Spatio-Temporal Clustering of Multi-Location Time Series to Model Seasonal Influenza Spread

Hootan Kamran, *Member, IEEE*, Dionne M. Aleman, *Member, IEEE*, Michael W. Carter, and Kieran M. Moore

***Abstract*—Although seasonal influenza disease spread is a spatio-temporal phenomenon, public surveillance systems aggregate data only spatially, and are rarely predictive. We develop a hierarchical clustering-based machine learning tool to anticipate flu spread patterns based on historical spatio-temporal flu activity, where we use historical influenza-related emergency department records as a proxy for flu prevalence. This analysis replaces conventional geographical hospital clustering with clusters based on both spatial and temporal distance between hospital flu peaks to generate a network illustrating whether flu spreads between pairs of clusters (direction) and how long that spread takes (magnitude). To overcome data sparsity, we take a model-free approach, treating hospital clusters as a fully-connected network, where arcs indicate flu transmission. We perform predictive analysis on the clusters' time series of flu ED visits to determine direction and magnitude of flu travel. Detection of recurrent spatio-temporal patterns may help policymakers and hospitals better prepare for outbreaks. We apply this tool to Ontario, Canada using a five-year historical dataset of daily flu-related ED visits, and find that in addition to expected flu spread between major cities/airport regions, we were able to illuminate previously unsuspected patterns of flu spread between non-major cities, providing new insights for public health officials. We showed that while a spatial clustering outperforms a temporal clustering in terms of the direction of the spread (81% spatial v. 71% temporal), the opposite is true in terms of the magnitude of the time lag (20% spatial v. 70% temporal).**

***Index Terms*—Clustering, influenza, spatial, spatiotemporal, surveillance, temporal.**

I. Introduction

INFLUENZA causes 3–5 million severe illnesses and 250,000–500,000 deaths globally each year [1]. It is among the 10 leading causes of death in the US [2] and Canada [3], with hundreds of deaths and hundreds of thousands of heath care utilization episodes in Ontario each year [4]. The influenza virus undergoes antigenic shifts (sudden changes) and drifts (gradual changes), which make permanent immunity through vaccination an elusive solution [5], with vaccine effectiveness averaging below 50% in the previous 10 years [6], and signify the importance of surveillance systems. Moreover, contagious diseases pose epidemic and pandemic risks with lower probability, but with much larger impact, than seasonal influenza [7], as recently observed during the COVID-19 pandemic.

To formulate national and provincial level health system responses to surveillance data, temporal data coming from multiple spatial sources must be aggregated for abstraction. Most current public surveillance systems gather temporal disease data from multiple sources, e.g., daily flu activity at multiple hospitals' emergency departments, then pool data coming from near hospitals into regional data, and analyze the pooled data individually for each region. Instead, we introduce a spatio-temporal aggregation scheme founded on the philosophy that space and time are interwoven [8], and spatio-temporally recorded data may, based on the objective, be better aggregated temporally or spatially. Our approach allows for the extraction of recurrent inter-cluster spatio-temporal patterns by modelling flu correlations as a network of inter-connected hospital clusters.

Surveillance systems collect, analyze, and interpret specific health data to detect disease spread patterns and, through intervention, minimize the future harm caused by the outbreaks [9]. Quantitative surveillance in the literature generally involves the analysis of disease time series. These time series are successive measurements of the disease prevalence in time. Pioneered by France in 1984 [10], computerized surveillance systems have since been implemented in many other developed countries to monitor the spread of communicable diseases.

The Influenza Division at the Center for Disease Control and Prevention in the US, for example, monitors the spread of flu on an ongoing basis and publishes the weekly FluView reports [11]. Similarly, the Public Health Agency of Canada (PHAC) prepares FluWatch, also a weekly influenza surveillance report that gathers data from a national network of participating labs, hospitals, doctor's offices and provincial and territorial ministries of health and provides an overview of flu activity for each province [12].

At the provincial level in Canada, public health agencies provide province-wide surveillance. For example, Ontario's

Manuscript received 26 August 2022; revised 4 December 2022; accepted 29 December 2022. Date of publication 6 January 2023; date of current version 5 April 2023. *(Corresponding author: Hootan Kamran.)*

Hootan Kamran, Dionne M. Aleman, and Michael W. Carter are with the Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, ON M5S 3G8, Canada (e-mail: hootan@mie.utoronto.ca; aleman@mie.utoronto.ca; carter@mie.utoronto.ca).

Kieran M. Moore is with the KFL&A Public Health Informatics, Kingston, ON K7M 1V5, Canada (e-mail: kieran.moore@kflaph.ca).

Digital Object Identifier 10.1109/JBHI.2023.3234818

ILIMapper [13] pools data from nearby hospitals regardless of their temporal similarities. Normal activity thresholds are then manually set by experience for each individual region regardless of inter-regional correlations, and activity levels beyond those thresholds are considered as outbreak in that region. However, flu activity at a particular time and location may depend not only on the previous measurements at that location, but also on the previous measurements at some other locations, such as neighbouring cities, or even non-neighbouring cities that are connected by population movement through, for example, airports [14] or other contact networks [15].

When reliable information about both the space and the time of certain events are available (e.g., positive flu cases recorded on certain days at certain hospitals), it is suboptimal to design a surveillance system that adopts only spatial distances. In fact, we show that spatial and temporal aggregation can each be more predictive than the other depending on the variable being predicted. The methodologies presented here, although designed for a flu surveillance system, are not limited to epidemeology, and may improve the analysis of other spatio-temporal systems such as in immunology [16], meteorology [17], and palaeoanthropology [18].

Traditionally, regression and time series models are used in the statistical surveillance literature, the latter being more suitable for explicit modelling of temporal correlations (e.g., autocorrelation [19], and seasonalities [20]). Such single-variate methods for time series analysis, however, are designed for individual time series, and cannot model the interactions between time series from multiple locations. There are many statistical surveillance systems [21], but we discuss here only the methods that are capable of multi-variate spatio-temporal analysis.

Shewhart control charts [22] use historical data to estimate statistical norms and raise a flag whenever a measurement of a variable falls outside its historical norms. The variable can be activity levels within a certain spatio-temporal frame. For example, if flu activity in a certain week at a certain location is higher than its historical norms, it is an indication of an outbreak in that particular spatio-temporal frame.

More sophisticated methods such as cumulative sums [23], exponentially-weighted moving average [24], and scan statistics [25] have been developed to allow for spatio-temporal aggregation in control charts. The basic assumption in such methods is that events happened at the same time in farther locations, and at the same location at earlier times are weighted less than the events that happened closer to the spatio-temporal reference point. But by aggregating the distribution of events into a single number to be used in control charts, these methods neglect temporal correlations between pairs of time series.

While standard control charts consider current values of the statistics independently from their previous values, an underlying stochastic structure may be assumed using Markovian models to estimate the probabilities of a system transitioning from one state to another state (e.g., non-outbreak to outbreak). Hidden Markov models (HMMs), for example, define a probability distribution over sequences of observations by invoking another sequence of latent (hidden) variables. The key assumptions are that the hidden variables have the Markov property, and that the observed variable at any time depends only on the hidden variable at that time. Unobserved process is in one of the two states: outbreak, and non-outbreak [26]. Prior probabilities can then be incorporated using a Bayesian approach [27].

An HMM with more than one time series is called a multivariate hidden Markov model (MHMM). Standard MHMMs assume that the multiple variates are independent [28]. Preconceived notions of spread patterns may be modelled with correlated latent variables between, for example, airports [14]. However, if those notions do not exist or must be further validified, a mechanism is needed to extract them from historical data.

We therefore introduce a spatio-temporal clustering method to replace the conventional geographic regions. Data clustering is a machine learning task that systematically partitions objects into a set of groups (called clusters) such that objects within a cluster are similar, and objects from different clusters are dissimilar [29]. The notion of similarity in data clustering is essentially linked to the notion of distance. Our model uses a set of predefined temporal distance measures, as well as the spatial distance measure, and finds the one that best serves each surveillance objective. By extracting recurrent inter-cluster spatio-temporal patterns in a network of hospital clusters, the introduced methodologies address the two limitations of current surveillance systems: (1) purely spatial or purely temporal aggregation of spatio-temporal data, and (2) implicit assumption of regional independence.

