

A Spatial Epidemiology Case Study of Coronavirus (COVID-19) Disease and Geospatial Technologies

Muditha K. Heenkenda

Department of Geography and the Environment, Lakehead University, Thunder Bay, Canada
Email: muditha.heenkenda@lakeheadu.ca

How to cite this paper: Heenkenda, M.K. (2023) A Spatial Epidemiology Case Study of Coronavirus (COVID-19) Disease and Geospatial Technologies. *Journal of Geographic Information System*, 15, 540-562. <https://doi.org/10.4236/jgis.2023.155027>

Received: September 8, 2023

Accepted: October 22, 2023

Published: October 25, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Spatiotemporal pattern analysis provides a new dimension for data interpretation due to new trends in computer vision and big data analysis. The main aim of this study was to explore the recent advances in geospatial technologies to examine the spatiotemporal pattern of COVID-19 at the Public Health Unit (PHU) level in Ontario, Canada. The spatial autocorrelation results showed that the incidence rate (no. of confirmed cases per 100,000 population–IR/100K) was clustered at the PHU level and found a tendency of clustering high values. Some PHUs in Southern Ontario were identified as hot spots, while Northern PHUs were cold spots. The space-time cube showed an overall trend with a 99% confidence level. Considerable spatial variability in incidence intensity at different times suggested that risk factors were unevenly distributed in space and time. The study also created a regression model that explains the correlation between IR/100K values and potential socioeconomic factors.

Keywords

Spatial Epidemiology, Spatiotemporal Analysis, Space-Time-Cube, Spatial Regression

1. Introduction

Spatial epidemiology examines disease and its geographic variations based on demographic, environmental, behavioural, socioeconomic, genetic, and infectious risk factors [1] [2]. Spatial epidemiology dates back to John Snow's Cholera map of Soho, London, in the 1800s [2]. The map showed a concentration of deaths around one water pump, serving as the closest water source for affected houses. Snow's map was a visual and iconic representation and confirmed his theory about

Cholera as a waterborne disease rather than the belief that Cholera was airborne [3]. Later, different aspects of spatial epidemiology were tested on this map. Although Snow's contribution is a remarkable milestone in spatial epidemiology, the study didn't incorporate the population at risk and disease surveillance or the spatiotemporal spread of the disease into the analysis [2] [4]. However, the population at risk and disease surveillance are two essential factors in understanding the causes influencing the spatial pattern and disease spread rate over time [2] [3] [4]. Over the decades, with the development of computer science and big data analysis, Geographic Information Systems (GIS) and geospatial technologies, new opportunities were created to improve traditional reporting of disease on a global, national, and local scale. Also, spatial epidemiology became a rapidly advancing field that pushed GIS and geospatial technologies to measure, monitor and map disease occurrences at spatiotemporal scales [5].

Spatial epidemiology can be divided into three main areas: 1) disease mapping, 2) clustering and surveillance, and 3) geographical correlation studies or spatial regression [1]. Compared to tabular data, *disease maps* provide underlying patterns and an easy-to-understand visual data summary in the data. Disease maps allow recognizing, understanding, and monitoring incidences and prevalence until outbreak patterns occur ([6], p. 5). They can be used for descriptive analysis, especially highlighting areas with apparent high risk, creating underlying hypotheses, and managing resources—medicine, food, other aids distribution, and studying specific clusters in-depth. For example, Shannon [7] studied several early maps (the yellow fever maps of Seaman in 1786 and 1797 and the Pascalis map of cases in 1819) that were produced to derive theories of local origin and the non-contagious nature of yellow fever. Meselson *et al.* [8] investigated the unusual anthrax epidemic in Sverdlovsk, Union of Soviet Socialist Republics, in April and May 1979. The investigation found that most victims worked or lived, and livestock died in a narrow zone extending from the military facility to the southern city limit. Based on the geographic locations and whereabouts of victims, the study confirmed the reason for the outbreak—the escape of an aerosol of anthrax pathogen at the military facility [8].

Disease clustering and surveillance studies—an unusual concentration of incidences and prevalence or greater than expected incidences identified in a group of people living or working in the same area is called a disease cluster ([9], p. 150). A common definition of a cluster is a “geographically bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance” [10]. Spatial cluster analysis answers many public health questions, such as whether there are unusual clusters of public health events, where an unusually high or low prevalence of diseases, where the risk is high and low, etc. ([9], p. 150). Moreover, cluster analysis includes the population at risk to conclude. Regardless of the geographic location, all health data are initially dealt with by individuals (as point data) and released to the public by aggregating to administrative units (as polygon data) due to privacy and confidentiality. Hence, the spatial

resolution of the cluster analysis is always limited to the aggregated unit such as a county, public health unit, census division, postal code, etc. ([9], Chapter 2).

Many epidemiology studies employing spatial cluster analysis can be found in the literature; for example, Acharya *et al.* [11] analyzed the spatiotemporal pattern of dengue fever in Nepal from 2010 to 2014. The study identified the Chitwan district as the most likely cluster and the Jhapa district as the first secondary dengue cluster. Another study in Ethiopia found the clustering of Smear-positive tuberculosis cases in certain parts of Dabat, Ethiopia only [12]. Sloan *et al.* [13] studied the spatial and temporal clustering of patients hospitalized with influenza across the United States and identified many local clusters. It was further confirmed that socioeconomic factors strongly impacted the patient's hospitalization. The study on HIV-associated mortality in rural West Kenya, specifically the Kombewa health and demographic surveillance systems' population, showed significant hot spots and clusters [14].

The third aspect of spatial epidemiology—*geographical correlation studies* (or spatial regression) aims to analyze the exposure measures between environmental (relates to air, water or soil), behavioural (smoking, diet, exercise; socioeconomic), demographic (race and income), genetic, infectious risk factors and health outcomes ([1] [10]). Regression analysis allows modelling and exploring spatial relations and explaining the factors behind those spatial patterns. For example, in the early days of COVID-19 spread in New York City, Cordes & Castro [15] identified clusters with less testing and a low proportion of positive tests among the majority of the white population with higher income and education. According to Sloan *et al.* [13], there is a relationship between census tract-level socioeconomic factors and clusters of laboratory confirmed influenza hospitalizations in the United States—areas of high poverty were the clusters with higher than expected cases.

Spatiotemporal pattern analysis—one of the emerging research areas with the development of computer sciences and big data analysis is spatiotemporal pattern analysis, which provides a new dimension for data interpretation. Clusters only occur relatively to specific space and time intervals; thus, time is essential in spatial cluster analysis. The most recent example is the COVID-19 pandemic—health professionals are investigating unusual concentrations of disease occurrences in space and time. A spatiotemporal analysis is widely used in other applications, such as in the domains of biology, ecology, meteorology, medicine, transportation and forestry. The most novel space-time analysis method, a space-time cube, shows how phenomena change over time within a given geographic space. During the global pandemic of COVID-19, the spatiotemporal pattern of COVID-19 spread in China was analyzed and evaluated using a space-time cube [16]. Cordes & Castro [15] identified clusters of testing rates, positivity rates and proportions of tests that were positive to understand test access and thus evaluated contextual factors associated with found COVID-19 clusters in New York City. Mattera [17] developed a fuzzy clustering model to identify the geographical clustering of COVID-

19 in Italy. According to Boudou *et al.* [18], there was an apparent clustering of laboratory confirmed symptomatic COVID-19 in the urban areas of the Republic of Ireland. The study demonstrated the level of risk in densely populated areas at the early stage of a pandemic. Yang *et al.* [19] investigated socioeconomic factors affecting the regional spread of COVID-19 in China and confirmed that different combinations of socioeconomic factors affect different stages of the epidemic. However, the spatiotemporal analysis is not limited to the COVID-19 studies. Gurjav *et al.* [20] analyzed spatial and temporal variations of multidrug-resistant tuberculosis (MDR-TB) in Mongolia. They indicated a border-cross spread of MDR-TB along the Trans-Siberian Railway line.

COVID-19 Infectious Disease—Social interactions between individuals in human and animal populations tend to spread various infectious diseases. These diseases are a threat to public health and socioeconomic status. The most recent public health threat is the Coronavirus (COVID-19) disease caused by the SARS-CoV-2 virus [21]. To prevent infection or slow transmission, the recommended way is to stay at least one meter apart from others, wear a properly fitted mask, and wash your hands or use an alcohol-based rub frequently [21]. In general, mild to moderate respiratory illnesses was experienced by most people, while some people were affected seriously and required special treatments. The later section was mostly older people with underlying medical conditions like cardiovascular disease, diabetes, etc. However, according to the World Health Organization, anyone can get sick and become seriously ill or die at any age [21]. Therefore, mapping spatial occurrences of the disease, exploring the variation in the geographic space and predicting future occurrences are important. As a result, since the beginning of the pandemic, there have been many resources for public awareness, from various news items to scientific research outcomes. Most of them are visually appealing and intuitive; however, it is also important to analyze such variations over geographical space. It is crucial to explore emerging technologies, especially geospatial technologies, at a two/three/four-dimensional scale within this context. Therefore, the main aim of this study was to utilize state-of-the-art spatiotemporal pattern analysis tools and examine the spread of COVID-19 at the Public Health Unit (PHU) level in Ontario province, Canada. The specific objectives were to 1) analyze the spatial variation of the total number of confirmed cases from April 4, 2020, to December 31, 2021; 2) explore the spatial outliers and clusters at the PHU level; 3) investigate the spatiotemporal trend of the pandemic using a space-time cube; and 4) employ spatial regression techniques to explore the socioeconomic and demographic factors that affect such trends.

