVID-19 Tracker Canada. 2021. Total Cases by Province. https://covid19tracker.ca/. (Accessed April 14, 2021).
Shereen, M.A., Khan, S., Kazmi, A., Bashir, N., Siddique, R., 2020. COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. J. Adv. Res. https://doi.org/10.1016/j.jare.2020.03.005. Published online.
Kang, D., Choi, H., Kim, J.H., Choi, J., 2020. Spatial epidemic dynamics of the COVID-19 outbreak in China. Int. J Infect. Dis. https://doi.org/10.1016/j.ijid.2020.03.076. Published online.
World Health Organization, 2020. WHO COVID-19 Case definition. https://www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance_Case_Definition-2020.2. (Accessed April 14, 2021).
City of Toronto. 2021. Toronto at a Glance. https://www.toronto.ca/city-government/data-research-maps/toronto-at-a-glance/#:~:text=Located%20on%20a%20broad%20sloping,perimeter%20is%20approximately%20180%20km. (Accessed April 14, 2021).
WHO. World Health Organization: COVID-19 dashboard. Published 2021. https://covid19.who.int/.
Cucinotta, D., Vanelli, M., 2020. WHO Declares COVID-19 a Pandemic. Acta Biomed. 91 (1), 157–160. https://doi.org/10.23750/abm.v91i1.9397.
Cordes, J., Castro, M.C., 2020a. Spatial analysis of COVID-19 clusters and contextual factors in New York City. Spat. Spatiotemporal Epidemiol. https://doi.org/10.1016/j.sste.2020.100355. Published online.
Sun, Y., Hu, X., Xie, J., 2020a. Spatial inequalities of COVID-19 mortality rate in relation to socioeconomic and environmental factors across England. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.143595. Published online.
Vaz, E., 2021. COVID-19 in Toronto: a spatial exploratory analysis. Sustainability 13 (2), 498. https://doi.org/10.3390/su13020498.
Miller, H.J., 2004. Tobler's first law and spatial analysis. Ann. Assoc. Am. Geogr. 94 (2), 284–289. https://doi.org/10.1111/j.1467-8306.2004.09402005.x.
Ali, M., Emch, M., Yunus, M., Clemens, J., 2009. Modeling spatial heterogeneity of disease risk and evaluation of the impact of vaccination. Vaccine. https://doi.org/10.1016/j.vaccine.2009.03.085. Published online.
Chowell, G., Rothenberg, R., 2018. Spatial infectious disease epidemiology: on the cusp. BMC Med. https://doi.org/10.1186/s12916-018-1184-6. Published online.
Acharya, B.K., Khanal, L., Mahyoub, A.S.M., et al., 2020. Execution of intervention matters more than strategy A lesson from the spatiotemporal assessment of COVID-19 clusters in Nepal. medRxiv. Published online.
Alkhamis, M.A., Al Youha, S., Khajah, M.M., et al., 2020. Spatiotemporal dynamics of the COVID-19 pandemic in the State of Kuwait. Int. J. Infect. Dis. https://doi.org/10.1016/j.ijid.2020.06.078. Published online.
Andersen, L.M., Harden, S.R., Sugg, M.M., Runkle, J.D., Lundquist, T.E., 2021a. Analyzing the spatial determinants of local Covid-19 transmission in the United States. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.142396. Published online.
Andrade, L.A., Gomes, D.S., Lima, S.V.M.A., et al., 2020. COVID-19 Mortality in an area of northeast Brazil: epidemiological characteristics and prospective spatiotemporal modeling. Epidemiol. Infect. https://doi.org/10.1017/S0950268820002915. Published online.
Arashi, M., Bekker, A., Salehi, M., et al., 2020. Spatial Analysis and Prediction of COVID-19 Spread in South Africa After Lockdown. arXiv. Published online.
Azmach N.N., Tesfahannes T.G., Abdulsemed S.A., Hamza T.A. Prospective time periodic geographical COVID-19 surveillance in ethiopia using a space-time scan statistics: detecting and evaluating emerging clusters. Published online 2020. 10.21203/rs.3.rs-76052/v1.