In current surveillance systems, regions are defined according to geography (local regions, cities, and countries), and surveillance control limits are calculated for each region using its own past data. There are two main contributions in this paper. As the first contribution, we consider a network of interacting regions instead of individual regions and find their pairwise distances based on historical data. The second contribution is the definition of pairwise distances for cluster-based aggregation. Calculating temporal distances, in addition to the conventional spatial distances, allows aggregation by temporal similarities between regions that are characteritically similar, but may be geographically distant. For example, we found that temporal distances are better predictors of directional analysis and that areas with major airports usually lead the other regions in outbreaks. For the magnitude of the travel time lag, however, spatial aggregation yields better predictive power.

The hidden Markov models are non-parametric models, where a hidden space is supposed to yield the probability of outbreak in the observable states. The temporal aspect has been considered as temporal autoregressive terms in the previous studies [30], [31], however, the temporal similarities between geographic regions are not addressed yet. We explicitly extract the empirical network with pairwise distances that are calculated spatio-temporally.

To address the first limitation, we extend the notion of distance between a pair of hospitals and include temporal distance between their corresponding time series in addition to their Euclidean distance in space. Despite the rich literature in time series distance measures and their applications in mining of sequenced data [32], little has been done to incorporate such measures in a

spatio-temporal surveillance system. Spatial distance D^{ℓ} is the Euclidean distance between hospitals; Temporal distance D^{f} is any member of a predefined set of common time series distance measures in the literature (e.g., correlation $D^{f:\mathrm{COR}}$).

To address the second limitation, i.e., the implicit assumption of regional independence, we consider the hospital clusters as a fully-connected network, with arcs indicating flu transmission between clusters as predicted by the time series. We consider two surveillance variables for each cluster pair, direction and magnitude of flu travel, to identify in which cluster(s) is a particular cluster's current outbreak preceded, and how long does it takes for the outbreak to travel from one cluster to another. We showed that depending on which surveillance variable is being studied, certain distance measures provide higher prediction capabilities than the others. Inter-cluster patterns so-obtained help policymakers gain insights into the spatial and temporal emergence patterns of outbreaks and produce accurate predictions of the activities of one cluster given the activities of the other clusters. To validate our model, we use daily hospital data from 2012–2016 in Ontario, Canada to make predictions about the most recent outbreaks in the data.

II. Data

Various variables may be chosen as the proxy for actual disease activity in quantitative surveillance. Newly available sources of data from such technologies as Google [33], and Twitter [34], [35], albeit helpful, cannot replace health utilization and laboratory data [36]. The main source of data in international [37], national [38], and provincial [13] surveillance systems is still obtained directly from health units.

We therefore use a flu syndromic hospital dataset. The data in this study has been collected from 129 hospitals in Ontario during 2004-2016, by our collaborator at KFL&A Public Health (Kingston, Ontario). We obtained Research Ethics Board (REB) approval from the University of Toronto. The protocol reference number is 28839, and the original approval date is 15 April 2013. It was renewed annually until study was concluded in 2017.

ln order to fulfill privacy requirements, the REB required us to de-link identifying information, encrypt personally identifiable information stored outside of a secure university server, and only allow personnel access to data on a need-to-know basis.

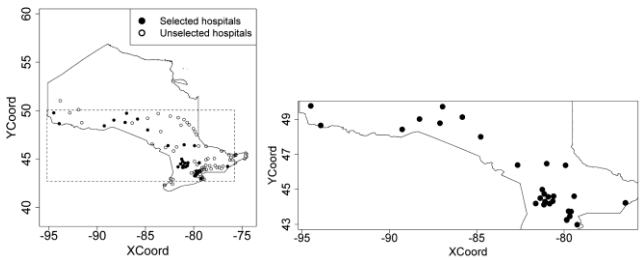
Each record in the dataset includes information about a visit to one of the participating hospitals (Table I). When a patient visits an emergency department, the triage personnel record the main reasons for the visit, including symptoms, diseases, mechanisms of injury, etc. This data is called the chief complaints (CCs) and is a popular data source in the literature of syndromic surveillances due to its availability and timeliness [39].

Symptoms are extracted from the chief complaints, and subsets of symptoms are classified into syndromes, which are then used for prognostic purposes. Two 0-1 valued flu-related syndromic variables, respiratory (RESP) and influenza-like illnesses (ILI), are included in this dataset, and we consider a record to be flu-positive if either of these variables are 1. The daily flu activity level at a hospital is estimated by the proportion of flu-positive visits to that hospital in that day.

TABLE I
FLU DATA FEATURES

Category	Feature	Details
Time	date	
Space	hospital name	
	hid	unique hospital identifier
	XCoord	hospital's longitude
	YCoord	hospital's latitude
Patient	symptomatic	chief complaints and CTAS[1] code
	syndromic	boolean (RESP or ILI)[2]

[1]Canadian triage & acuity scale is a symptom severity scale [40]
[2]Respiratory syndrome or influenza-like illnesses syndrome [41]



(a) All hospitals (b) Selected hospitals: Reported everyday for the last five years of data

Fig. 1. Participating hospitals in Ontario.

The original ED dataset was in a comma separated values (csv) file of size 1.46 GB with 18,503,184 rows, each representing a single visit's record. Regular expressions were used to handle formatting errors in the csv file, mainly missing and invalid entries. Invalid entries violated either data type constraints (e.g., an alphabetic character for age) or data range constraints (e.g., a negative integer for age).

Missing and invalid entries were allowed in some fields, e.g., symptoms. But, syndromic information features are critical because they are used to create and cluster the flu time series. Thus, rows with invalid syndromatic entries (less than 1% of all rows) were removed from the database. We also suppressed the text for symptoms of non-flu visits (89% of records) to reduce the file size. As a result, the csv file shrank by 35% to 0.94 GB. The number of participating hospitals grew from 3 hospitals in 2004 to 129 hospitals in 2016. We chose the subset of hospitals that reported every day for at least the most recent five years of data (2012-2016). Thirty hospitals met this condition and were selected to be included in the analysis (Fig. 1). We also applied other preprocessing treatments to each hospital's time series, including a weekly moving average filter to smooth out weekly patterns, and a linear normalization to zero mean and unit variance.

III. Methods

Let $\mathbf{x}_h$ denote the daily time series of the proportion of flu-positive visits at hospital $h \in H$, where H is the set of all hospitals. The geographical notion of "region" in a spatial

hospital aggregation is replaced by "spatio-temporal cluster" in our model. The set of all distance measures between pairs of hospitals, denoted D, includes a spatial $d^{£}$ and multiple temporal d^{f} distance measures. For each distance measure $d \in D$, a hierarchical clustering C^d is created. To create clusterings with desired numbers of total clusters, then, the hierarchy C^d can be cut at $n \in N$ to obtain C_n^d, where N is the set of all numbers that is desirable as the total number of clusters; It can reflect expectations or needs of public health agencies. Then, the time series for each cluster C in each of the clusterings C_n^d is calculated by averaging the time series of the hospitals that belong to that cluster.

We use a fully connected network to model flu travel between pairs of those cluster-aggregated time series in each clustering $(\mathbf{x}^C, \mathbf{x}^{C'}$, where $C, C' \in \mathsf{C}^d)_n$ The clustering is then assigned a predictiveness score for direction and magnitude of flu travel between pairs of its clusters, and the highest-scoring clustering will be used to make predictions about the future outbreaks. For example, if a certain clustering of cities yields significant values of lagged correlations, or low values of residual errors with respect to the lag magnitude, they receive a high score.

A. Hospital Clustering

Hierarchical clustering has been successful in clustering both spatial [42] and temporal [43] data, and is therefore chosen as the clustering algorithm in this research. Our model is an agglomerative bottom up clustering, that starts by putting each object in its own cluster. Then, at each step, the nearest two clusters are combined into a new higher-level cluster until all objects are in one cluster, creating a hierarchy of clusterings with different numbers of clusters.

Let $d_{hh'}$ denote the distance measured by $d \in D$ between hospitals $h, h' \in H$. Then, define $d^{CC'}$, the distance between clusters C and C' in clustering C as the average of all pairwise hospital distances:

$$d^{CC'} = \frac{1}{|C| \times |C'|} \sum_{h \in C} \sum_{h' \in C'} d_{hh'} \quad C, C' \in \mathsf{C},\ C \neq C' \qquad (1)$$

We now briefly present an example of bottom-up hierarchical clustering, where clusters are merged iteratively to form bigger clusters. The opposite is also possible, and called top-down clustering, where larger clusters are iteratively split into smaller ones.

