COVID-19 Disease Mapping Based on Poisson Kriging Model and Bayesian Spatial Statistical Model

Jingrui Mu

Thesis submitted in partial fulfillment of the requirements for the Master in Philosophy degree in Mathematics and Statistics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The M.Sc. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

Since the start of the COVID-19 pandemic in December 2019, much research has been done to develop the spatial-temporal methods to track it and to predict the spread of the virus. In this thesis, a COVID-19 dataset containing the number of bi-weekly infected cases registered in Ontario since the start of the pandemic to the end of June 2021 is analysed using Bayesian Spatial-temporal models and Area-to-area (Area-to-point) Poisson Kriging models. With the Bayesian models, spatial-temporal effects on infected risk will be checked and ATP Poisson Kriging models will show how the virus spreads over the space and the spatial clustering feature. According to these models, a Shinyapp website `https://mujingrui.shinyapps.io/covid19` is developed to present the results.

Keywords COVID-19, Bayesian Spatial-temporal Models, Area-to-area Poisson Kriging, Integrated nested Laplace approximation.

Acknowledgment

First of all I would like to thank my supervisor Professor Mayer Alvo. He introduced me to COVID-19 data analysis with spatial statistical methods. He provided me with an interesting research topic where I could contribute due to its novelty. He also helped me with providing much insight into the topic, while still keeping an open mind when I found something interesting.

Special thanks would go to my beloved family for their loving considerations and great confidence in me. I also owe my sincere gratitude to my friends and my fellow classmates who gave me the help and time in listening to me during the difficult course of the thesis.

Contents

List of Figures

List of Tables

List of Acronyms

MLE: Maximum Likelihood Estimator

GMRF: Gaussian Markov Random Field

ATP: Area-to-Point

ATA: Area-to-Area

INLA: Integrated Nested Laplace Approximation

BYM: Besag-York-Mollie

MAEP: Mean Absolute Error of Prediction

PK: Poisson Kriging

SLA: Simplified Laplace Approximation

MAEE: Mean Absolute Error of Estimation

List of Nomenclature

$\boldsymbol{X} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12})$

$\boldsymbol{\theta}_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_K)$

$\boldsymbol{Y} = (y_1, y_2, y_3, y_4, ..., y_N)$

$\boldsymbol{\psi} = (\tau_u, \tau_v, \sigma_\gamma, \sigma_\phi)$

$k \sim i$: Regions k and i are neighbors.

Chapter 1

Introduction

The pandemic caused by the coronavirus disease 2019 (COVID-2019) has led to an unprecedented number of related scientific outcomes. Many of the studies on COVID-19 focus on the evolution of viral transmission, or the clinical factors that increase the risk of contagion, among other relevant topics. But few studies make use of geostatistical methods to analyze it. Kriging methods are often used to develope isopleth maps. However, simple kriging methods to create isopleth maps make use of data measured at each point in space, which is based on areal aggregation. In the public health area, we often get only areal aggregated data. When performing point kriging of areal data, the user makes the practical assumption that all the inhabitants of the administrative unit live at the same location and the measured rate thus refers to this specific location. This assumption is reasonable whenever the units of aggregation are small with respect to the spacing of the interpolation grid. However, the size of public health units in Ontario are not relatively small. Therefore, the assumption of point measurement support becomes clearly inappropriate. There is a need to develop specific methods to incorporate the shape and size of those units in the analysis. Area-to-Point (ATP) Poisson Kriging and Area-to-Area (ATA) Poisson Kriging incorporate the size and shape of administrative units, as well as the population density into the mapping of the corresponding risk at a fine scale (Goovaerts P,2005; Goovaerts P,2006). These kriging methods can be used to see how the virus spreads over the space and downscale the areal infected risks into point ones, which can show the spatial clustering feature of COVID-19 spreading well.

It is noteworthy that Bayesian statistical techniques have been developed to estimate areal risk by borrowing information from neighboring entities (Best N et al., 2005; Pickle LW et al., 2000; Chrietensen OF et al., 2002; Besag J et al., 1991; Waller LA et al., 2004). However, the main challenge in Bayesian statistics still resides in the computational aspects. Models have been implemented in software such as WinBUGS (Spiegelhalter DJ et al., 2002). Markov Chain Monte Carlo (MCMC) methods

(Brooks et al.,2011; Robert and Casella, 2004) are normally used for Bayesian computation. Yet, the estimation of model parameters are computer intensive and require fine-tuning, which makes the iteration process time-consuming. Recently, the Integrated Nested Laplace Approximation (INLA; Rue et al., 2009) approach has been developed as a computationally efficient alternative to MCMC. INLA is designed for latent Gaussian models, a very widely used and flexible class of models ranging from generalized linear mixed to spatial and spatial-temporal models. For this reason, INLA has been successfully used in a great variety of applications (eg. Li et al., 2012; Riebler et al., 2012; Ruiz-Cardenas et al.,2012; Martino et al., 2011; Roos and Held, 2011). With Bayesian spatial-temporal models, it can be seen that the spatial and temporal effects on the infected risk can be incorporated.

The objective of this thesis is to exploit the use of Bayesian Spatial-temporal Models, ATA and ATP Poisson Kriging to track the spread of the virus over Ontario and to compare these two different methods in prediction performance. Moreover, we assess the effectiveness of policy factors into Bayesian Spatial-temporal models with different intervention variables in order to determine how these policies alter the spread of the virus. The thesis is structured as follows: Chapter 2 introduces the datasets used. As well, COVID-19 relevant indicators and intervention factors are defined. In Chapter 3, Bayesian Spatial-temporal models with INLA are presented. The spatial, temporal effects along with covariates, intervention factors are defined and the use of Bayesian Spatial-temporal models is to evaluate policy decisions. In Chapter 4, we discuss the prediction performance of the models and compare them in short-term prediction and longer-term prediction. In Chapter 5, we review ATA and ATP Poisson Kriging and provide the clear algorithm to implement them in use of professional QGIS software and Python under Linux system. A presentation of smoothed temporal point risk maps is followed after downscaling areal risk maps into 15km*15km cell risk ones. In Chapter 6, conclusions are made based on the previous discussions.

The contribution of the thesis are as follows: First, we develop a model to take into account spatial and temporal effects on the spread of the virus in Ontario. Second, this model is used to assess policy decisions whose objective includes limiting the number of people attending gatherings. The model can also be used to test for the significance of auxiliary variables. Third, we implemented Poisson Kriging in order to provide another approach to create spatial maps which takes into account the sizes of the units used in aggregation of the data. The point support semivariogram from ATP Poisson Kriging can show the spreading spatial feature well. This initial setting re-programmed is different from the python package pyinterpolate to provide a better result. Fourth, based on these methods, an interactive website `https://mujingrui.shinyapps.io/covid19` was developed to show these tracking maps with the use of Shiny package in R.

Chapter 2

Data Preparation

2.1 COVID-19 Datasets

Statistics Canada provides the official datasets for each province, which include the number of daily confirmed cases, mortality cases, recovered cases and testing cases. These also include aggregated data in different official health regions. **Note**: The date is recorded as case reported date.

Table 2.1: Provincial Level Dataset

—	Daily	Cumulative	Health Region
Confirmed Cases	☑	☑	☑
Mortality	☑	☑	☑
Recovered Cases	☑	☑	☑
Testing Cases	☑	☑	☑

Table 2.2: Confirmed Positive Cases in Ontario

Variable Name	Definition
ID	Identifier for each individual row/record within the dataset.
Accurate Episode Date	The field uses a number of dates entered in the Integrated Public Health Information System (iPHIS) to provide an approximation of onset date.
Case Reported Date	The date that the case was reported to the local public health unit (PHU).
Test Reported Date	The test reported date as indicated on the laboratory slip.
Specimen Date	Set to the earliest specimen date on record for the case, as indicated on the laboratory slip.
Age Group	Age group of the patient.
Client Gender	Gender information of the patient.
Outcome	Patient outcome.
Reporting PHU ID	Public Health Unit (PHU) ID where confirmed positive case occurred.
Reporting PHU	Public Health Unit (PHU) where confirmed positive case occurred.
Reporting PHU Address	Official physical street address of Public Health Unit (PHU).
Reporting PHU Latitude	Latitude of Public Health Unit (PHU) physical address for mapping purposes.
Reporting PHU Longitude	Longitude of Public Health Unit (PHU) physical address for mapping purposes.

The above dataset contains information licensed under the Open Government License Ontario. Age group information of each case can be used to calculate age-adjusted infection rate or age-adjusted mortality rate. More details will be discussed in Section 2.3.

2.2 Population Datasets

Table 2.3: Population Size of Health Regions

Geography	Age Group	Population Size
☑	☑	☑

Statistics Canada, Demography Division, provides population estimates of 2019 in different health regions by age and sex. Individuals living in a geographic area are divided into nine age groups, which are < 20, 20s, 30s, 40s, 50s, 60s, 70s, 80s, and 90+.

Statistic Canada published 2016 Census data, which is accessible in a number of different aggregation levels including:

Table 2.4: 2016 Census Data Aggregation Levels

Code	Description	Count in Census 2016
C	Canada(total)	1
PR	Provinces(Territories)	13
CMA	Census Metropolitan Area	35
CA	Census Agglomeration	14
CD	Census Division	287
CSD	Census Subdivision	713
CT	Census Tracts	5621
DA	Dissemination Area	56589
DB	Dissemination Block	489676

The population of Canada by age group in the 2016 Census can be chosen as the standard population. We use the 2016 Canadian census data because the census data is not available in 2019. The age-adjusted rates are computed through direct standardization or indirect standardization method based on the 2019 population estimates. Otherwise, the relative proportion of the public health region level population at risk within each cell (more details about how to calculate these proportions will be seen in Section 5.5) can be retrieved from the readily available 2016 census dissemination block level data. (The population at risk is back-calculated from the age-adjusted rate and counting cases). The population at risk in Ontario is shown as follows (each cell is in 15×15 = 225 km^2):

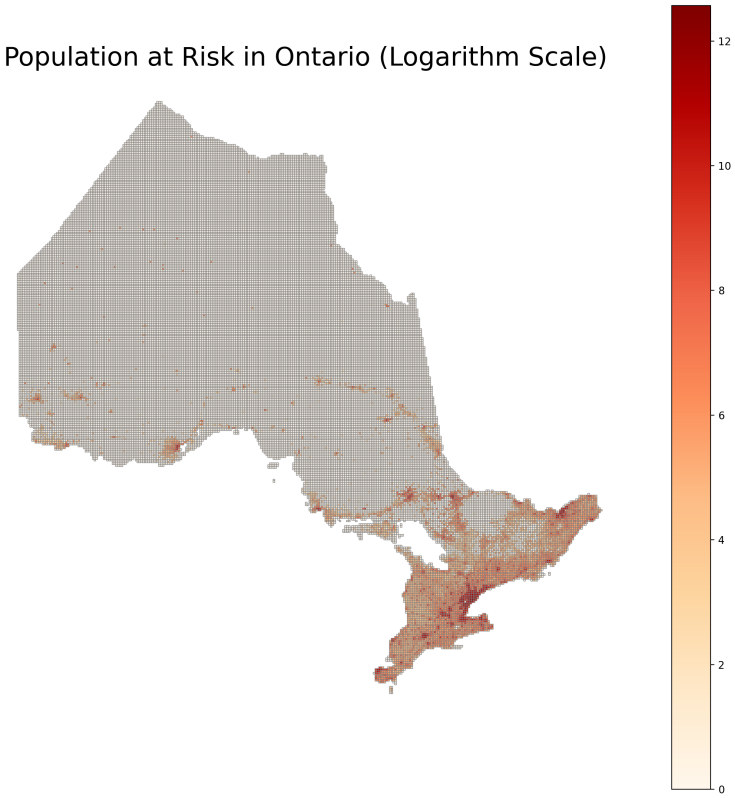


Figure 2.1: Population at Risk in Ontario

2.3 Age-Adjusted Rate

The data that will be used in ATA(ATP) Poisson Kriging Models is standardized in order to account for differences in age composition of the population. Without the age-adjusted rate, it would be difficult to derive a single conclusion of comparative rate, when the relative rate of different age groups differ.

We make use of a method of adjusting called "direct standardization." The process of "direct standardization" consists of applying different age-specific rates to a standard population structure. Three major components are needed to perform adjusted infection rate calculations: the number of cases, the 2019 population, and a "standard"

population. The standard million age distribution of the total population of Ontario according to the 2016 Canadian Census is chosen as the "standard" one. Following is a step-by-step calculation of age-adjusted infection rates:

1) List infected cases and the 2019 population by the age categories for each public health unit

2) Calculate the age-specific rates by dividing the infected cases by the 2019 population in each age group

3) Multiply each age-specific rate by the number in each of the corresponding age groups of the "standard" population. The results are the number of cases to be "expected" in each age category if the specific infected rates had prevailed for one year within the "standard" population.

4) Sum the expected cases and standard population for all age categories and simply divide the total expected cases by the total standard population. Then multiply by 100,000 to get the age-adjusted infection rate per 100,000 people.

Note: These same steps using the same standard million population distribution must be followed when calculating the age-adjusted rates for any other units to achieve comparability.

2.4 Biweekly Infected Rates

The infected rates used in the thesis are biweekly infected ones. It can be calculated as follows:

$$\textbf{Biweekly Infected Rates}_{it} = \frac{\text{Infected Cases}_{it^*} - \text{Infected Cases}_{it^-}}{\text{Population Size}_i}$$

Where i refers to the different public health unit in Ontario, t^* and t^- represent the end day of t biweek since 2020 March and the start day of t biweek, respectively. In Bayesian Spatial-temporal Models, raw biweekly infected rates are used. For ATA(ATP) Poisson Kriging Models, age-adjusted biweekly infected rates are in use. (Infected Cases_{it^*} – Infected Cases_{it}) are used to replace the infected cases in Section 2.3 when computing age-adjusted rates.

2.5 COVID-19 Relevant Indicators

Statistics Canada released COVID-19 relevant indicators from different characteristics based on the 2016 Canadian census:

1) **Population and dwellings characteristics**: population density per square kilometer;

2) **Age characteristics**: distribution of the population by broad age groups (65 years and over)

3) **Household and dwelling characteristics**:
Total - Occupied private dwellings by structural type of dwelling;
Apartment in a building that has five or more storeys;
Other Attached dwelling;
Apartment in a building that has fewer than five storeys in other attached dwelling;
Average household size.

4) **Low income in 2015**:
Prevalence of low income based on the Low-income measure, after tax (LIM-AT) (%);
0 to 17 years (%);
18 to 64 years (%);
65 years and over (%)

5) **Class of worker**:
Proportion of self-employed workers in all classes of workers aged 15 years and over

6) **Occupation - National Occupational Classification (NOC) 2016**:
Proportion of health occupations in all occupations aged 15 years and over

Note: the data in items 5 and 6 are from a sample of approximately 25% Canadian workers according to the 2016 Canadian census. The data "0 to 17 years (%)", "18 to 64 years (%)", and "65 years and over (%)" in item 4 mean the ratio of the number of people (in 0-17 years age group, 18-64 years age group, 65 years and over age group respectively) whose income falls below the low income line.

2.6 COVID-19 Intervention Timeline

Since March 2020, the government published different policies to stop the spread of the virus. Among all policies implemented, there are 3 categories labeled as follows: Indoor Gathering, Outdoor Gathering and Non-essential Services, for a more in-depth analysis. As various restrictions and rules were imposed at different levels of

enforcement across the different public health units or provinces, it is difficult to compare them directly. The levels within these three categories according to the common features of the restrictions are:

1) **Indoor Gathering**
   level 0: No upper limit on the number of people allowed;
   level 1: The upper limit was less than 50;
   level 2: The upper limit was less than 10;
   level 3: The upper limit was less than 5.

2) **Outdoor Gathering**
   level 0: No upper limit on the number of people allowed;
   level 1: The upper limit was less than 100;
   level 2: The upper limit was less than 25;
   level 3: The upper limit was less than 10;
   level 4: The upper limit was less than 5.

3) **Non-essential Services**
   level 0: No restrictions were imposed;
   level 1: Sit-down dining (with varying upper limit and some limitations, like wearing facial mask) at cafes, restaurants was allowed.
   level 2: Access to some non-essential and leisure services allowed, such as outdoor and non-contact activities.
   level 3: Restaurants/bars/cafes (except takeout/delivery) closed, retail services restricted and personal services closed.
   level 4: Non-essential businesses closed.

Table 2.5: Indoor Gathering Intervention Timeline Under Level in Ontario

Date	Levels
2020-03-28	level 3
2020-06-12	level 2
2020-07-17	level 1 (24 public health units that entered stage 3)
2020-07-24	level 1 (7 public health units that entered stage 3)
2020-07-31	level 1 (Toronto, Peel public health units)
2020-09-18	level 2 (Toronto, Peel and Ottawa public health units)
2020-09-19	level 2
2020-11-07	level 3 (Peel public health unit)
2020-11-14	level 3 (Toronto public health unit)
2020-11-16	level 3 (Hamilton, Halton, York public health units)
2020-11-23	level 3 (Durham, Waterloo public health units)
2020-11-23	level 2 (5 regions entered orange level and 6 regions entered yellow level)
2020-11-28	level 3 (Windsor Essex County health unit)
2020-11-28	level 2 (Haldimand Norfolk, and 3 regions entered yellow level)
2020-12-07	level 2 (Middlesex London, Thunder Bay health units and 3 regions entered yellow level)
2020-12-14	level 3 (Windsor-Essex, York Region health units and 3 regions entered red level)
2020-12-14	level 2 (Eastern, Leeds, Grenville and Lanark District health units)
2020-12-21	level 3 (Hamilton, Brant County, Niagara Region health units)
2020-12-21	level 2 (Kingston, Frontenac and Lennox and Addington health unit, Timiskaming and Sudbury Districts health units)
2020-12-26	level 3
2021-02-10	level 2 (Kingston, Frontenac and Lennox and Addington health unit, Renfrew County and District health unit, Hastings and Prince Edward Counties health unit)
2021-02-16	level 2 (Leeds, Grenville and Lanark District health unit, Timiskaming health unit, 9 regions entered orange level, 4 regions entered yellow level)
2021-02-22	level 3 (Lambton health unit)
2021-03-01	level 2 (Grey Bruce health unit, 2 regions entered yellow level, 3 regions entered orange level)
2021-03-08	level 3 (Peterborough, Sudbury health units)
2021-03-08	level 2 (Renfrew County health unit)
2021-03-19	level 3 (Ottawa health unit)
2021-03-22	level 2 (North Bay Parry Sound District, Wellington Dufferin Guelph health units)
2021-03-22	level 3 (Brant, Chatham-Kent, Leeds, Grenville and Lanark District health units)
2021-03-26	level 3 (Timiskaming health unit)
2021-03-30	level 3 (Middlesex London health unit)
2021-04-03	level 3

Table 2.6: Outdoor Gathering Intervention Timeline Under Level in Ontario

Date	Levels
2020-03-28	level 4
2020-06-12	level 3
2020-07-17	level 1 (24 public health units that entered stage 3)
2020-07-24	level 1 (7 public health units that entered stage 3)
2020-07-31	level 1 (Toronto, Peel public health units)
2020-09-18	level 2 (Toronto, Peel and Ottawa public health units)
2020-09-19	level 2
2020-11-23	level 3 (Toronto, Peel health units)
2020-12-14	level 3 (Windsor Essex, York Region health units)
2021-01-13	level 4
2021-02-10	level 2 (Kingston, Frontenac and Lennox and Addington health unit, Renfrew County and District health unit, Hastings and Prince Edward Counties health unit)
2021-02-16	level 2 (Leeds, Grenville and Lanark District health unit, Timiskaming health unit, 9 regions entered orange level, 4 regions entered yellow level, 11 regions entered red level)
2021-02-16	level 3 (Niagara Region health unit)
2021-02-22	level 2 (York Region health unit)
2021-03-01	level 3 (Simcoe Muskoka District, Thunder Bay District health units)
2021-03-01	level 2 (Niagara Region health unit)
2021-03-08	level 3 (Toronto, Peel health units)
2021-03-08	level 2 (North Bay Parry Sound District, Sudbury and District health units)
2021-03-12	level 3 (Sudbury and District health unit)
2021-04-03	level 4
2021-05-20	level 3

Table 2.7: Non-essential Services Intervention Timeline Under Level in Ontario

Date	Levels
2020-03-25	level 4
2020-05-19	level 3
2020-06-12	level 2 (24 public health unit regions entering stage 3)
2020-06-19	level 2 (7 public health unit regions entering stage 3)
2020-06-24	level 2 (Toronto, Peel public health units)
2020-07-17	level 1 (24 public health units that have entered stage 3)
2020-07-24	level 1 (7 public health units that have entered stage 3)
2020-07-31	level 1 (Toronto, Peel public health units)
2020-10-10	level 2 (Ottawa, Toronto, Peel public health units)
2020-10-19	level 2 (York region public health unit)
2020-11-23	level 3 (Peel, Toronto public health units)
2020-12-14	level 3 (Windsor Essex County, York Region public health units)
2020-12-21	level 3 (Hamilton health unit)
2020-12-26	level 3
2021-02-10	level 1 (3 regions that entered green-prevent level)
2021-02-16	level 1 (Leeds, Grenville and Lanark District health unit, Timiskaming health unit)
2021-02-16	level 1 (11 regions entered red-control level, 9 regions entered orange level, 4 regions entered yellow level)
2021-02-22	level 1 (York Region, Lambton health units)
2021-03-01	level 1 (Niagara Region public health unit)
2021-03-08	level 1 (North Bay Parry Sound District health unit)
2021-03-19	level 1 (Ottawa public health unit)
2021-04-03	level 2

Chapter 3

Bayesian Spatial Statistic Models

In this chapter, we define the Bayesian spatial model, and the Bayesian spatial-temporal Model. We make use of the biweekly infected raw rates data to estimate the posterior distribution of different coefficients of covariates and effects (including spatial, temporal, and spatial-temporal effect). Spatial data can be defined as realizations of a stochastic process indexed by

$$Y(s) \equiv \{y(s), s \in D\}$$

where D is a fixed subset of $\mathcal{R}^2$. The actual data can be formally represented by a collection of observations $\boldsymbol{y} = \{y(s_1), ..., y(s_n)\}$, where the set $(s_1, ..., s_n)$ represents the spatial units at which measurements are taken (here we consider these units are public health units in Ontario). Here, $\boldsymbol{y}$ represents the number of infected cases observed in each public health unit.