2. Data and Method

2.1. Study Area

The study was conducted in the province of Ontario, Canada. The smallest spatial unit considered for analysis was a Public Health Unit (PHU), formed by

grouping several urban and rural municipalities [22]. **Figure 1** shows 34 PHUs in Ontario, color-coded based on the total population and the total number of confirmed positive cases from April 4, 2020, to December 31, 2021. A higher number of cases appeared in the densely populated PHUs; in contrast, most larger spatial units have fewer confirmed cases.

2.2. Data

This study used daily COVID-19 pandemic data published by the Government of Ontario from April 4, 2020, to December 31, 2021, for individual PHUs [23]. The end of the period was based on Ontario’s updated eligibility for Polymerase Chain Reaction (PCR) testing and contact management guidance [24]. The downloaded health data file (*.csv) included active cases, resolved cases, and deaths as daily cumulative values; thus, the daily confirmed number of positive cases was calculated. When calculating daily cases, there were some issues with a few PHUs. For example, Simcoe Muskoka, Thunder Bay, Toronto, Windsor-Essex got minus values for daily confirmed cases for several days. These days were counted as

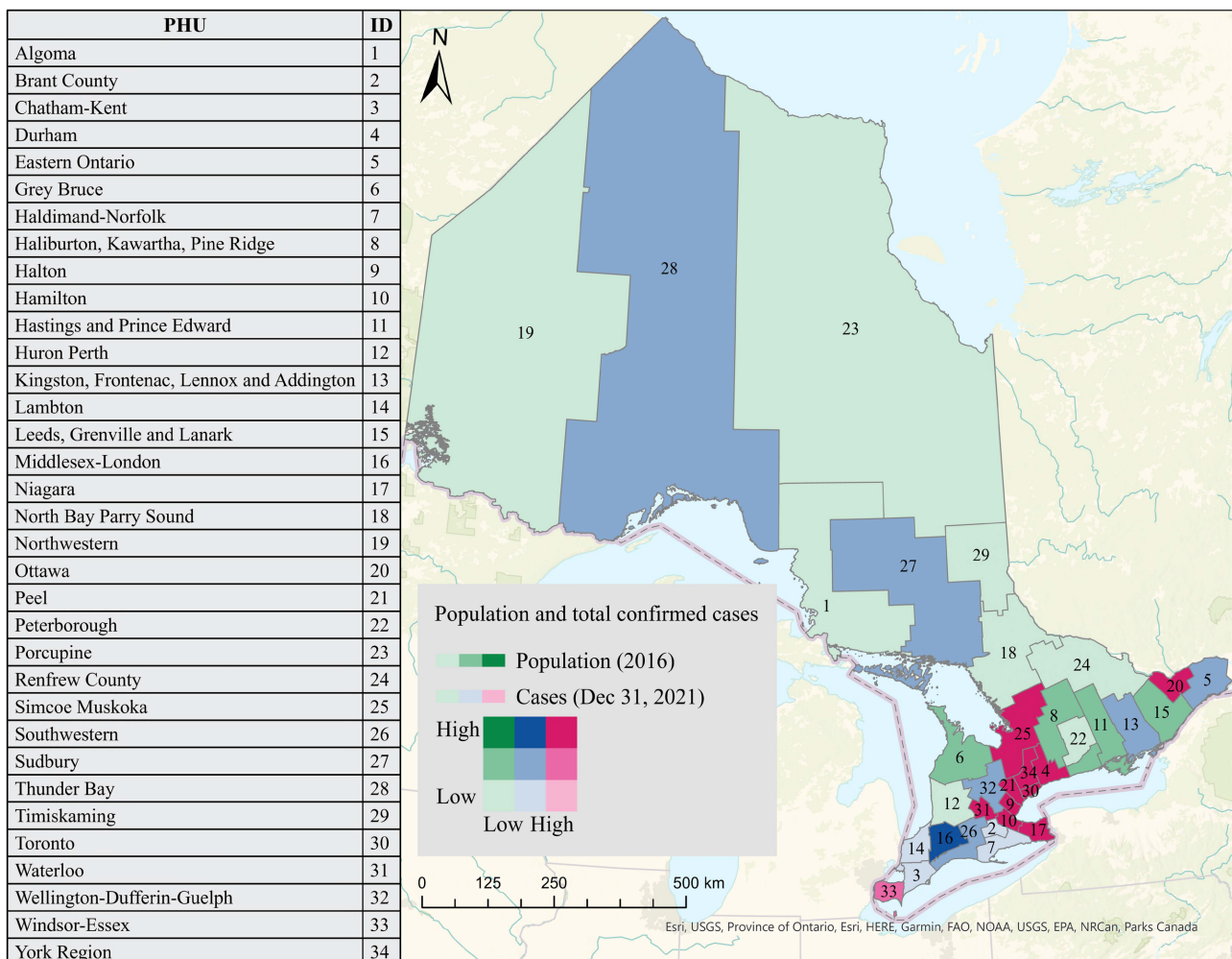


Figure 1. The total population and total number of confirmed positive cases in each Public Health Unit (PHU) in Ontario.

zero. Some PHUs had very accurate data sets, for example, Peel Region PHU got minus values for two days only.

The demographic and socioeconomic data were extracted from the Census Profile 2016 published by Statistics Canada [25]. At that time, there were 36 PHUs in Ontario. In 2018, Elgin-St Thomas Health Unit and Oxford County Health Unit were amalgamated to create Southwestern PHU [25]. In 2020, Huron County Health Unit and Perth District Health Unit were combined as Huron Perth PHU [26]. Census data (2016) were manipulated accordingly to represent the status of PHUs [25]. The vaccine data was downloaded from the Ontario Data Catalogue and calculated the number of people having the first dose only and the number of fully vaccinated (double dose) people for each PHU [28].

2.3. Method

The study explored all three categories of spatial epidemiology (disease mapping, clustering and surveillance, and geographic correlation). **Figure 2** explains the overall workflow of the analysis.

2.3.1. Data Exploration and Disease Mapping

The incidence rates per 100,000 population (hereafter denoted as “IR/100K”) were calculated for each PHU using Equation (1). Although age-adjusted standardized rates are most suitable for comparing different populations [29], age-specific data were unavailable under the Ontario Open Data license agreement.

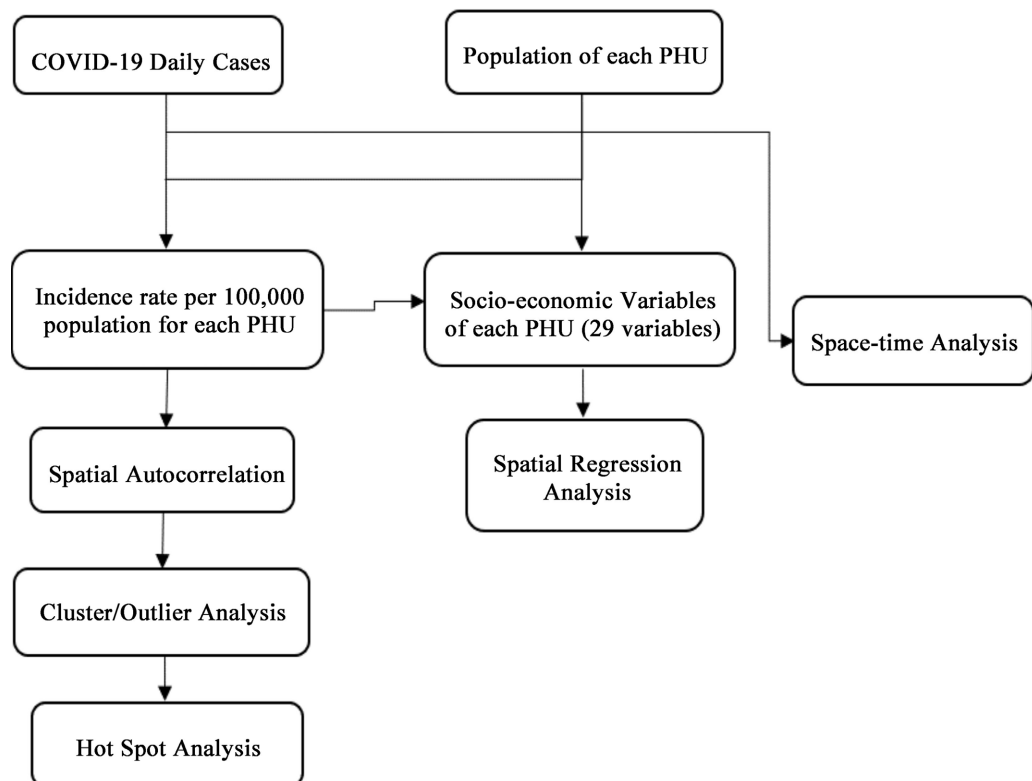


Figure 2. The workflow diagram of spatial epidemiology analysis in the province of Ontario, Canada.

$$(\text{IR}/100\text{K})_{\text{PHU}(x)} = \left(\frac{\text{Total number of confirmed cases}_{\text{PHU}(x)}}{\text{Total Population of PHU}(x)} \right) * 100,000 \quad (1)$$

According to the literature and data availability, 12 categories that influenced the spread of COVID-19 were identified. Each category was divided into sub-categories and used as exploratory variables in the spatial regression analysis (geographic correlation) (Table 1). All values were converted into “cases per 100,000 population” to bring all data into one scale. The IR/100K and exploratory variables were linked to each PHU polygon in the GIS environment for further analysis.