Ballesteros, P., Salazar, E., Sánchez, D., Bolanos, C., 2020. Spatial and spatiotemporal clustering of the COVID-19 pandemic in Ecuador. Rev. Fac. Med. https://doi.org/10.15446/revfacmed.v69n1.86476. Published online.
Benita, F., Gasca-Sanchez, F., Rosenbaum-Feldbruegge, M., et al., 2020. On the Main Factors Influencing COVID-19 Spread and Deaths in Mexico A comparison Between Phase I and II. medRxiv. Published online.
Chakraborty, J., 2021. Social inequities in the distribution of COVID-19: an intra-categorical analysis of people with disabilities in the U.S. Disabil. Health J. https://doi.org/10.1016/j.dhjo.2020.101007. Published online.
Chow, T.E., Choi, Y., Yang, M., Mills, D., Yue, R., 2020. Geographic pattern of human mobility and COVID-19 before and after Hubei lockdown. Ann. GIS. https://doi.org/10.1080/19475683.2020.1841828. Published online.
Cordes, J., Castro, M.C., 2020b. Spatial analysis of COVID-19 clusters and contextual factors in New York City. Spat. Spatiotemporal Epidemiol. https://doi.org/10.1016/j.sste.2020.100355. Published online.
Da Silveira Moreira, R., 2020. COVID-19: intensive care units, mechanical ventilators, and latent mortality profiles associated with case-fatality in Brazil. Cad. Saude Publica. https://doi.org/10.1590/0102-311X00080020. Published online.
Desjardins, M.R., Hohl, A., Delmelle, E.M., 2020. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. Appl. Geogr. https://doi.org/10.1016/j.apgeog.2020.102202. Published online.
Ferreira, M.C., 2020. Spatial association between the incidence rate of Covid-19 and poverty in the São Paulo municipality, Brazil. Geospat. Health. https://doi.org/10.4081/gh.2020.921. Published online.
Ferreira, R.V., Martines, M.R., Toppa, R.H., Assunção, L.M., Desjardins, M.R., Delmelle, E.M., 2020. Applying a prospective space-time scan statistic to examine the evolution of COVID-19 clusters in the State of São Paulo, Brazil. medRxiv. https://doi.org/10.1101/2020.06.04.20122770. Published online.
Gomes, D.S., Andrade, L.A., Ribeiro, C.J.N., et al., 2020. Risk clusters of COVID-19 transmission in Northeastern Brazil: prospective space-time modeling. Epidemiol. Infect. https://doi.org/10.1017/S0950268820001843. Published online.
Greene, S.K., Peterson, E.R., Balan, D., et al., 2020. Detecting emerging COVID-19 community outbreaks at high spatiotemporal resolution – New York City, June–July 2020. medRxiv. https://doi.org/10.1101/2020.07.18.20156901. Published online.
Han, Y., Yang, L., Jia, K., et al., 2021. Spatial distribution characteristics of the COVID-19 pandemic in Beijing and its relationship with environmental factors. Sci. Total Environ. 761, 144257 https://doi.org/10.1016/j.scitotenv.2020.144257.
Hohl, A., Delmelle, E.M., Desjardins, M.R., Lan, Y., 2020. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. Spat Spatiotemporal Epidemiol. https://doi.org/10.1016/j.sste.2020.100354. Published online.
Islam, A., Sayeed, MDA, Rahman, MDK, Ferdous, J., Islam, S., Hassan, M.M., 2021. Geospatial dynamics of COVID-19 clusters and hotspots in Bangladesh. Transbound. Emerg. Dis. https://doi.org/10.1111/tbed.13973. Published online.
Kim, S., Castro, M.C., 2020. Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). Int. J. Infect. Dis. https://doi.org/10.1016/j.ijid.2020.07.004. Published online.
Leal-Neto, O.B., Santos, F.A.S., Lee, J.Y., Albuquerque, J.O., Souza, W.V., 2020. Prioritizing COVID-19 tests based on participatory surveillance and spatial scanning. Int. J. Med. Inform. https://doi.org/10.1016/j.ijmedinf.2020.104263. Published online.
Masrur, A., Yu, M., Luo, W., Dewan, A., 2020. Space-time patterns, change, and propagation of covid-19 risk relative to the intervention scenarios in Bangladesh. Int. J. Environ. Res. Public Health. https://doi.org/10.3390/ijerph17165911. Published online.