A three-stage hierarchical process in Bayesian spatial statistic models has been widely used (Best et al., 2005). The first stage consists of the likelihood model where we assume. $\boldsymbol{Y} = (y_1, y_2, ..., y_N)$ and $y_i \sim \text{Poisson}(\mu_i)$ for the second stage we place a prior distribution on $log(\mu_i)$. The logarithm of μ_i is often expressed as a regression equation and typically includes an overall fixed effect (intercept, denoted α), covariate effects and spatial random, temporal random, spatial-temporal interaction effects.The third stage consists of the prior distributions on each of the unknown parameters, which are usually specified as weakly informative with Gaussian distributions having zero mean and some large variance. If the parameters are unknown, then the hyperpriors represent a fourth stage of the hierarchy.

3.1 Bayesian Spatial Model

A pixel-based image analysis from a Bayesian perspective was proposed by (Besag et al.,1991). Let θ_i denote the unknown log relative risk in zone i (i = 1,2,...,n) and y_i

the corresponding observed number of cases of the disease during the study period. Here, the y_i's are assumed to have a Poisson distribution with means $e_i e^{\theta_i}$, where e_i is the expected number of cases in zone i. The formulation is given by

$$\theta_i = t + u_i + v_i$$

Here t is a term associated with measured covariates that are known or suspected to be relevant to the disease, and is usually in the guise of a linear model $t = \mathbf{X}\beta$. The additional term, u_i's, can represent variables that display substantial spatial structure in that the values for a pair of contiguous zones would be generally much more alike than for two arbitrary zones, whereas the v_i's represent unstructured variables.

In the case of areal level data, it is reasonable to reformulate the problem in terms of the neighborhood structure. Under the Markovian property, the generic element θ_i of the parameters vector $\boldsymbol{\theta}_i$ is independent of any other element and is denoted given the set of its neighbors $\mathcal{N}(i)$

$$\theta_i \perp\!\!\!\perp \boldsymbol{\theta}_{-i} | \boldsymbol{\theta}_{\mathcal{N}(i)}$$

where $\boldsymbol{\theta}_{-i}$ indicates all the elements in $\boldsymbol{\theta}$ but the i-th. The specification is known as a Gaussian Markov Random Field (GMRF) (Rue and Held, 2005).The general model is formulated as follows:

$$\begin{aligned} y_i &\sim Poisson(\mu_i) \\ log(\mu_i) &= log(e_i) + \theta_i \\ \theta_i &= \alpha + \beta X + u_i + v_i \end{aligned}$$

where y_i is the biweekly infected number of COVID-19 cases in $i = 1, ..., I$ areas; e_i and θ_i are, respectively, the population size in area i and the log infected risk of COVID-19; Here, $\beta = (\beta_1, \beta_2, ..., \beta_p)$ represents the vector of covariate coefficients; $X = (x_1, x_2, ..., x_p)$ is the COVID-19 relevant covariate data vector to be discussed in Section 3.3; u_i is a spatially structured random effect under GMRF and v_i is a spatially unstructured random effect with mean zero and variance τ_v^2. We assume as well:

$$\begin{aligned} (u_i | u_k, k \neq i, \tau_u^2) &\sim N\left(\frac{\sum\limits_{k\sim i} u_i \omega_{ki}}{\sum\limits_{k\sim i} \omega_{ki}}, \frac{\tau_u^2}{\sum\limits_{k\sim i} \omega_{ki}} \right) \\ v_i &\sim N(0, \tau_v^2) \end{aligned}$$

Here u follows a Markov Random Field Model on the irregular lattice of regions, with $k \sim i$ referring to neighbor regions i and j sharing a common boundary line. The sum $\sum\limits_{k\sim i} \omega_{ki}$ can never be zero since$\omega_{ki} = 1$ if k,i are adjacent, $\omega_{ki} = 0$ otherwise.

3.2 Integrated Nested Laplace Approximation (INLA)

The model in Section 3.1.1 can be fitted using a general Bayesian Hierarchical Modeling framework defined as follows:

$$y \sim \prod_{i=1}^{N} f(y_i|\mu_i)$$
$$g(\mu_i) = \eta_i = \alpha + \sum_{m=1}^{M} \beta_m x_{mi} + \sum_{l=1}^{L} f_l(z_{li})$$

where α is a scalar representing the intercept; the coefficients $\boldsymbol{\beta} = (\beta_1, ..., \beta_M)$ quantify the effect of some covariates $\boldsymbol{x} = (x_1, ..., x_M)$ on the response; and $\boldsymbol{f} = \{f_1(.), ..., f_L(.)\}$ are a set of functions defined in terms of some covariates $z = (z_1, ..., z_L)$, such as random (iid) effects, spatially or temporally correlated effects; $\boldsymbol{y} = (y_1, y_2, ..., y_N)$ represents the vector of COVID-19 cases and N is the number of public health units in Ontario. For the Bayesian Spatial model in Section 3.1.1, we identify $f_1(.) \sim N\left(\frac{\sum_i u_i \omega_{ki}}{\sum_i \omega_{ki}}, \frac{\tau_u^2}{\sum_i \omega_{ki}}\right)$ and $f_2(.) \sim N(0, \tau_v^2)$. Upon varying the form of the functions $f_l(.)$, this formulation can accommodate a wide range of models, from standard and hierarchical regression, to spatial and spatial-temporal models (Rue et al., 2009).

Given the specification, we may represent the vector of parameters by $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{f})$. In line with the discussion in Section 3.1.1, we can assume a GMRF prior on $\boldsymbol{\theta}$, with mean 0 and a sparse precision matrix $\boldsymbol{Q}$. In other words, for any pair of elements (i,j)

$$\theta_i \perp\!\!\!\perp \boldsymbol{\theta}_{-ij} \iff \boldsymbol{Q}_{ij} = 0$$

Thus, $\boldsymbol{Q}_{ij} \neq 0$ only if $j \in \{i, \mathcal{N}(i)\}$. The specification is know as a *Gaussian Markov Random Field* (GMRF) (Rue and Held,2005). (Note: $\boldsymbol{Q}$ is defined as the inverse of variance covariance of matrix)

The objectives of the Bayesian computation consist of calculating the marginal posterior distributions for each of the elements of the parameters vector and possibly for each element of the hyper-parameters vector $\boldsymbol{\psi} = (\tau_u, \tau_v)$:

$$p(\theta_i|\boldsymbol{y}) = \int p(\boldsymbol{\psi}|\boldsymbol{y}) p(\theta_i|\boldsymbol{\psi}, \boldsymbol{y}) d\boldsymbol{\psi}$$
$$p(\psi_k|\boldsymbol{y}) = \int p(\boldsymbol{\psi}|\boldsymbol{y}) d\boldsymbol{\psi_{-k}}$$

Thus, we need to approximate $p(\boldsymbol{\psi}|\boldsymbol{y})$, from which also all the relevant marginals $p(\psi_k|\boldsymbol{y})$ can be obtained; and $p(\theta_i|\boldsymbol{\psi}, \boldsymbol{y})$, which is needed to compute the marginal posterior for the parameters. The INLA approach exploits the assumptions of the

model to produce a numerical approximation to the posteriors of interest, based on the *Laplace approximation* (Tierney and Kadane, 1986).

The first item we need compute is an approximation to the posterior marginal distribution of the hyper-parameters as

$$\begin{aligned} p(\boldsymbol{\psi}|\boldsymbol{y}) &= \frac{p(\boldsymbol{\psi},\boldsymbol{y})}{p(\boldsymbol{y})} \\ &= \frac{p(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}{p(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})} \\ &\propto \frac{p(\boldsymbol{\psi},\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\psi},\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})} \\ &= \frac{p(\boldsymbol{\psi},\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})} \\ &= \frac{p(\boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})} \\ &\approx \left.\frac{p(\boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{y}|\boldsymbol{\theta})}{\widetilde{p}(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*(\boldsymbol{\psi})} =: \widetilde{p}(\boldsymbol{\psi}|\boldsymbol{y}) \end{aligned}$$

Here we expand the numerator with the simplification

$$p(\boldsymbol{\psi},\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\psi},\boldsymbol{\theta}) = p(\boldsymbol{\psi},\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})$$

holding because the data distribution, $p(\boldsymbol{y}|\boldsymbol{\psi},\boldsymbol{\theta})$, depends only on $\boldsymbol{\theta}$; the hyperparameters $\boldsymbol{\psi}$ affect $\boldsymbol{y}$ only through $\boldsymbol{\theta}$. The Laplace approximation is used in the denominator, where $\widetilde{p}(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})$ is the Gaussian approximation of $p(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})$ and $\boldsymbol{\theta}^*(\boldsymbol{\psi})$ is its mode.

Next, we need to approximate $p(\theta_i|\boldsymbol{\psi},\boldsymbol{y})$, and it is possible to re-write the vector of parameters as $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}_{-i})$ and use the Laplace approximation again to obtain:

$$\begin{aligned} p(\theta_i|\boldsymbol{\psi},\boldsymbol{y}) &= \frac{p((\theta_i,\boldsymbol{\theta}_{-i})|\boldsymbol{\psi},\boldsymbol{y})}{p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})} \\ &\propto \frac{p(\boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})} \\ &\approx \left.\frac{p(\boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{y}|\boldsymbol{\theta})}{\widetilde{p}(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})}\right|_{\boldsymbol{\theta}_{-i}=\boldsymbol{\theta}^*_{-i}(\theta_i,\boldsymbol{\psi})} =: \widetilde{p}(\theta_i|\boldsymbol{\psi},\boldsymbol{y}) \end{aligned}$$

$\widetilde{p}(\theta_i|\boldsymbol{\psi},\boldsymbol{y})$ represents the Gaussian approximation to $p(\theta_i|\boldsymbol{\psi},\boldsymbol{y})$ and $\boldsymbol{\theta}_{-i} = \boldsymbol{\theta}^*_{-i}(\theta_i,\boldsymbol{\psi})$ is its mode. The approximation typically works very well, but it can be very expensive

in computational terms. Consequently, the Simplified Laplace Approximation has been proposed (Rue et al., 2009)

$$p_{SLA}(\theta_i|\boldsymbol{\psi},\boldsymbol{y}) \propto N(\theta_i|\mu_i(\boldsymbol{\psi}),\sigma_i^2(\boldsymbol{\psi}))(\exp(\text{spline}(\theta_i))$$

$$\text{spline}(\theta_i) = \text{constant} - \frac{1}{2}(\theta_i)^2 + \frac{1}{6}(\theta_i)^3$$

The Simplified Laplace Approximation is also the default method in the INLA package. Each marginal posterior $\widetilde{p}(\psi_k|\boldsymbol{y})$ can be obtained using an interpolation based on the computed values corrected for skewness. For each ψ_k, the conditional posteriors $\widetilde{p}(\theta_i|\psi_k,\boldsymbol{y})$ are then evaluated on a grid of selected values for θ_i and the marginal posteriors $\widetilde{p}(\theta_i|\boldsymbol{y})$ are obtained by numerical integration (Rue et al.,2009; Martins et al.,2012; Blangiardo and Cameletti, 2013)

$$\widetilde{p}(\theta_i|\boldsymbol{y}) \approx \sum_{k=1}^{K} \widetilde{p}(\theta_i|\psi_k,\boldsymbol{y})\widetilde{p}(\psi_k|\boldsymbol{y})\triangle_k$$

3.3 Spatial-temporal Model

Spatial models have been discussed in Section 3.1. In order to incorporate time, we can build spatial-temporal models that include spatial and temporal random effects, as well as interaction effects between space and time. For temporal effect, there is a dynamic formulation in the linear predictor:

$$log(\mu_{it}) = \eta_{it} = \alpha + \sum_{m=1}^{M} \beta_m x_{mi} + u_i + v_i + \gamma_t + \phi_t$$

Here α, $\boldsymbol{\beta} = (\beta_1,...,\beta_m)$, u_i, and v_i have the same parametrization as in Section 3.1. The term γ_t represents the temporally structured effect, modeled dynamically through a neighboring structure using a first-order random walk

$$\gamma_t|\gamma_{t-1} \sim N(\gamma_{t-1},\sigma_\gamma^2)$$

where σ_γ^2 is the variance component. Finally, an independent Gaussian prior is chosen for ϕ_t: $\phi_t \sim N(0,\sigma_\phi^2)$. In the study, the random temporal effects are set on a bi-weekly basis. Now in this formulation, the parameters and hyper-parameters are $\boldsymbol{\theta} = \{\alpha,\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{v},\boldsymbol{\gamma},\boldsymbol{\phi}\}$ and $\boldsymbol{\psi} = \{\tau_u,\tau_v,\tau_\gamma,\tau_\phi\}$ respectively.

To allow for an interaction between space and time, in order to explain differences in the time trend of infected risk for different areas, we may specify:

$$\eta_{it} = \alpha + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$$

There are four ways to define the interaction term (Knorr-Held,2000). Each of these structures consists of specifying the non-separable spatial-temporal term of the model according to a concrete combination of a structured/unstructured spatial effect with a structured/unstructured temporal effect. The combination of these effects is carried out through the Kronecker product of the two matrices that represent the spatial and temporal effect chosen, respectively. The following table indicates the four types of spatial-temporal interactions considered.

Table 3.1: Specification of the Four Types of Spatial-temporal Interaction

Type of sptaio-temporal interaction	R_δ
I	$I_s \otimes I_t$
II	$I_s \otimes R_t$
III	$R_s \otimes I_t$
IV	$R_s \otimes R_t$

the identity matrices $I_s(I_t)$ correspond to the unstructured spatial (temporal) effect respectively, whereas $R_t(R_s)$ represent non-identity matrices that correspond to a specific structured temporal (spatial) effect as follows:

$$R_t = \begin{bmatrix} 1 & -1 & & & & & & & \\ -1 & 2 & -1 & & & & & & \\ & -1 & 2 & -1 & & & & & \\ & & -1 & 2 & -1 & & & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 & 0 & 0 & \\ & & & & & -1 & 2 & -1 & 0 \\ & & & & & & -1 & 2 & -1 \\ & & & & & & & -1 & 1 \end{bmatrix}$$

$$R_s = \begin{cases} \sum_{k \sim i} \omega_{ki}, & i = k \\ -1, & i \sim k \\ 0, & otherwise \end{cases}$$

In summary, the general spatial-temporal model including interactions is:

$$
\begin{aligned}
&y_{it} \sim Poisson(\mu_{it}) \\
&log(\mu_{it}) = log(e_i) + \theta_{it} \\
&\theta_{it} = \alpha + \beta X + u_i + v_i + \gamma_t + \phi_t + \delta_{it} \\
&(u_k | u_i, k \neq i, \tau_u^2) \sim N\left(\frac{\sum_i u_i \omega_{ki}}{\sum_i \omega_{ki}}, \frac{\tau_u^2}{\sum_i \omega_{ki}}\right) \\
&v_i \sim N(0, \tau_v^2) \\
&\gamma_t | \gamma_{t-1} \sim N(\gamma_{t-1}, \sigma_\gamma^2) \\
&\phi_t \sim N(0, \sigma_\phi^2)
\end{aligned}
$$

As discussed in Section 2.5, there are three different policies selected and categorized as three different variables. The variable Indoor Gathering (IG) will be defined as 3 indicator variables according to the different restrictions level: $IG_i = 1$ if the i-th level gathering restriction was in place, and 0 otherwise, for $i = 1, 2, 3$. The variable Outdoor Gathering (OG) also can be defined as 4 indicator variables according to the different restrictions level: $OG_i = 1$ if the i-th level gathering restriction was in place, and 0 otherwise, for $i = 1, 2, 3, 4$. The variable Non-essential service (E) is based on the restrictions imposed on business, accordingly. There are four indicator variables defined for different limitations level: $E_i = 1$ if the i-th level non-essential service restriction was in place, and 0 otherwise, for $i = 1, 2, 3, 4$.

All of these variables can be included into Bayesian Spatial-temporal Models to determine how they affect the infected risk:

$$
\begin{aligned}
&y_{it} \sim Poisson(\mu_{it}) \\
&log(\mu_{it}) = log(e_i) + \theta_{it} \\
&\theta_{it} = \alpha + \beta X_i + \sum_{j=1}^{3} \beta_{IGj} G_{ijt} + \sum_{j=1}^{4} \beta_{OGj} G_{ijt} + \sum_{j=1}^{4} \beta_{Ej} E_{ijt} + u_i + v_i + \gamma_t + \phi_t + \delta_{it} \\
&(u_k | u_i, k \neq i, \tau_u^2) \sim N\left(\frac{\sum_i u_i \omega_{ki}}{\sum_i \omega_{ki}}, \frac{\tau_u^2}{\sum_i \omega_{ki}}\right) \\
&v_i \sim N(0, \tau_v^2) \\
&\gamma_t | \gamma_{t-1} \sim N(\gamma_{t-1}, \sigma_\gamma^2) \\
&\phi_t \sim N(0, \sigma_\phi^2)
\end{aligned}
$$

3.4 Model Selection

According to the discussion in Section 3.1 and 3.2, there are 11 models proposed as follows:

Table 3.2: Model Specification

Models	Components
Model 1	$\alpha + u_i + v_i$
Model 2	$\alpha + \boldsymbol{\beta} X_i + u_i + v_i$
Model 3	$\alpha + \boldsymbol{\beta} X_i + u_i + v_i + \gamma_t + \phi_t$
Model 4	$\alpha + \boldsymbol{\beta} X_i + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (Type I)
Model 5	$\alpha + \boldsymbol{\beta} X_i + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (Type II)
Model 6	$\alpha + \boldsymbol{\beta} X_i + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (Type III)
Model 7	$\alpha + \boldsymbol{\beta} X_i + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (Type IV)
Model 8	$\alpha + \sum_{j=1}^{3} \beta_{IGj} G_{ijt} + \sum_{j=1}^{4} \beta_{Ej} E_{ijt} + \sum_{j=1}^{4} \beta_{OGj} B_{ijt} + \boldsymbol{\beta} X_i + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (Type I)
Model 9	$\alpha + \sum_{j=1}^{3} \beta_{IGj} G_{ijt} + \sum_{j=1}^{4} \beta_{Ej} E_{ijt} + \sum_{j=1}^{4} \beta_{OGj} B_{ijt} + \boldsymbol{\beta} X_i + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (Type II)
Model 10	$\alpha + \sum_{j=1}^{3} \beta_{IGj} G_{ijt} + \sum_{j=1}^{4} \beta_{Ej} E_{ijt} + \sum_{j=1}^{4} \beta_{OGj} B_{ijt} + \boldsymbol{\beta} X_i + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (Type III)
Model 11	$\alpha + \sum_{j=1}^{3} \beta_{IGj} G_{ijt} + \sum_{j=1}^{4} \beta_{Ej} E_{ijt} + \sum_{j=1}^{4} \beta_{OGj} B_{ijt} + \boldsymbol{\beta} X_i + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (Type IV)

Here X represents the covariates discussed in Section 2.4 as follows:

Table 3.3: Variables Representation in the Models

Variables	Meaning
x_1	proportion of apartment in a building that has five or more storeys in the total occupied private dwellings for different areas
x_2	proportion of other attached dwelling in the total occupied private dwellings
x_3	proportion of apartment in a building that has fewer than five storeys
x_4	average household size
x_5	prevalence of low income based on the Low-income measure (\%, after tax)
x_6	prevalence of low income in 0-17 years age group(\%, after tax)
x_7	prevalence of low income in 18-64 years age group(\%, after tax)

x_8	prevalence of low income in over 65 years age group(\%, after tax)
x_9	proportion of self-employed workers in all class workers
x_{10}	population density
x_{11}	proportion of health occupations in all occupations
x_{12}	65 years and over distribution (%) of the population by broad age groups

The models are assessed using the Deviance Information Criterion (DIC) with lower values indicating a better fit (Spiegelhalter et al.,2002). The criterion takes into account the goodness-of-fit as well as a penalty term that is based on the complexity of the model via the estimated effective number of parameters. The DIC is defined as:

$$DIC = D(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) + 2p_D$$

Here, D(.) is the deviance computed by $-2logp(y|\hat{\theta}, \hat{\psi})$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\psi}}$ are the posterior expectations of the latent effects and hyperparameters, respectively, and p_D is the effective number of parameters. This can be computed as follows:

$$p_D = E[D(.)] - D(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$$

The Watanabe-Akaike information criterion (WAIC), also known as widely applicable Bayesian information criterion, is similar to the DIC but the effective number of parameters is computed in a different way (Watanabe and Opper,2010; Watanabe and Gelman,2013; Hwang and Vehtani,2014). The following table shows the DIC value and WAIC value for different models, and then the model with better performance will be chosen accordingly.