2.3.2. Disease Mapping and Spatial Autocorrelation

The spatial distribution of the IR/100K map was prepared, and the distribution trends were noted at the PHU level. The mean, median and standard deviation of the distribution were calculated. ArcGIS Pro software was used to analyze further the IR/100K data [30].

The statistical concept used to evaluate the clustering pattern is spatial autocorrelation, which considers attributes and locations. Spatial autocorrelation refers to the dependency or a measure of similarities between nearby observations or spatial units ([31] [32], p. 3). Statistics generally rely on observations being independent of one another; however, in spatial statistics, if autocorrelation exists in a dataset, this violates the fact that observations are independent of one another and indicates evidence of an important underlying spatial process or clustering [32]. Hence, the conceptualization of spatial relationships is essential.

Table 1. Demographic, socioeconomic and behavioural variables for each public health unit.

Criteria	Exploratory variable
Age characteristics	Less than 18 yrs. old; 18 - 39 yrs., 40 - 69 yrs., 70 - 79 yrs., over 80 yrs.
Household income	Less than 5K, 5K - 20K, 40K - 60K, 60K - 80K, 80K - 100K, over 100K
Education	No certificate, up to College, Bachelor or higher
Vaccination	One dose over 18 yrs. old, fully vaccinated over 18 yrs. old
Family composition	Two persons or less; two-five persons; more than five
Household composition	Less than two persons living in one house; two to five persons living in one house; five or more
Marital status	Single and never married; married or common; divorced, widowed, separated
Health conditions	Overweight, obese, high blood pressure
Health behaviours	Occasionally smoking, daily smoking, heavy drinking
Well-being	Perceived good physical health, good mental health, life stress
Citizenship	Canadian, non-Canadian
Confirmed case density for each PHU	Number of confirmed cases per square km

The measure of spatial autocorrelation depends on the method used to conceptualize the spatial relation of units and the degree of dependence on observations. The spatial relation is distance-dependent; thus, the strength reduces as distance increases ([33], p. 265). This study conceptualized the spatial relationship (the interaction between different PHUs) based on K nearest neighbours. This option was selected because some features are far from other features, and thus feature densities vary across the province. By choosing this option, it can be ensured that each PHU is evaluated based on at least its nearest eight neighbours [34].

Moran [35] developed a method to decide whether events in neighbouring counties are independent and the number of counties possessing a random arrangement. This statistical measure is called Moran's Index, which is positive when nearby counties are similar in attributes (clustered), negative when they are dissimilar (dispersed), and approximately zero when attributes are arranged independently of location (random) ([34] p. 16; [35]). In this study, Moran's Index calculated the spatial autocorrelation—the degree of influence employed by confirmed cases of one PHU over its neighbouring PHUs. K nearest neighbour method was used to select neighbouring polygons for PHUs.

2.3.3. Cluster and Outlier Analysis

Once the spatial pattern was identified, Local Indicators of Spatial Association (LISA) was performed to determine the spatial cluster of PHUs with high and low IR/100K. According to Anselin [36], LISA provides the strength of clustering similar values (positive autocorrelation-Local Moran's I (Equation (2) & (3)) of each observation and identifies outliers. For example, positive I of a PHU indicates the PHU is similar to neighbouring PHUs that form a cluster. When the PHU is an outlier, the I is negative, meaning dissimilarity [37]. A p-value should be statistically significant to decide on a cluster or outlier.

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{i,j} (x_j - \bar{X}) \quad (2)$$

where x_i is an attribute for feature i , \bar{X} is the mean of the corresponding attribute, $w_{i,j}$ is the spatial weight between feature i and j , and;

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n (x_j - \bar{X})^2}{n-1} \quad (3)$$

where n is the total number of features.

2.3.4. Hot and Cold Spot Analysis

Getis-Ord G_i^* statistics was used to find out statistically significant spatial clusters of PHUs with either high (hot spots—statistically significant positive z-score) or low (cold spots—statistically significant negative z-score) values ([38] [39]). For example, a hot spot is a PHU with a high IR/100K value surrounded by other PHUs with high values. The spatial relationship was conceptualized based on K nearest neighbors.

2.3.5. Space-Time Cluster Analysis

The most common way to analyze spatial and temporal data trends is to divide the data set into time snapshots and explore each separately. Although this works well for trend analysis, the question is how to decide on breakpoints, for example, daily, weekly or monthly. Space-time results are traditionally visualized using heat maps [40]. However, the incorporation of information about the temporal dimension is questionable. 3D visualization that uses time as the third dimension is more suitable for point features which can be extruded to reflect temporal progression [40].

When there are time-stamped data, they can be aggregated spatially (aggregation polygons) as a grid cube structure to understand spatiotemporal patterns. The grid cube structure contains rows (x) and columns (y) that define the spatial extent of the cube and time step (t) to determine the temporal extent (Figure 3). Each bin has a fixed location in space (x, y) and time (t). In each cube bin, number of occurrences will be counted, and summary statistics and the trend for bin values across time at each location will be calculated [16] [40] [41].

A space-time cube was prepared by aggregating the number of positive cases for one week (time-step interval) of space-time bins and converting it into a netCDF data structure to be compatible with ArcGIS Pro software [30]. This grid cube contains row, column and time steps (Figure 3(a)). Each grid cube is called a “spatial bin,” which has a fixed location in space (X, Y) and in time (t) (Figure 3). In this study, the centre of each PHU was selected as the (X, Y) of the bin. Then, a unique “LOCATION_ID”, time-step-ID, and COUNT were assigned to each PHU [42]. Since the LOCATION_ID is unique to each bin, the value of “ t ” will change as time progresses at each time-step interval (Figure 3(b)).

Once the time-space-cube was created, significant clusters and outliers were identified. These results provided time periods with unexpectedly high rates of

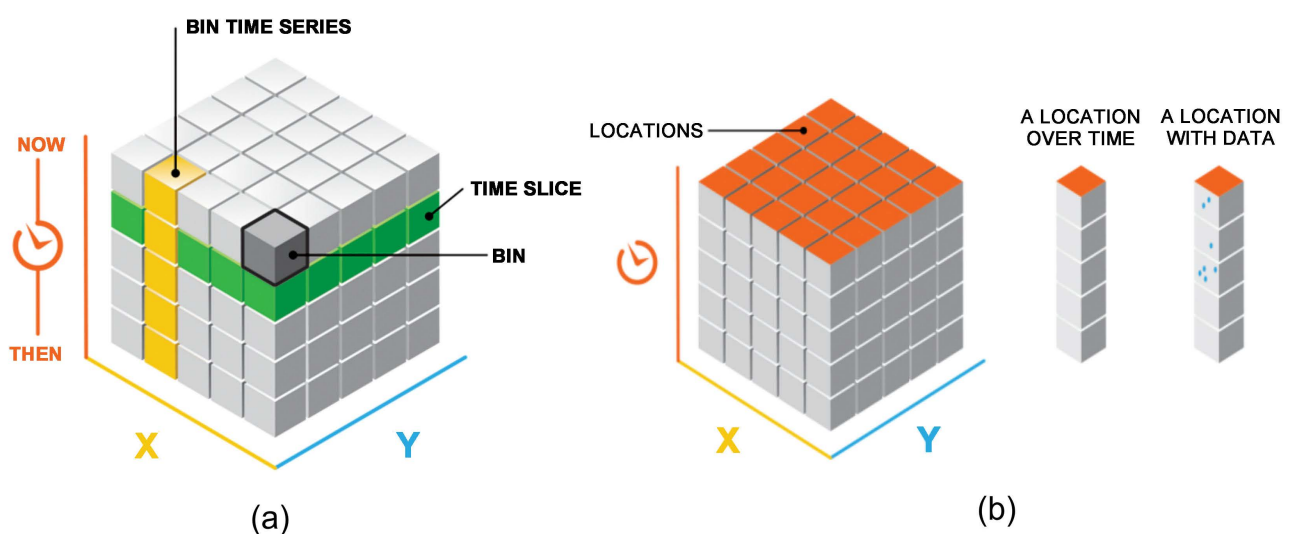


Figure 3. (a) The grid cube structure of the space-time cube; (b) each bin has a location in space (X, Y) and in time (t) and LOCATION_ID, time_step_ID, COUNT value and other values as calculated (source: [22]).

disease outbreaks across Ontario. After that, statistically significant hot and cold spots over time were determined. The “Emerging Hot Spot Analysis” tool in ArcGIS Pro software identifies trends at four stages—new, intensifying, diminishing, and sporadic hot and cold spots [43]. The conceptualization of spatial relationships was used as the earlier steps. The emerging hot spots for each month were analyzed and compared with Public Health Ontario data [44].

