

Dynamic Poisson Autoregression for Influenza-Like-Illness Case Count Prediction

Zheng Wang¹, Prithwish Chakraborty², Sumiko R. Mekaru³, John S. Brownstein³
Jieping Ye^{1,4}, Naren Ramakrishnan²

¹Department of Computational Medicine and Bioinformatics,
University of Michigan, Ann Arbor, MI

²Discovery Analytics Center, Department of Computer Science, Virginia Tech, Arlington, VA

³Children’s Hospital Informatics Program, Boston Children’s Hospital, Boston, MA

⁴Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI

ABSTRACT

Influenza-like-illness (ILI) is among of the most common diseases worldwide, and reliable forecasting of the same can have significant public health benefits. Recently, new forms of disease surveillance based upon digital data sources have been proposed and are continuing to attract attention over traditional surveillance methods. In this paper, we focus on short-term ILI case count prediction and develop a dynamic Poisson autoregressive model with exogenous inputs variables (DPARX) for flu forecasting. In this model, we allow the autoregressive model to change over time. In order to control the variation in the model, we construct a model similarity graph to specify the relationship between pairs of models at two time points and embed prior knowledge in terms of the structure of the graph. We formulate ILI case count forecasting as a convex optimization problem, whose objective balances the autoregressive loss and the model similarity regularization induced by the structure of the similarity graph. We then propose an efficient algorithm to solve this problem by block coordinate descent. We apply our model and the corresponding learning method on historical ILI records for 15 countries around the world using a variety of syndromic surveillance data sources. Our approach provides consistently better forecasting results than state-of-the-art models available for short-term ILI case count forecasting.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications—
Data Mining

General Terms

Algorithms

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

KDD’15, August 11–14, 2015, Sydney, NSW, Australia.

© 2015 ACM. ISBN 978-1-4503-3664-2/15/08...\$15.00

DOI: <http://dx.doi.org/10.1145/2783258.2783291>.

Keywords

Time series methods, autoregressive models, flu forecasting

1. INTRODUCTION

Forecasting influenza-like-illness (ILI) case counts has been of great interest to epidemiologists for decades. Seasonal influenza regularly affects the global population and improvements in forecasting capability can directly translate into tangible measures of public health.

Traditionally, surveillance reports have been used as reference data and epidemiologists aim to predict several characteristics about ILI from such reports. Such characteristics of interest can be broadly classified into: (a) seasonal characteristics and (b) short-term characteristics. *Seasonal characteristics* are concerned with the overall shape of ILI counts for the particular season. Such methods are generally trained by assigning greater importance to statistics of the ILI curve such as peak value and the peak size. Conversely, short-term characteristics are concerned with accurately predicting the next few data points in absolute value rather than aiming for an overall fit for the season. In this paper we are motivated by the second problem, i.e. the short-term forecasting challenge (but we also evaluate our methods w.r.t. seasonal characteristics).

There are several challenges to ILI case count forecasting, one of the most important being the fact that surveillance reports are often delayed by a number of weeks and therefore estimating the current on-ground scenario is a crucial problem. The case count estimates for a given week can be delayed anywhere from 1 week to 4 weeks, depending on the quality of the surveillance apparatus in a given country. Thus in this paper we aim to provide *reliable* short-term forecasts from the last available surveillance data such that we can estimate the on-ground case counts.

In traditional epidemiology, several models such as SEIR and SIRS [12], have been proposed to model the temporal profile of infectious diseases. In modern computational epidemiology, more advanced methods have been used. One of the currently popular methods is to fit prediction models by matching observational data against a large library of simulated curves [4, 13, 21]. The curve simulations are generated by using different epidemiological parameters and assumptions. Sometimes network-based models are used to generate the curves [2]. Partially observed influenza counts

for a particular year can then be matched to a library of curves to produce the best set of predictions [13]. Closely related to such curve matching methods are filtering-based methods that dynamically fit epidemic models onto observed data by letting the base epidemic parameters vary over time. Yang et al. [22] provide an excellent survey of filtering-based methods used for influenza forecasting and also present comparative analysis of such methods.