Table 3.5: DIC and WAIC Values Corresponding to Models1 to 11

Models	DIC	WAIC
Model 1*	354.50	345.07
Model 2*	354.34	344.73
Model 3	55201.30	73679.47
Model 4	8435.38	8208.90
Model 5	8411.24	8267.15
Model 6	8448.78	8239.42
Model 7	8449.13	8342.63
Model 8	8434.35	8211.21
Model 9	8410.87	8266.66
Model 10	8448.85	8239.20
Model 11	8442.71	8331.01

In Model 1* and Model 2*, the DIC value is the mean value of all different models in different time under the study.
Models 1 and 2 are fitted and are independent of time, but all of the data should be taken into account. As for Model3, the DIC value is much higher than Model4 - Model11, so it is important to consider the space and time interaction effect. Comparing Model4 - Model7, Model5 is best; Comparing Model8 - Model11, Model9 is best. Model9 is better than Model5.

3.5 Covariates effect

In Section 3.3, we evaluated all of the models. Regarding the effect of each covariate on the spread of COVID-2019, the posterior distribution of each coefficient on the covariates is reflected as follows:

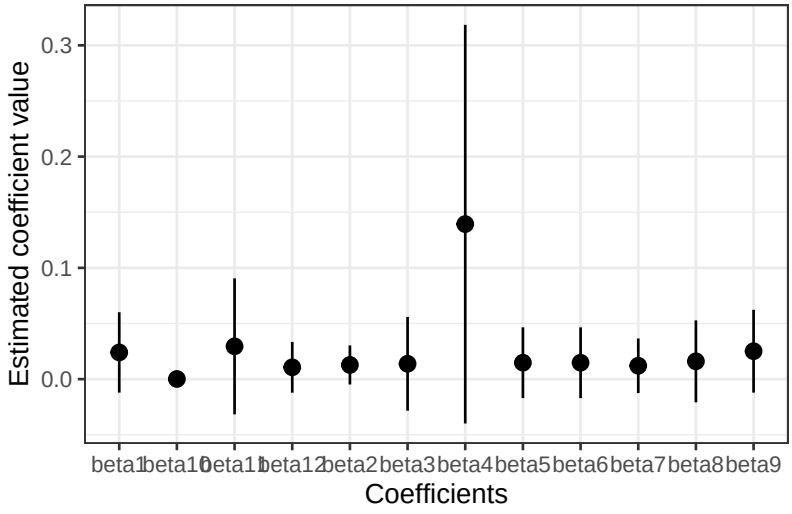


Figure 3.1: Summary of the Estimates Obtained for the Coefficients Associated with Covariates for Model 5

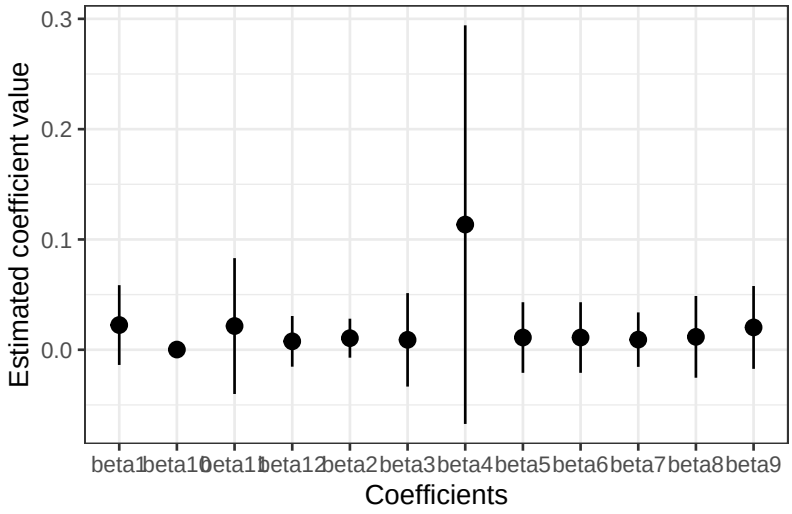


Figure 3.2: Summary of the Estimates Obtained for the Coefficients Associated with Covariates for Model 9

Comparing Model 5 and Model 9, it can be seen that almost all these covariates have significant positive associations with infected risk. Covariate x_4 has more statistically significant positive associations compared with the others. It represents average household size. Also x_9, the proportion of self-employed workers in all class workers, shows a clear association with infected risk.

3.6 Policy Effect

For the policy analysis, regarding the effect of each intervention factor on the spread of COVID-2019, the posterior distribution of each coefficient on the intervention factor is reflected as follows:

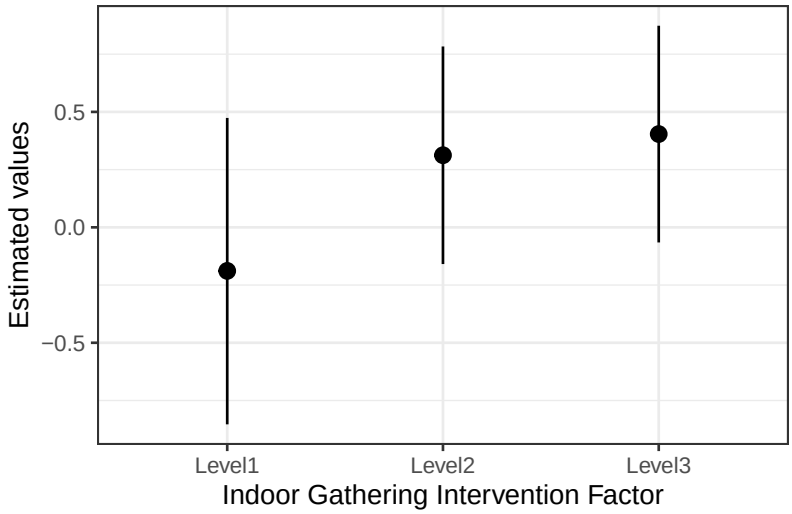


Figure 3.3: Summary of the Estimates Obtained for the Intervention Factors Associated with Covariates for Model 9

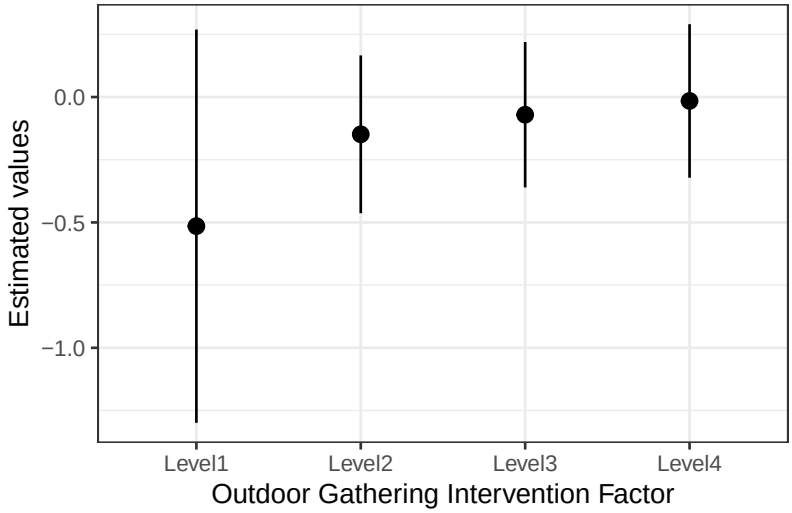


Figure 3.4: Summary of the Estimates Obtained for the Intervention Factors Associated with Covariates for Model 9

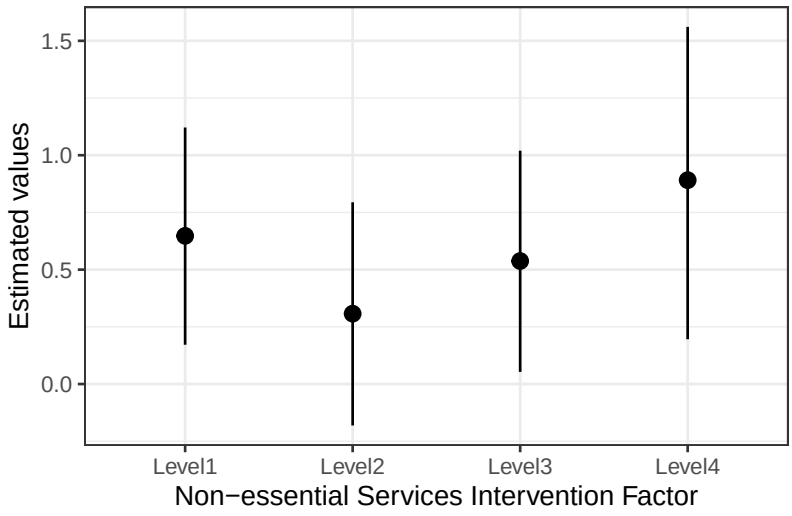


Figure 3.5: Summary of the Estimates Obtained for the Intervention Factors Associated with Covariates for Model 9

It can be seen that the level 2 and level 3 in Indoor Gathering Intervention Factor are significant, and the level 1, level 3 and level 4 in Non-essential Services Intervention Factor could be significant as well. Significance is determined by whether or not the confidence interval includes zero.

3.7 Spatial-temporal Effect

The inclusion of spatial-temporal effects helps to understand how the disease has spread throughout Ontario. Specifically, the estimates of spatial and temporal effects and their interaction assign an infected risk to each spatial, temporal, spatial-temporal unit analyzed. These infected risks are obtained by exponentiating the space, time or space-time parameters that can be used to describe the $log(\mu_{it})$ expression in each model.

The following figure shows the infected risks over time in terms of the random temporal effects estimated through Model 3, which are selected in Section 3.3. The infected risk represented by the structured component of the random temporal effect ($exp(\gamma_t)$) captures the evolution of the pandemic in Ontario:

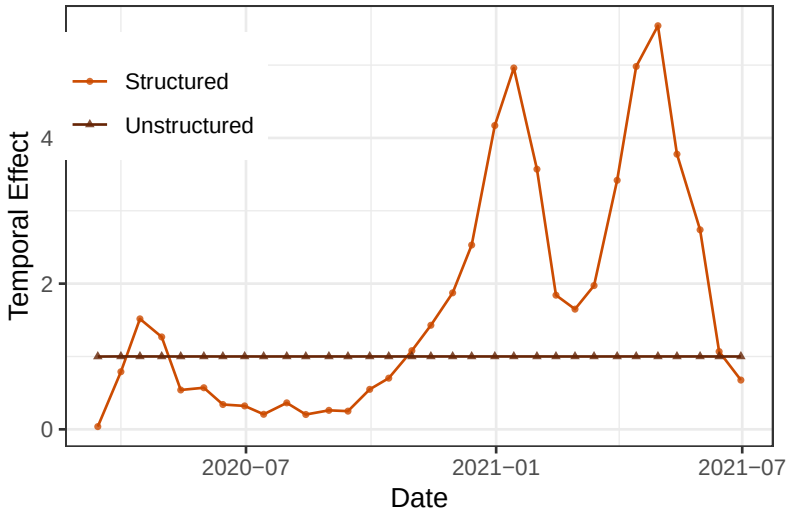


Figure 3.6: The Infected Risk Represented by the Temporal Structured Component in Model 3

Here, the infected risk corresponding to the structured component is computed as $exp(\gamma_t)$, whereas the corresponding unstructured component is computed as $exp(\phi_t)$. From Figure 3.6, we can see that the risk was nearly 0 in March 2020. The second wave started from November 2020, and then decreased from February, 2021. However, the third wave began again in April, 2021 and the peak of infected risk is larger than these two previous waves. The risk associated with the unstructured temporal component in Model 3 barely fluctuates around 1, suggesting that there were no notable overall changes in the risk during the period of study attributable only to single days. This fact can also be verified by comparing the estimates of the precision parameters associated with each of the random effects included in the models shown in the following table:

Table 3.6: Posterior Precision of Parameters Associated with Each Spatial, Temporal, and Spatial-temporal Random Effect Included in Models 3-11

Model	θ_u	θ_v	τ_λ	τ_ϕ	τ_δ
Model 3	1832.53	5.56	2.13	20619.17	-
Model 4	1145.38	6.32	1.85	4148.08	1.96
Model 5	1910	1840	2.08	16600	0.683
Model 6	2520	2.64	18700	17800	0.861
Model 7	1837.637	1532.830	2.649	33.864	0.212
Model 8	$1 * 10^8$	7.19	6.37	836	2.06
Model 9	2250	2570	3.99	32500	0.685
Model 10	1687.655	3.188	1751.203	3544.643	0.867
Model 11	4.895	1156.541	Inf	2720.174	0.194

Values in the table are the posterior precision of each random component. Each precision parameter represents the inverse of the variance of the corresponding random effect. For example, $\tau_\lambda = \frac{1}{\sigma_\lambda^2}$ is the precision parameter associated with the temporally structured effect, γ_t. A smaller precision parameter indicates a larger variance from the corresponding random effect, which also reflects that such effect has a greater contribution to risk variations. With the exception of models 6 and 11, τ_γ is clearly smaller than τ_ϕ, which confirms the larger contribution of the temporally-structured effect to risks along with time.

With respect to the random spatial effects, the contribution to risks, as measured by the unstructured spatial effect, v_i, captures most of the spatial variation since $\theta_v < \theta_u$. This indicates that there has been a not so strong spatial dependence between the regions studied in terms of their COVID-19 risks.

Interpreting the precision parameters of spatial-temporal interaction terms is more challenging, but we can derive some general outcomes. The precision of the interaction parameter δ_{it} is very relatively small in all models, which indicates that the space-time interaction highly contributes to risks. Figure 3.7 exhibits for Model 9 the spatial-temporal risks computed as $exp(u_i + v_i + \gamma_t + \phi_t + \delta_{it})$.

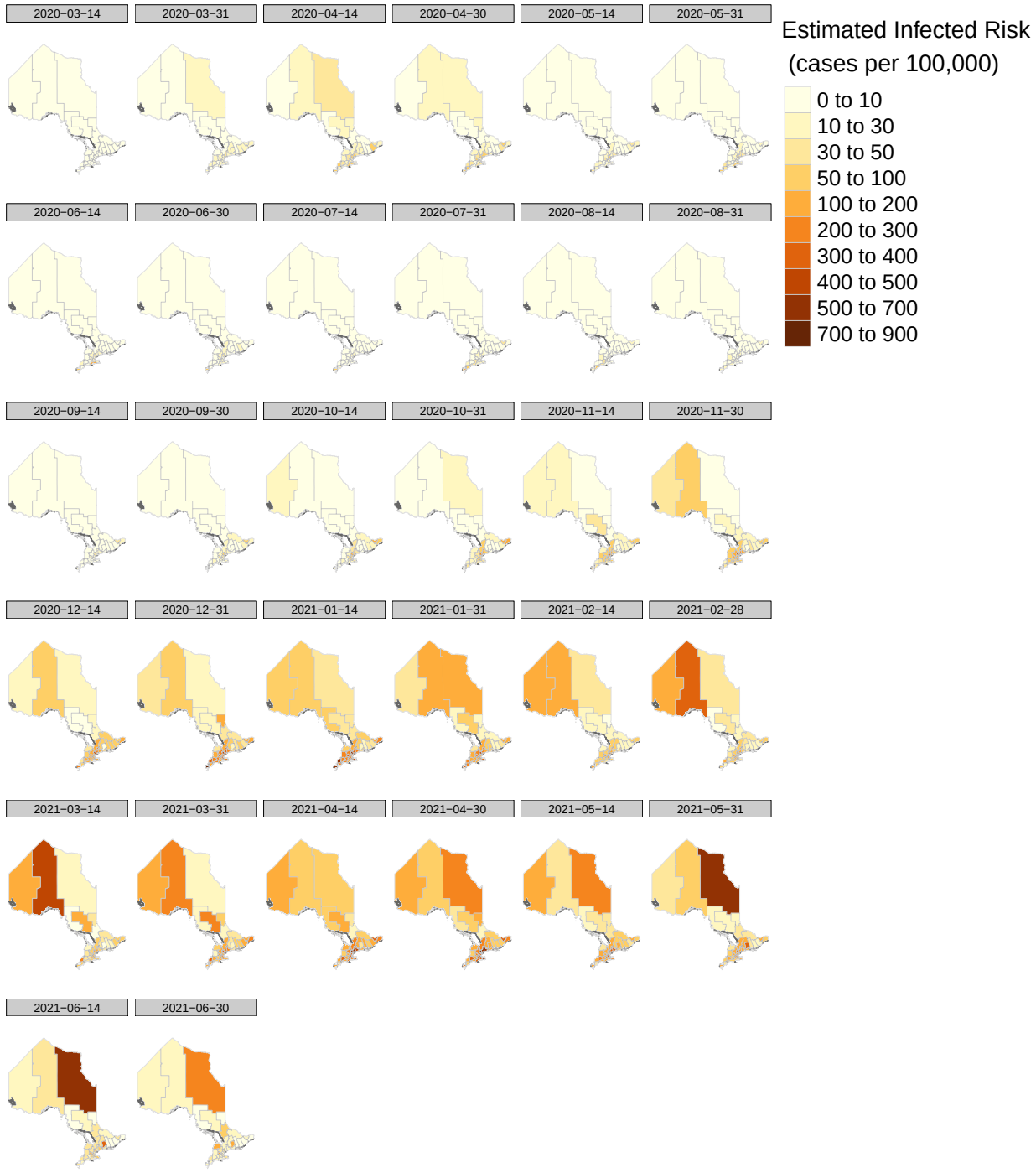


Figure 3.7: Evolution of the Infected Risks at the Public Health Unit Level (Model 9)

By observing the evolution of the infected risks across regions and time in Figure 3.7, it can be seen how certain regions in the southern zone of Ontario presented larger risks over time. At the beginning, the risks in the Northern area are relatively small, but they increase after January, 2021, when the second wave started. Since the third wave, the infected current cases in Northern Ontario are more than in Southern Ontario.

Chapter 4

Discussion

4.1 Detection of Zones of Low and High Risks

Infected risk maps can be used by public health officials to identify areas of excess and to guide surveillance and control activities. It is thus important that the prediction method lead to a correct ranking of geographical units in terms of infected risk. The Spearman rank correlation coefficient measures the strength of the monotonic relation between two variables. The correlation between the rank of observed actual infected rates and estimated risk values can be computed as:

$$\rho_{\text{rank}} = \frac{1}{N} \frac{\sum_{\alpha=1}^{N} [\gamma_P(\boldsymbol{u}_\alpha) - \bar{\gamma}_P][\gamma(\boldsymbol{u}_\alpha) - \bar{\gamma}]}{s_P s}$$

where $\gamma_P(\boldsymbol{u}_\alpha)$ and $\gamma(\boldsymbol{u}_\alpha)$ are the rank of estimated and observed risk rates. The corresponding mean and standard deviation are denoted $\bar{\gamma}$ and s. The rank correlation was averaged over all different times under the study (T = 31). The following table shows the performance of different methods using the correlation coefficient between estimators and observed infected rates.

Table 4.1: Performance Comparison of Different Methods: Rank Correlation Coefficient between Estimates and Observed Infected Rates

Models	Rank Correlation Coefficients
Bayesian Spatio-temporal Model 5	0.9944
Bayesian Spatio-temporal Model 9	0.9985

It's obvious that the rank correlation coefficients between estimates from Bayesian methods and observed infected rates are very close to 1, which represents the rank from Bayesian estimates is closer to the observed infected rates. We can use Table 4.2 to see the prediction performance of different methods.

4.2 Prediction Performance

We examine the prediction performance of our proposed models. We take the data until time T^* and predict the infected risk within q bi-weeks ahead, i.e., y^*_{it} for $t = T^* + 1,...,T^* + q$. The prediction performance is evaluated using the absolute mean error of prediction (AMEP) given by

$$AMEP = \frac{1}{Iq} \sum_{T^*+1}^{T^*+q} \sum_{i=1}^{I} \|exp(\hat{\theta_{it}} - exp(\theta_{it})\|$$

The results for q=1 and q=2, i.e. predicting 1 and 2 time periods ahead, are given in the following tables. We choose $T^* = 30$ or $T^* = 31$ to give a sufficient observed data for prediction. For each T^* and q, the AMEP is reported for the Model 8-Model 11. The Model 9 (corresponding to Type II interaction) performs better by giving the smaller AMEP value when $T^* = 30$ and q=2. The Model 10 (corresponding to Type III interaction) gives the smallest AMEP value when $T^* = 31$ and q=1. Generally, Model 9 (Type II interaction) has a better performance when doing longer-term predictions.