Suppose we want to create a hierarchical clustering for the set of points $\{a, b, c, d, e\}$ with coordinates $a = (1, 5)$, $b = (2.5, 5)$, $c = (3, 1)$, $d = (3, 2)$, $e = (5, 1)$. The algorithm iteratively merges closest clusters to form bigger clusters starting by merging the closest pair (c, d) into a single cluster. The resulting hierarchy of clusters can be visualized using a dendogram whose height represents the average pairwise distance defined in Equation 1. The dendogram can then be cut at different heights to create clusterings with different sizes (Fig. 2).

The spatial distance between hospitals h and h' is denoted $d^{£}_{hh'}$ and measures the Euclidean distance using their longitude and latitude coordinates, denoted XC_h and YC_h, respectively:

$$d^{£}_{hh'} = \sqrt{(\mathrm{XC}_h - \mathrm{XC}_{h'})^2 + (\mathrm{YC}_h - \mathrm{YC}_{h'})^2} \quad h, h' \in H \qquad (2)$$

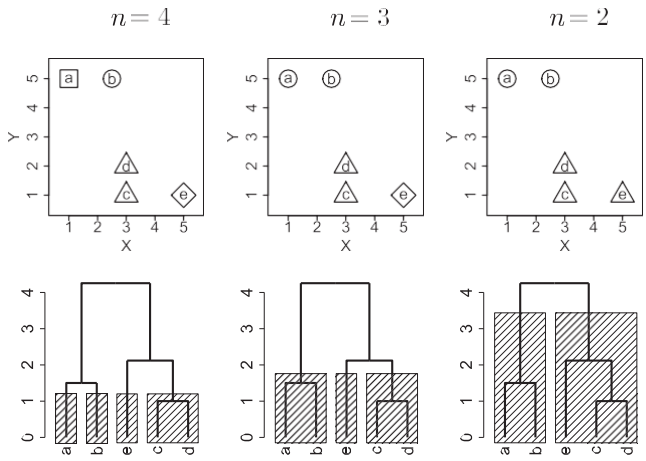


Fig. 2. An example of hierarchical clustering dendogram cut at different heights. n is the number of clusters. Shapes in the top row and shaded areas in the bottom row indicate cluster membership.

The temporal distance between hospitals h and h', denoted $d^{f}_{hh'}$, is the distance between their corresponding time series. Unlike the Euclidean distance for spatial measure, there is no consensus in the literature on how to measure the temporal distance, mainly due to the fact that different measures are designed to capture different types of similarity in temporal behaviour such as collinearity and shape.

For example, consider two large cities with similar urban characteristics, but a large geographical distance. Although they may exhibit similar transmission patterns, in a purely geographical aggregation, such cities will not end up in the same cluster. Calculation of temporal distances allows for clustering of cities with similar spread patterns, and therefore profiling of regions, and identification of influential urban characteristics, such as the presence of airports, or financial hubs.

We compare the performance of spatial clustering with the performance of clustering based on five common time series distance measures to find the most predictive clustering in terms of predicting certain surveillance variables. The five time series distance measures are dissimilarity based on autocorrelation and partial autocorrelation functions $d^{f:\mathrm{ACF}}$ and $d^{f:\mathrm{PACF}}$ [44], correlation-based distance $d^{f:\mathrm{COR}}$ [45], dynamic time warping $d^{f:\mathrm{DTW}}$ [46], and Euclidean distance $d^{f:\mathrm{EUC}}$.

We now briefly describe how each of the five common temporal distance measures (ACF, PACF, COR, DTW, and EUC) are calculated for a pair of hospital time series for hospitals h and h', defined as:

$$\mathbf{x}_h = (x_h[1], \ldots, x_h[T]) \quad h \in H \qquad (3)$$

$$\mathbf{x}_{h'} = (x_{h'}[1], \ldots, x_{h'}[T]) \quad h' \in H \qquad (4)$$

The autocorrelation-based distance $d^{f:\mathrm{ACF}}$ measures the distance between estimated autocorrelation functions. Let $\hat{\varphi}_{\mathbf{x}} = (\hat{\varphi}_{\mathbf{x}}[1], \ldots, \hat{\varphi}_{\mathbf{x}}[M])$ be the estimated autocorrelation vector of $\mathbf{x}$, found by regression [44], for some M such that $\hat{\varphi}_{\mathbf{x}}[m] \approx 0$ for $m > M$. The autocorrelation-based distance between $\mathbf{x}_h$ and

$\mathbf{x}_{h'}$ is

$$d_{hh'}^{\mathrm{f:ACF}} = \sqrt{(\hat{\varphi}_{\mathbf{x}_h} - \hat{\varphi}_{\mathbf{x}_{h'}})^{\mathsf{T}}(\hat{\varphi}_{\mathbf{x}_h} - \hat{\varphi}_{\mathbf{x}_{h'}})} \quad h, h' \in H$$

and measures how similarly the two time series regress to their previous values. The partial autocorrelation-based distance $d_{ij}^{\mathrm{f:PACF}}$ is similarly calculated, but instead, based on the partial autocorrelation coefficients that are obtained by removing the dependencies on intermediate lag values.

The correlation-based distance $d^{\mathrm{f:COR}}$ is calculated using the Pearson correlation coefficient:

$$p(\mathbf{x}_h, \mathbf{x}_{h'}) = \frac{\sum_{t=1}^{T}(x_h[t] - \bar{\mathbf{x}}_h)(x_{h'}[t] - \bar{\mathbf{x}}_{h'})}{\sqrt{\sum_{t=1}^{T}(x_h[t] - \bar{\mathbf{x}}_h)^2}\sqrt{\sum_{t=1}^{T}(x_{h'}[t] - \bar{\mathbf{x}}_{h'})^2}} \quad h, h' \in H \tag{5}$$

where $\bar{\mathbf{x}}_h$ is the average value of $\mathbf{x}_h$. Recall that we normalized the series such that they have a mean of zero; therefore $\bar{\mathbf{x}}_h = 0$ for $h \in H$. Correlation-based distance $d^{\mathrm{f:COR}}$ is then calculated as follows:

$$d_{hh'}^{\mathrm{f:COR}} = \sqrt{2(1 - p(\mathbf{x}_h, \mathbf{x}_{h'}))} \quad h, h' \in H \tag{6}$$

Dynamic time warping finds a mapping between $\mathbf{x}_h$ and $\mathbf{x}_{h'}$ such that a specific distance measure between the coupled observations is minimized. The minimum distance is denoted $d^{\mathrm{f:DTW}}$ and is given by

$$d_{hh'}^{\mathrm{f:DTW}} = \min_{(a_k, b_k) \in R_{hh'}} \sum_{k=1,\ldots,\tau} |x_h[a_k] - x_{h'}[b_k]| \quad h, h' \in H \tag{7}$$

where $R_{hh'}$ is the set of all warping paths (a_k, b_k), $k \in \{1 \ldots, \tau\}$ between $\mathbf{x}_h$ and $\mathbf{x}_{h'}$ that satisfy three conditions:

$$\begin{aligned} R_{hh'} = {} & (a_k, b_k),\ k \in \{1 \ldots, \tau\} \mid a_k, b_k \in \{1, \ldots, T\}, \\ & a_1, b_1 = 1,\ a_\tau, b_\tau = T, \\ & (a_k - a_{k-1}, b_k - b_{k-1}) \in \{(1, 0), (0, 1), (1, 1)\} \end{aligned} \tag{8}$$

The first condition ensures that (a_k, b_k) is a mapping between $\mathbf{x}_h$ and $\mathbf{x}_{h'}$ both with length T. The second condition ensures that we use all of the information in both sequences. The third condition allows the index of one sequence to be stopped while the other continues, which also implies that we cannot go back in time (a_k and b_k are monotonically increasing). DTW recognizes similarly-shaped time series in the presence of shifting or scaling, i.e., when two similarly-shaped time series are stretched or compressed, their DTW distance remains small.

Finally, the Euclidean distance $d^{\mathrm{f:EUC}}$ is calculated as

$$d_{hh'}^{\mathrm{f:EUC}} = \sqrt{\sum_{t=1}^{T}(x_h[t] - x_{h'}[t])^2} \tag{9}$$

Once the distance matrix for hospital pairs is calculated for each distance measure $d \in D$, hierarchical clusterings C_n^d are found for $n = 1, \ldots, N$, and cluster-aggregated time series $\mathbf{x}^C$, which are used in modelling inter-cluster flu movement, are defined as

$$\mathbf{x}^C = \frac{1}{|C|}\sum_{h \in C} \mathbf{x}_h \quad C \in \mathsf{C}_n^d,\ n \in N,\ d \in D \tag{10}$$

B. Flu Movement

From an infectious disease epidemeology point of view, we hypothesize that there are established population and geography factors that may affect the spread patterns in a certain way that is recurrent from outbreak to outbreak. Note that in this work, we do not distinguish between dominant strains and A/B types, therefore, the findings are not specific to certain outbreak behaviours. Learning such patterns can help policymakers make informed resource management decisions to contain the outbreak more efficiently. For example, the knowledge that cluster C usually peaks a few days earlier than cluster C' gives policymakers a few days advanced notice for planning in C' as soon as an outbreak starts in C. We therefore investigate two questions for each cluster pair. The first question involves the direction of flu travel: Which cluster precedes the other in the outbreak? The second question involves the magnitude: How many days does it take for the outbreak to travel from one cluster to another?