2.3.6. Spatial Regression

A global model of regression—Ordinary Least Squares (OLS), a multiple linear regression method that fits a linear equation to a model, is the starting point of most spatial regression analyses [45]. It provides reliable statistics for examining linear relationships at the point level or regions (polygons) representing spatial units [33]. This study used an OLS regression for the dependent variable (IR/100K) and 40 explanatory variables (Table 1) to create an equation that estimates values for the dependent variable. All possible combinations of the model were evaluated first using exploratory regression. The report included redundancy, completeness, significance, bias and performance combinations. Variables exhibiting multicollinearity were selected based on their significance to the model. All variables with less than 10% significance to the model were eliminated from the second iteration. That means there were 24 variables only for the second iteration. The process was repeated and obtained 14 variables only for the third iteration. Finally, seven (7) selected variables were used for OLS regression. The model performance was evaluated against the six statistical tests, and the spatial autocorrelation of model residuals was tested.

Ordinary Least Squares (OLS) Regression

When estimating the model with two independent variables, Equation (4) shows the OLS equation used. The OLS calculates its estimates to minimize the sum of squared residuals as described in Equation (5) ([33] [45]). Generally, Equation (4) can be extended to k independent variables and estimates $k + 1$ will be selected to minimize the sum of squared residuals.

$$\tilde{Y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 \quad (4)$$

where \tilde{Y} is the dependent variable; x_i ; $i = 1, 2$ are independent variables; $\tilde{\beta}_i$; $i = 0, 1, 2$ are the estimates of β_i .

$$\sum_{i=1}^n (Y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \tilde{\beta}_2 x_{i2})^2 \quad (5)$$

where given n observations on y , x_1 and x_2 , $\{(x_{i1}, x_{i2}, Y_i) : i = 1, 2, \dots, n\}$; $\tilde{\beta}_i$; $i = 0, 1, 2$ are the estimates of β_i .

Exploratory Regression

Finding the best OLS regression model is iterative, mainly when many potential exploratory variables exist. Exploratory regression, similar to the stepwise regression that iteratively examines the statistical significance of exploratory variables in the linear regression model, can be used to find the best possible combination of the potential exploratory variable best to explain the dependent va-

reliable [46]. The advantage of using exploratory regression over stepwise regression is that exploratory regression evaluates the OLS model performance in six different ways than just adjusted R^2 value [46].

3. Results

3.1. Data Exploration and Disease Mapping

Figure 1 shows each PHU's total population and number of confirmed cases. Most Southern Ontario PHUs showed a high population and high no. of cases, while Northern Ontario had a low population and low no. of cases. However, the spatial extent of Northern Ontario PHUs is higher than others.

Figure 4 shows the IR/100K values of each PHU. North Bay Parry Sound PHU had the lowest IR/100K value (1408.4962). Peel Region PHU had the highest IR/100K value (9367.003). The mean, median and standard deviations were 3960.3, 4013.416 and 1816.19, respectively.

3.2. Spatial Autocorrelation

The spatial pattern indicated was unlikely to have occurred due to a random chance—Moran's Index (0.2935) and z-score (7.2779) were positive; the p -value was almost zero. Therefore, it was worth investigating the underlying clustering pattern at a local scale.

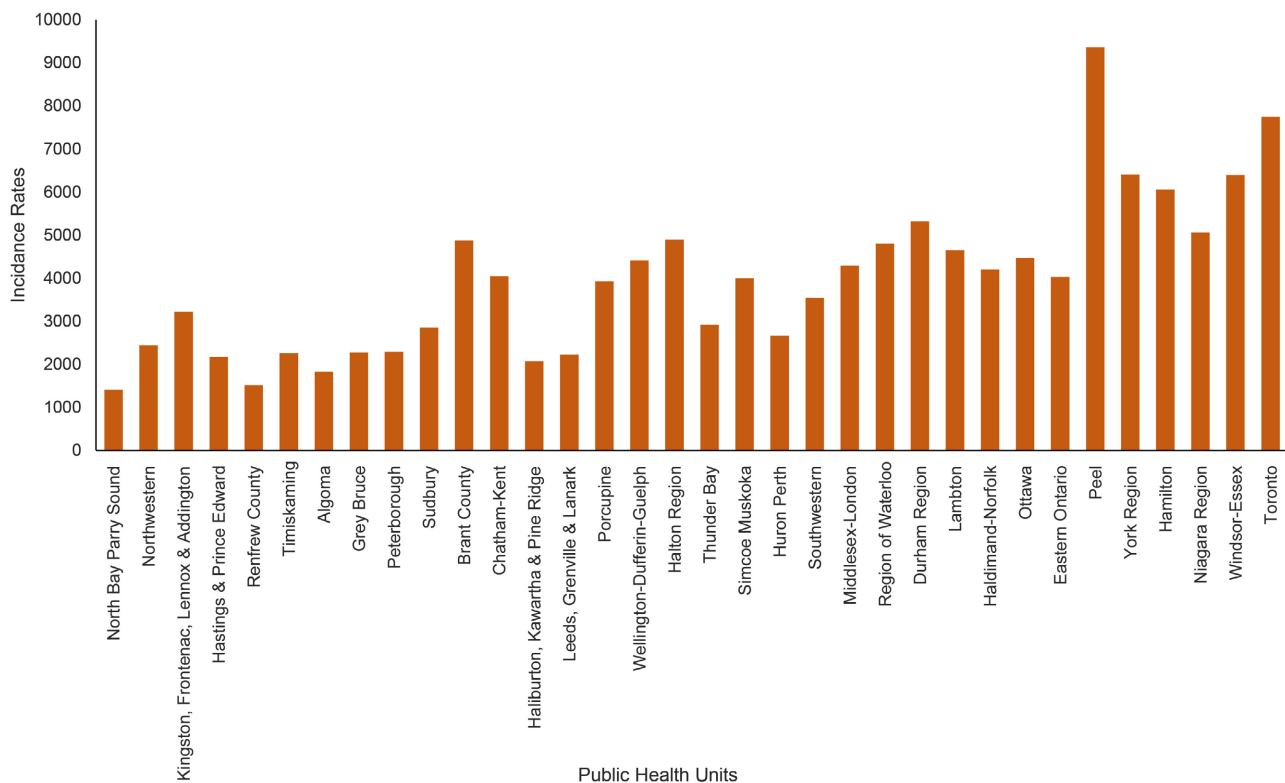


Figure 4. Incidence rates (IR/100K) of the confirmed cases in each Public Health Unit (PHU) ((IR/100K = Total no. of confirmed cases/total population) * 100,000).

3.3. Cluster and Outlier Analysis

These results were focused on the neighbourhood-level relationship between IR/100K values. The Observed General G value was greater than the Expected G value, and the z-score (4.062) was positive (High/Low clustering report). This indicated that PHUs with high IR/100K tend to cluster together. However, this statistic didn't show spatial locations of concentration of values, and only high values tend to occur near each other. The next step was to find locations in the study area that were statistically different from their neighbours in space (cluster and outlier analysis). Most PHUs showed a statistically insignificant pattern (**Figure 5(a)**). For example, Northwestern, Thunder Bay, Porcupine and Algoma PHUs were statistically not significant. Timiskaming and Sudbury PHUs that showed low values were clustered with previously mentioned insignificant PHUs (**Figure 5(a)**). High values were clustered with another high value for 13 PHUs (Chatham-Kent, Middlesex-London, Haldimand-Norfolk, Niagara Region, Hamilton, Brant County, Region of Waterloo, Halton Region, Wellington-Dufferin-Guelph, Peel, Toronto, York Region, Durham Region PHUs) in the southern Ontario. Three PHUs (Grey Bruce, Huron Perth, and Southwestern) with low values were clustered with surrounding high values (**Figure 5(a)**).