Table 4.2: Results of 2-biweeks Ahead Prediction for Models 8-11

Models	AMEP
Model 8 (Type I)	89.0046
Model 9 (Type II)	16.4302
Model 10 (Type III)	16.9821
Model 11 (Type IV)	166.7995

Table 4.3: Results of 1-biweek Ahead Prediction for Models 8-11

Models	AMEP
Model 8 (Type I)	63.3313
Model 9 (Type II)	30.8323
Model 10 (Type III)	27.5105
Model 11 (Type IV)	76.0943

Chapter 5

Poisson Kriging Models

Ordinary Kriging is a commonly used geostatistical approach for spatial interpolation, which relies on a spatial model to predict attribute values at unsampled locations. The basic idea of Ordinary Kriging is to predict the value of a function at a given point by computing a weighted average of the known values of the function in the neighborhood of the point. However, what we have in the COVID-19 dataset is areal polygonal data with discrete counts. Even though Ordinary Kriging can accept polygonal input by associating the polygon's value with its centroid and treat it as a point, the sizes of the polygons are not taken into account. Therefore, the method is inappropriate for this type of count data.

The use of Poisson Kriging for the analysis of COVID-19 data will be illustrated using directly age-adjusted infection rates. The population at risk was set at:

$$\frac{100,000 \times \text{the biweekly number of infected cases}}{\text{the corresponding biweekly age-adjusted infection rates}}$$

For a given number N of geographical units v_α (e.g. counties, public health units), denote the number of recorded infection cases by $d(v_\alpha)$. Let $\mathbf{u}_\alpha = (\mathbf{x}_\alpha, \mathbf{y}_\alpha)$ be the vector of spatial coordinates for the centroid of the unit v_α. The observed infection rate is then denoted as $z(\mathbf{u}_\alpha) = d(\mathbf{u}_\alpha)/n(\mathbf{u}_\alpha)$, where $n(\mathbf{u}_\alpha)$ is the population size in unit v_α. At each location $\mathbf{u}_\alpha$, the disease counts $d(\mathbf{u}_\alpha)$ can be interpreted as realizations of random variables $D(\mathbf{u}_\alpha)$ that follow a Poisson distribution with parameter (expected number of counts) that is the product of the population size $n(\mathbf{u}_\alpha)$ by the local risk $R(\mathbf{u}_\alpha)$:

$$D(\mathbf{u}_\alpha)|R(\mathbf{u}_\alpha) = \text{Poisson}(n(\mathbf{u}_\alpha)R(\mathbf{u}_\alpha)) \quad \alpha = 1, ..., \text{N}$$

Given the risk $R(\mathbf{u}_\alpha)$, the infection count variables $D(\mathbf{u}_\alpha)$ are assumed to be conditionally independent. Hence any spatial correlation among the counts is influenced

by spatial trends in either the population sizes or in the local risks. The risk variable $R(\mathbf{u}_\alpha)$ itself can be modeled as a stationary random field with mean m, variance σ_R^2 and variance function $C_R(\mathbf{h})$ (Goovaerts, P., 2006). The conditional mean and variance of the rate variable $Z(\mathbf{u}_\alpha)$ are defined respectively as:

$$\begin{aligned} E[Z(\mathbf{u}_\alpha)|R(\mathbf{u}_\alpha)] &= E[\frac{D(\mathbf{u}_\alpha)}{n(\mathbf{u}_\alpha)}|R(\mathbf{u}_\alpha)] \\ &= \frac{1}{n(\mathbf{u}_\alpha)} E[D(\mathbf{u}_\alpha)|R(\mathbf{u}_\alpha)] \\ &= R(\mathbf{u}_\alpha) \end{aligned}$$

$$\begin{aligned} \mathrm{Var}[Z(\mathbf{u}_\alpha)|R(\mathbf{u}_\alpha)] &= \mathrm{Var}[\frac{D(\mathbf{u}_\alpha)}{n(\mathbf{u}_\alpha)}|R(\mathbf{u}_\alpha)] \\ &= \frac{1}{n(\mathbf{u}_\alpha)^2} \mathrm{Var}[D(\mathbf{u}_\alpha)|R(\mathbf{u}_\alpha)] \\ &= \frac{R(\mathbf{u}_\alpha)}{n(\mathbf{u}_\alpha)} \end{aligned}$$

For the covariance expression, the conditional independence of observations, infection cases, at different sites leads to (Monestiez P. etal., 2004):

$$\begin{aligned} E[D_i D_j|R] &= \mathrm{Cov}[D_i, D_j|R] + E[D_i|R_i]E[D_j|R_j] \\ &= \delta_{ij} n_i R_i + n_i n_j R_i R_j \end{aligned}$$

To find the unconditional mean rate observed in region α, $Z(\mathbf{u}_\alpha)$, we need to take the expectation over $R(\mathbf{u}_\alpha)$ of the conditional expectation:

$$E_R[Z(\mathbf{u}_\alpha)] = E_R E[Z(\mathbf{u}_\alpha)|R(\mathbf{u}_\alpha)] = E_R[R(\mathbf{u}_\alpha)] = m,$$

where E_Z and E_R denote expectation with respect to the marginal distributions of Z and R, respectively (Waller LA and Gotway CA, 2004). The unconditional variance of $Z(\mathbf{u}_\alpha)$ equals the sum of the variance of the conditional mean with respect to $R(\mathbf{u}_\alpha)$ and the expectation of the conditional variance:

$$\begin{aligned} \mathrm{Var}[Z(\mathbf{u}_\alpha)] &= E_R[\mathrm{Var}((Z(\mathbf{u}_\alpha)|R(\mathbf{u}_\alpha))] + \mathrm{Var}_R[E(Z(\mathbf{u}_\alpha)|R(\mathbf{u}_\alpha))] \\ &= E_R[R(\mathbf{u}_\alpha)/n(\mathbf{u}_\alpha)] + \mathrm{Var}_R[R(\mathbf{u}_\alpha)] \\ &= m/n(\mathbf{u}_\alpha) + \sigma_R^2 \end{aligned}$$

Different methods are available to estimate the risk over a given geostatistical unit with centroid $\mathbf{u}_\alpha$ from the set of observed rates, $\{z(\mathbf{u}_\alpha), \alpha = 1, ..., \mathrm{N}\}$. The estimator can be formulated as a linear combination of K neighboring rates or functions of those rates:

$$\hat{r}_{PK}(\mathbf{u}_\alpha) = \sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) z(\mathbf{u}_i),$$

where $\lambda_i(\mathbf{u}_\alpha)$ is the weight assigned to the rate $z(\mathbf{u_i})$ when estimating the risk at $\mathbf{u}_\alpha$, the geographical centroid of the unit $\mathbf{v}_\alpha$.

Neighbors can be selected in a number of ways. For example, either areas having centroids within specified distance of the estimated unit's centroid $\mathbf{u}_\alpha$ may be chosen (Kafadar K, 1994), or those that share a border with the area to be smoothed (Marshall RJ,1991) may be selected. In this thesis, the search strategy allows the user to select a maximum number of neighbors that fall within a fixed distance from the centroid of the area to be smoothed (Goovaerts P.,2005). To account for the shape of geographical units and their heterogeneous population density, the distance between any two units is estimated as a population-weighted average of Euclidean distances between points discretizing the pair of units:

$$\text{Dist}(v_\alpha, v_\beta) = \frac{1}{\sum_{s=1}^{P_\alpha}\sum_{s'=1}^{P_\beta} n(\mathbf{u}_s)n(\mathbf{u}_{s'})} \sum_{s=1}^{P_\alpha}\sum_{s'=1}^{P_\beta} n(\mathbf{u}_s)n(\mathbf{u}_{s'}) \left\|\mathbf{u}_s - \mathbf{u}_{s'}\right\|, \tag{5.0.1}$$

where P_α and P_β are the number of points $\mathbf{u}_s$ and $\mathbf{u}_{s'}$ used to discretize the two units v_α and v_β, respectively. To compute $n(\mathbf{u}_s)$ and $n(\mathbf{u}_{s'})$, there is a need to draw high-resolution population maps, which are produced by allocating the unit-level or county-level population at risk estimates to a set of hexagons discretizing each study area. The relative proportion of the county-level or unit-level population within each hexagon was retrieved from the readily available Canadian 2016 census block level data. So the quantity $n(\mathbf{u}_s)$ represents the population size within each hexagon centered on the discretizing point $\mathbf{u}_s$.

5.1 Point Poisson Kriging of Areal Data

This section provides a brief recall of the centroid-based implementation of Poisson kriging (PK) for prediction of aggregated risk values (Goovaerts,P.,2006). Assume that all units $\mathbf{v}_\alpha$ have similar shapes and sizes, with a uniform population density and that each unit or measurement support is a single point. The risk over a given unit v_α is estimated as a linear combination of the rate observed for that unit, $z(\mathbf{u}_\alpha)$, in K neighboring units:

$$\hat{r}_{PK}(\mathbf{u}_\alpha) = \sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) z(\mathbf{u}_i),$$

In order to assure its unbiasedness, we compute its expectation:

$$\begin{aligned} E[\hat{r}_{PK}(\mathbf{u}_\alpha)|R] &= \sum_{i=1}^{K} \frac{\lambda_i(\mathbf{u}_\alpha)}{n_i(\mathbf{u}_\alpha)} E[D(\mathbf{u}_i)|R(\mathbf{u}_i)] \\ &= \sum_{i=1}^{K} \frac{\lambda_i(\mathbf{u}_\alpha)}{n(\mathbf{u}_i)} n_i(\mathbf{u}_i) R(\mathbf{u}_i) \\ &= \sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) R(\mathbf{u}_i) \end{aligned} \tag{5.1.1}$$

$$\begin{aligned} E[\hat{r}_{PK}(\mathbf{u}_\alpha)] &= \sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) E[R(\mathbf{u}_i)] \\ &= m \sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) \end{aligned} \tag{5.1.2}$$

So the condition for unbiasedness requires the usual one:

$$\sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) = 1 \tag{5.1.3}$$

In the same way the mean square error of prediction - MSEP - if unbiased, can be obtained by application of equation (4.2) and (4.3) to the kriging estimate:

$$\begin{aligned} E[(\hat{r}_{PK}(\mathbf{u}_\alpha) - R(\mathbf{u}_\alpha))^2|R] &= E[(\sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) \frac{D(\mathbf{u}_i)}{n(\mathbf{u}_i)} - R(\mathbf{u}_\alpha))^2|R] \\ &= \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{\lambda_i(\mathbf{u}_\alpha)}{n(\mathbf{u}_i)} \frac{\lambda_j(\mathbf{u}_\alpha)}{n(\mathbf{u}_j)} E[D(\mathbf{u}_i) D(\mathbf{u}_j)|R] + R^2(\mathbf{u}_\alpha) \\ &\quad - 2R(\mathbf{u}_\alpha) \sum_{i=1}^{K} \frac{\lambda_i(\mathbf{u}_\alpha)}{n(\mathbf{u}_i)} E[D(\mathbf{u}_i)|R] \\ &= \sum_{i=1}^{K} \sum_{j=1}^{K} \lambda_i(\mathbf{u}_\alpha) \lambda_j(\mathbf{u}_\alpha) R(\mathbf{u}_i) R(\mathbf{u}_j) + \sum_{i=1}^{K} \frac{\lambda_i^2(\mathbf{u}_\alpha)}{n(\mathbf{u}_i)} R(\mathbf{u}_\alpha) + R^2(\mathbf{u}_\alpha) \\ &\quad - 2R(\mathbf{u}_\alpha) \sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) R(\mathbf{u}_i) \end{aligned} \tag{5.1.4}$$

Then,

$$
\begin{aligned}
E[(\hat{r}_{PK}(\mathbf{u}_\alpha) - R(\mathbf{u}_\alpha))^2] &= E_R E[(\hat{r}_{PK}(\mathbf{u}_\alpha) - R(\mathbf{u}_\alpha))^2 | R] \\
&= E_R[\sum_{i=1}^{K}\sum_{j=1}^{K}\lambda_i(\mathbf{u}_\alpha)\lambda_j(\mathbf{u}_\alpha)R(\mathbf{u}_i)R(\mathbf{u}_j) + \sum_{i=1}^{K}\frac{\lambda_i^2(\mathbf{u}_\alpha)}{n(\mathbf{u}_i)}R(\mathbf{u}_\alpha) + R^2(\mathbf{u}_\alpha) \\
&- 2R(\mathbf{u}_\alpha)\sum_{i=1}^{K}\lambda_i(\mathbf{u}_\alpha)R(\mathbf{u}_i)] \\
&= \sum_{i=1}^{K}\sum_{j=1}^{K}\lambda_i(\mathbf{u}_\alpha)\lambda_j(\mathbf{u}_\alpha)C_R(\mathbf{u}_i,\mathbf{u}_j) + \sum_{i=1}^{K}\frac{\lambda_i^2(\mathbf{u}_\alpha)}{n(\mathbf{u}_i)}m + \sigma_R^2 \\
&- 2\sum_{i=1}^{K}\lambda_i(\mathbf{u}_\alpha)C_R(\mathbf{u}_i,\mathbf{u}_\alpha) + 2m^2 - 2m^2 \\
&= \sigma_R^2 + \sum_{i=1}^{K}\frac{\lambda_i^2(\mathbf{u}_\alpha)}{n(\mathbf{u}_i)}m + \sum_{i=1}^{K}\sum_{j=1}^{K}\lambda_i(\mathbf{u}_\alpha)\lambda_j(\mathbf{u}_\alpha)C_R(\mathbf{u}_i,\mathbf{u}_j) \\
&- 2\sum_{i=1}^{K}\lambda_i(\mathbf{u}_\alpha)C_R(\mathbf{u}_i,\mathbf{u}_\alpha)
\end{aligned}
\tag{5.1.5}
$$

By minimizing this expression with respect to the $\lambda_i(\mathbf{u}_\alpha)$'s with the unbiasedness constraint, the Kriging system of (n+1) equations will be derived as the Ordinary Kriging optimization process, where μ is the Lagrange multiplier (see Appendix A for details):

$$
\begin{aligned}
&\sum_{j=1}^{K}\lambda_j(\mathbf{u}_\alpha)[C_R(\mathbf{u}_i,\mathbf{u}_j) + \delta_{ij}\frac{m*}{n(\mathbf{u}_i)}] + \mu(\mathbf{u}_\alpha) = C_R(\mathbf{u}_i,\mathbf{u}_\alpha) \quad \text{for} \quad i = 1,2,...,K, \\
&\sum_{i=1}^{K}\lambda_i(\mathbf{u}_\alpha) = 1
\end{aligned}
$$

and $\delta_{ij} = 1$ if $\mathbf{u}_i = \mathbf{u}_j$ and 0 otherwise. Here, m* is the population-weighted mean of the N rates. The prediction variance associated with the estimator $\hat{r}_{PK}(\mathbf{u}_\alpha)$ is computed as (further details are given in Appendix A):

$$
\begin{aligned}
\mathrm{Var}(\hat{r}_{PK} - R(\mathbf{u}_\alpha)) &= \sigma_{PK}^2(\mathbf{u}_\alpha) \\
&= \sigma_R^2 - \sum_{i=1}^{K}\lambda_i(\mathbf{u}_\alpha)C_R(\mathbf{u}_i,\mathbf{u}_\alpha) - \mu(\mathbf{u}_\alpha)
\end{aligned}
$$

To solve the Poisson kriging system, we need a model for the covariance of the risk, $C_R(\mathbf{u}_i,\mathbf{u}_j)$ ($C_R(\mathbf{h})$), or can use a semivariogram to calculate it as $\gamma_R(\mathbf{h}) = C_R(0) - C_R(\mathbf{h})$ (see in Appendix A). The semivariogram of the risk is estimated as (Monestiez

et al.,2005):

$$\hat{\gamma}_R(\mathbf{h}) = \frac{1}{2\sum_{\alpha=1}^{N(\mathbf{h})} \frac{n(\mathbf{u}_\alpha)n(\mathbf{u}_{\alpha+h})}{n(\mathbf{u}_\alpha)+n(\mathbf{u}_{\alpha+h})}} \sum_{\alpha=1}^{N(\mathbf{h})} \{\frac{n(\mathbf{u}_\alpha)n(\mathbf{u}_{\alpha+h})}{n(\mathbf{u}_\alpha)+n(\mathbf{u}_{\alpha+h})}[z(\mathbf{u}_\alpha)-z(\mathbf{u}_{\alpha+h})]^2 - m*\} \quad (5.1.6)$$

5.2 Area-to-Area Poisson Kriging

It is overly simplistic to assimilate each unit v_α to its geographic centroid $\mathbf{u}_\alpha$ because the geographical units have very different shapes and sizes. The spatial support of each unit needs to be accounted for in both the semivariogram estimate and in kriging system. Area-to-Area kriging refers to the case where both the prediction and measurement supports are areas instead of points and areal supports are disjointed (Kyriakidis,2004), i.e. $v_k \cap v_l = \varnothing$. The Poisson kriging estimate for the areal risk value $r(v_\alpha)$ thus can be expressed as a weighted linear combination of the K available areal data:

$$\hat{r}_{PK}(v_\alpha) = \sum_{i=1}^{K} \lambda_i(v_\alpha) z(v_i)$$

$$\hat{r}_{PK}(v_\alpha) = \frac{1}{|v_\alpha|}\int_{s\in v_\alpha} z(\mathbf{s})d\mathbf{s} \simeq \frac{1}{P_k}\sum_{i=1}^{P_k} z(\mathbf{s}_i), \quad \mathbf{s}_i \in v_\alpha$$

The weights $\lambda_i(v_\alpha)$ are computed by solving the following kriging system (the process is the same as Section 4.1):

$$\sum_{j=1}^{K} \lambda_j(v_\alpha)[\bar{C}_R(v_i,v_j) + \delta_{ij}\frac{m*}{n(v_i)}] + \mu(v_\alpha) = \bar{C}_R(v_i,v_\alpha) \quad i = 1,...,K$$

$$\sum_{j=1}^{K} \lambda_j(v_\alpha) = 1$$

where $\bar{C}_R(v_i,v_j) = \text{Cov}\{Z(v_i), Z(v_j)\}$.