Concurrently, there has been a lot of interest in using indicator data sources to predict seasonal influenza. In [8], Ginsberg et al. presented a method of estimating weekly influenza counts based on search query volumes (Google Flu Trend). Following this seminal work, researchers have investigated a wide-variety of data sources such as Wikipedia [9], Twitter [5, 10, 16], and online restaurant reservations [14]. Weather has been found to be a significant indicator of seasonal influenza [17, 18, 19, 20]. In [5], different indicator sources are contrasted to understand their relative influence on short-term forecasting quality.

As rich and varied as the above approaches are, most approaches in the literature aim to use the same model to predict for the entire influenza season. This is not entirely desirable as ‘in-season’ ILI characteristics may vary significantly from the ‘out-of-season’ characteristics (see Section 3.5). While researchers appreciate the need for dynamic models (e.g., [5]), constraints on temporal consistency are not explicitly imposed in current models. Thus in this paper we aim to propose a general purpose time series prediction model allowing external factors from indicator sources to produce robust short-term forecasts in a consistent manner.

A popular model for analyzing time series data is the autoregressive exogenous (ARX) model [3, 11]. The ARX model has also been adopted by Paul et al. [16] to predict ILI case counts by using Twitter and Google Flu Trends (GFT) as the indicator sources. However, the underlying static autoregressive model may not be suitable for flu trend forecasting, as the activity of the disease and the human living environment evolve over time. Ohlsson et al. [15] have designed a more flexible ARX model for time-varying systems based on model segmentation. It allows the weight of the autoregressive model to be temporally piecewise constant. In this paper, we further relax this requirement. We build separate models for each time point, but we constrain the models to share common characteristics. To capture such characteristics, we build a graph over models at different time points and embed the prior knowledge on model similarity in terms of the structure of the graph. Then we formulate the dynamic ARX model learning problem as a convex optimization problem, whose objective balances the autoregressive loss and the model similarity regularization induced by the graph structure. In this optimization problem, the variables have a natural block structure. Thus we apply a block coordinate descent method to solve this problem. We further extend our dynamic ARX modeling to the Poisson regression model for a better fitting of the count data [3, 6], as is relevant for ILI case counts forecasting. We perform extensive experimental studies to evaluate the effectiveness of the proposed model and the corresponding learning algorithm. We use various real world datasets in the experiments, including different types of indicator data sources from 15 countries around the world. Our experimental studies illustrate that the dynamic modeling of the linear Poisson autoregressive model captures well the under-

lying progression of disease counts. Further, our results also show that our proposed method outperforms state-of-the-art ILI case counts forecast methods.

Our main contributions are summarized as follows:

- We propose a new dynamic ARX model for the task of ILI case count forecasting. This approach incorporates a linear Poisson regression model with non-negativity constraints into an ARX model, ideal for case counts modeling.
- Prior domain knowledge can be encoded as structural relationships among different time points in a graph, which is embedded into the objective as a regularization term while still ensuring that the optimization problem is convex.
- We evaluate the proposed method using various real world datasets, including different types of indicator data sources from the USA and 14 Latin American countries.

2. CASE COUNT FORECASTING

Let us denote $\mathbf{y} = [y_1, y_2 \dots, y_N]$ as the known total weekly ILI case counts for a particular country under consideration. $y_t \in \mathbb{N}$ denotes the case count for time point t and N denotes the time point till which the ILI case count is known. Let us denote the available surrogate information for the same country by $\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_S^T]$, where S is the time point till which the surrogate information (i.e., indicator values) is available and $\mathbf{x}_t \in \mathbb{R}^m$ denotes the surrogate attributes for time point t . We aim to predict the number of ILI case counts at time point $t > T$.

2.1 ARX Model for Time Series Prediction

The ILI case count prediction is a time series modeling problem and thus autoregressive models are natural candidates for such problem. The order p AR(p) model can be written as

$$y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \varepsilon_t + c,$$

where the prediction is a linear combination of the target values at previous p time points with weight $\alpha_1, \dots, \alpha_p$. ε_t is a small random noise and c is a constant. However, AR models are only dependent on target signal itself and can be extended to include the effects of exogenous input variables to get ARX models as follows:

$$y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=0}^b \beta_i^T \mathbf{x}_{t-d-i} + \varepsilon_t + c,$$

where β_1, \dots, β_b are the weight parameters of the exogenous input variables \mathbf{x}_t with $\beta_i \in \mathbb{R}^m$, and d is the time delay between the input features and the output target.