Ladoy, A., Opota, O., Carron, P.N., et al., 2021. Size and duration of COVID-19 clusters go along with a high SARS-CoV-2 viral load: a spatio-temporal investigation in Vaud state, Switzerland. Sci. Total Environ. 787, 147483 https://doi.org/10.1016/j.scitotenv.2021.147483.
Bilal, U., Barber, S., Diez-Roux, A.V., 2020. Spatial Inequities in COVID-19 Outcomes in 3 US Cities. medRxiv. Published online.
Li, H., Li, H., Ding, Z., et al., 2020a. Spatial statistical analysis of coronavirus disease 2019 (COVID-19) in China. Geospat. Health. https://doi.org/10.4081/gh.2020.867. Published online.
Saffary, T., Adegboye, O.A., Gayawan, E., Elfaki, F., Kuddus, M.A., Saffary, R., 2020. Analysis of COVID-19 cases' spatial dependence in US counties reveals health inequalities. Front. Public Health. https://doi.org/10.3389/fpubh.2020.579190. Published online.
Xie, Z., Qin, Y., Li, Y., Shen, W., Zheng, Z., Liu, S., 2020. Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.140929. Published online.
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. Int. J. Environ. Res. Public Health. https://doi.org/10.3390/ijerph17113903. Published online.
Ye, L., Hu, L., 2020. Spatiotemporal distribution and trend of COVID-19 in the Yangtze river delta region of the People's Republic of China. Geospat. Health. https://doi.org/10.4081/gh.2020.889. Published online.
Ridder, D.D., Sandoval, J., Vuilleumier, N., et al., 2020. Geospatial digital monitoring of COVID-19 cases at high spatiotemporal resolution. Lancet Digit. Health 2 (8), e393–e394. https://doi.org/10.1016/S2589-7500(20)30139-4.
De Ridder, D., Sandoval, J., Vuilleumier, N., et al., 2021. Socioeconomically disadvantaged neighborhoods face increased persistence of SARS-CoV-2 clusters. Front. Public Health 8, 626090. https://doi.org/10.3389/fpubh.2020.626090.
Curtis, A., Ajayakumar, J., Curtis, J., et al., 2020. Geographic monitoring for early disease detection (GeoMEDD). Sci. Rep. 10 (1), 21753. https://doi.org/10.1038/s41598-020-78704-5.
Cao, Y., Hiyoshi, A., Montgomery, S., 2020. COVID-19 case-fatality rate and demographic and socioeconomic influencers: worldwide spatial regression analysis based on country-level data. BMJ Open. https://doi.org/10.1136/bmjopen-2020-043560. Published online.
Demenech, L.M., Dumith S de, C., Vieira, M.E.C.D., Neiva-Silva, L., 2020. Income inequality and risk of infection and death by COVID-19 in Brazil. Rev. Bras. Epidemiol. https://doi.org/10.1590/1980-549720200095. Published online.

Feinhandler, I., Cilento, B., Beauvais, B., Harrop, J., Fulton, L., 2020. Predictors of Death Rate during the COVID-19 Pandemic. Healthcare. https://doi.org/10.3390/healthcare8030339. Published online.
Sannigrahi, S., Pilla, F., Basu, B., Basu, A.S., Molter, A., 2020a. Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. Sustain. Cities Soc. https://doi.org/10.1016/j.scs.2020.102418. Published online.
Sun, Y., Hu, X., Xie, J., 2020b. Spatial inequalities of COVID-19 mortality rate in relation to socioeconomic and environmental factors across England. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.143595. Published online.
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. Int. J. Environ. Res. Public Health. https://doi.org/10.3390/ijerph17103417. Published online.
Das, A., Ghosh, S., Das, K., Basu, T., Dutta, I., Das, M., 2020. Living environment matters: unravelling the spatial clustering of COVID-19 hotspots in Kolkata megacity, India. Sustain. Cities Soc. https://doi.org/10.1016/j.scs.2020.102577. Published online.