The covariances are numerically approximated by averaging the point-support covariance $C(\mathbf{h})$ computed between two locations discretizing the areas v_i and v_j:

$$\bar{C}_R(v_i,v_j) = \frac{1}{\sum_{s=1}^{P_i}\sum_{s'=1}^{P_j} w_{ss'}} \sum_{s=1}^{P_i}\sum_{s'=1}^{P_j} w_{ss'} C(\mathbf{u}_s,\mathbf{u}_{s'})$$

where P_i and P_j denote the respective number of points discretizing the two supports v_i and v_j. The weights $w_{ss'}$ are computed as the product of population sizes within the cells centered on the discretizing point u_s and $u_{s'}$:

$$w_{ss'} = n(\mathbf{u}_s) \times n(\mathbf{u}_{s'}) \quad \text{with} \sum_{s=1}^{P_i} n(\mathbf{u}_s) = n(v_i) \text{and} \sum_{s'=1}^{P_j} n(\mathbf{u}_{s'}) = n(v_j)$$

The kriging variance for the Area-to-Area kriging estimator is computed as in Section 4.1:

$$\sigma^2_{PK}(v_\alpha) = \bar{C}_R(v_\alpha, v_\alpha) - \sum_{i=1}^{K} \lambda_i(v_\alpha)\bar{C}_R(v_i, v_\alpha) - \mu(v_\alpha)$$

where $\bar{C}_R(v_\alpha, v_\alpha)$ is the within-area covariance:

$$\begin{aligned}\bar{C}_R(v_\alpha, v_\alpha) &= \frac{1}{\sum_{s=1}^{P_\alpha}\sum_{s'=1}^{P_\alpha} w_{ss'}} \sum_{s=1}^{P_\alpha}\sum_{s'=1}^{P_\alpha} w_{ss'} C(\mathbf{u}_s, \mathbf{u}_{s'}) \\ &= \frac{1}{n^2(v_\alpha)}[\sum_{s=1}^{P_\alpha} n^2(\mathbf{u}_s)C(\mathbf{0}) + \sum_{s=1}^{P_\alpha}\sum_{s'=1}^{P_\alpha} w_{ss'} C(\mathbf{u}_s, \mathbf{u}_{s'})\delta_{ss'}(s \neq s')]\end{aligned}$$

5.3 Area-to-Point Poisson Kriging

The objective of Area-to-Point spatial interpolation is to predict any unknown point value $z(\mathbf{u}_s)$ using K areal data $\{z(v_i), i = 1, ..., K\}$ (Kyriakidis,2004). There is an assumption for areal supports, which is $v_i \cap v_j = \varnothing$. The predicted point value $\hat{r}_{PK}(\mathbf{u}_s)$ is expressed as a weighted linear combination of the available K areal data values.

$$\hat{r}_{PK}(\mathbf{u}_s) = \sum_{i=1}^{K} \lambda_i(\mathbf{u}_s) z(v_i) \tag{5.3.1}$$

The system of linear equations that are used to compute the kriging weights and Lagrange parameter $\mu(\mathbf{u}_s)$ are derived by minimizing the mean square error of prediction in Section 4.1.

$$\begin{aligned}&\sum_{j=1}^{K} \lambda_j(\mathbf{u}_s)[\bar{C}_R(v_j, v_j) + \delta_{ij}\frac{m*}{n(v_i)}] + \mu(\mathbf{u}_s) = \bar{C}_R(v_i, \mathbf{u}_s), \quad i = 1, ..., K \\ &\sum_{i=1}^{K} \lambda_i(\mathbf{u}_s) = 1\end{aligned}$$

The only difference between Area-to-Area Poisson kriging system and Area-to-Point Poisson kriging system is the right-hand-side term where the covariances $\bar{C}_R(v_i, v_\alpha)$ are replaced by Area-to-Point covariances $\bar{C}_R(v_i, \mathbf{u}_s)$ approximated as follows:

$$\bar{C}_R(v_i, \mathbf{u}_s) = \frac{1}{\sum_{s'=1}^{P_i} w_{s's}} \sum_{s'=1}^{P_i} w_{s's} C(\mathbf{u}_{s'}, \mathbf{u}_s)$$

Here P_i is the number of points used to discretize the area v_i and the weights $w_{s's}$ are computed as:

$$w_{s's} = n(\mathbf{u}_{s'}) \times n(\mathbf{u}_s)$$

In view of the unbiasedness condition, the Area-to-Point kriging estimator variance is given as for the process is the same as in Section 4.1:

$$\sigma^2_{PK}(\mathbf{u}_s) = C_R(0) - \sum_{i=1}^{K} \lambda_i(\mathbf{u}_s)\bar{C}_R(v_i, \mathbf{u}_s) - \mu(\mathbf{u}_s)$$

Coherence is an interesting property of the Area-to-Point kriging estimator (Goovaerts P., 2006). It means that the population-weighted average of the risk values estimated at the P_α points $\mathbf{u}_s$ in a given unit v_α yields the Area-to-Area risk estimate for this unit:

$$\hat{r}_{PK}(v_\alpha) = \frac{1}{|v_\alpha|}\int_{s\in v_\alpha} \hat{r}_{PK}(\mathbf{u}_s)ds \simeq \frac{1}{n(v_\alpha)}\sum_{s=1}^{P_\alpha} n(\mathbf{u}_s)\hat{r}_{PK}(\mathbf{u}_s) \tag{5.3.2}$$

The above constraint is satisfied if the same K areal data are used for the ATP kriging of the P_α risk values. Indeed, the population-weighted average of the right-hand-side covariances of the K Area-to-Point Poisson kriging systems is equal to the right-hand-side covariance of the single Area-to-Area kriging Poisson system:

$$\begin{aligned}\frac{1}{n(v_\alpha)}\sum_{s=1}^{P_\alpha} n(\mathbf{u}_s)\bar{C}_R(v_i, \mathbf{u}_s) &= \frac{1}{n(v_\alpha)}\sum_{s=1}^{P_\alpha} n(\mathbf{u}_s)[\frac{1}{n(v_i)}\sum_{s'=1}^{P_i} n(\mathbf{u}'_s)C(\mathbf{u}_s, \mathbf{u}_{s'})] \\ &= \bar{C}_R(v_i, v_\alpha)\end{aligned}$$

and then using equation (4.7) and (4.8), the relationship between the sets of Area-to-Area and Area-to-Point kriging weights is:

$$\hat{r}_{PK}(v_\alpha) = \frac{1}{n(v_\alpha)}\sum_{s=1}^{P_\alpha} n(\mathbf{u}_s)\hat{r}_{PK}(\mathbf{u}_s) = \frac{1}{n(v_\alpha)}\sum_{s=1}^{P_\alpha} n(\mathbf{u}_s)\sum_{i=1}^{K}\lambda_i(\mathbf{u}_s)z(v_i)$$

therefore, $\lambda_i(v_\alpha) = \frac{1}{n(v_\alpha)}\sum_{s=1}^{P_\alpha} n(\mathbf{u}_s)\lambda_i(\mathbf{u}_s) \quad i = 1, ..., K$

5.4 Deconvolution of the Semivariogram of the Risk

In this section, we describe how the semivariograms are estimated (Goovaerts P., 2008). The theoretically regularized and point support semivariograms are related by the general formula (Journel and Huijbregts, 1978):

$$2\gamma_v(h) = 2\bar{\gamma}_v(v(\mathbf{u}), v(\mathbf{u}+\mathbf{h})) - \bar{\gamma}(v(\mathbf{u}), v(\mathbf{u})) - \bar{\gamma}(v(\mathbf{u}+\mathbf{h}), v(\mathbf{u}+\mathbf{h})),$$

Under the assumption of stationarity, it becomes:

$$\gamma_v(\mathbf{h}) = \bar{\gamma}(v, v_\mathbf{h}) - \bar{\gamma}(v, v).$$

In irregular geographical units, the size of the areas used in semivariogram computation varies as a function of the distance between them. Therefore, there is a more general expression for the regularization which is proposed as:

$$\gamma_v(\mathbf{h}) = \bar{\gamma}(v, v_{\mathbf{h}}) - \bar{\gamma}_{\mathbf{h}}(v, v) \tag{5.4.1}$$

The second term, within-area semivariogram value, is estimated as the arithmetical average of within-area semivariogram values for any pair of areas separated by a given distance $\mathbf{h}$:

$$\bar{\gamma}_{\mathbf{h}}(v, v) = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [\bar{\gamma}(v_\alpha, v_\alpha) + \bar{\gamma}(v_{\alpha+\mathbf{h}}, v_{\alpha+\mathbf{h}})],$$

where $\bar{\gamma}(v_\alpha, v_\alpha)$ and $\bar{\gamma}(v_{\alpha+\mathbf{h}}, v_{\alpha+\mathbf{h}})$ are estimated as:

$$\bar{\gamma}(v_\alpha, v_\alpha) = \frac{1}{P_\alpha^2} \sum_{s=1}^{P_\alpha} \sum_{s'=1}^{P_\alpha} \gamma(\mathbf{u}_s, \mathbf{u}_{s'})$$

$$\bar{\gamma}(v_{\alpha+\mathbf{h}}, v_{\alpha+\mathbf{h}}) = \frac{1}{P_{\alpha+\mathbf{h}}^2} \sum_{s=1}^{P_{\alpha+\mathbf{h}}} \sum_{s'=1}^{P_{\alpha+\mathbf{h}}} \gamma(\mathbf{u}_s, \mathbf{u}_{s'})$$

The Area-to-Area semivariogram value, $\bar{\gamma}(v, v_{\mathbf{h}})$, represents the mean value of the point support semivariogram between an arbitrary point in the geographical support v and another in the translated support $v_{\mathbf{h}}$. Similarly, it is estimated as:

$$\bar{\gamma}(v, v_{\mathbf{h}}) = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \bar{\gamma}(v_\alpha, v_{\alpha+\mathbf{h}})$$

The semivariogram value between any two areas, v_α and $v_{\alpha+\mathbf{h}}$, separated by the population-based distance $\mathbf{h}$ is computed as:

$$\bar{\gamma}(v_\alpha, v_{\alpha+\mathbf{h}}) = \frac{1}{P_\alpha P_{\alpha+\mathbf{h}}} \sum_{s=1}^{P_\alpha} \sum_{s'=1}^{P_{\alpha+\mathbf{h}}} \gamma(\mathbf{u}_s, \mathbf{u}_{s'})$$

where P_α and $P_{\alpha+\mathbf{h}}$ are the number of points used to discretize the two areas v_α and $v_{\alpha+\mathbf{h}}$, respectively.

The deconvolution procedure starts with the choice of an initial point-support model $\gamma^{(0)}(\mathbf{h})$ and like most inverse problems, the deconvolution is best tackled using an iterative procedure:

1) Compute the experimental semivariogram of areal data $\gamma_v(\mathbf{h})$ and fit a model $\gamma_v^{exp}(\mathbf{h})$ using weighted least-square regression (Pardo-Iguzquiza 1999). Three types of semivariogram models (spherical, exponential, and cubic) are tried (further details are provided in Appendix B) and the one that yields the smallest deviation between the experimental and modeled curves is selected. Each lag is weighted by $\sqrt{N(\mathbf{h})}/\gamma_v^{exp}(\mathbf{h})$ to assign more importance to the fitting of semivariogram values at short distances.

2) As an initial point support model $\gamma^{(0)}(\mathbf{h})$, use the model (type of semivariogram function and parameters) fitted to areal data, $\gamma_v^{exp}(\mathbf{h})$.

3) Regularize the model $\gamma^{(0)}(\mathbf{h})$ according to expression (4.9).

$$\gamma_v^{(0)}(\mathbf{h}) = \bar{\gamma}^{(0)}(v, v_{\mathbf{h}}) - \bar{\gamma}_{\mathbf{h}}^{(0)}(v, v)$$

4) Quantify the deviation between the 'data-based' ($\gamma_v^{exp}(\mathbf{h})$) and the theoretically regularized ($\gamma_v^{(0)}(\mathbf{h})$) semivariogram models using the average relative difference between these two curves over L lags $\mathbf{h}_l$

$$D^{(0)} = \frac{1}{L}\sum_{l=1}^{L}\frac{|\gamma_v^{(0)})(h_l) - \gamma_v^{exp}(h_l)|}{\gamma_v^{exp}(h_l)} \tag{5.4.2}$$

5) Consider the initial point support model $\gamma^{(0)}(\mathbf{h})$, the regularized model $\gamma_v^{(0)}(\mathbf{h})$, and the associated difference statistic $D^{(0)}$ as 'optimal' at this stage

$$\gamma^{(opt)}(\mathbf{h}) = \gamma^{(0)}(\mathbf{h}), \gamma_v^{(opt)}(\mathbf{h}) = \gamma_v^{(0)}(\mathbf{h}), \text{and} D^{(opt)} = D^{(0)}$$

6) For each lag $\mathbf{h}_l$, compute experimental values for the new point support semivariogram through a rescaling of the optimal point support model $\gamma^{(opt)}(\mathbf{h})$

$$\hat{\gamma}^{(1)}(h_l) = \gamma^{(opt)}(h_l) \times w^{(1)}(h_l) \quad \text{with} \quad w^{(1)}(h_l) = 1 + \frac{(\gamma_v^{exp}(h_l) - \gamma_v^{(opt)}(h_l))}{s_{exp}^2\sqrt{iter}}$$

7) Fit a model $\gamma^{(1)}(\mathbf{h})$ to the scaled values using weighted least-square regression (same procedure as in step 1)

8) Regularize the model $\gamma^{1}(\mathbf{h})$ according to expression (4.10)

$$\gamma_v^{(1)}(\mathbf{h}) = \bar{\gamma}^{(1)}(v, v_{\mathbf{h}}) - \bar{\gamma}_{\mathbf{h}}^{(1)}(v, v)$$

9) Compute the difference statistic (4.4.2) for the new regularized model $\gamma_v^{(1)}(\mathbf{h})$

   – If $D^{(1)} < D^{opt}$, use the point support model $\gamma^{(1)}(\mathbf{h})$ and the associated statistic $D^{(1)}$. Repeat steps 6 through 8.

   – If $D^{(1)} \geq D^{opt}$, repeat steps 6 through 8 using the same optimal model but the new rescaling coefficients computed as

$$w^{(2)}(h_l) = 1 + \frac{w^{(1)}(h_l) - 1)}{2}$$

10) The procedure stops when the maximum number of allowed iterations has been tried (e.g. iter > 25) or the decrease in the D statistic becomes negligible from one iteration to the next.

5.5 Algorithm for ATA and ATP Poisson Kriging

In this section, we describe the algorithm for estimating the semivariogram for Poisson Kriging. We use the python package pyinterpolate to implement the deconvolution process in Section 4.4 (Molinski, S.,2021). The initial point semivariogram is important in the procedure. In order to get a better result, another choice of initial setting is added in the algorithm.

1) Input the 2016 Canadian Census Dissemination Block level data through API.
2) Obtain the $15 \times 15 = 225$ km^2 cell polygon discretizing each public health unit in Ontario with QGIS software.
3) Count the number of points in each cell and mark each one with a unique ID.
4) Calculate the proportion of population in each Dissemination Block given the 2016 Canadian Census.
5) Compute the number of population at risk in public health unit using back-calculation method from the age-adjusted rate and the count data. And calculate the number of people at risk in each Dissemination Block given the proportion in the previous step.
6) Sum the number of people at risk in each cell given the ID with Python. And obtain the center centroid of each cell using the QGIS software.
7) Count how many points included in each public health unit. And each unit are discretized through these points.
8) Calculate the experimental areal-weighted semivariogram and experimental areal semivariogram to be the initial point support semivariogram options.
9) Repeat the deconvolution process in Section 4.4 until the iteration stops.

5.6 Results and Analysis

Figure 5.1 (red curve) shows the omnidirectional semivariograms of COVID-19 infected risk from March 14 2020 to March 31 2020, which is computed from unit-level rates using estimator (5.6) and the distance measure (5.1). (For other semivariograms in different time, see Figure 5.4). The experimental semivariogram is fitted using a exponential model with range of 240km for COVID-19 infected cases in Ontario. The model is deconvoluted using the iterative procedure and the high-resolution population map is displayed in Figure 2.1. The black dotted curve represents the theoretical

point support semivariogram model. The deconvoluted semivariogram models were used to estimate aggregated risk values at the public health unit level in Ontario (ATA Poisson kriging), see Figure 5.2. Figure 5.3 shows the risk values estimated at the nodes of 15 km grid using ATP Poisson Kriging with theoretical point support semivariograms. The ATP kriged map displays much more details, with the presence of clearly delineated areas of lower and higher rates.

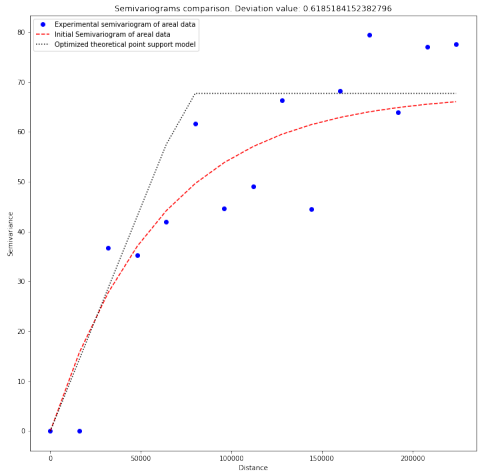


Figure 5.1: Semivariogram Model March 14-March 31 2020 Used for Mapping Infected Risk in Ontario