The ARX model can be written in a compact form by rearranging the features as:

$$\begin{aligned} \mathbf{z}_t &= [\mathbf{x}_{t-d}, \dots, \mathbf{x}_{t-d-b}, y_{t-1}, \dots, y_{t-p}, 1] \\ y_t &= \mathbf{w} \mathbf{z}_t^T + \varepsilon_t \end{aligned}$$

where $\mathbf{w} = [\beta_{t-d}, \dots, \beta_{t-d-b}, \alpha_{t-1}, \dots, \alpha_{t-p}, c]$. This is a standard linear model, which can be fitted by minimizing

the least squares loss as

$$\min_{\mathbf{w}} \sum_t l(\mathbf{z}_t, y_t) = \sum_t (y_t - \mathbf{w}\mathbf{z}_t^T)^2,$$

where $l(\mathbf{z}_t, y_t)$ is a general loss function, and here we use the least squares approximation residual as the loss function.

2.2 Dynamic Model

The standard ARX model assumes that the model parameters do not change with time. However, in the case count forecasting problem, the behavior and the spread of the flu can vary over time, especially as a result of human interventions. Thus the prediction model should adapt in a similar manner. Ohlsson et al. [15] relaxed the linear model to piecewise constant linear model in ARX but the piecewise constant model still requires the nearby model to be the same, and thus a limitation for ILI case count forecasting. In this paper, we further relax this requirement by building a separate model for each time point. We assume that each model is similar to a group of others and formulate a dynamic model as:

$$y_t = \sum_{i=1}^p \alpha_i^{(t)} y_{t-i} + \sum_{i=0}^b (\beta_i^{(t)})^T \mathbf{x}_{t-d-i} + \varepsilon_t + c^{(t)}.$$

This dynamic autoregressive model, with time dependent parameters, can be compactly written as:

$$y_t = \mathbf{w}_t \mathbf{z}_t^T + \varepsilon_t,$$

where $\mathbf{w}_t = [\beta_{t-d}^{(t)}, \dots, \beta_{t-d-b}^{(t)}, \alpha_{t-1}^{(t)}, \dots, \alpha_{t-p}^{(t)}, c^{(t)}]$ are the time-dependent weights.

Without any constraints, the dynamic ARX model has too many parameters to be learnt effectively. We constrain this complexity by enforcing that, even in the presence of model changes, the models at different time points should still share some common structure. We model our prior knowledge of such model similarity as a graph structure, which is shown in Figure 1. Given a series of models within a period of time, we define a graph $\mathcal{G} = \{V, E\}$, where V is the node set composed of all models $\{\mathbf{w}_t\}$ at different time points and E is the edge set, which represents the similarity of the corresponding pair of nodes. For simplicity, we assign binary values of $S_{ij} = \{0, 1\}$ for each edge. $S_{ij} = 1$ indicates that the corresponding two nodes \mathbf{w}_i and \mathbf{w}_j should be similar to each other, otherwise $S_{ij} = 0$ to denote that there are no constraint between the corresponding nodes.

There are three types of useful graph structure for ILI case count forecasting (see Figure 2) which can be described as:

- *Fully connected graph*: All elements in \mathbf{S} are 1, except the diagonal elements. This is similar to the standard ARX model. However, it allows a tolerance of model variance.
- *Nearest neighbor graph*: All nearby models have an edge with element 1 in a small neighborhood area, which mean $S_{t,t+k} = 1, \forall k \leq K$. This is similar to piecewise constant model or fused lasso problem if we choose $K = 1$.
- *Seasonal nearest neighbor graph*: All nearby models in the same season have edge connections as 1 and the models at similar time points from different seasons have edge connections as 1. This is more useful for flu forecasting, where the trends are highly periodic.