To answer these questions, we first manually extract the outbreak days from the cluster-aggregated time series to keep the focus away from no-risk days. Seasonal flu, albeit variable in timing and duration, often starts as early as October and can last as late as April with outbreaks between December and February [47]. For years with multiple outbreaks, each one is isolated and treated separately.

Consider the full time series $\mathbf{x} = \{x[1], \ldots, x[T]\}$. Since the data outside of October of each year and April of the next year is irrelevant, we set October 1st as index $t = 1$, and April 31st as index $t = T$, that is 212. For an isolated outbreak $o \in O$, let α_o and β_o, extracted from the average time series, denote the start and end dates, respectively. Note that the outbreak starts and ends between October 1st and April 31st, that is, $1 < \alpha < \beta < T$. Then, for the cluster pair $C, C' \in \mathsf{C}$, the lagged cross correlation is defined as

$$\rho_{\alpha\beta}^{CC'}(l) = \sum_{t=\alpha}^{\beta}\left(x^C[t+l] - \bar{\mathbf{x}}_{\alpha\beta}^{C}(l)\right)\left(x^{C'}[t] - \bar{\mathbf{x}}_{\alpha\beta}^{C'}(0)\right) \times \left(\sum_{t=\alpha}^{\beta}\left(x^C[t+l] - \bar{\mathbf{x}}_{\alpha\beta}^{C'}(l)\right)^2\right)^{-\frac{1}{2}} \left(\sum_{t=\alpha}^{\beta}\left(x^{C'}[t] - \bar{\mathbf{x}}_{\alpha\beta}^{C'}(0)\right)^2\right)^{-\frac{1}{2}} \quad l \in \{-L, \ldots, +L\} \tag{11}$$

where $\bar{\mathbf{x}}_{\alpha\beta}^{C}(l)$ is the average of the lagged values

$$\bar{\mathbf{x}}_{\alpha\beta}^{C}(l) = \frac{1}{\beta - \alpha + 1}\sum_{t=\alpha}^{\beta} x^C[t+l] \tag{12}$$

and L is the maximum suggested lag. We now use the values of ρ to identify the travel direction and distance magnitude.

Magnitude: To extract recurrent lag magnitudes between cluster pairs, we consider all clusterings $\mathbf{C}_n^d$ for $d \in D$ and $n \in N$ for aggregation, and choose the one whose inter-cluster magnitudes are least variant. That is, the clustering between cluster pairs in which flu travels least variantly in terms of the magnitude.

To measure the variance of magnitude for clustering $\mathbf{C}$, we first extract the lag by which the two time series most highly correlate as follows:

$$l_o^{CC'} = \arg \max_{A \in \{-L,\ldots,L\}} \rho_{\alpha_o \beta_o}^{CC'}(l) \quad o \in O,\ C, C' \in \mathbf{C},\ C \neq C' \tag{13}$$

A small value for $l_o^{CC'}$ indicates that it took a short time for outbreak o to travel between C and C', and a large value means the reverse.

Lower variance in the values of $l_o^{CC'}$, $o \in O$ means more consistent distance magnitudes from outbreak to outbreak, hence less uncertain future predictions. The average magnitude score is the reciprocal of the average of pairwise yearly variances (lower variances are represented by higher scores):

$$s_{\text{mag}}(d, n) = \left(\frac{1}{\binom{n}{2}} \sum_{\substack{C,C' \in \mathbf{C}_n^d \\ C \neq C'}} \frac{1}{|O|} \sum_{o \in O} \left(l_o^{CC'} - \bar{l}^{CC'} \right)^2 \right)^{-1} \tag{14}$$

where $\bar{l}^{CC'}$ is the average of lag magnitudes across all outbreaks, defined as

$$\bar{l}^{CC'} = \frac{1}{|O|} \sum_{o \in O} l_o^{CC'} \tag{15}$$

Among the clusterings with different sizes based on different distance measures, the one that achieves the highest magnitude score will be used for magnitude prediction:

$$(d^*_{\text{mag}}, n^*_{\text{mag}}) = \arg \max_{\substack{d \in D \\ n \in N}} (s_{\text{mag}}(d, n)) \tag{16}$$

For brevity, let $\mathbf{C}^*(\text{mag})$ denote $\mathbf{C}_{n^*_{\text{mag}}}^{d^*_{\text{mag}}}$ as the winning clustering. For any given cluster pair $C, C' \in \mathbf{C}^*(\text{mag})$, future magnitude is predicted as $\bar{l}^{CC'}$.

Direction: To extract recurrent directions between cluster pairs, we consider all clusterings $\mathbf{C}_n^d$ for $d \in D$ and $n \in N$ for aggregation, and choose the pair whose inter-cluster directionalities are most significantly consistent. The start and end date of each outbreak $o \in O$ are denoted α_o, β_o and are identified by isolating the peaks from the average time series from all hospitals. Each outbreak, once isolated, has then the same length $\beta_o - \alpha_o + 1$ for all hospitals.

Consider outbreak $o \in O$. We use the first moment of the lagged cross-correlation function between each pair $C, C' \in \mathbf{C}$ to measure directionality between the pair:

$$m_o^{CC'} = \sum_{A=-L} l \times \rho_{\alpha_o \beta_o}^{CC'}(l) \tag{17}$$

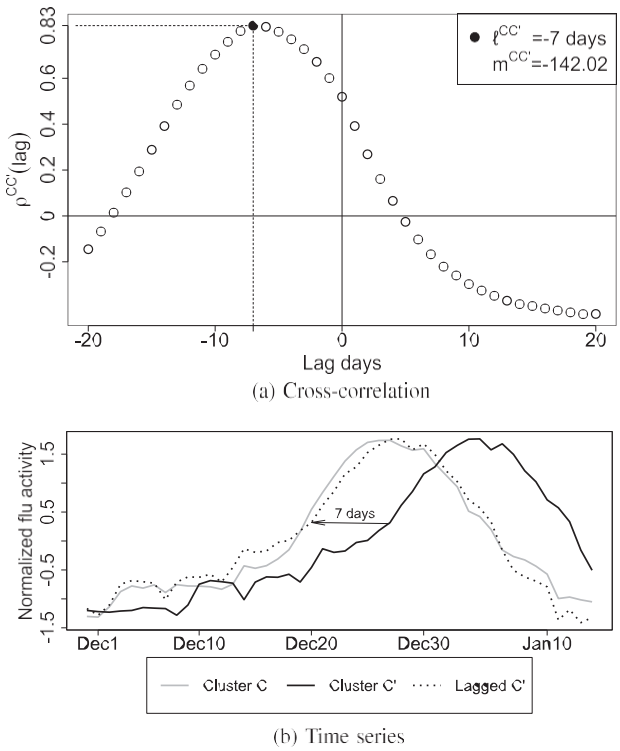


Fig. 3. An example of finding the lag moment and optimal lag for a pair of time series.

where $\rho_{\alpha_o \beta_o}^{CC'}(l)$ is defined in (11). The more moment accumulates on either side of the positive lags or the negative lags, the stronger directionality can be inferred. A negative value for $m^{CC'}$ indicates that negatively-lagged $\mathbf{x}_o^{C'}$ (starts earlier) correlates more highly with $\mathbf{x}_o^{C}$ than positively-lagged, and therefore outbreak o in cluster C precedes C' (Fig. 3). A positive $m_o^{CC'}$ means the reverse. Note that, because of the momentum, the higher-magnitude lags achieve high ρ values, the higher the magnitude of lag moment function will be, indicating stronger directionality.