3.4. Hot Spot Analysis

Significant spatial clusters of high and low values of PHUs in the geographic space were identified. Most of the Northern PHUs showed cold spots with a 99% or 90% confidence level. Five PHUs in southern Ontario (Eastern Ontario, Ottawa, Leeds, Grenville and Lanark District Health Unit, Kingston, Frontenac and

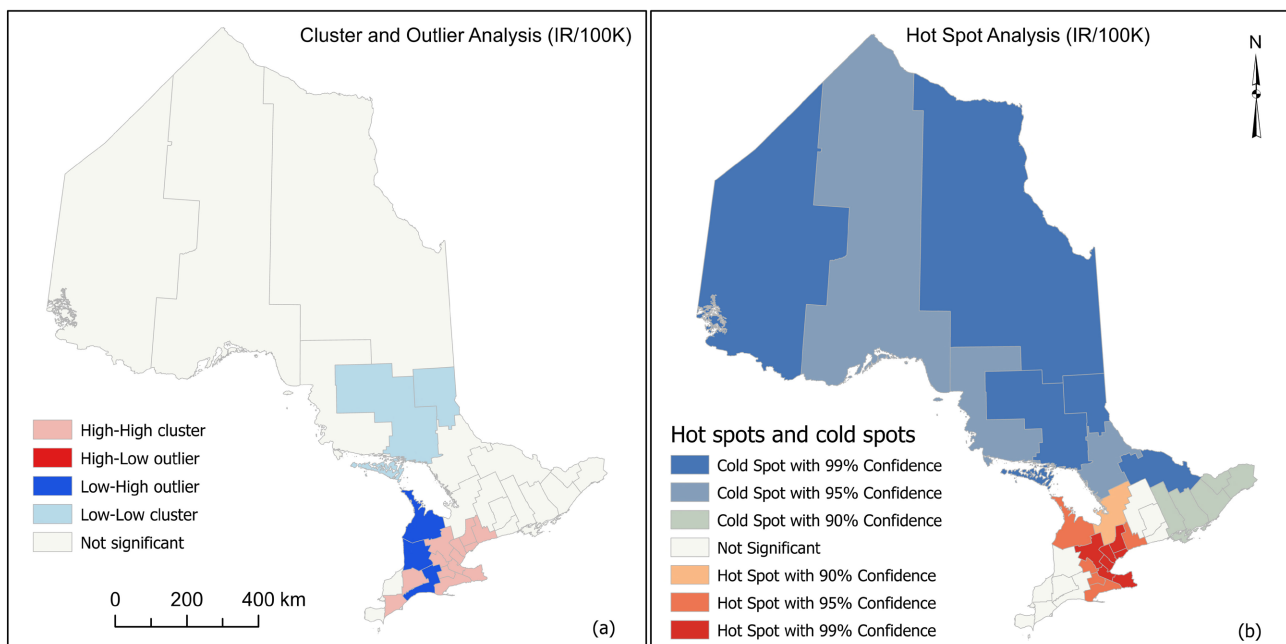


Figure 5. (a) Cluster and Outlier analysis results; (b) Hot spot analysis (PHU names are available in **Figure 2**) (please refer to **Figure 1** for PHU names).

Lennox and Addington, and Hastings and Prince Edwards Counties) had a cold spot with a 90% confidence level (**Figure 5(b)**). A few PHUs had statistically insignificant IR/100K values. A hot spot with 99% confidence level included seven PHUs (York Region, Toronto, Peel, Wellington-Dufferin-Geulph, Halton Region, Hamilton and Niagara Region). North Bay Parry Sound PHU got the lowest IR/100K value and was part of a cold spot. The highest value was from Peel Region PHU and was part of a hot spot identified at 99% confidence level (**Figure 5(b)**).

3.5. Temporal Pattern Analysis

The space-time cube was created with 92 time steps (weekly from April 1, 2020 to December 31, 2021). The Mann-Kendall trend test was performed on each PHU as an independent bin time-series test [22]. The overall increasing data trend was identified. **Figure 6(a)** showed a statistically significant trend for all PHUs except Peel and Toronto. However, the temporal aggregation didn't find an overall trend direction; the overall trend direction was statistically not significant ($p\text{-value} = 0.2181$).

Unlike local cluster and outlier analysis in space (**Figure 5(a)**), the local clusters and outliers in space and time were analyzed. All northern Ontario PHUs showed a similar pattern—low confirmed no. of cases clustered with neighbouring low confirmed no. of cases. However, southern Ontario PHUs showed multiple clustering patterns in space and time. The dates of different waves published by Public Health Ontario aligned with the change in mean shift dates of each PHUs. For example, **Figure 7** and **Figure 8** show that the mean value of

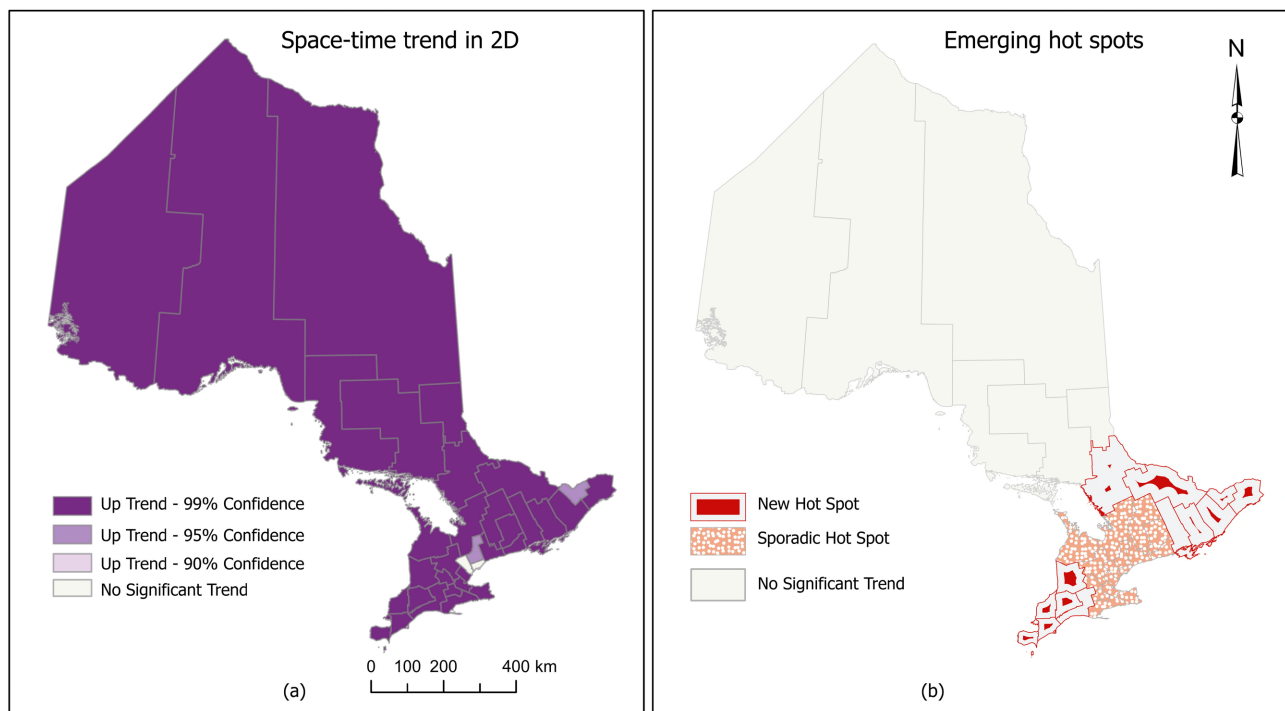


Figure 6. (a) Visualizing space-time trend in 2D; (b) Emerging hot spots.

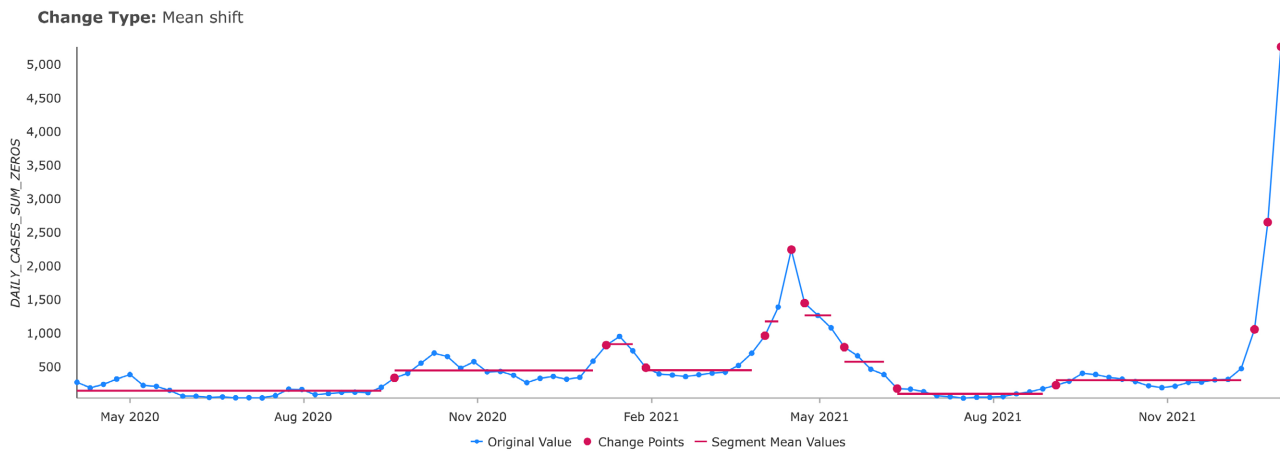


Figure 7. Change in the mean shift for each week: Ottawa PHU.