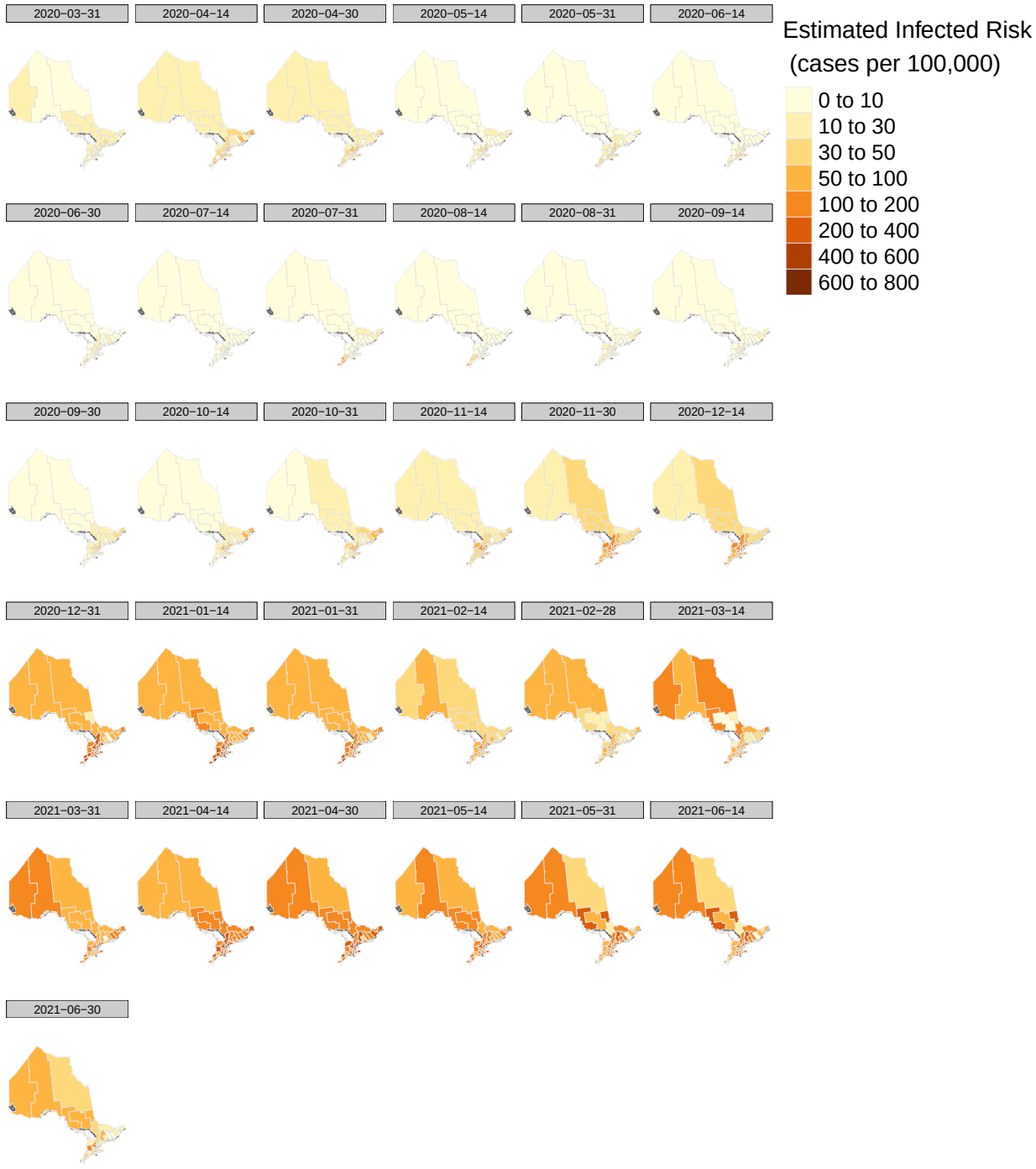
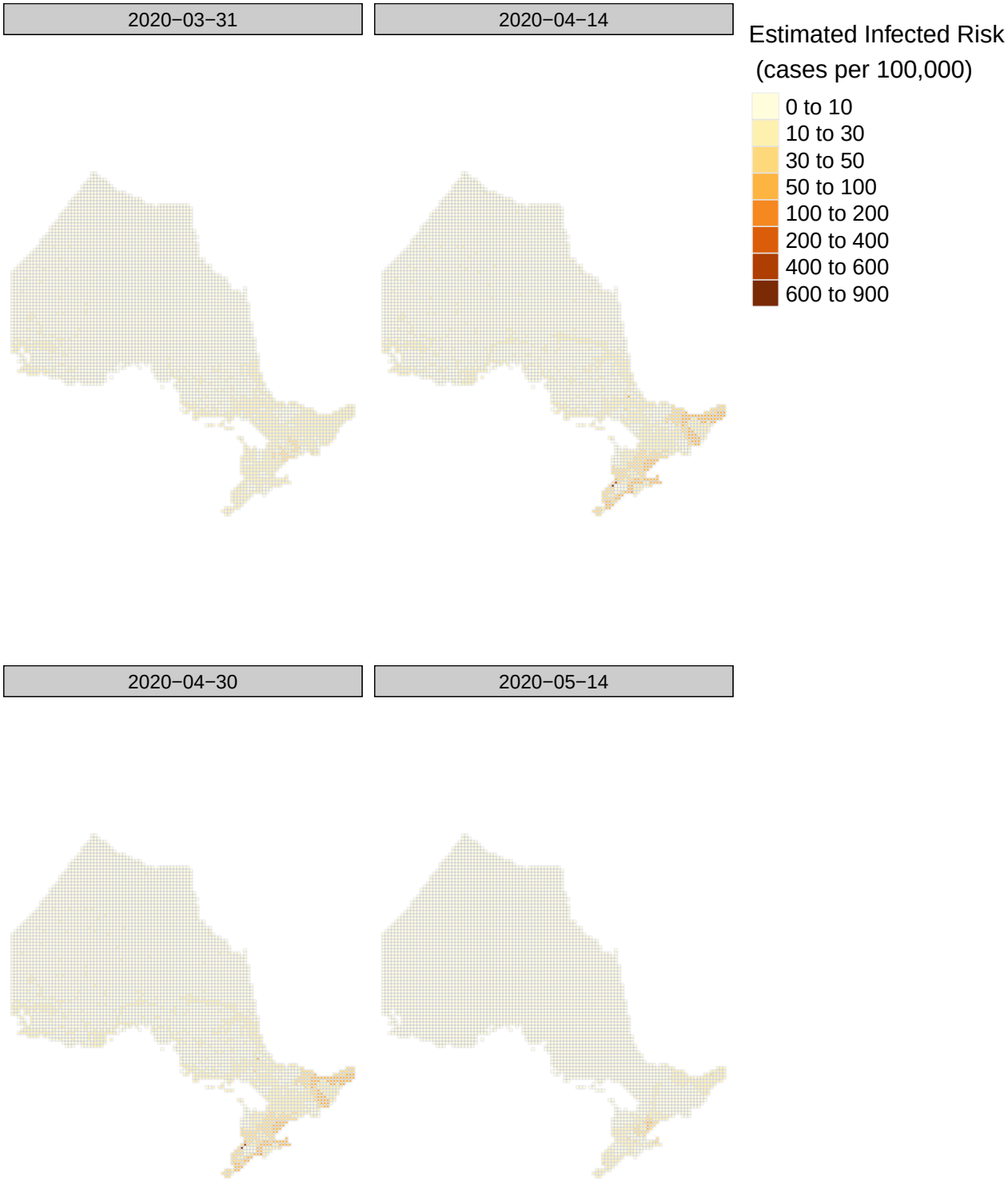
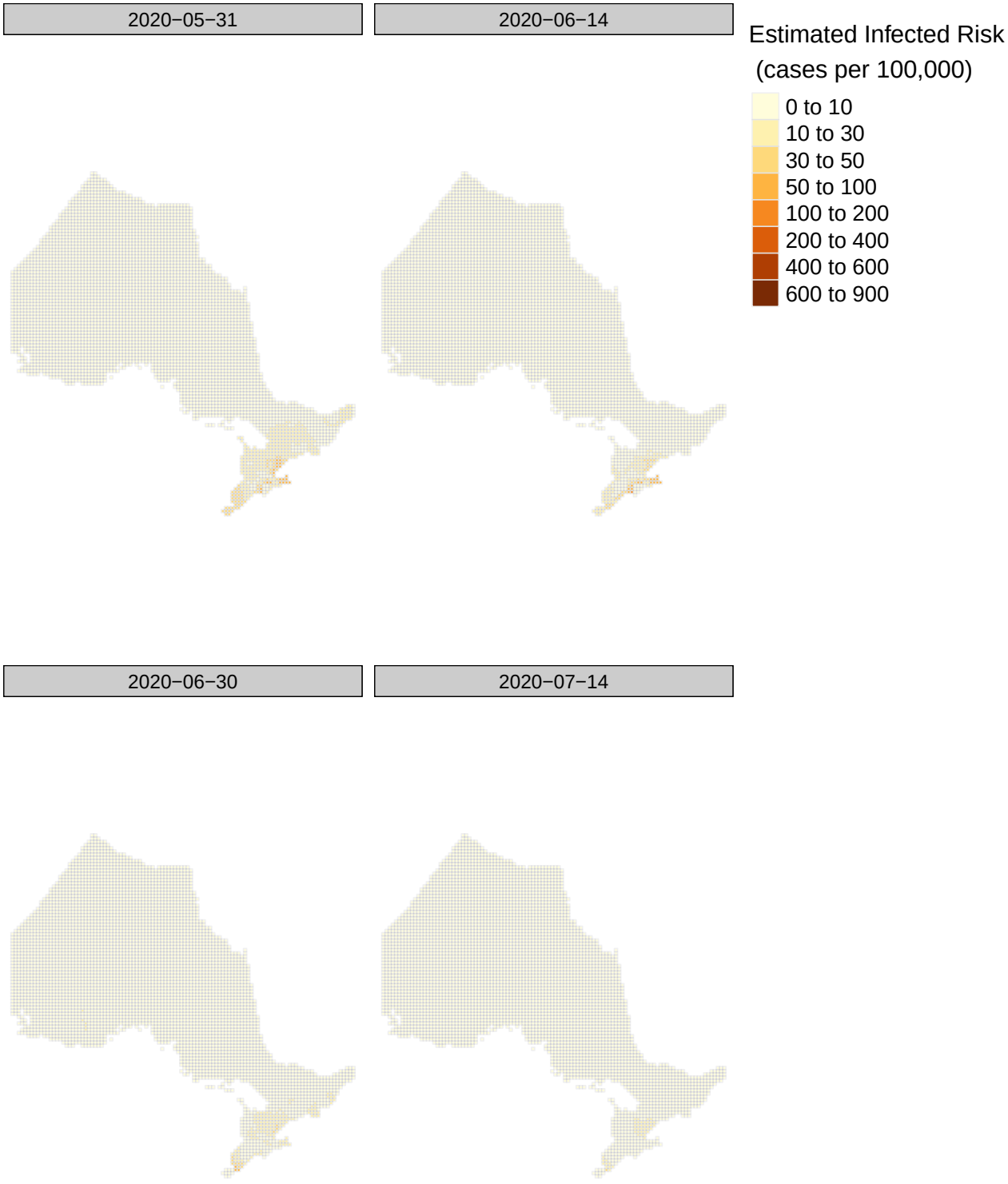
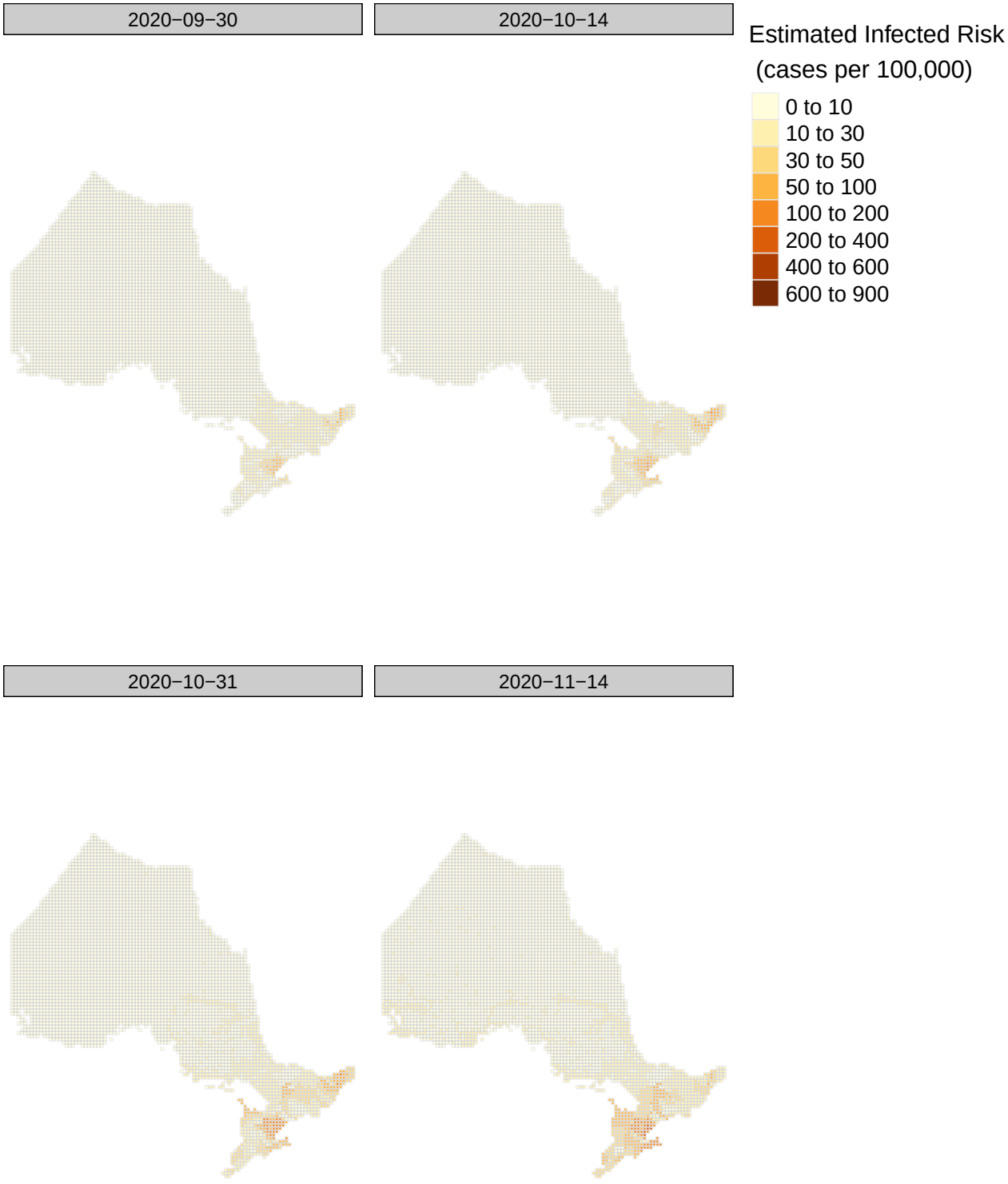


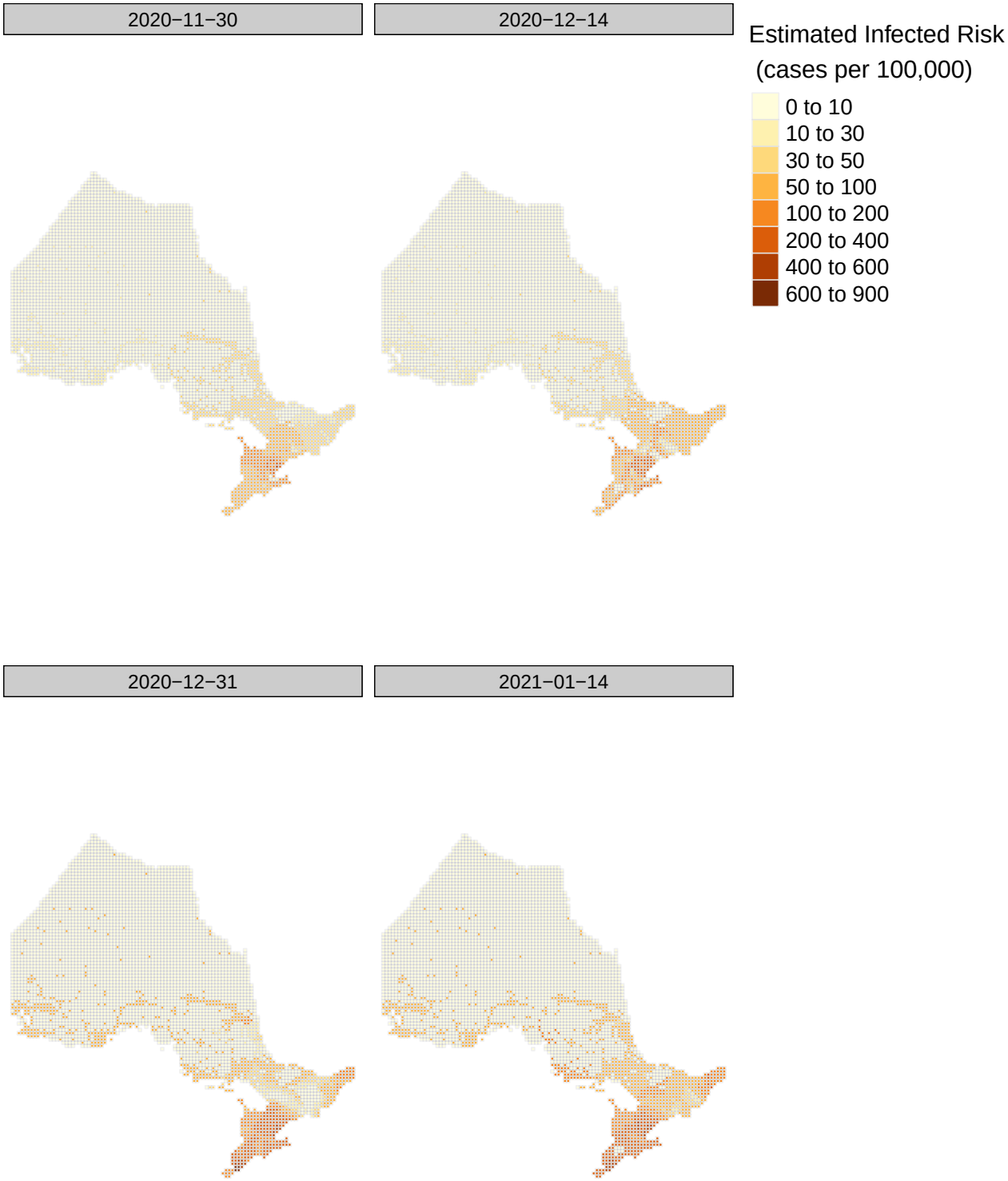
Figure 5.2: Evolution of Infected Risk at the Public Health Unit Level (ATA Poisson Kriging)

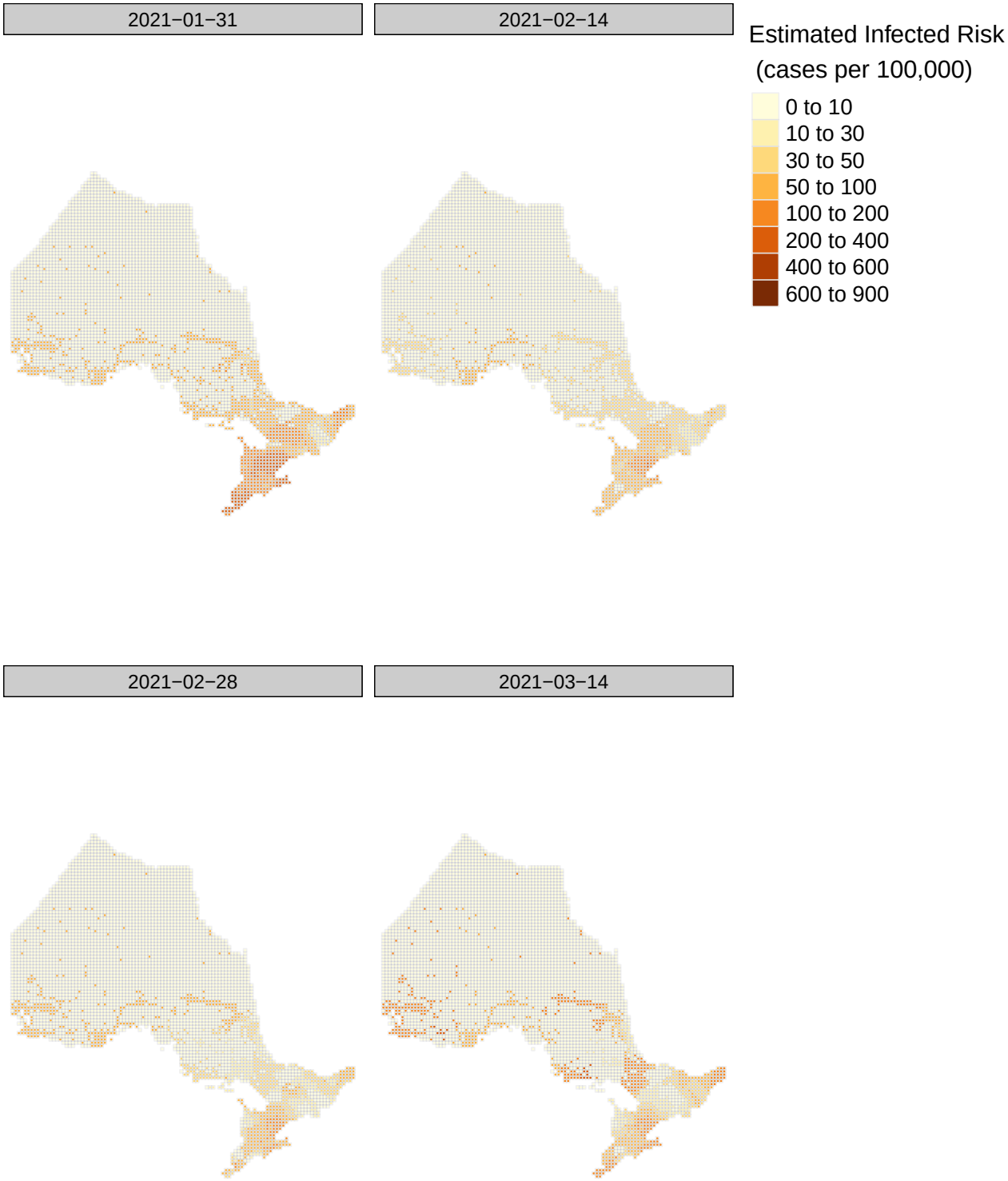

2020−03−31
2020−04−14
2020−04−30
2020−05−14
Estimated Infected Risk
(cases per 100,000)
0 to 10
10 to 30
30 to 50
50 to 100
100 to 200
200 to 400
400 to 600
600 to 900


2020−05−31
2020−06−14
2020−06−30
2020−07−14
Estimated Infected Risk
(cases per 100,000)
0 to 10
10 to 30
30 to 50
50 to 100
100 to 200
200 to 400
400 to 600
600 to 900


2020−07−31
2020−08−14
2020−08−31
2020−09−14
Estimated Infected Risk
(cases per 100,000)
0 to 10
10 to 30
30 to 50
50 to 100
100 to 200
200 to 400
400 to 600
600 to 900


2020−09−30
2020−10−14
2020−10−31
2020−11−14
Estimated Infected Risk
(cases per 100,000)
0 to 10
10 to 30
30 to 50
50 to 100
100 to 200
200 to 400
400 to 600
600 to 900


2020−11−30
2020−12−14
2020−12−31
2021−01−14
Estimated Infected Risk
(cases per 100,000)
0 to 10
10 to 30
30 to 50
50 to 100
100 to 200
200 to 400
400 to 600
600 to 900


2021−01−31
2021−02−14
2021−02−28
2021−03−14
Estimated Infected Risk
(cases per 100,000)
0 to 10
10 to 30
30 to 50
50 to 100
100 to 200
200 to 400
400 to 600
600 to 900

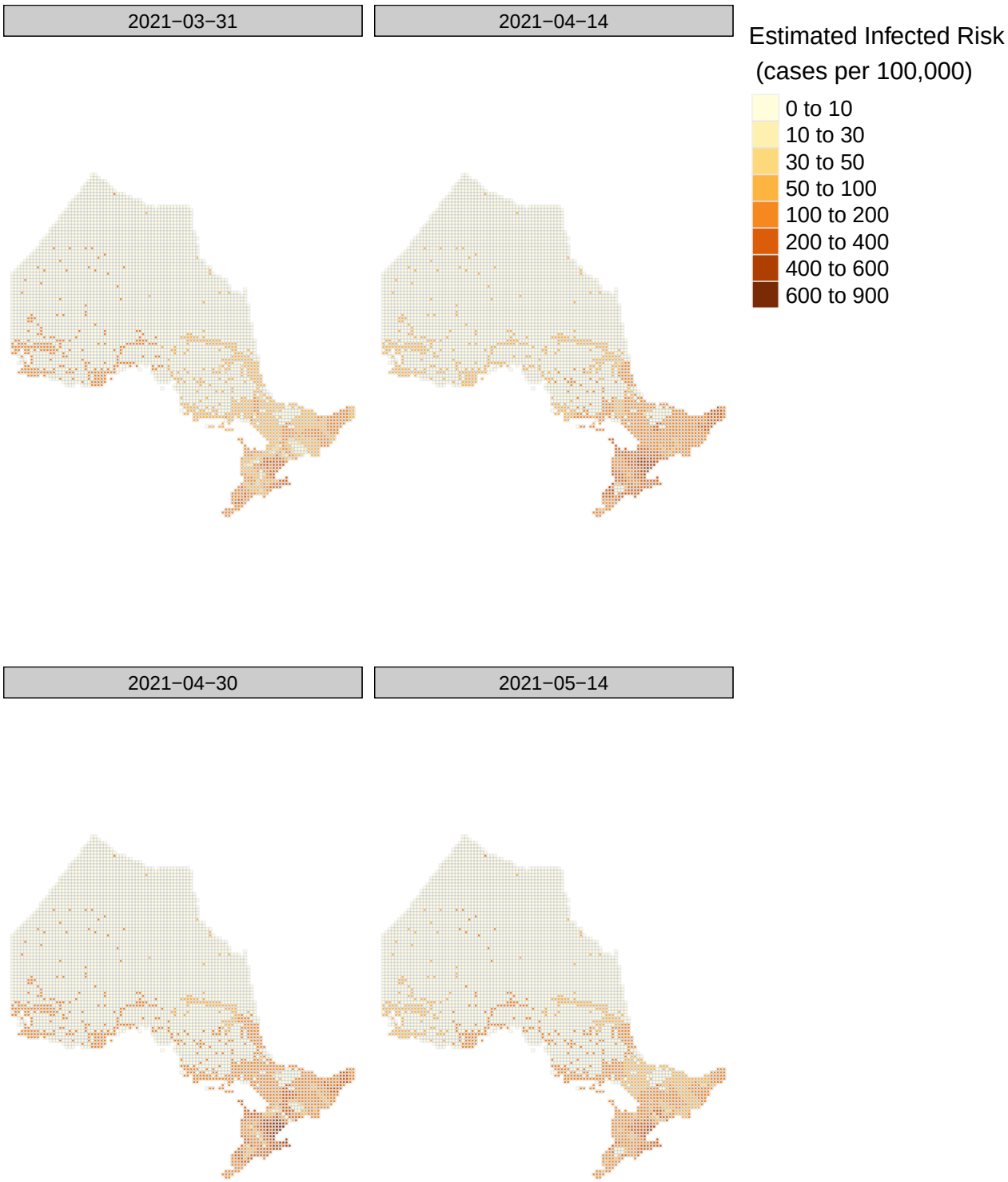


Figure 5.3: Evolution of Infected Risk at the Public Health Unit Level (ATP Poisson Kriging)