The moments from multiple outbreaks $o \in O$ are then averaged, and the average's absolute value represents the strength of directionality between that pair. The average of directionalities of all pairs is then the direction score:

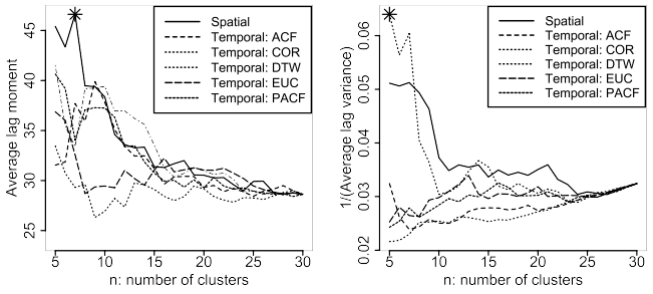
$$s_{\text{dir}}(d, n) = \frac{1}{\binom{n}{2}} \sum_{\substack{C,C' \in \mathbf{C}^d_n \\ C \neq C'}} \bar{m}^{CC'} \quad d \in D,\ n \in N \tag{18}$$

where where $\bar{m}^{CC'}$ is the average of moments across all outbreaks:

$$\bar{m}^{CC'} = \frac{1}{|O|} \sum_{o \in O} m_o^{CC'} \tag{19}$$

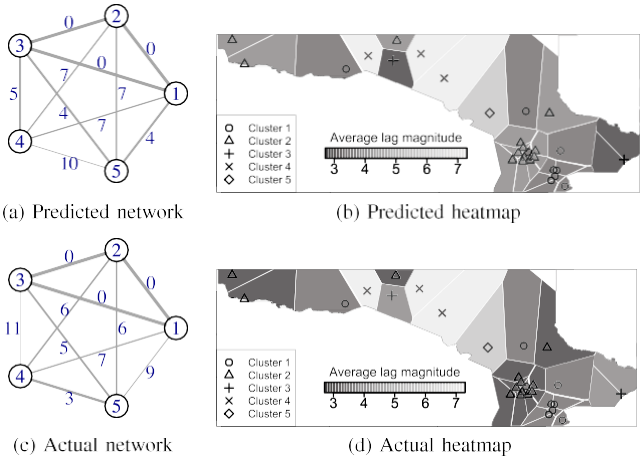
Among the clusterings with different sizes based on different distance measures, the one that achieves the highest direction score is the most suitable to be used for direction prediction:

$$(d^*_{\text{dir}}, n^*_{\text{dir}}) = \arg \max_{\substack{d \in D \\ n \in N}} (s_{\text{dir}}(d, n)) \tag{20}$$



(a) Direction score (Equation 18) (b) Magnitude score (Equation 14)

Fig. 4. The direction and magnitude scores for different clusterings C^{d}_{n}. The star indicates the highest score and the winner clustering.



(a) Predicted network (b) Predicted heatmap

(c) Actual network (d) Actual heatmap

Fig. 5. Pairwise lag magnitude networks and individual average magnitude heatmaps. Edge thickness in the graph is inversely proportional to the lag magnitude.

For brevity, let $\mathsf{C}^{*}(\text{dir})$ denote $\mathsf{C}^{d^{*}_{\text{dir}}}_{n^{*}_{\text{dir}}}$ as the winning clustering that will be used for direction prediction. For any given cluster pair $C, C' \in \mathsf{C}^{*}(\text{dir})$, directionality in testing outbreak is predicted using training outbreaks. If $\bar{m}^{CC'}$ for the past outbreaks is negative, we predict that C will precede C' in the future outbreak and vice versa.

Fig. 3 shows an example of how lag moment ($m^{CC'}$ from (17)) and optimal lag ($l^{CC'}$ from (13)) are calculated for one outbreak and one cluster pair.

IV. Results

Computations were performed on a machine with 2.4 GHz processor and 8 GB memory. Seven outbreaks were isolated from 2012-2016 (two in 2012, one in each year from 2013 to 2015, and two in 2016) for each of the 30 selected hospitals. The daily flu-positive proportion of ED visits was extracted from the SQL database and imported into R as a time series object for each hospital during each outbreak. From a public health point of view, to represent at least the five population-dense regions in Ontario (southern Ontario, Kingston, Sudbury, Thunder Bay, and Rainy River), the set of acceptable values for the total number of clusters is set to $N = \{5, \ldots, 30\}$.

Once the cluster-aggregated time series are extracted, for each clustering, the direction and magnitude scores are calculated using (18) and (14), respectively. For the direction score, spatial clustering with seven clusters had the highest score ($\mathsf{C}^{*}(\text{dir}) = \mathsf{C}^{\text{f}}_{7}$), and for the magnitude score, DTW temporal clustering with five clusters had the highest score $\mathsf{C}^{*}(\text{mag}) = \mathsf{C}^{\text{f:DTW}}_{5}$ (Fig. 4). We now analyze each of the two surveillance variables (direction and magnitude) by using its respective winning clustering for aggregation.

A. Magnitude

Consider the winning magnitude clustering $\mathsf{C}^{*}(\text{mag})$, that is, the DTW temporal clustering with five clusters. For each cluster pair $C, C' \in \mathsf{C}^{\text{f:DTW}}_{5}$, we compare the predicted magnitude based on the first six outbreaks Z with the actual magnitude from the seventh outbreak $Z^{CC'}$:

$$\hat{Z}^{CC'} = \frac{1}{6}\sum_{o=1}^{6} l_o^{CC'} \qquad C, C' \in \mathsf{C}^{\text{f}}_{7},\ C \neq C' \tag{21}$$

$$Z^{CC'} = l_{o=7}^{CC'} \qquad C, C' \in \mathsf{C}^{\text{f:DTW}}_{5},\ C \neq C' \tag{22}$$

Then, the reported magnitude accuracy is the root mean squared error (RMSE):

$$\text{Accuracy}^{\text{mag}} = \sqrt{\frac{1}{\binom{5}{2}}\sum_{C, C' \in \mathsf{C}^{\text{f:DTW}}_{5},\ C \neq C'} \left(Z^{CC'} - \hat{Z}^{CC'}\right)^2} \tag{23}$$

The RMSE for $\mathsf{C}^{\text{f:DTW}}_{5}$ is 3.16 days. Out of the total $\binom{5}{2} = 10$ cluster pairs, we are able to predict the magnitude for 7 of 10 cluster pairs within 1 d (Fig. 5(a) and (c)).

In addition to the pairwise magnitude, we examined the overall average magnitude for each cluster to identify the clusters that are generally within a short temporal distance of the other clusters. For each clustering C,

$$I(C) = \frac{1}{5-1}\sum_{C' \in \mathsf{C}^{\text{f:DTW}}_{5},\ C' \neq C} \bar{l}_o^{CC'} \qquad C \in \mathsf{C}^{\text{f:DTW}}_{5} \tag{24}$$

indicates how closely cluster C is connected to the other clusters in terms of the lag magnitude. Fig. 5(b) and (d) show the five clusters of $\mathsf{C}^{\text{f:DTW}}_{5}$, and their predicted and actual overall lag magnitude. Dark regions indicate regions with small $I(C)$ and are expected to be more closely connected to other regions than the lighter regions that have large $I(C)$.

B. Direction

Consider the winning direction clustering $\mathsf{C}^{*}(\text{dir})$, that is, the spatial clustering with seven clusters. For each cluster pair $C, C' \in \mathsf{C}^{\text{f}}_{7}$, we compare the predicted direction $\hat{Y}^{CC'}$ based on the first six outbreaks with the actual direction for the seventh outbreak $Y^{CC'}$:

$$\hat{Y}^{CC'} = \begin{cases} 1 & \text{if } \frac{1}{|\bar{O}|}\sum_{o=1}^{6} m_o^{CC'} < 0 \quad C, C' \in \mathsf{C}^{\text{f}}_{7}, C \neq C' \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

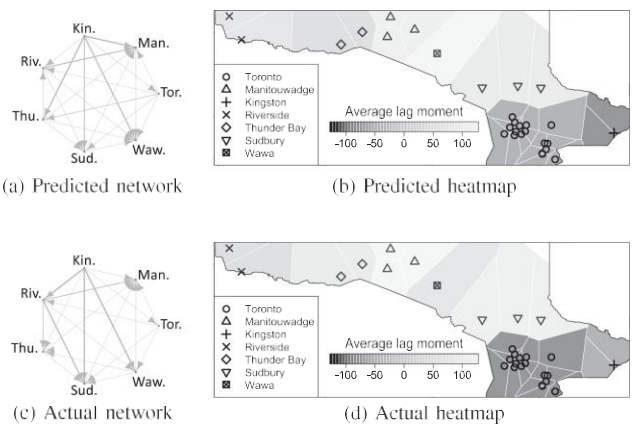


Fig. 6. Pairwise lag direction moments and cluster-wise heatmap. Thickness in the network indicates magnitude of moment for each pair.

$$Y^{CC'} = \begin{cases} 1 & \text{if } m_{o=7}^{CC'} < 0 \\ 0 & \text{otherwise} \end{cases} \qquad C, C' \in \mathsf{C}_7^{\mathsf{f}}, C \neq C' \tag{26}$$

Then, the reported direction accuracy indicates the proportion of the cluster pairs for which the predicted direction matches the actual direction:

$$\text{Accuracy}^{\text{dir}} = 1 - \frac{1}{2} \sum_{\substack{C,C' \in \mathsf{C}_7^{\mathsf{f}} \\ C \neq C'}} \left| Y^{CC'} - \hat{Y}^{CC'} \right|^2 \tag{27}$$

Out of the total $\binom{7}{2} = 21$ cluster pairs in $\mathsf{C}_7^{\mathsf{s}}$, we were able to correctly predict the direction for 17 cluster pairs, achieving a direction accuracy of 81% (Fig. 6(a) and (c)).