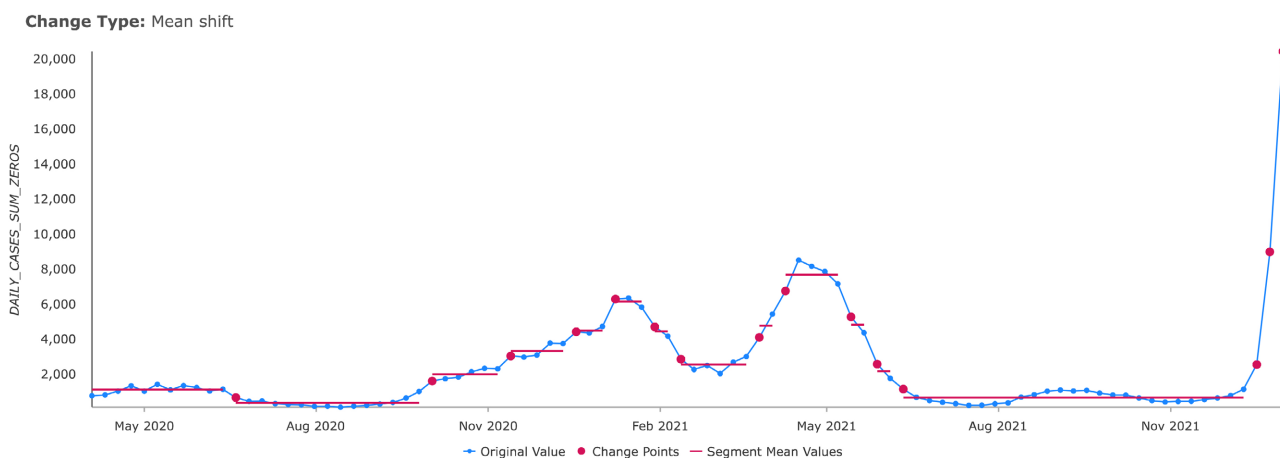


Figure 8. Change in the mean shift for each week: Toronto PHU, during this period, the mean was shifted 15 times.

IR/100K shifted several times from September 1, 2020 to February 28, 2021, the official second wave published by Public Health Ontario.

During the second wave, Toronto PHU shifted the mean five times. Ottawa and York Region PHUs indicated the uptrend with 95% confidence level. After that, the shift in the mean of confirmed no. of cases for each week was detected. Ottawa PHU showed the following pattern (Figure 7). There were ten times the mean value shifted and number of cases increased or decreased.

Only three PHUs (North Bay Parry Sound, Haliburton, Kawartha, Pine Ridge, and Peterborough) had 5 - 6 mean change points, while others had more time locations where the mean shifted. Toronto, Timiskaming and Leeds, Grenville and Lanark PHUs had the highest change points (14 - 16). For example, Figure 8 shows Toronto PHU mean shift variation.

A Space-time cluster indicates an elevated confirmed number of cases presented inside a PHU in a specific time period compared to the PHUs in the same period. All northern PHUs didn't show a statistically significant pattern when analyzing the emerging hot spots. However, there were many new hot spots and sporadic hotspots in the Southern Ontario (Figure 6(b)). Hastings and Prince

Edward Counties and Southwestern PHUs were the only two PHUs that didn't show statistically significant emerging hot spots in Southern Ontario.

3.6. Spatial Regression

After exploratory regression (testing several combinations of potential exploratory variables to model the IR/100K (dependent variable)), the following model was identified as the best model explaining the IR/100K variation over the geographic space (Table 2). The household income, vaccination, citizenship, and health behaviour criteria were not statistically significant to explain the dependent variable.

OLS Model Performance

The model was assessed against six different criteria. First, the strength and type of relationship that the explanatory variable has to the dependent variable was checked. The T test was used to identify statistically significant explanatory variables. For example, if one of the model coefficients is zero, the associated explanatory variable is not representing the model. Most of the exploratory variables were statistically significant, indicating an excellent contribution to the model except population 18 - 39 years old and bachelor's degree or higher education levels. Next, the model stationarity was tested using Koenker statistics. The null hypothesis for this test was that the model was stationarity. That means the explanatory variables in the model have a consistent relationship to the dependent variable throughout the study area. Since the Koenker statistics were not statistically significant, the null hypothesis cannot be rejected and provided a stationary model.

The second test is to check the signs of the model coefficients. The sign indicates the positive or negative relation to the model, that is, an increase or decrease of the influence of the variable, respectively. Most of these signs were as expected; for instance, the IR/100K values density per sq.km was positive, while the population 60 - 80 yrs old and married or common law status showed a negative relationship.

Next, the redundancy of the explanatory variables was checked. The Variance

Table 2. Exploratory variables and their OLS model coefficients.

Exploratory variable	Coefficient
<i>Intercept</i>	8611.8
Population (18 - 39 yrs old)	0.067
Population (60 - 80 yrs old)	-0.621
Married or common law	-0.210
Family of five or more	1.378
Bachelors degree or higher	0.021
Life stress	0.210
Density of IR/100K per square km	4.301

Inflation Factor (VIF) calculates the variance of the estimated regression coefficient because of collinearity. VIF of all explanatory variables were less than 7.5, and thus there were no redundant variables in the model [47]. When the model is defined correctly, the model residuals should be normally distributed. The degree of model biasness was assessed using Jarque-Bera statistics. The null hypothesis for this test was the model residuals were normally distributed. At a 90% confidence level, the Jarque-Bera statistics were not statistically significant. The model is not biased.

The fifth test was to look at any evidence of missing key explanatory variables. If there are missing variables, then the model residuals should exhibit a spatial pattern; therefore, the spatial autocorrelation of model residuals were checked to see whether they were randomly distributed than clustering. The Moran's Index (0.027539), positive z score (0.519773) and p-value (0.6032) indicated a statistically not significant performance. Therefore, the model residuals were randomly distributed over the study area, and there were no missing key variables.

Finally, the model's performance was assessed using the adjusted coefficient of determination (R^2). This explains the percentage of the variation of the dependent variable explained by the explanatory variables. The adjusted R^2 value (0.76) showed a strong model.

4. Discussion

4.1. Data Exploration and Disease Mapping

Spatial epidemiology studies are susceptible to the type of data available, their quality and the spatial relationships between data. For instance, Ontario data were aggregated to PHUs due to privacy concerns rather than their actual spatial locations as point data. Point data provides a localized exposure on a disease map, while area data will generalize or smooth away some of the true variability. Small areas may capture the underlying pattern of the incidences at fine-grained variation over space, especially when there are variable sizes of spatial aggregation units ([11] [46]). However, when the extent of aggregation units differs in the risk population, the data might suffer small number problem ([11] [46]). The province of Ontario also suffers from the small number problem; for instance, Thunder Bay District Health Unit is larger in size and less in population; in contrast, Toronto PHU has a smaller extent and larger population. Hence, although **Figure 1** compares the number of confirmed cases and risk population, the underlying spatial pattern cannot be uncovered correctly. For instance, **Figure 1** shows Thunder Bay and Sudbury had moderate population and no. of confirmed cases, but cluster and outlier analysis identified Thunder Bay as statistically not significant and Sudbury as a part of a low-low cluster. This study considered the incidence rate per 100,000 population (Equation (1)) instead of the confirmed number of cases for further analysis to minimize these effects. However, the main issue of the spatial distribution of this rate map is that when the population of the spatial unit is smaller, the influence from random variability is

higher [10]. Therefore, instead of disease maps, this study statistically evaluated the spatial pattern of crude rates statistically. Although the best solution is to assess age-adjusted standardized rates, it was difficult due to a lack of data [10] [29] [48].

The mean and median IR/100K values were approximately equal, and IR/100K data were normally distributed (Figure 4). Normally distributed data will provide unbiased results when conceptualizing the spatial relations between neighbouring PHUs. When calculating daily confirmed cases from the COVID-19 data, there were some discrepancies, and some PHUs got negative values, which were considered zero in the analysis (Simcoe Muskoka, Toronto, Thunder Bay, and Windsor-Essex PHUs). The other issue is that the recent publicly available demographic and socioeconomic data related to the Census 2016 were outdated. Although we wouldn't expect huge differences between recent statistics, some variations might exist.