Huang, J., Kwan, M.P., Kan, Z., Wong, M.S., Kwok, C.Y.T., Yu, X., 2020. Investigating the Relationship between the Built Environment and Relative Risk of COVID-19 in Hong Kong. ISPRS Int. J. Geoinf. 9 (11), 624. https://doi.org/10.3390/ijgi9110624.
Iyanda, A.E., Adeleke, R., Lu, Y., et al., 2020. A retrospective cross-national examination of COVID-19 outbreak in 175 countries: a multiscale geographically weighted regression analysis. J. Infect. Public Health. https://doi.org/10.1016/j.jiph.2020.07.006. Published online 2020.
Mollalo, A., Vahedi, B., Rivera, K.M., 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.138884. Published online.
Shariati, M., Jahangiri-rad, M., Mahmud Muhammad, F., Shariati, J, 2020a. Spatial analysis of COVID-19 and exploration of its environmental and socio-demographic risk factors using spatial statistical methods: a case study of Iran. TT -. HDQIR 5 (3), 145–154. https://doi.org/10.32598/hdq.5.3.358.1.
Snyder, B.F., Parks, V., 2020. Spatial variation in socio-ecological vulnerability to Covid-19 in the contiguous United States. Health Place. https://doi.org/10.1016/j.healthplace.2020.102471. Published online.
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.
Maiti, A., Zhang, Q., Sannigrahi, S., Pramanik, S., Chakraborti, S., Pilla, F., 2020a. Spatiotemporal Effects of the Causal Factors On COVID-19 Incidences in the Contiguous United States. arXiv preprint arXiv:201015754.. Published online.
Middya, A.I., Roy, S., 2021. Geographically varying relationships of COVID-19 mortality with different factors in India. Sci. Rep. 11 (1), 7890. https://doi.org/10.1038/s41598-021-86987-5.
Ma, J., Zhu, H., Li, P., et al., 2022. Spatial patterns of the spread of COVID-19 in Singapore and the influencing factors. ISPRS Int. J. Geoinf. 11 (3), 152. https://doi.org/10.3390/ijgi11030152.
Sannigrahi, S., Pilla, F., Basu, B., Basu, A.S., Molter, A., 2020b. Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. Sustain. Cities Soc. https://doi.org/10.1016/j.scs.2020.102418. Published online.
Chaudhry, R., Dranitsaris, G., Mubashir, T., Bartoszko, J., Riazi, S., 2020. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. EClinicalMedicine. https://doi.org/10.1016/j.eclinm.2020.100464. Published online.
Abedi, V., Olulana, O., Avula, V., et al., 2020. Racial, economic, and health inequality and COVID-19 infection in the United States. J. Racial. Ethn. Health Disparities. https://doi.org/10.1007/s40615-020-00833-4. Published online.
Maiti, A., Zhang, Q., Sannigrahi, S., Pramanik, S., Chakraborti, S., Pilla, F., 2020b. Spatiotemporal Effects of the Causal Factors On COVID-19 Incidences in the Contiguous United States. arXiv preprint arXiv:201015754.. Published online.
Goutte, S., Péran, T., Porcher, T., 2020. The role of economic structural factors in determining pandemic mortality rates: evidence from the COVID-19 outbreak in France. Res. Int. Bus. Finance. https://doi.org/10.1016/j.ribaf.2020.101281. Published online.
Chen, Y., Jiao, J., 2020. Relationship between socio-demographics and COVID-19: a case study in three Texas regions. SSRN. Published online.
Fielding-Miller, R., Sundaram, M., Brouwer, K., 2020. Social determinants of COVID-19 mortality at the county level. PLoS ONE. https://doi.org/10.1101/2020.05.03.20089698. Published online.
Richmond, H.L., Tome, J., Rochani, H., Fung, I.C.H., Shah, G.H., Schwind, J.S., 2020. The use of penalized regression analysis to identify county-level demographic and socioeconomic variables predictive of increased COVID-19 cumulative case rates in the state of Georgia. Int J. Environ. Res. Public Health. https://doi.org/10.3390/ijerph17218036. Published online.
Wu, Y., Yan, X., Zhao, S., et al., 2020. Association of time to diagnosis with socioeconomic position and geographical accessibility to healthcare among symptomatic COVID-19 patients: a retrospective study in Hong Kong. Health Place. https://doi.org/10.1016/j.healthplace.2020.102465. Published online.
