

# Integrated spatiotemporal surveillance system: Data, Analysis and Visualization

Lennon Li<sup>\*1,2</sup>, Reuben Pererita<sup>2</sup>, Steven Johnson<sup>1</sup> and Ian Johnson<sup>1,2</sup>

<sup>1</sup>Public Health Ontario, Toronto, ON, Canada; <sup>2</sup>University of Toronto, Toronto, ON, Canada

## Objective

To build an open source spatiotemporal system that integrates analysis and visualization for disease surveillance

## Introduction

Most surveillance methods in the literature focus on temporal aberration detections with data aggregated to certain geographical boundaries. SaTScan has been widely used for spatiotemporal aberration detection due to its user friendly software interface. However, the software is limited to spatial scan statistics and suffers from location imprecision and heterogeneity of population. R Surveillance has a collection of spatiotemporal methods that focus more on research instead of surveillance

## Methods

Based in Ontario, Canada, we used postal codes for determining the location of cases of reportable infectious diseases. The variation in geographic sizes and shapes of the case and census geographies created challenges for developing a uniform temporal spatial surveillance system, including:

- Linking case and population data due to misclassification errors,
- Distance based correlations due to irregularly shaped areas (e.g. FSA's), and
- Visualization bias due to variation in population density, e.g. large area with little population.

To overcome these challenges, we developed the Ontario Hybrid Information Map (OHIM) boundary, which is a combination of Public Health Unit boundaries (rural areas), census subdivisions (rural urban mixed) and regular grid cells (urban). The goal is to capture population details in urban areas without losing information in rural areas. OHIM has around 4600 geographies with more than half located in urban centers. Population distribution by gender and age group was calculated for each OHIM geography. A lookup file was also created to link all Ontario postal codes to OHIM geography.

To create baselines, historical data for influenza A were used to model the seasonality and calculate expected case count for each OHIM geography for each week. Standardized incident ratios (SIR) were calculated as exploratory statistics, and a spatiotemporal Besag-York-Mollie (BYM) model was used to calculate the probability that the risk is higher than a pre-specified threshold. Integrated Nested Laplace Approximation (R-INLA) was used in R to explore different types of spatiotemporal interactions and for fast Bayesian inference. The ability to apply the models was verified by examining previous outbreaks and seeking the opinion of staff that routinely perform surveillance on influenza.

To ensure the visualization integrates with the analysis, R package Shiny was used to build an interactive spatiotemporal visualization on OHIM boundary utilizing Open Street Map and html5. The application not only allows users to pan and zoom in space and time to explore the results and locate high risk areas, it also gives users the flexibility to change algorithm parameters for instant feedback. Figure 1 demonstrates a zoomed-in OHIM boundary with pointers signal for “high risk” area at user specified statistics exceeds a threshold (e.g.,  $SIR > 2$ ). Using the algorithms and visualization tools, surveillance experts pick the optimal time and place to be notified

based on historical data and therefore the optimal threshold, which will be verified by prospectively running the algorithms.

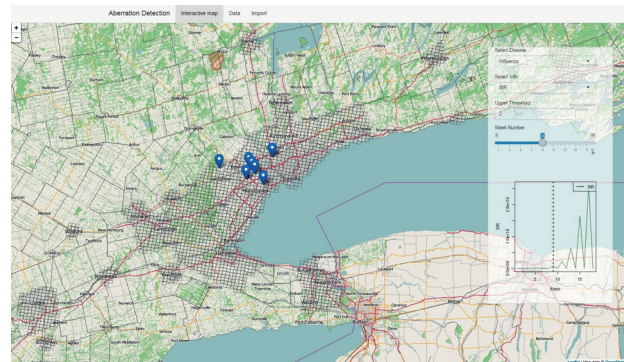
## Results

The OHIM boundaries build the foundation for efficient spatial modelling and visualization for public health surveillance in Ontario. Together with the integrated modelling and visualization system, staff are able to interactively optimize the aberration thresholds and identify potential outbreaks in real time. Staff reported preference of SIR due to its faster computations and easier interpretation.

One major challenge was scalability: the ability to handle high resolutions of spatiotemporal data. When the system was applied on 4600 polygons by 200 weeks, significant delays were encountered in both analysis and visualization. Difficulties in computational time, memory requirement and visualization interactivity created delays and freezing, thereby limited user experience. This problem was partially addressed by optimizing parameters for fast computations

## Conclusions

This work shows the “proof of concept” for an open source, customizable spatiotemporal surveillance system that overcomes existing data challenges in Ontario. However, more work is required to make this fully operational and efficient in production.



## Keywords

spatiotemporal; surveillance; interactive visualization; R; shiny

\*Lennon Li

E-mail: lennon.li@oahpp.ca