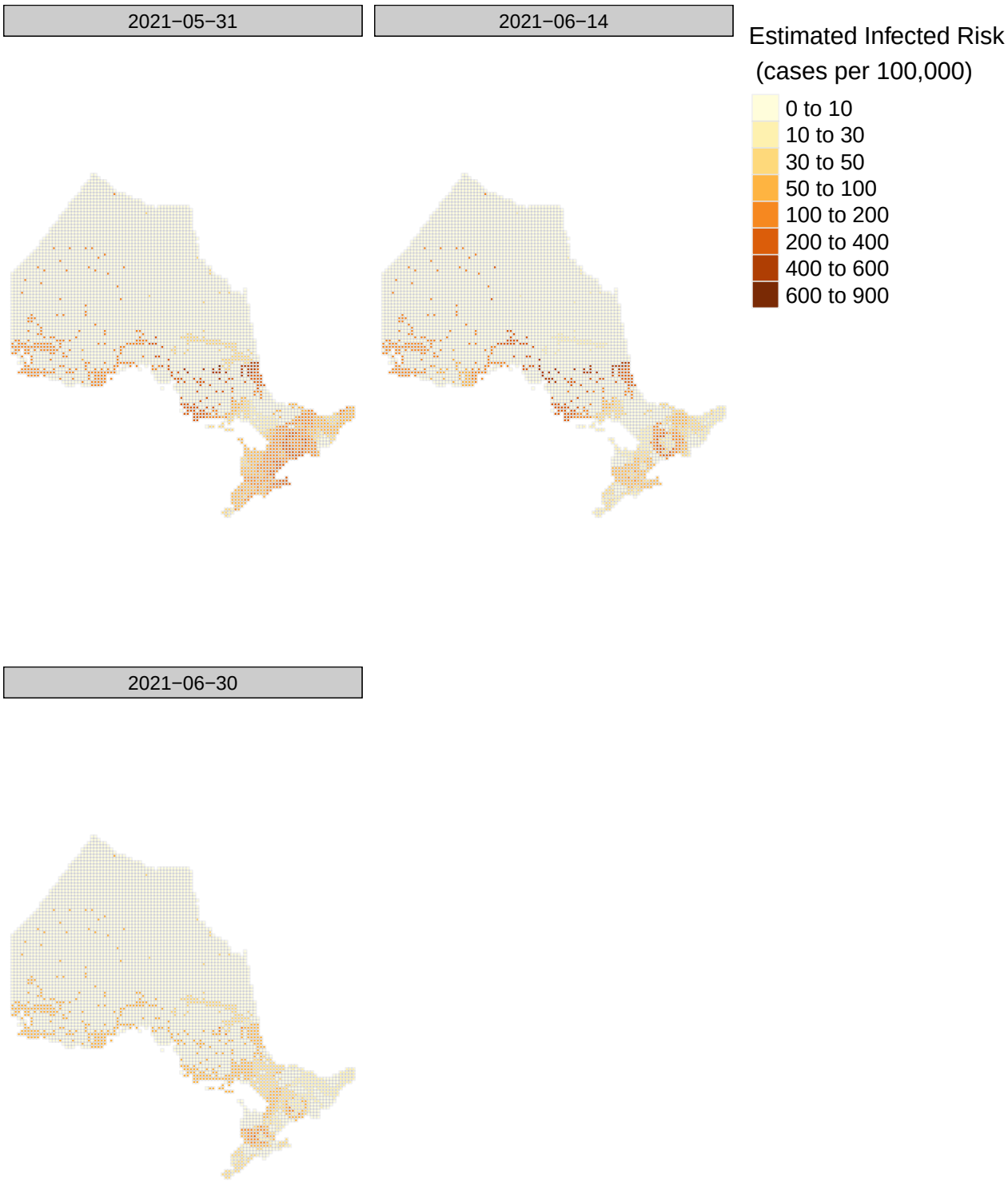


Figure 5.4: Evolution of Infected Risk at the Public Health Unit Level (ATP Poisson Kriging)

We examine the estimation performance of Area-to-area (ATA) Poisson Kriging Model in use of Mean Absolute Error of Estimation (MAEE).

$$MAEE = \frac{1}{I}\sum_{i=1}^{34}(\|\hat{r}_{PK}(v_\alpha) - z(\mathbf{v}_\alpha)\|)$$

Table 5.1: Estimation Performance with ATA Poisson Kriging Model

Date	MAEE	Date	MAEE
2020-03-31	9.1583	2020-04-14	27.5731
2020-04-30	18.1985	2020-05-14	9.9172
2020-05-31	14.4495	2020-06-14	14.0699
2020-06-30	10.9806	2020-07-14	4.9548
2020-07-31	9.8015	2020-08-14	9.1887
2020-08-31	5.9091	2020-09-14	6.1367
2020-09-30	12.4405	2020-10-14	15.2833
2020-10-31	23.9789	2020-11-14	31.6626
2020-11-30	45.2185	2020-12-14	51.2377
2020-12-31	81.5218	2021-01-14	75.8301
2021-01-31	56.2901	2021-02-14	27.8829
2021-02-28	37.8338	2021-03-14	54.8017
2021-03-31	73.2944	2021-04-14	76.0162
2021-04-30	72.6466	2021-05-14	78.9172
2021-05-31	95.2153	2021-06-31	73.0008
2021-06-31	49.7580	-	-

It can be seen that generally the estimation performance is good. It means the range in point support semivariograms from ATP Poisson Kriging Model can show the spatial clustering feature of COVID-19 spreading reliably. The following table shows different ranges in each time:

Table 5.2: Ranges in Point Support Semivariograms

Date	Range	Date	Range
2020-03-31	75.4km	2020-04-14	114.9km
2020-04-30	28.0km	2020-05-14	123.7km
2020-05-31	240.0km	2020-06-14	108.4km
2020-06-30	126.8km	2020-07-14	80.5km
2020-07-31	116.7km	2020-08-14	133.8km
2020-08-31	102.0km	2020-09-14	53.0km
2020-09-30	87.6km	2020-10-14	104.8km
2020-10-31	104.4km	2020-11-14	72.8km
2020-11-30	44.0km	2020-12-14	73.6km
2020-12-31	151.0km	2021-01-14	136km
2021-01-31	106.0km	2020-02-14	101.6km
2021-02-28	151km	2021-03-14	247.5km
2021-03-31	76km	2021-04-14	121km
2021-04-30	84.8km	2021-05-14	84.4km
2021-05-31	355.9km	2021-06-14	272.2km
2021-06-30	74.3km	-	-

In the case of semivariogram, for sample points with close distances, the difference in values between points tends to be small. In other words, the semivariance is small. But when sample point distances are farther away, they are less likely to be similar. This means that the semivariance becomes large. As the distance increases away from sample points, there is no longer a relationship between the sample points. Their variance begins to flatten out, and sample values are not related to one another. The range represents the distance at which the model first flattens out. Therefore, range value can represent the spatial clustering feature. It can be seen that the largest range value is 355km, which means the infected cases are clustered spatially within 355km at most. Public health office can pay more attention to monitor the infected cases around each hotspot under 355km.

Chapter 6

Conclusions

After the use of Bayesian methods and non-parametric geostatistical ATA(ATP) Poisson Models, there are some conclusions made as follows:

1) Comparing ATA Poisson Kriging with Bayesian Spatial-temporal methods applied in the COVID-19 data, it can be seen that ATA Poisson Kriging provides a smoother surface and ATP Poisson Kriging exhibits how the virus spreads more clearly. However, the performance of Kriging methods on prediction is not as good as Bayesian Spatial-temporal methods. The Kriging method tends to get a much smoother surface, leading to overestimated predictions at non-hot spot regions. Bayesian Spatial-temporal methods can offer a more precise result.

2) ATP Poisson Kriging Model is a good way to downscale areal risk maps into point risk maps. The range values from point support semivariograms can show the spatial clustering feature well. The largest range value is 355 km. It can be a good reference for government to monitor the infected cases around each hotspot.

3) The dataset used in this thesis is current biweekly infected cases, not cumulative ones. Intervention influence can be shown more clearly with current datasets using Bayesian Spatial-temporal Models. Cumulative datasets have been tried, but the influence of the policies is not clear.

4) The following are significant: the level 1, level 3 and level 4 in Non-essential Services are sit-down dining (with varying upper limit and some limitations, like wearing facial mask) at cafes and restaurants, restaurants/bars/cafes (except takeout/delivery) closed and retail services restricted and personal services closed, Non-essential businesses closed, respectively; Levels 2 and 3 in Indoor Gathering Intervention are the upper limit was less than 10 and 5, respectively.

5) The significant COVID-19 auxiliary variables are Average Household Size and the Proportion of Self-employed Workers. Increases in these variables increases the infection rate.

6) The risk was nearly 0 in March 2020. The second wave started from November 2020, and then decreased from February, 2021. However, the third wave began again in April, 2021. At the beginning, the risks in the Northern area are relatively small, but they increase after January, 2021, when the second wave started. Since the third wave, the infected current cases in Northern Ontario are more than in Southern Ontario.

7) The Bayesian Spatial-temporal Models 9 and 10 which incorporate interactions between space and time were used for predictions on infection rates two weeks and four weeks ahead. The performance for two weeks ahead was better for both Models. Model 9 includes interaction between unstructured spatial effect and structured temporal effect. Model 10 includes interaction between structured spatial effect and unstructured temporal effect.

8) The input to the interactive website we developed is the date and the choice of the model, Bayesian Spatial-temporal or Kriging. Then the output consists of the estimated risk maps. It is clear to compare the estimated risk maps with the observed maps.

9) In Bayesian Spatial-temporal Models, there is not only one way to define the spatial relationship with BYM specification, Leroux CAR specification can be used as well. How to define the spatial association can be a subject for the future research.

Appendix A

Calculation of the optimal weights is an example of constrained optimization: the objective function is the mean square error of prediction (MSEP) in expression (5.6), and the constraint is the restriction on the weights in equation (5.4).

The Lagrange method of multipliers allows the problem to be solved by reducing it to an unconstrained one by means of a new objective function called the Lagrangian function (Hillier and Lieberman, 1995).

Definition 1. *Let* $E[(\hat{r}_{PK}(\mathbf{u}_\alpha) - R(\mathbf{u}_\alpha))^2]$ *be the mean square error of prediction in (5.6) for an estimate at* $\mathbf{u}_\alpha$, $\lambda_i(\mathbf{u}_\alpha)$ *be the weights in (5.4), and let* $\mu(\mathbf{u}_\alpha)$ *be a Lagrange multiplier. Then:*

$$L(\lambda_1(\mathbf{u}_\alpha), \lambda_2(\mathbf{u}_\alpha), ..., \lambda_K(\mathbf{u}_\alpha); \mu) = E[(\hat{r}_{PK}(\mathbf{u}_\alpha) - R(\mathbf{u}_\alpha))^2] + 2\mu(\mathbf{u}_\alpha)(\sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) - 1)$$

is the Lagrangian function for Poisson kriging.

Theorem 1. *Let* $\lambda_i(\mathbf{u}_\alpha)$ *be the weights in (5.4), and* $\mu(\mathbf{u}_\alpha)$ *be the Lagrange multiplier in Definition 1. Function* $L(\lambda_1(\mathbf{u}_\alpha), \lambda_2(\mathbf{u}_\alpha), ..., \lambda_K(\mathbf{u}_\alpha); \mu)$ *is the new objective function. Then the weights that produce the unique minimum mean square error of prediction are the solution to:*

$$\sum_{j=1}^{K} \lambda_j(\mathbf{u}_\alpha) C_R(\mathbf{u}_i, \mathbf{u}_j) + \lambda_i(\mathbf{u}_\alpha) \frac{m}{n(\mathbf{u}_i)} + \mu(\mathbf{u}_\alpha) = C_R(\mathbf{u}_i, \mathbf{u}_\alpha) \quad for \quad i = 1, 2, ..., K,$$
$$\sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) = 1$$

or simplify it as:

$$\sum_{j=1}^{K} \lambda_j(\mathbf{u}_\alpha)[C_R(\mathbf{u}_i, \mathbf{u}_j) + \delta_{ij} \frac{m*}{n(\mathbf{u}_i)}] + \mu(\mathbf{u}_\alpha) = C_R(\mathbf{u}_i, \mathbf{u}_\alpha) \quad for \quad i = 1, 2, ..., K,$$
$$\sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) = 1$$

where $\delta_{ij} = 1$ *if* $\mathbf{u}_i = \mathbf{u}_j$ *and 0 otherwise, and* m^* *is the population-weighted mean of the N rates.*

Proof 1. *Replacing the mean square error of prediction in* $L(\lambda_1(\mathbf{u}_\alpha), \lambda_2(\mathbf{u}_\alpha), ..., \lambda_K(\mathbf{u}_\alpha); \mu)$,

$$L(\lambda_1(\mathbf{u}_\alpha), \lambda_2(\mathbf{u}_\alpha), ..., \lambda_K(\mathbf{u}_\alpha); \mu) = \sigma_R^2 + \sum_{i=1}^{K} \frac{\lambda_i^2(\mathbf{u}_\alpha)}{n(\mathbf{u}_i)} m + \sum_{i=1}^{K}\sum_{j=1}^{K} \lambda_i(\mathbf{u}_\alpha)\lambda_j(\mathbf{u}_\alpha) C_R(\mathbf{u}_i, \mathbf{u}_j) - 2\sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) C_R(\mathbf{u}_i, \mathbf{u}_\alpha) + 2\mu(\mathbf{u}_\alpha)(\sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) - 1)$$

One can see that the Lagrangian function is a quadratic expression in the unknown weights and the constraint on the weights is linear. In such a case, the necessary and sufficient condition to have a unique global minimum is that all second derivatives of the Lagrangian function with respect to the weights are larger than zero. These second derivative functions will be calculated as:

$$\frac{\partial L(\lambda_1(\mathbf{u}_\alpha), \lambda_2(\mathbf{u}_\alpha), ..., \lambda_K(\mathbf{u}_\alpha); \mu)}{\partial \lambda_i(\mathbf{u}_\alpha)} = \frac{2m}{n(\mathbf{u}_i)} \lambda_i(\mathbf{u}_\alpha) + 2\sum_{j=1}^{K} \lambda_j(\mathbf{u}_\alpha) C_R(\mathbf{u}_i, \mathbf{u}_j) - 2C_R(\mathbf{u}_i, \mathbf{u}_\alpha) + 2\mu(\mathbf{u}_\alpha), \quad for \quad i = 1, ..., K$$

$$\frac{\partial^2 L(\lambda_1(\mathbf{u}_\alpha), \lambda_2(\mathbf{u}_\alpha), ..., \lambda_K(\mathbf{u}_\alpha); \mu)}{\partial^2 \lambda_i(\mathbf{u}_\alpha)} = \frac{2m}{n(\mathbf{u}_i)} + 2C_R(\mathbf{u}_i, \mathbf{u}_i) \quad for \quad i = 1, ..., K$$

The second derivative function must be larger than or equal to zero by definition of $C_R(\mathbf{u}_i, \mathbf{u}_j)$. *So the minimum mean square error of prediction is given by those weights that make zero all first derivatives of the Lagrangian function:*

$$\sum_{j=1}^{K} \lambda_j(\mathbf{u}_\alpha) C_R(\mathbf{u}_i, \mathbf{u}_j) + \lambda_i(\mathbf{u}_\alpha) \frac{m}{n(\mathbf{u}_i)} + \mu(\mathbf{u}_\alpha) = C_R(\mathbf{u}_i, \mathbf{u}_\alpha) \quad for \quad i = 1, 2, ..., K,$$

$$\sum_{i=1}^{K} \lambda_i(\mathbf{u}_\alpha) = 1$$

Assumption 1. *The random function honors the intrinsic hypothesis over the sampling domain, which implies that*

$$E[R(\mathbf{x})] = m$$

$$Var[R(\mathbf{x}) - R(\mathbf{x} + \mathbf{h})] = 2\gamma(\mathbf{h})$$

where $\gamma(\cdot)$ *is the semivariogram of the random variable.*

Lemma 1. *If* R *is an intrinsic random function. Then*

$$2\gamma(\mathbf{h}) = Var[R(\mathbf{x})] + Var[R(\mathbf{x} + \mathbf{h})] - 2Cov(\mathbf{x}, \mathbf{x} + \mathbf{h}),$$

where $\mathbf{x}$ *and* $\mathbf{x} + \mathbf{h}$ *are any two locations in the sample domain.*

Proof 2. *Because the mean is a constant, from the definition of variance,*

$$\begin{aligned} Var[R(\mathbf{x}) - R(\mathbf{x}+\mathbf{h})] &= E[\{R(\mathbf{x}) - R(\mathbf{x}+\mathbf{h})\}^2] - (E[R(\mathbf{x}) - R(\mathbf{x}+\mathbf{h})])^2 \\ &= E[\{R(\mathbf{x}) - R(\mathbf{x}+\mathbf{h})\}^2] \end{aligned}$$

By its definition in Lemma1,

$$2\gamma(\mathbf{h}) = E[\{R(\mathbf{x}) - R(\mathbf{x}+\mathbf{h})\}^2]$$

Expanding and then adding and subtracting the term $2m^2$*,*

$$\begin{aligned} 2\gamma(\mathbf{h}) &= E[R^2(\mathbf{x})] - m^2 + E[R^2(\mathbf{x}+\mathbf{h})] - m^2 - 2E[R(\mathbf{x})R(\mathbf{x}+\mathbf{h})] + 2m^2 \\ &= Var[R(\mathbf{x})] + Var[R(\mathbf{x}+\mathbf{h})] - 2Cov(\mathbf{x}, \mathbf{x}+\mathbf{h}) \end{aligned}$$

Corollary 1. *If* R *is a second order stationary random function, then*

$$\gamma(\mathbf{h}) = Cov(0) - Cov(\mathbf{h})$$

Proof 3. *By Lemma 1,*

$$2\gamma(\mathbf{h}) = Var[R(\mathbf{x})] + Var[R(\mathbf{x}+\mathbf{h})] - 2Cov(\mathbf{x}, \mathbf{x}+\mathbf{h}),$$

which is equivalent to

$$2\gamma(\mathbf{h}) = Cov(\mathbf{x}, \mathbf{x}) + Cov(\mathbf{x}+\mathbf{h}, \mathbf{x}+\mathbf{h}) - xCov(\mathbf{x}, \mathbf{x}+\mathbf{h}).$$