Niedzwiedz, C.L., O'Donnell, C.A., Jani, B.D., et al., 2020. Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank. BMC Med. https://doi.org/10.1186/s12916-020-01640-8. Published online.
Sun, F., Matthews, S.A., Yang, T.C., Hu, M.H., 2020c. A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters? Ann. Epidemiol. https://doi.org/10.1016/j.annepidem.2020.07.014. Published online.
Andersen, L.M., Harden, S.R., Sugg, M.M., Runkle, J.D., Lundquist, T.E., 2021b. Analyzing the spatial determinants of local COVID-19 transmission in the United States. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.142396. Published online.
Kathe, N., Wani, R., 2020. Determinants of COVID-19 incidence and mortality in the US: spatial analysis. medRxiv. Published online.
Borjas, G.J., 2020. Demographic Determinants of Testing Incidence and COVID-19 Infections in New York City Neighborhoods. SSRN Electron. J. https://doi.org/10.2139/ssrn.3572329. Published online.
Statistics Canada. 2022. Census tract (CT) - Census Dictionary. https://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo013-eng.cfm. (Accessed July 4, 2022).
City of Toronto. 2020. City of Toronto Neighbourhood Profiles. https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/ (Accessed April 16, 2021).
Canada Government. 2021. National case definition: coronavirus disease (COVID-19). Government of Canada. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals/national-case-definition.html. Published February 3, 2021. (Accessed March 10, 2022).
City of Toronto. 2020. Coronavirus disease 2019 (COVID-19): Epidemiology update. https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/ (Accessed April 16, 2021).
Government of Canada SC. 2021. Focus on Geography Series, 2016 Census - Canada. https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Facts-can-eng.cfm?Lang=Eng&GK=CAN&GC=01&TOPIC=7. (Accessed April 23, 2021).
Government of Canada. 2021. Prevalence of low income – National Household Survey (NHS) Dictionary. https://www12.statcan.gc.ca/nhs-enm/2011/ref/dict/fam025-eng.cfm. (Accessed April 23, 2021).
Li, H., Wang, S., Zhong, F., et al., 2020b. Age-dependent risks of incidence and mortality of COVID-19 in Hubei Province and Other Parts of China. Front. Med. (Lausanne). https://doi.org/10.3389/fmed.2020.00190. Published online.
Anselin, L., 2022. GeoDa: An Introduction to Spatial Data Analysis: Moran Scatter Plot. GeoDa WorkBook. GeoDa.https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html. (Accessed 5 January 2022).
Shariati, M., Mesgari, T., Kasraee, M., Jahangiri-rad, M, 2020b. Spatiotemporal analysis and hotspots detection of COVID-19 using geographic information system (March and April 2020). J. Environ. Health Sci. Eng. https://doi.org/10.1007/s40201-020-00565-x. Published online.
Anselin, L., Syabri, I., Kho, Y., 2010. GeoDa: an introduction to spatial data analysis. In: Fischer, MM, Getis, A (Eds.), Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Springer, pp. 73–89. https://doi.org/10.1007/978-3-642-03647-7_5.
Anselin, L., 1995. Local indicators of spatial association—LISA. Geogr. Anal. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x. Published online.
Kulldorff, M., 2009a. Information Management Services Inc. SaTScan v9.3: Software for the Spatial and Space-Time Scan Statistics. StatScan, Boston, USA. Published online.
Rao, H., Shi, X., Zhang, X., 2017. Using the kulldorff's scan statistical analysis to detect spatio-temporal clusters of tuberculosis in Qinghai Province. China, 2009-2016 BMC Infect. Dis.. https://doi.org/10.1186/s12879-017-2643-y. Published online.
Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale geographically weighted regression (MGWR). Ann. Am. Assoc. Geogr. 107 (6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480.
Yu, D., Peterson, N.A., Sheffer, M.A., Reid, R.J., Schnieder, J.E., 2010. Tobacco outlet density and demographics: analysing the relationships with a spatial regression approach. Public Health. https://doi.org/10.1016/j.puhe.2010.03.024. Published online.