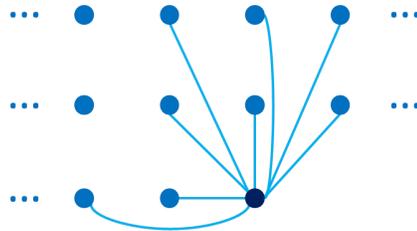


Figure 1: Modeling similarity as graph structures. In this graph, there are three consecutive years (rows). Nodes in each row denote the models in each year, which are ordered in time. The nodes from different years are aligned in the same order. For ILI forecasting, the last node should be similar to the nearby nodes in the same year as well as nodes in previous years it is connected to.

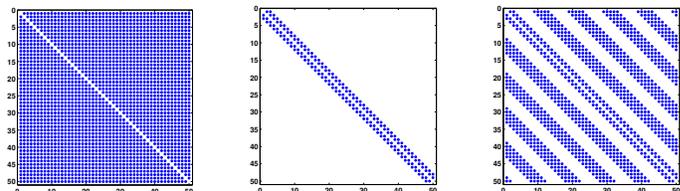


Figure 2: The graph structure for three different types of prior knowledge (fully connected, nearest neighbors, seasonal nearest neighbors).

Using Euclidean distance as a similarity measure, we obtain the following formulation of the problem:

$$\min_{\mathbf{W}} l(\mathbf{z}_t, y_t) + \eta \sum_{i,j} S_{ij} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2 + \gamma \|\mathbf{W}\|_F^2 \quad (1)$$

where, $\mathbf{W} = [\mathbf{w}_1^T, \dots, \mathbf{w}_N^T]$. In the objective function, the second term constrains the model similarity and the third term constrains the complexity of the models.

Reformulating the Frobenius norm term as $\sum_i 1 \|\mathbf{w}_i - \mathbf{0}\|_2^2$, we find that it to be similar to the other regularization term. Adding a null model node, $\mathbf{w}_0 = \mathbf{0}$, to the graph, and assuming it to be connected to all the other nodes with $S_{0i} = 1 \forall i > 0$, the regularization term can be simplified as:

$$\sum_{i,j=0,\dots,N} S_{ij} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2 \quad (2)$$

A simple form of the optimization problem can be then stated as:

$$\min_{\mathbf{W}} l(\mathbf{z}_t, y_t) + \eta \sum_{i,j} S_{ij} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2 \quad (3)$$

where, $\mathbf{w}_0 = \mathbf{0}$

2.3 Dynamic ARX Model

Using the least squares loss in (3), the objective function can be given as:

$$\sum_{i=1}^N (y_i - \mathbf{w}_i \mathbf{z}_i^T)^2 + \eta \sum_{i,j} S_{ij} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2. \quad (4)$$

The overall problem is then a convex optimization problem. We apply the block coordinate descent method (see

Algorithm 1) to optimize (4), as the weight \mathbf{w} naturally preserves a block structure. In block coordinate descent, we iteratively optimize until convergence each \mathbf{w}_i keeping all other blocks fixed. In each step, we solve a simple regression problem

$$\min_{\mathbf{w}_i} (y_i - \mathbf{w}_i \mathbf{z}_i^T)^2 + \sum_{t \neq i}^N (y_t - \mathbf{w}_t \mathbf{z}_t^T)^2 + \eta \sum_j S_{ij} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2 + \eta \sum_{t \neq i, j} S_{tj} \|\mathbf{w}_t - \mathbf{w}_j\|_2^2. \quad (5)$$

It can be simplified as

$$\min_{\mathbf{w}_i} \left(y_i - \mathbf{w}_i \mathbf{z}_i^T \right)^2 + \eta \sum_{j \in B_i} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2,$$

where B_i is the set of nodes connected with node i . The gradient of this objective function is:

$$2 \left(\mathbf{w}_i \mathbf{z}_i^T - y_i \right) \mathbf{z}_i + 2\eta \sum_{j \in B_i} (\mathbf{w}_i - \mathbf{w}_j).$$

Under first order optimality condition, this problem has a closed-form solution as

$$\mathbf{w}_i = \left(\mathbf{z}_i^T \mathbf{z}_i + \eta K_i \mathbf{I} \right)^{-1} \left(y_i \mathbf{z}_i^T + \eta \sum_{j \in B_i} \mathbf{w}_j \right),$$

where K_i is the number of the connected nodes for node i .