In addition to the pairwise directionalities, we examined the overall outgoing (incoming if positive) moment for each cluster to identify the clusters that generally lead (follow) the other clusters. For each clustering C,

$$m(C) = \frac{1}{7-1} \sum_{\substack{C' \in \mathsf{C}_7^{\mathsf{f}} \\ C' \neq C}} \overline{m}_o^{CC'} \qquad C \in \mathsf{C}_7^{\mathsf{f}} \tag{28}$$

indicates how likely the cluster C is to be a leader or a follower in the future outbreaks and is used to create a heatmap.Fig. 6(b) and (d) show the seven regions of $\mathsf{C}_7^{\mathsf{s}}$, and the predicted and actual direction heatmap for those regions. Dark regions in the heatmap (e.g., Kingston and Toronto) indicate regions with negative $m(C)$ and are therefore expected to lead, while light regions (e.g., Manitouwadge) indicate regions with positive $m(C)$ and are expected to follow. Also, note that the larger the absolute value of the moment is, the stronger directionality can be inferred.

To put the reported accuracies into perspective for comparison of spatial and temporal clusterings for public policymaking, we compared magnitude and direction accuracies for each of the winning clusterings $\mathsf{C}_7^{\mathsf{f}}$ and $\mathsf{C}_5^{\mathsf{f:DTW}}$ with clusterings of the same size but with the other distance measure (Fig. 7).

The direction winner $\mathsf{C}_7^{\mathsf{f}}$ predicts the correct direction for 81% of the cluster pairs, while the best temporal clustering with the same size predicts the correct direction for only 71% of the cluster pairs. On the other hand, while the RMSE for the magnitude winner $\mathsf{C}_5^{\mathsf{f:DTW}}$ is 3.16 days, and it predicts the lag magnitude within 1 d of the actual value for 70% of the cluster pairs, the spatial $\mathsf{C}_7^{\mathsf{f}}$ achieves an RMSE of 5.84 days, being able to predict the actual lag magnitude within 1 d for only 20% of the cluster pairs.

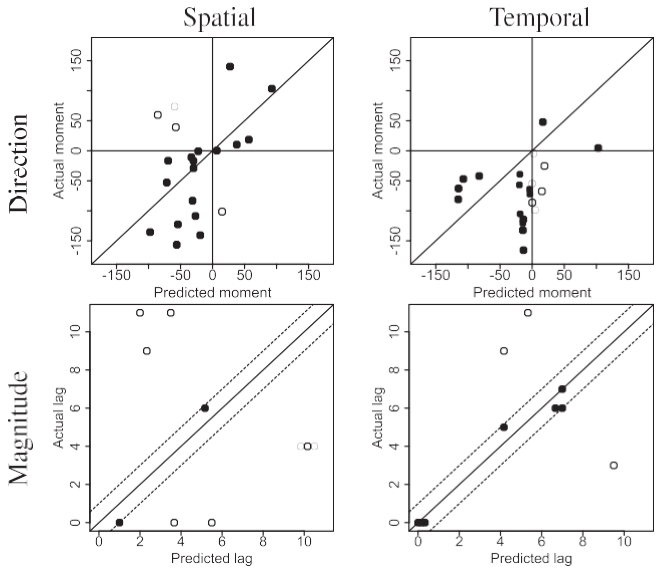


Fig. 7. Direction and magnitude accuracies for spatial and temporal clusterings. The filled circles indicate the hits (same-sign moments and within 1 d lag magnitudes), and empty circles the misses.

V. Discussion

Spread patterns of influenza are studied in the literature on scales ranging from wearable sensors and contact diaries in a high school [48] to work commute patterns in cities [49] and even global transmission patterns that were studied on both the pathogen and the host sides [50], [51], [52]. These approaches, however, assume an underlying model a priori to accommodate preconceived notions in the assumptions of spread models. However, those preconceived notions of transmission, although heavily relied upon in the literature, are not always right. For example, contrary to the widely-accepted notion that influenza epidemics depends on the climatic variables [53], [54], [55], there is recent evidence [56] for cities with vast intercontinental climactic variations and more than 3,000 km of geographical distance that exhibit significant synchronicity in terms of the onset of epidemics.

Extensions of the SIR model have also been proposed to consider spatio-temporal data [57]; however, each country has their own SIR coefficients. Therefore, the temporal aspect is only in the temporal differentials of the SIR system with country-specific parameters. None of the above papers aggregate data based on temporal distances.

We also found evidence for rapid transmission in non-populous regions like Rainy River and Thunder Bay [58], contrary to the paradigm that influenza spreads more rapidly

between more populous areas [59]. Our model requires no prior assumptions on spread patterns or other complex covariates (e.g., bird migration patterns [60]) or weather [61], and instead finds recurrent inter-regional patterns directly from the historical data. By virtue of ignoring the preconceived notions on the spread patterns, then, our method is immune to the complexities of the underlying structure and is suitable for the analysis of overly complicated systems with limited historical data for training.

By focusing on the surveillance objectives as early as the aggregation stage, our model allows for the extraction of actionable insights for health system response. We showed that geographical clustering is more suitable than temporal for predicting the spatial direction of the flu transmission. Although pairs such as Toronto and Ottawa, both population-dense regions [58], have a small directionality between them (i.e., the outbreak equally likely went either way in the past) some region pairs such as Kingston to Wawa and Manitouwadge exhibit a strong directionalities, which is consistent with previous findings [62] that influenza tends to start in the larger cities and then spread to the smaller ones.

Other than spatial closeness, work commute and air traffic have also been connected to spread of influenza [63]. We found that temporal aggregation provided more accurate predictions for the pairwise temporal distances, that are the lag magnitudes. Our findings that outbreaks take between three to seven days to travel between pairs of Ontario hospital clusters is consistent with previous findings [64] that influenza takes about a week to travel between the metropolitan areas and the smaller cities. Once a metropolitan cluster faces an outbreak, remote clusters can receive an advanced notice and incorporate our pairwise lag magnitudes into their planning horizons.

Our findings indicate that regions such as Rainy River and Thunder Bay, albeit not populous, are tightly connected to other regions. This observation is in contrast with the paradigm that influenza is marked by rapid spread between populous centers followed by subsequent spread to less populated areas [59], and suggests that there must be other factors than population to determine the rapidness of transmission between regions. It is worth mentioning that the dark regions in the magnitude heatmap (Fig. 5) often have a major airport near them (Toronto, Ottawa, and Thunder Bay all have international airports), while there is no major airport near the light regions like Manitouwadge. Although further analysis is required before drawing the conclusion that airports cause a faster flu travel between regions, there are studies that link transmission dynamics to air travel patterns [65], [66].

From a public health applicability point of view, this analysis allows policymakers to create advance alarms to certain regions based on the current situation. For example, when an outbreak starts in Toronto, hospitals in Manitouwadge may be warned about expecting an increase in flu-related emergency visits and plan to allocate their resources accordingly. Although we showed quantitative improvement in the accuracy of magnitude prediction by introducing a temporal aggregagtion, the direct impact may not necessarily be obtained through earlier or more accurate prediction of the outbreaks, but rather through novel insights such as the quantitative demonstration that major airports are flu hubs. In terms of prediction accuracy, a temporal aggregation proved more accurate than spatial aggregation in predicting the magnitude of spread. The improved accuracy may be capitalized in resource management such as bed planning.

Our spatio-temporal model may improve studies of other infectious diseases, as well as other health informatics problems where spatio-temporal patterns are important. Examples include applications in mapping brain activity, biosensors, health monitoring systems, and public health surveillance.