If the random function is second order stationary, the covariance is independent of location and depends only on the distance $\mathbf{h}$ *between two variates:*

$$2\gamma(\mathbf{h}) = 2Cov(0) - 2Cov(\mathbf{h})$$

Appendix B

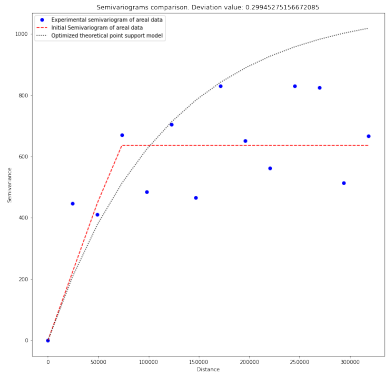


Figure 1: Semivariogram Model from 2020-03-31 to 2020-04-14

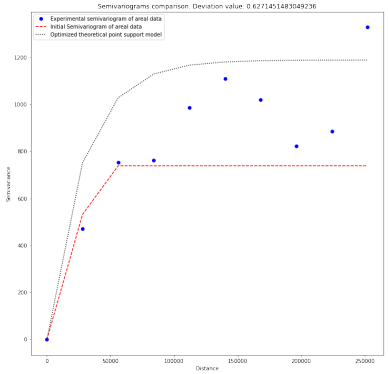


Figure 2: Semivariogram Model from 2020-04-14 to 2020-04-30

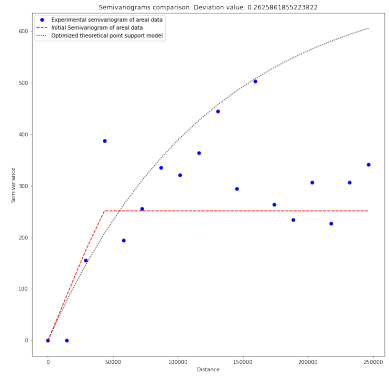


Figure 3: Semivariogram Model from 2020-04-30 to 2020-05-14

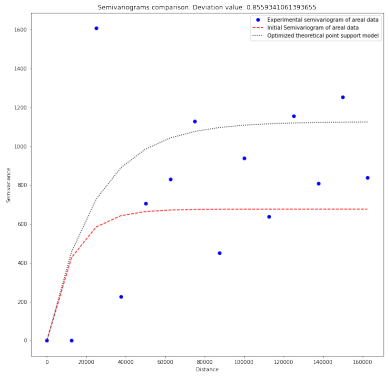


Figure 4: Semivariogram Model from 2020-05-14 to 2020-05-30

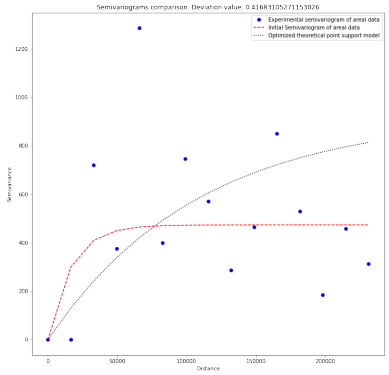


Figure 5: Semivariogram Model from 2020-05-30 to 2020-06-14

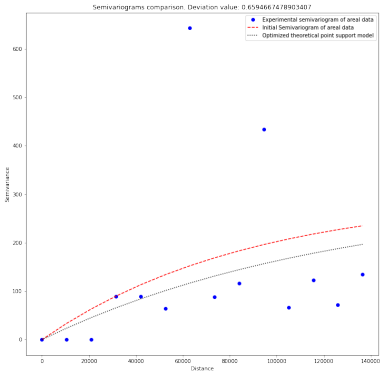


Figure 6: Semivariogram Model from 2020-06-14 to 2020-06-30

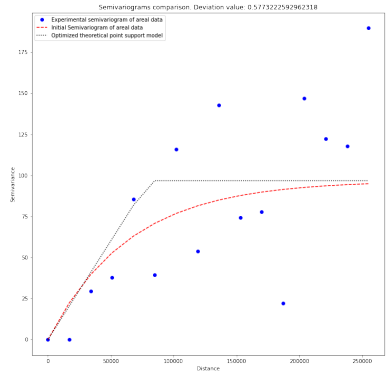


Figure 7: Semivariogram Model from 2020-06-30 to 2020-07-14

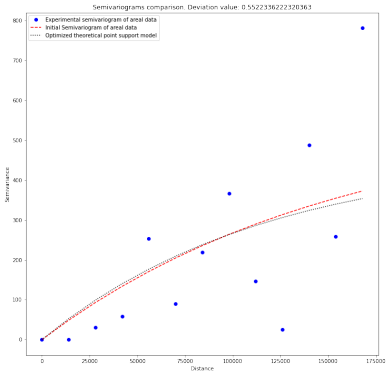


Figure 8: Semivariogram Model from 2020-07-14 to 2020-07-31

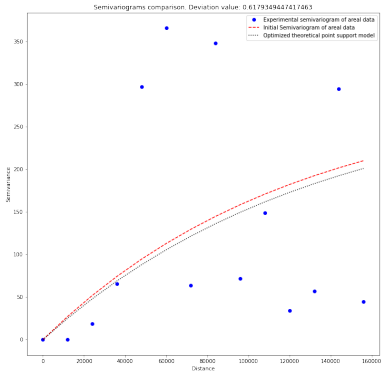


Figure 9: Semivariogram Model from 2020-07-31 to 2020-08-14

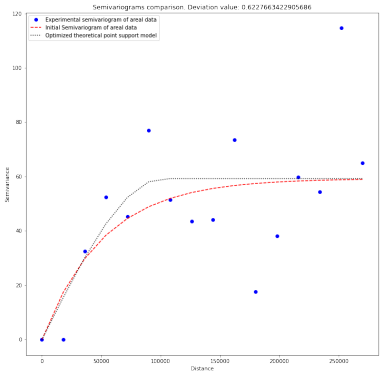


Figure 10: Semivariogram Model from 2020-08-14 to 2020-08-31

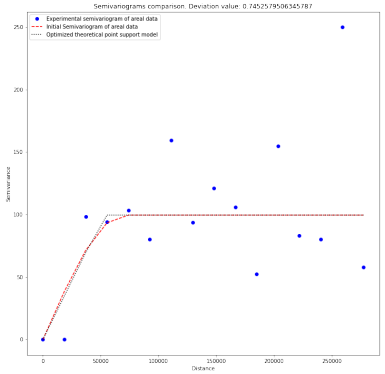


Figure 11: Semivariogram Model from 2020-08-31 to 2020-09-14

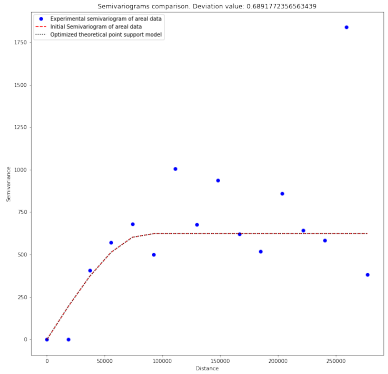


Figure 12: Semivariogram Model from 2020-09-14 to 2020-09-30

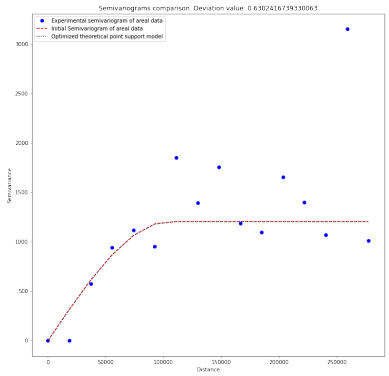


Figure 13: Semivariogram Model from 2020-09-30 to 2020-10-14

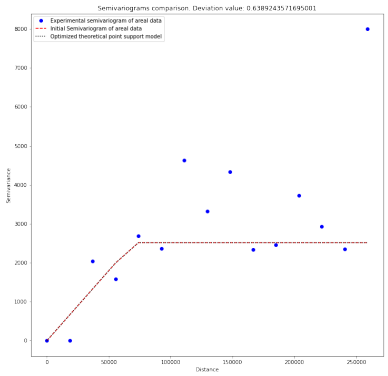


Figure 14: Semivariogram Model from 2020-10-14 to 2020-10-31

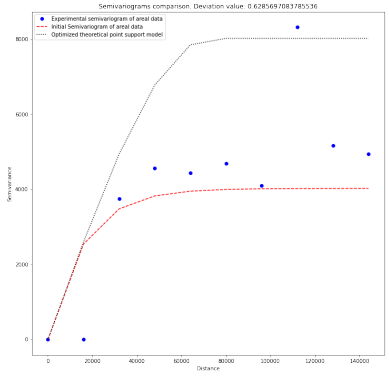


Figure 15: Semivariogram Model from 2020-10-31 to 2020-11-14

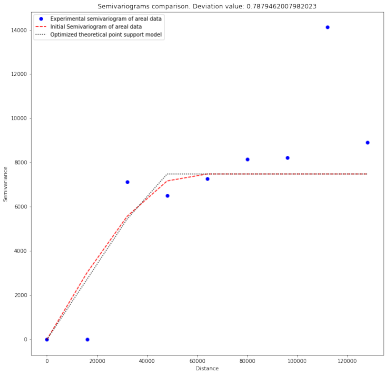


Figure 16: Semivariogram Model from 2020-11-14 to 2020-11-30

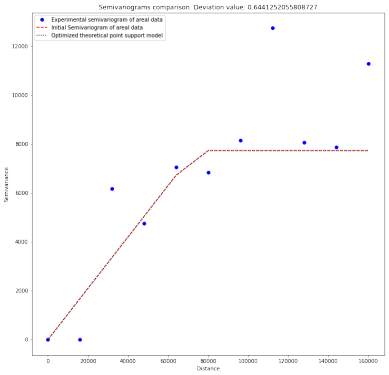


Figure 17: Semivariogram Model from 2020-11-30 to 2020-12-14

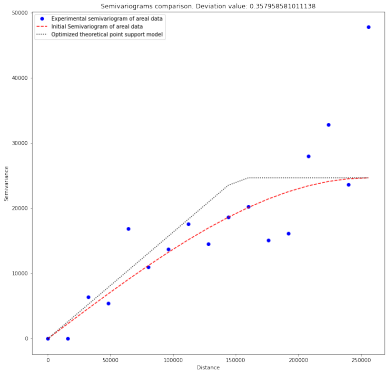


Figure 18: Semivariogram Model from 2020-12-14 to 2020-12-31

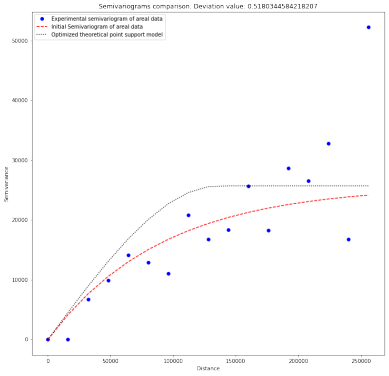


Figure 19: Semivariogram Model from 2021-12-31 to 2021-01-14

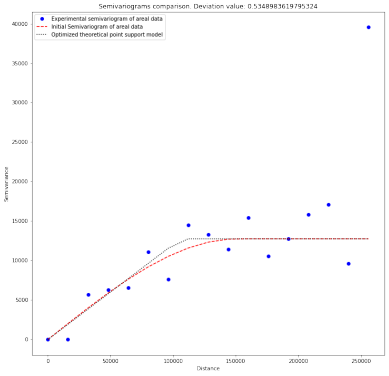


Figure 20: Semivariogram Model from 2021-01-14 to 2021-01-31

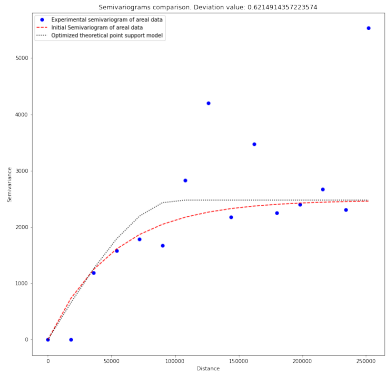


Figure 21: Semivariogram Model from 2021-01-31 to 2021-02-14

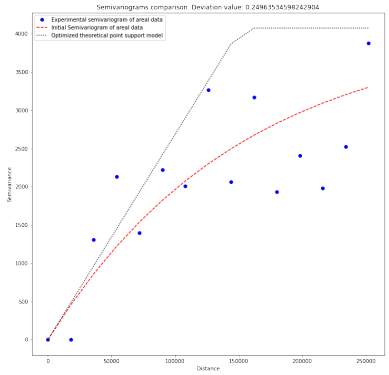


Figure 22: Semivariogram Model from 2021-02-14 to 2021-02-28

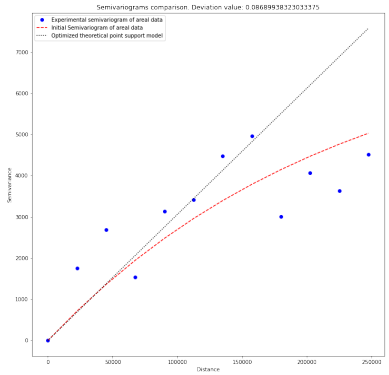


Figure 23: Semivariogram Model from 2021-02-28 to 2021-03-14

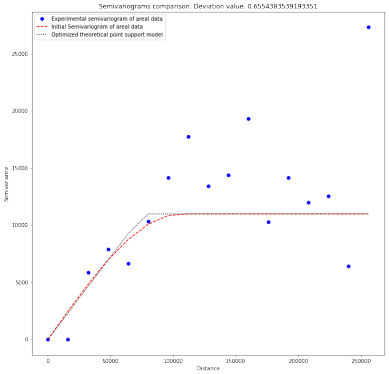


Figure 24: Semivariogram Model from 2021-03-14 to 2021-03-31

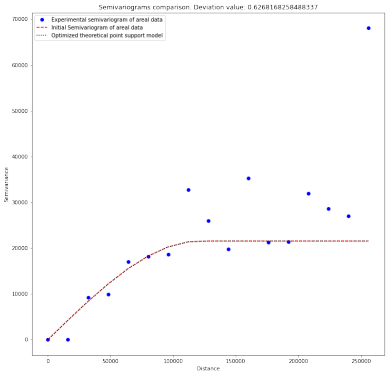


Figure 25: Semivariogram Model from 2021-03-31 to 2021-04-14

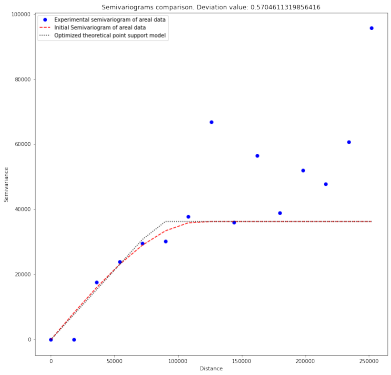


Figure 26: Semivariogram Model from 2021-04-14 to 2021-04-30

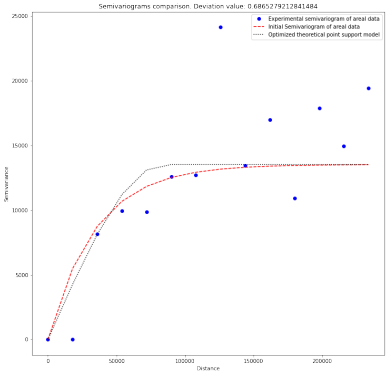


Figure 27: Semivariogram Model from 2021-04-30 to 2021-05-14

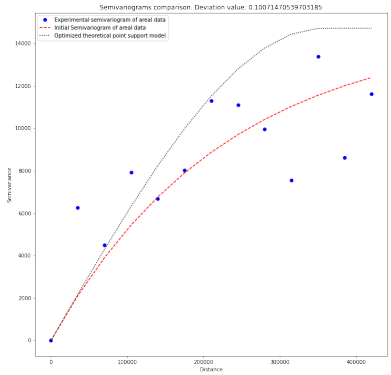


Figure 28: Semivariogram Model from 2021-05-14 to 2021-05-31

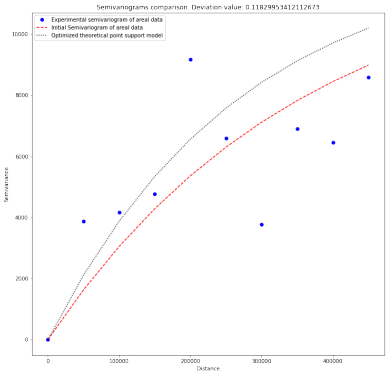


Figure 29: Semivariogram Model from 2021-05-31 to 2021-06-14

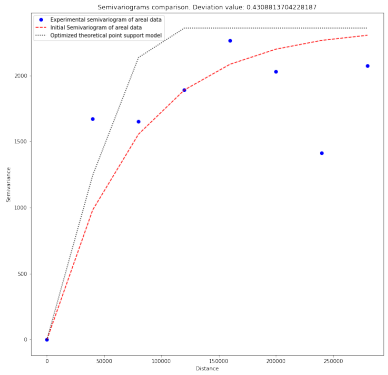


Figure 30: Semivariogram Model from 2021-06-14 to 2021-06-30

Bibliography

[1] Goovaerts, P. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. International Journal of Health Geographics. 2006, 5(52): 1-31.

[2] Kelsall J, Wakefield J: Modeling spatial variation in disease risk: a geostatistical approach. Journal of the American Statistical Association. 2002, 97 (459): 692-701.

[3] Choi K-M, Serre ML, Christakos G: Efficient mapping of California mortality fields at different spatial scales. Journal of Exposure Analysis and Environmental Epidemiology. 2003, 13: 120-133.

[4] Goovaerts, P. Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units. Mathmatical Geosciences. 2008, 40: 101-128.

[5] Kyriakidis P: A geostatistical framework for area-to-point spatial interpolation. Geographical Analysis. 2004, 36 (3): 259-289.

[6] Goovaerts P: Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. International Journal of Health Geographics. 2005, 4: 31.

[7] Goovaerts P: Geostatistical analysis of disease data: visualization and propagation of spatial uncertainty in cancer mortality risk using Poisson kriging and p-field simulation. International Journal of Health Geographics. 2006, 5: 7.

[8] Best N, Richardson S, Thomson A: A comparison of Bayesian spatial models for disease mapping. Statistical Methods in Medical Research. 2005, 14: 35-59.

[9] Besag J, York J, Mollie A: Bayesian image restoration with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics. 1991, 43: 1-59.

[10] Johnson GD: Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modeling. International Journal of Health Geographics. 2004, 3: 29.

[11] Best NG, Arnold RA, Thomas A, Waller LA, Conlon EM: Bayesian models for spatially correlated disease and exposure data. Bayesian Statistics 6. Edited by: Bernardo JM, Berger JO, Dawid AP, Smith AFM. 1999, Oxford, UK, Oxford University Press, 131-156.

[12] Monestiez P, Dubroca L, Bonnin E, Durbec JP, Guinet C: Comparison of model based geostatistical methods in ecology: application to fin whale spatial distribution in northwestern Mediterranean Sea. Geostatistics Banff 2004. Edited by: Leuangthong O, Deutsch CV. 2005, Dordrecht, The Netherlands, Kluwer Academic Publishers, 2: 777-786.

[13] Monestiez P, Dubroca L, Bonnin E, Durbec JP, Guinet C: Geostatistical modelling of spatial distribution of Balenoptera physalus in the northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. Ecological Modeling. 2006, 193: 615-628.

[14] Grauman DJ, Tarone RE, Devesa SS, Fraumeni JF: Alternate ranging methods for cancer mortality maps. Journal of the National Cancer Institute. 2000, 92 (7): 534-543.

[15] Brewer CA, Pickle L: Evaluation of methods for classifying epidemiological data on choropleth maps in series. Annals of the Association of American Geographers. 2002, 92 (4): 662-681.

[16] Waller LA, Gotway CA: Applied Spatial Statistics for Public Health Data. 2004, New Jersey: John Wiley and Sons.

[17] Greenland, S.: Bayesian perspectives for epidemiological research: I.foundations and basic methods. International journal of Epidemiology.2006, 35: 765-775.

[18] Jewell, C., Kypraios, T., Neal, P., Roberts, G.: Bayesian analysis for emerging infectious diseases. Bayesian Analysis. 2009, 4 (3): 465-496.

[19] Rue, H., Held, L., 2005. Gaussian Markov Random Fields. Theory and Applications. Chapman & Hall.

[20] Rue, H., Martino, S., Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society Series B. 2009, 71 (2): 1-35.

[21] Ruiz-C´ardenas, R., Krainski, E., Rue, H.: Direct fitting of dynamic models using integrated nested Laplace approximations inla. Computational Statistics & Data Analysis .2012, 56 (6): 1808-1828.

[22] Schr¨odle, B., Held, L., Riebler, A., Danuser, J.: Using integrated nested laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study. Journal of the Royal Statistical Society Series C (Applied Statistics). 2011, 60 (2): 261-279.

[23] Blangiardo M, Cameletti M, Baio G, et al. Spatial and spatio-temporal models with R-INLA[J]. Spatial and spatio-temporal epidemiology. 2013, 4: 33-49.

[24] Pierre Goovaerts, Samson Gebreab: How does Poisson kriging compare to the popular BYM model for mapping disease risks?. International Journal of Health Geographics. 2008, 1(7): 1-25.

[25] Kerry, R., Goovaerts, P., Smit, I. P. J., Ingram, B. R.: A comparison of multiple indicator kriging and area-to-point Poisson kriging for mapping patterns of herbivore species abundance in Kruger national park, South Africa. International Journal of Geographical Information Science. 2013, 27: 47-67.

[26] Murphy, B., Müller, S., Yurchak, R.: GeoStat-frameworkPyKrige v1.5.1 (Version301 v1.5.1) [Computer software]. Zenodo. 2002

[27] Molinski, S.: Pyinterpolate: Spatial Interpolation in Python for point measurements and aggregated datasets. Journal of Open Source. 2021.

[28] Aswi A, Cramb S M, Moraga P, et al. Bayesian spatial and spatio-temporal approaches to modeling dengue fever: a systematic review[J]. Epidemiology & Infection, 2019, 147.

[29] Martínez-Bello D, Lopez-Quilez A and Alexander Torres P: Relative risk estimation of dengue disease at small spatial scale. International Journal of Health Geographics. 2017, 16(1): 1-15.

[30] Martínez-Bello D, Lopez-Quilez A, Prieto A T.: Spatiotemporal modeling of relative risk of dengue disease in Colombia. Stochastic environmental research and risk assessment, 2018, 32(6): 1587-1601.

[31] MacNab Y C.: On Gaussian Markov random fields and Bayesian disease mapping. Statistical Methods in Medical Research, 2011, 20(1): 49-68.

[32] Kafadar K. Choosing among two-dimensional smoothers in practice[J]. Computational statistics and data analysis, 1994, 18(4): 419-439.

[33] Marshall R J. Mapping disease and mortality rates using empirical Bayes estimators[J]. Journal of the Royal Statistical Society: Series C (Applied Statistics), 1991, 40(2): 283-294.