4.2. Spatial Pattern Analysis

Spatial analysis in GIS considers two distinct types of data: attributes of spatial features and their locations ([32], p. 03). Although many GIS analysis tasks are limited to attribute analysis only, the influence from the location is indirect; for example, in its simplest form, the locations define whether the attributes are related to the study area. Spatial relations sometimes misperceive the understanding of spatial data. According to the first law of geography—“Everything is related to everything else, but near things are more related than distant things,” spatial effects have an interrelated meaning [49]. Hence, the analysis of both attributes and locations is essential. Once the spatial resolution of pattern analysis is defined, the conceptualization of spatial relationships based on the modeling of how features interact in space should be determined [34]. Results may vary with the selection of the spatial conceptualization method. Contiguity, or adjacency that describes whether the geographic units share boundaries and proximity and the distance between spatial units, are the common approaches for defining spatial weights in GIS ([9], p. 159). The adjacency finds spatial units that share common boundaries as “neighbours” using inverse distance, travel time, fixed distance, K nearest neighbours and contiguity [36]. This study used K nearest neighbours to ensure that each PHU is evaluated based on at least eight neighbours due to the varying PHU densities across the province [36].

Spatial autocorrelation provides similarities between attributes and locations and is unavailable from another form of statistical analysis ([32], p. 5). An insight into Ontario's pandemic spatial distribution was noticed—IR/100K values were clustered at the PHU level; however, the tendency was to cluster high values together. The most Northern PHUs with less population and enormous land mass show a low-low clustering pattern. Huron Perth, Grey Bruce and South-western had low values but clustered with surrounding PHUs with high values. Most Southern Ontario PHUs had high population densities and IR/100K values

and exhibited hot spots in the analysis. Here, we would notice the influence of neighbouring PHUs well; however, when the distance between PHUs increased, the influence diminished as explained in the Tobler's first law of geography [49].

4.3. Spatiotemporal Pattern Analysis

With the recent advancement of computer technology and big data analysis algorithms, GIS has changed the map-making process and plays an essential role in public health and epidemiology [50]. Traditionally maps were used to communicate information; however, today's mapping process includes data visualization in two-dimensional (2D) or three-dimensional (3D) space, exploration, and analysis. Given that GIS transforms how we describe and study the earth, rapid diffusion of this technology to various applications is evident. For example, there is a growing interest in applying GIS to public health; thus, spatiotemporal pattern analysis of disease.

The pandemic trend and distribution of the outbreak were tested using thematic maps, line graphs and space-time-cube. The mean shift line graphs showed the temporal variation, but spatial variation cannot be pinpointed since the data was aggregated into polygons. Main shift line graphs are useful for longitudinal analysis—to differentiate different waves and their slopes, although a comprehensive space-time analysis was difficult. The spatiotemporal weekly cluster and outlier analysis at PHU level showed an uneven distribution. All northern PHUs (six) and Lambton PHU exhibited a low-low clustering pattern only throughout the period, while the remaining southern PHUs showed multiple-type clustering. The weekly emerging hot spot analysis identified IR/100K value trends based on the space-time cube. There were several new hot spots emerged toward the end of the period (December 31, 2021) (Figure 6(b)). This is comparable with the "Weekly COVID-19 intervention timeline," published by the Canadian Institute for Health Information [51]. Also, the overall space-time positive up trend indicates the same results (Figure 6(a)).

4.4. Spatial Regression

The study selected 40 exploratory variables from 12 categories representing demographic, socioeconomic, and behavioural variations. This comprehensive selection was entirely based on factors positively affecting disease transmission and data availability. The performance of the regression model was relatively high, and the adjusted R^2 value was 0.76. The regression model's most contributing explanatory variables were the IR/100K density per square km and family of five or more people living in one household. These results matched the COVID-19 prevention guidelines published by various health authorities during the pandemic. The 60 to 80-year-old population and the social status of married or common-law partners negatively affected the model as these groups can be more socially responsible, taking more precautions and following guidelines to minimize the spread. The population group—older than 60 is identified as more

vulnerable to the disease than other age groups [52]. Persons who perceived life stress (well-being) and those with a bachelor's or higher degree (education) also positively contributed to this regression model. There was no evidence that the model results varied from PHU to PHU, as model results were stationary.

Vaccination statistics were included as explanatory variables (one dose over 18 years old and fully vaccinated over 18 years old) initially; however, they didn't contribute to the model performance. Health condition, health behaviour and citizenship data didn't significantly influence the model. Although tobacco use increases the risk of many pulmonary diseases, there is no reliable evidence that COVID-19 infection among them is high. The study conducted by Jordan et al [53] didn't find a statistically significant correlation between the high risk of severe COVID-19 disease progression and former smokers. Further, although the amount of alcohol consumption weakens the immune system, no studies confirm the correlation between alcohol use and COVID-19 infection rates [53]. The impact on those in different population groups might vary.

The spatial aggregation of IR/100K values at the PHU level leads to the loss of fine-grained spatial variation in the number of confirmed cases. Hence, the study recommends a sensitivity analysis to assess how different levels of spatial aggregation impact the investigation, for example, at the census division and census sub-division levels; however, this might not be possible for the entire province due to a lack of data.

5. Conclusions

Spatiotemporal pattern analysis is one of the emerging trends in Geographic Information Systems due to the recent development of computer vision and big data analysis. The spatiotemporal analysis provides a new dimension for data interpretation. The main aim of this study was to explore the recent advances in geospatial technologies to examine the spatiotemporal pattern of COVID-19 at the Public Health Unit (PHU) level in Ontario province, Canada. To minimize the small number problem, the confirmed number of cases was converted into the incidence rates (no. of confirmed cases per 100,000 population—IR/100K). The spatial autocorrelation results indicated that IR/100K was clustered at the PHU level. The high/low clustering results showed that the high IR/100K values have a higher tendency to cluster high IR/100K values. A few PHUs in Southern Ontario were identified as hot spots, while northern PHUs were cold spots.

The weekly space-time-cube showed an overall up trend—individual PHU trend graphs matched with different COVID-19 waves published by Public Health Ontario. The COVID-19 distribution was uneven throughout the province. Some PHUs shifted the IR/100K mean value very frequently during the study period (e.g. Ottawa and Toronto), indicating intermediate waves. Considerable spatial variability in incidence intensity at different times suggests that risk factors were unevenly distributed in space and time. Hence, the study also explored various socioeconomic factors influencing COVID-19 distribution, and a regression model

was established. The study demonstrated the high-risk level in densely populated areas in Ontario. Although the study showed a complete spatial analysis of confirmed no. of COVID-19 case distribution within a geographic space and time, these results should be used with caution due to limitations associated with the data. For example, incident data (points) were aggregated into polygons, and some variations might be smoothed away. There are some large PHUs with low populations and small PHUs with high populations. The results can be used as a foundation to pursue further studies on COVID-19 distribution in Ontario and improve them to make management decisions. Also, the study demonstrated the details that geospatial technologies could uncover in spatial epidemiology. For example, the space-time-cube can be used for longitudinal analysis of disease incidences—to divide data into different waves and identify their slopes, etc. Geospatial technologies can also be useful for uncovering different phenomena associated with data such as understanding the spatial variations based on per population density or per area density.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Elliott, P. and Wartenberg, D. (2004) Spatial Epidemiology: Current Approaches and Future Challenges. *Environmental Health Perspectives*, **112**, 998-1006. <https://doi.org/10.1289/ehp.6735>
- [2] Shiode, N., Shiode, S., Rod-Thatcher, E., Rana, S. and Vinten-Johansen, P. (2015) The Mortality Rates and the Space-Time Patterns of John Snow's Cholera Epidemic Map. *International Journal of Health Geographics*, **14**, Article No. 21. <https://doi.org/10.1186/s12942-015-0011-y>
- [3] Snow, J. (1855) On the Mode of Communication of Cholera. 2nd Edition, John Churchill, London.
- [4] Ostfeld, R.S., Glass, G.E. and Keesing, F. (2005) Spatial Epidemiology: An Emerging (or Re-Emerging) Discipline. *Trends in Ecology & Evolution*, **20**, 328-336. <https://doi.org/10.1016/j.tree.2005.03.009>
- [5] Tatem, A.J. (2018) Innovation to Impact in Spatial Epidemiology. *BMC Medicine*, **16**, Article No. 209. <https://doi.org/10.1186/s12916-018-1205-5>
- [6] Collins Kelley, L. and Breeze, R.G. (2005) Investigation of Suspicious Disease Outbreaks. In: Breeze, R.G., Budowle, B. and Schutzer, S.E., Eds., *Microbial Forensics*, Academic Press, Cambridge, 187-212. <https://doi.org/10.1016/B978-012088483-4/50013-5>
- [7] Shannon, G.W. (1981) Disease Mapping and Early Theories of Yellow Fever. *The Professional Geographer*, **33**, 221-227. <https://doi.org/10.1111/j.0033-0124.1981.00221.x>
- [8] Meselson, M., Guillemin, J., Hugh-Jones, M., *et al.* (1994) The Sverdlovsk Anthrax Outbreak of 1979. *Science*, **266**, 1202-1208. <https://doi.org/10.1126/science.7973702>
- [9] Cromley, E.K. and McLafferty, S.L. (2012) GIS and Public Health. 2nd Edition,

Guilford Press, New York.