Our work has two limitations that may be addressed through further research: The first limitation is that it does not provide an explicit learning mechanism in terms of maximum likelihood. A Bayesian extension may provide the capability to learn causal structures from conditional distributions and identify disease hubs that may cause outbreaks in other regions.

The second limitation is that, due to data availability, our dataset is small given the scope of infectious disease spread and our focus on a single Canadian province. A more comprehensive dataset with longer history and bigger geographic areas containing more major airports could provide further validation of our findings. Further insights could be gained through the incorporation of relevant external factors, e.g., weather (as influenza viruses are more stable in the cold [67]) and bird migrations (as influenza can be carried by birds, e.g., H5N1 avian flu [60]), as well as examination of various other influenza subtypes and the impact of flu vaccinations.

VI. Conclusion

We showed that the spatial direction and temporal magnitude of flu travel between clusters of hospitals can be predicted using historical data. Exploiting the spatial and the temporal similarities of spatio-temporally recorded data can each be more useful than the other depending on the variable being studied. The methodologies presented here may improve the analysis of other problems where a set of time series are recorded from multiple related locations.

Our current analysis does not distinguish between flu virus types A and B. However, based on their genetic properties they may exhibit different outbreak behaviours. A possible future direction for this work is to obtain laboratory results on the types of circulating virus and differentiate between outbreaks with different types of dominant viral strains. For example, influenza type A that goes through antigenic shifts may cause a new strain against which the population has no immunity, resulting in a pandemic that rapidly effects large population with significant health system impact.

FluWatch reports by PHAC in Canada [12] during the past five years suggest that a type A outbreak usually precedes a type B outbreak and has a larger magnitude. Confirmed laboratory results may enable quantitative verification of these hypotheses.

Additionally, future research can differentiate between outbreaks happening at different times of the year. Seasonality of transmission rates is known to have a significant influence on the temporal patterns of epidemics of infectious diseases [68]. An outbreak peaking during holiday travel periods may be

faster transmitted between airport-connected cities, while it may be more slowly spread into work places due to the holiday closures. The detection of pandemic behaviour may then be turned into action-based prediction of impact to plan for effective containment. For example, knowing the importance of regions with major airports, the policymakers could have prioritized first containment and then tracking efforts in those regions; or even planned for bed capacities more efficiently, knowing how long the peaks take to travel betweeen regions.

Finally, similar analysis can be repeated for COVID-19 to determine if emerging global pandemics share the same spread properties as seasonal influenza when applied to specific regions, i.e., a state or province, or even an entire country with enough participating hospitals providing both historical flu and current pandemic data.

By analyzing data from only one province, we found an interesting potential player: the major airports. The analysis of global pandemics like COVID-19 can verify the role of the airports [69], [70], and potentially reveal even more interesting urban patterns (R2C2), such as the role of big cities, tourist attractions, and financial hubs [71], [72].

Acknowledgment

Hootan Kamran would like to thank Bahman Radjabalipour for helping with the project's database.

References

[1] World Health Organization, "Global influenza programme," 2015, Accessed: 2018. [Online]. Available: http://www.who.int/influenza/en/
[2] Centers for Disease Control and Prevention, "Leading causes of death," 2013, Accessed: 2021. [Online]. Available: http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm
[3] Statistics Canada, "The 10 leading causes of death, 2011," 2011, Accessed: 2021. [Online]. Available: http://www.statcan.gc.ca/pub/82-625-x/2014001/article/11896-eng.htm
[4] Institute for Clinical Evaluative Sciences, "Ontario burden of infectious disease study," 2010.
[5] Centers for Disease Control and Prevention, "Types of influenza viruses," 2013, Accessed: 2018. [Online]. Available: http://www.cdc.gov/flu/about/viruses/types.htm
[6] Centers for Disease Control and Prevention, "CDC seasonal flu vaccine effectiveness studies," 2021, Accessed: 2021. [Online]. Available: https://www.cdc.gov/flu/vaccines-work/effectiveness-studies.htm
[7] P. Cirillo and N. N. Taleb, "Tail risk of contagious diseases," *Nature Phys.*, vol. 16, no. 6, pp. 606–613, 2020.
[8] H. Minkowski, "Space and time, lecture given at the 80th meeting of natural scientists in Cologne, Germany," 1908.
[9] R. A. Jajosky and S. L. Groseclose, "Evaluation of reporting timeliness of public health surveillance systems for infectious diseases," *BMC Public Health*, vol. 4, no. 1, 2004, Art. no. 29.
[10] A.-J . Valleron et al., "A computer network for the surveillance of communicable diseases: The French experiment," *Amer. J. Public Health*, vol. 76, no. 11, pp. 1289–1292, 1986.
[11] Centers for Disease Control and Prevention, "FluView," 2013, Accessed: 2018. [Online]. Available: http://www.cdc.gov/flu/weekly/
[12] Public Health Agency of Canada, "FluWatch," 2022. [Online]. Available: http://www.phac-aspc.gc.ca/fluwatch/
[13] KFL&A, "ILIMapper," 2018, Accessed: 2018. [Online]. Available: http://mapper.kflaphi.ca/ilimapper/
[14] X. Zhou et al., "A spatial-temporal method to detect global influenza epidemics using heterogeneous data collected from the internet," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 3, pp. 802–812, May/Jun. 2018.
[15] D. M. Cornforth, T. C. Reluga, E. Shim, C. T. Bauch, A. P. Galvani, and L. A. Meyers, "Erratic flu vaccination emerges from short-sighted behavior in contact networks," *PLoS Comput. Biol.*, vol. 7, no. 1, 2011, Art. no. e1001062.
[16] G. Bindea et al., "Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer," *Immunity*, vol. 39, no. 4, pp. 782–795, 2013.
[17] A. Diehl et al., "Visual analysis of spatio-temporal data: Applications in weather forecasting," in *Computer Graphics Forum*, vol. 34. Hoboken, NJ, USA: Wiley, 2015, pp. 381–390.
[18] T. Higham et al., "The timing and spatiotemporal patterning of neanderthal disappearance," *Nature*, vol. 512, no. 7514, pp. 306–309, 2014.
[19] B. Y. Reis and K. D. Mandl, "Time series modeling for syndromic surveillance," *BMC Med. Informat. Decis. Mak.*, vol. 3, no. 1, 2003, Art. no. 2.
[20] L. Ngo, I. B. Tager, and D. Hadley, "Application of exponential smoothing for nosocomial infection surveillance," *Amer. J. Epidemiol.*, vol. 143, no. 6, pp. 637–647, 1996.
[21] S. Unkel, C. Farrington, P. Garthwaite, C. Robertson, and N. Andrews, "Statistical methods for the prospective detection of infectious disease outbreaks: A review," *J. Roy. Stat. Soc.: Ser. A*, vol. 175, no. 1, pp. 49–82, 2012.
[22] W. A. Shewhart, *Economic Control of Quality of Manufactured Product*. New York, NY, USA: D. Van Nostrand Company, Incorporated, 1931.
[23] P. A. Rogerson, "Monitoring point patterns for the development of space–time clusters," *J. Roy. Stat. Soc.: Ser. A*, vol. 164, no. 1, pp. 87–96, 2001.
[24] C.-J. Lin, K.-L. Tsui, and C.-Y. Lin, "A spatial-EWMA framework for detecting clustering," *Qual. Rel. Eng. Int.*, vol. 30, no. 2, pp. 181–189, 2014.
[25] S. K. Greene, E. R. Peterson, D. Kapell, A. D. Fine, and M. Kulldorff, "Daily reportable disease spatiotemporal cluster detection, New York City, New York, USA, 2014–2015," *Emerg. Infect. Dis.*, vol. 22, no. 10, pp. 1808–1812,2016.
[26] Y. L. Strat and F. Carrat, "Monitoring epidemiologic surveillance data using hidden Markov models," *Statist. Med.*, vol. 18, no. 24, pp. 3463–3478, 1999.
[27] D. Conesa, M. Martínez-Beneito, R. Amorós, and A. López-Quílez, "Bayesian hierarchical poisson models with a hidden Markov structure for the detection of influenza epidemic outbreaks," *Stat. Methods Med. Res.*, vol. 24, no. 2, pp. 206–223, 2015.
[28] M. Frisén, "Principles for multivariate surveillance," *Front. Stat. Qual. Control*, vol. 9, pp. 133–144, 2010.
[29] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, vol. 20. Philadelphia, PA, USA: SIAM, 2007.
[30] R. Amorós, D. Conesa, A. López-Quílez, and M.-A. Martinez-Beneito, "A spatio-temporal hierarchical Markov switching model for the early detection of influenza outbreaks," *Stochastic Environ. Res. Risk Assessment*, vol. 34, no. 2, pp. 275–292, 2020.
[31] M. A. Martinez-Beneito, P. Botella-Rocamora, and O. Zurriaga, "A kernel-based spatio-temporal surveillance system for monitoring influenza-like illness incidence," *Stat. Methods Med. Res.*, vol. 20, no. 2, pp. 103–118, 2011.
[32] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental comparison of representation methods and distance measures for time series data," *Data Mining Knowl. Discov.*, vol. 26, pp. 275–309, 2013.
[33] H. A. Carneiro and E. Mylonakis, "Google trends: A web-based tool for real-time surveillance of disease outbreaks," *Clin. Infect. Dis.*, vol. 49, no. 10, pp. 1557–1564, 2009.
[34] H. Achrekar, A. Gandhe, R. Lazarus, S.-H. Yu, and B. Liu, "Predicting flu trends using twitter data," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2011, pp. 702–707.
[35] K. Lee, A. Agrawal, and A. Choudhary, "Real-time disease surveillance using twitter data: Demonstration on flu and cancer," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2013, pp. 1474–1477.
[36] D. Butler, "When Google got flu wrong," *Nature*, vol. 494, no. 7436, pp. 155–156, 2013.
[37] World Health Organization, *A Manual for Estimating Disease Burden Associated With Seasonal Influenza*. Geneva, Switzerland: World Health Organization, 2015.
[38] Public Health Agency of Canada, "Fluwatch program," 2015. [Online]. Available: http://www.phac-aspc.gc.ca/fluwatch/
[39] H.-M. Lu, D. Zeng, L. Trujillo, K. Komatsu, and H. Chen, "Ontology-enhanced automatic chief complaint classification for syndromic surveillance," *J. Biomed. Informat.*, vol. 41, no. 2, pp. 340–356, 2008.
[40] Canadian Association of Emergency Physicians, "Prehospital Canadian triage and acuity scale," 2016, Accessed: 2018. [Online]. Available: http://caep.ca/resources/ctas