Yandell, B.S., Anselin, L, 1990. Spatial econometrics: methods and models. J. Am. Stat. Assoc. https://doi.org/10.2307/2290042. Published online.
Lin, C.H., Wen, T.H., 2011. Using geographically weighted regression (GWR) to explorespatial varying relationships of immature mosquitoes andhuman densities with the incidence of dengue. Int. J. Environ. Res. Public Health. https://doi.org/10.3390/ijerph8072798. Published online.
Maiti, A., Zhang, Q., Sannigrahi, S., et al., 2021. Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States. Sustain. Cities Soc. 68, 102784 https://doi.org/10.1016/j.scs.2021.102784.
Government of Canada SC. 2021. Labour force characteristics by age group, monthly, seasonally adjusted. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410028702. (Accessed April 23, 2021).
Nakaya T. 2016. *GWR4.09 User Manual.* Windows Application for Geographically Weighted Regression Modelling. https://raw.githubusercontent.com/gwrtools/gwr4/master/GWR4manual_409.pdf. Accessed April 15. 2021.
Iyanda, A.E., Osayomi, T., 2021. Is there a relationship between economic indicators and road fatalities in Texas? A multiscale geographically weighted regression analysis. GeoJournal 86 (6), 2787–2807. https://doi.org/10.1007/s10708-020-10232-1.
Yu, H., Fotheringham, A.S., Li, Z., Oshan, T., Kang, W., Wolf, L.J., 2020. Inference in Multiscale geographically weighted regression. Geogr. Anal. 52 (1), 87–106. https://doi.org/10.1111/gean.12189.
Leong, Y.Y., Yue, J.C., 2017. A modification to geographically weighted regression. Int. J. Health Geogr. 16 (1), 11. https://doi.org/10.1186/s12942-017-0085-9.
da Silva, A.R., Fotheringham, A.S., 2016. The multiple testing Issue in geographically weighted regression. Geogr. Anal. 48 (3), 233–247. https://doi.org/10.1111/gean.12084.
Oshan, T., Wolf, L.J., Fotheringham, A.S., Kang, W., Li, Z., Yu, H., 2019a. A comment on geographically weighted regression with parameter-specific distance metrics. Int. J.

Geogr. Inform. Sci. 33 (7), 1289–1299. https://doi.org/10.1080/13658816.2019.1572895.
Anselin, L., Syabri, I., Kho, Y., 2006. GeoDa: an introduction to spatial data analysis. Geogr. Anal. https://doi.org/10.1111/j.0016-7363.2005.00671.x. Published online.
Laohasiriwong, W., Puttanapong, N., Luenam, A., 2018. A comparison of spatial heterogeneity with local cluster detection methods for chronic respiratory diseases in Thailand. F1000Res. 6 https://doi.org/10.12688/f1000research.12128.2.
COVID-19 disproportionately impacted immigrants and refugees in Ontario, New report finds | CTV News. Accessed May 26, 2021. https://toronto.ctvnews.ca/covid-19-disproportionately-impacted-immigrants-and-refugees-in-ontario-new-report-finds-1.5097363.
Choi K.H., Denice P., Haan M., Zajacova A. Studying the social determinants of COVID-19 in a data vacuum. Published online May 13, 2020. 10.31235/osf.io/yq8vu.
Thayn, J.B., Simanis, J.M., 2013. Accounting for spatial autocorrelation in linear regression models using spatial filtering with eigenvectors. Ann. Assoc. Am. Geogr. https://doi.org/10.1080/00045608.2012.685048. Published online.
Oshan, T.M., Li, Z., Kang, W., Wolf, L.J., Fotheringham, A.S., 2019b. mgwr: a python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. ISPRS Int. J. Geoinf. 8 (6), 269. https://doi.org/10.3390/ijgi8060269.
Kulldorff, M., 2009b. Information Management Services Inc. SaTScan v9.3: Software for the Spatial and Space-Time Scan Statistics. StatScan, Boston, USA. Published online.
Hughes, G.J., Gorton, R., 2013. An evaluation of SaTScan for the prospective detection of space-time Campylobacter clusters in the North East of England. Epidemiol. Infect. 141 (11), 2354–2364. https://doi.org/10.1017/S0950268812003135.