- [10] Olsen, S.F., Martuzzi, M. and Elliott, P. (1996) Cluster Analysis and Disease Mapping—Why, When, and How? A Step by Step Guide. *The BMJ*, **313**, 863-866. <https://doi.org/10.1136/bmj.313.7061.863>
- [11] Acharya, B.K., Cao, C., Lakes, T., *et al.* (2016) Spatiotemporal Analysis of Dengue Fever in Nepal from 2010 to 2014. *BMC Public Health*, **16**, Article No. 849. <https://doi.org/10.1186/s12889-016-3432-z>
- [12] Tadesse, T., Demissie, M., Berhane, Y., Kebede, Y. and Abebe, M. (2013) The Clustering of Smear-Positive Tuberculosis in Dabat, Ethiopia: A Population Based Cross Sectional Study. *PLOS ONE*, **8**, e65022. <https://doi.org/10.1371/journal.pone.0065022>
- [13] Sloan, C., Chandrasekhar, R., Mitchel, E., *et al.* (2020) Spatial and Temporal Clustering of Patients Hospitalized with Laboratory-Confirmed Influenza in the United States. *Epidemics*, **31**, Article ID: 100387. <https://doi.org/10.1016/j.epidem.2020.100387>
- [14] Sifuna, P., Otieno, L., Andagalu, B., *et al.* (2018) A Spatiotemporal Analysis of HIV-Associated Mortality in Rural Western Kenya 2011-2015. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, **78**, 483-490. https://journals.lww.com/jaids/Fulltext/2018/08150/A_Spatiotemporal_Analysis_of_HIV_Associated.1.aspx <https://doi.org/10.1097/QAI.0000000000001710>
- [15] Cordes, J. and Castro, M.C. (2020) Spatial Analysis of COVID-19 Clusters and Contextual Factors in New York City. *Spatial and Spatio-Temporal Epidemiology*, **34**, Article ID: 100355. <https://doi.org/10.1016/j.sste.2020.100355>
- [16] Mo, C., Tan, D., Mai, T., *et al.* (2020) An Analysis of Spatiotemporal Pattern for COVID-19 in China Based on Space-Time Cube. *Journal of Medical Virology*, **92**, 1587-1595. <https://doi.org/10.1002/jmv.25834>
- [17] Mattera, R. (2022) A Weighted Approach for Spatio-Temporal Clustering of COVID-19 Spread in Italy. *Spatial and Spatio-Temporal Epidemiology*, **41**, Article ID: 100500. <https://doi.org/10.1016/j.sste.2022.100500>
- [18] Boudou, M., Khandelwal, S., ÓhAiseadha, C., *et al.* (2023) Spatio-Temporal Evolution of COVID-19 in the Republic of Ireland and the Greater Dublin Area (March to November 2020): A Space-Time Cluster Frequency Approach. *Spatial and Spatio-Temporal Epidemiology*, **45**, Article ID: 100565. <https://doi.org/10.1016/j.sste.2023.100565>
- [19] Yang, L., Qi, C., Yang, Z., *et al.* (2021) Socio-Economic Factors Affecting the Regional Spread and Outbreak of COVID-19 in China. *Iranian Journal of Public Health*, **50**, 1324-1333. <https://doi.org/10.18502/ijph.v50i7.6620>
- [20] Gurjav, U., Burneebaatar, B., Narmandakh, E., *et al.* (2015) Spatiotemporal Evidence for Cross-Border Spread of MDR-TB along the Trans-Siberian Railway Line. *The International Journal of Tuberculosis and Lung Disease*, **19**, 1376-1382. <https://doi.org/10.5588/ijtld.15.0361>
- [21] World Health Organization (2020) Coronavirus Disease (COVID-19). https://www.who.int/health-topics/coronavirus#tab=tab_1
- [22] Ontario Ministry of Health (2021) Public Health Units. https://geohub.lio.gov.on.ca/datasets/c2fa5249b0c2404ea8132c051d934224_0/about
- [23] Ontario Ministry of Health (2020) Status of COVID-19 Cases in Ontario by Public Health Unit (PHU).

- <https://data.ontario.ca/en/dataset/status-of-covid-19-cases-in-ontario-by-public-health-unit-phu/resource/d1bfe1ad-6575-4352-8302-09ca81f7ddfc>
- [24] Ontario Ministry of Health (2021) Updated Eligibility for PCR Testing and Case and Contact Management Guidance in Ontario. <https://news.ontario.ca/en/backgrounder/1001387/updated-eligibility-for-pcr-testing-and-case-and-contact-management-guidance-in-ontario>
- [25] Statistics Canada (2016) Census Profile, 2016 Census. <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E>
- [26] Southwestern PHU (2018) Southwestern Public Health Unit. <https://www.swpublichealth.ca/en/about-us/about-us.aspx>
- [27] Association of Local Public Health Agencies Ontario (2020) Milestones and History. <https://www.alphaweb.org/page/milestones?&hhsearchterms=%22elgin-st+and+thomas+and+health+and+unit%22>
- [28] Ontario Ministry of Health (2022) COVID-19 Vaccine Data in Ontario. <https://data.ontario.ca/en/dataset/covid-19-vaccine-data-in-ontario>
- [29] Statistics Canada (2021) Age-Standardized Rates. <https://www.statcan.gc.ca/en/dai/btd/asr>
- [30] ESRI (2021) ArcGIS Pro Software. Software Details. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>
- [31] Anselin, L. and Getis, A. (1992) Spatial Statistical Analysis and Geographic Information Systems. *The Annals of Regional Science*, **26**, 19-33. <https://doi.org/10.1007/BF01581478>
- [32] Goodchild, M. (1986) Spatial Autocorrelation. Geo Books, Norwich. <https://alexsingleton.files.wordpress.com/2014/09/47-spatial-aurocorrelation.pdf>
- [33] Fischer, M.M. and Getis, A. (2010) Handbook of Applied Spatial Analysis. Springer, Berlin. <https://doi.org/10.1007/978-3-642-03647-7>
- [34] ESRI (2021) Modeling Spatial Relationships. <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/modeling-spatial-relationships.htm>
- [35] Moran, P.A.P. (1948) The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, **10**, 243-251. <https://doi.org/10.1111/j.2517-6161.1948.tb00012.x>
- [36] Anselin, L. (1995) Local Indicators of Spatial Association—LISA. *Geographical Analysis*, **27**, 93-115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- [37] ESRI (2020) How Cluster and Outlier Analysis (Anselin Local Moran's I) Works. <https://pro.arcgis.com/en/pro-app/2.9/tool-reference/spatial-statistics/h-how-cluster-and-outlier-analysis-anselin-local-m.htm>
- [38] Getis, A. and Ord, J.K. (1992) The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, **24**, 189-206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- [39] Ord, J.K. and Getis, A. (1995) Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, **27**, 286-306. <https://doi.org/10.1111/j.1538-4632.1995.tb00912.x>
- [40] ESRI (2021) Space-Time Cluster Analysis. <https://pro.arcgis.com/en/pro-app/2.9/tool-reference/spatial-statistics/space-time-analysis.htm>

- [41] ESRI (2020) How Create Space Time Cube Works.
<https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/learnmorecreatecube.htm>
- [42] ESRI (2021) Create Space Time Cube By Aggregating Points (Space Time Pattern Mining).
<https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/create-space-time-cube.htm>
- [43] ESRI (2021) How Emerging Hot Spot Analysis Works.
<https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/learnmoreemerging.htm>
- [44] Public Health Ontario (2022) COVID-19 Data and Surveillance.
<https://www.publichealthontario.ca/en/data-and-analysis/infectious-disease/covid-19-data-surveillance>
- [45] Wooldridge, J.M. (2012) Introductory Econometrics—A Modern Approach. 5th Edition, Cengage Learning, Boston.
- [46] ESRI (2021) Exploratory Regression (Spatial Statistics).
<https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/exploratory-regression.htm>
- [47] ESRI (2022) What They Don't Tell You about Regression Analysis.
https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/what-they-don-t-tell-you-about-regression-analysis.htm#ESRI_SECTION1_B35ECA427435493793C7A7598577A08C
- [48] Pringle, D.G. (1996) Mapping Disease Risk Estimates Based on Small Numbers: An Assessment of Empirical Bayes Techniques. *The Economic and Social Review*, **27**, 341-363.
- [49] Tobler, W.R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, **46**, 234-240.
- [50] Cromley, E.K., McLafferty, S.L. and MacLafferty, S.L. (2012) GIS and Public Health. 2nd Edition, The Guilford Press, New York.
<https://www.routledge.com/GIS-and-Public-Health/Cromley-McLafferty-Rushton-Matthews-Hanchette/p/book/9781609187507#>
- [51] Canadian Institute for Health Information (2022) Canadian COVID-19 Intervention Timeline.
<https://www.cihi.ca/en/canadian-covid-19-intervention-timeline>
- [52] World Health Organization (2020) COVID-19: Vulnerable and High Risk Groups.
<https://www.who.int/westernpacific/emergencies/covid-19/information/high-risk-groups>
- [53] Jordan, S., Starker, A., Krug, S., *et al.* (2020) Health Behaviour and COVID-19: Initial Findings on the Pandemic. *Journal of Health Monitoring*, **5**, 2-14.