[41] N. Marsden-Haug, V. B. Foster, P. L. Gould, E. Elbert, H. Wang, and J. A. Pavlin, "Code-based syndromic surveillance for influenzalike illness by international classification of diseases, ninth revision," *Emerg. Infect. Dis.*, vol. 13, no. 2, pp. 207–216, 2007.
[42] X. Zhu and D. Guo, "Mapping large spatial flow data with hierarchical clustering," *Trans. GIS*, vol. 18, no. 3, pp. 421–435, 2014.
[43] F. Zhou, F. De la Torre, and J. K. Hodgins, "Hierarchical aligned cluster analysis for temporal clustering of human motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 582–596, Mar. 2013.
[44] P. Montero and J. A. Vilar, "TSclust: An R package for time series clustering," *J. Stat. Softw.*, vol. 62, no. 1, pp. 1–43, 2014.
[45] X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, and P. Boesiger, "A new correlation-based fuzzy logic clustering algorithm for FMRI," *Magn. Reson. Med.*, vol. 40, no. 2, pp. 249–260, 1998.
[46] T. Giorgino et al., "Computing and visualizing dynamic time warping alignments in R: The DTW package," *J. Stat. Softw.*, vol. 31, no. 7, pp. 1–24, 2009.
[47] Centers for Disease Control and Prevention, "The flu season," 2017, Accessed: 2018. [Online]. Available: https://www.cdc.gov/flu/about/season/flu-season.htm
[48] R. Mastrandrea, J. Fournet, and A. Barrat, "Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys," *PLoS One*, vol. 10, no. 9, 2015, Art. no. e0136497.
[49] B. A. Bozick and L. A. Real, "The role of human transportation networks in mediating the genetic structure of seasonal influenza in the United States," *PLoS Pathogens*, vol. 11, no. 6, 2015, Art. no. e1004898.
[50] T. Bedford et al., "Global circulation patterns of seasonal influenza viruses vary with antigenic drift," *Nature*, vol. 523, no. 7559, pp. 217–220, 2015.
[51] E. Kenah, D. L. Chao, L. Matrajt, M. E. Halloran, and I. M. Longini Jr, "The global transmission and control of influenza," *PLoS One*, vol. 6, no. 5, 2011, Art. no. e19515.
[52] P. Lemey et al., "Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2," *PLoS Pathogens*, vol. 10, no. 2, 2014, Art. no. e1003932.
[53] Z. P. Ross et al., "Inter-seasonal influenza is characterized by extended virus transmission and persistence," *PLoS Pathogens*, vol. 11, no. 6, 2015, Art. no. e1004991.
[54] E. Azziz Baumgartner et al., "Seasonality, timing, and climate drivers of influenza activity worldwide," *J. Infect. Dis.*, vol. 206, no. 6, pp. 838–846, 2012.
[55] C. Viboud et al., "Association of influenza epidemics with global climate variability," *Eur. J. Epidemiol.*, vol. 19, no. 11, pp. 1055–1059, 2004.
[56] J. L. Geoghegan, A. F. Saavedra, S. Duchêne, S. Sullivan, I. Barr, and E. C. Holmes, "Continental synchronicity of human influenza virus epidemics despite climactic variation," *PLoS Pathogens*, vol. 14, no. 1, 2018, Art. no. e1006780.
[57] D. He, R. Lui, L. Wang, C. K. Tse, L. Yang, and L. Stone, "Global spatio-temporal patterns of influenza in the post-pandemic era," *Sci. Rep.*, vol. 5, no. 1, pp. 1–11, 2015.
[58] Statistics Canada, "Canada census," 2011, Accessed: 2018. [Online]. Available: http://www12.statcan.gc.ca/census-recensement/2011/as-sa/fogs-spg/Facts-csd-eng.cfm?Lang=eng&GK=CSD&GC=3520005
[59] C. Viboud, O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell, "Synchrony, waves, and spatial hierarchies in the spread of influenza," *Science*, vol. 312, no. 5772, pp. 447–451, 2006.
[60] H. Tian et al., "Avian influenza H5N1 viral and bird migration networks in Asia," *Proc. Nat. Acad. Sci.*, vol. 112, no. 1, pp. 172–177, 2015.
[61] F. A. Saad and V. K. Mansinghka, "A Bayesian nonparametric method for clustering imputation, and forecasting in multivariate time series," 2017, *arXiv:1710.06900*.
[62] L. Schiöler, "Characterisation of influenza outbreaks in Sweden," *Scand. J. Social Med.*, vol. 39, no. 4, pp. 427–436, 2011.
[63] V. Charu et al., "Human mobility and the spatial transmission of influenza in the United States," *PLoS Comput. Biol.*, vol. 13, no. 2, 2017, Art. no. e1005382.
[64] M. Frisén, "Spatial outbreak detection based on inference principles for multivariate surveillance," *IIE Trans.*, vol. 46, no. 8, pp. 759–769, 2014.
[65] D. Brockmann and D. Helbing, "The hidden geometry of complex, network-driven contagion phenomena," *Science*, vol. 342, no. 6164, pp. 1337–1342, 2013.
[66] R. F. Grais, J. H. Ellis, and G. E. Glass, "Assessing the impact of airline travel on the geographic spread of pandemic influenza," *Eur. J. Epidemiol.*, vol. 18, no. 11, pp. 1065–1072, 2003.
[67] A. C. Lowen and J. Steel, "Roles of humidity and temperature in shaping influenza seasonality," *J. Virol.*, vol. 88, no. 14, pp. 7692–7695, 2014.
[68] R. M. Anderson, R. M. May, and B. Anderson, *Infectious Diseases of Humans: Dynamics and Control*, vol. 28. Hoboken, NJ, USA: Wiley, 1992.
[69] S. P. Ribeiro et al., "Severe airport sanitarian control could slow down the spreading of COVID-19 pandemics in Brazil," *PeerJ*, vol. 8, 2020, Art. no. e9446.
[70] M. Norizuki et al., "Effective screening strategies for detection of asymptomatic COVID-19 travelers at airport quarantine stations: Exploratory findings in Japan," *Glob. Health Med.*, vol. 3, pp. 107–111, 2021.
[71] A. Sahasranaman and H. J. Jensen, "Spread of COVID-19 in urban neighbourhoods and slums of the developing world," *J. Roy. Soc. Interface*, vol. 18, no. 174, 2021, Art. no. 20200599.
[72] A. Sharifi and A. R. Khavarian-Garmsir, "The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management," *Sci. Total Environ.*, vol. 749, 2020, Art. no. 142391.