Algorithm 1 Dynamic Autoregressive Model with Exogenous Variables (DARX)

input data source \mathbf{X} , historical target \mathbf{y} .

1: Build the samples \mathbf{Z} , initial weight $\mathbf{W}^{(0)}$

2: **repeat**

3: **for** $i = 1, \dots, N$ **do**

4: Solve sub-problem (5) by

$$\left(\mathbf{z}_i^T \mathbf{z}_i + \eta K_i \mathbf{I} \right)^{-1} \left(y_i \mathbf{z}_i^T + \eta \sum_{j \in B_i} \mathbf{w}_j \right)$$

5: **end for**

6: **until** Terminated

output Regression weight \mathbf{W} .

2.4 Dynamic Poisson ARX for ILI Forecasting

Since ILI case count forecasting aims to predict the number of the infected people at different time points, it is natural to apply a Poisson regression model. We thus model the probability of the value of the response variable as a Poisson distribution:

$$\Pr(y) = \frac{\lambda^y e^{-\lambda}}{y!},$$

where λ is the expected count or the mean parameter and y is the count of events.

Given the case counts at different time points as $\{y_1, \dots, y_N\}$ and the associated input features as $\{\mathbf{z}_1, \dots, \mathbf{z}_N\}$, the likelihood and hence the log-likelihood can be expressed as:

$$\begin{aligned} \Pr(\{y_i\}|\{\mathbf{z}_i\}) &= \prod_{i=1}^N \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \\ \Rightarrow \log(\Pr(\{y_i\}|\{\mathbf{z}_i\})) &= \sum_{i=1}^N (y_i \log(\lambda) - \lambda - \log(y_i!)) \end{aligned}$$

Poisson regression model is a generalized linear model with the logarithm as the canonical link function:

$$\log(\mathbb{E}[y|\mathbf{z}]) = \log(\lambda) = \mathbf{w} \mathbf{z}^T.$$

Thus $\lambda = e^{(\mathbf{w} \mathbf{z}^T)}$, and the log-likelihood can be written as

$$\sum_{i=1}^N \left(y_i (\mathbf{w} \mathbf{z}_i^T) - e^{(\mathbf{w} \mathbf{z}_i^T)} - \log(y_i!) \right).$$

The weight \mathbf{W} is learnt by maximizing the log-likelihood:

$$\max_{\mathbf{w}} \sum_{i=1}^N \left(y_i (\mathbf{w} \mathbf{z}_i^T) - e^{(\mathbf{w} \mathbf{z}_i^T)} \right).$$

Alternately, we can use the identity link function to get $\lambda = (\mathbf{z} \mathbf{w}^T)$, which is called linear Poisson regression [6] and has more computational benefits. Under the identity link function, the negative log-likelihood of the training samples can be written as

$$l(\mathbf{X}, \mathbf{y}, \mathbf{W}) = - \sum_{i=1}^N \left(y_i \log(\mathbf{w} \mathbf{z}_i^T) - (\mathbf{w} \mathbf{z}_i^T) - \log(y_i!) \right).$$

2.5 Block Coordinate Descent Optimization

Applying the linear Poisson regression loss function described above in (3), we obtain a convex optimization problem as:

$$\begin{aligned} \min_{\mathbf{W}} \quad & \sum_i (\mathbf{w}_i \mathbf{z}_i^T - y_i \log(\mathbf{w}_i \mathbf{z}_i^T)) + \eta \sum_{i,j} S_{ij} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2 \\ \text{s.t.} \quad & \mathbf{w}_i \mathbf{z}_i^T \geq 0, \quad \forall i. \end{aligned}$$

The non-negative constraints are naturally satisfied as $\mathbf{w}_i \mathbf{z}_i^T$ is in the logarithm form in the objective.

Similar to the least squares loss case, we apply block coordinate descent to solve the above formulation (see Algorithm 2). In block coordinate descent, we iteratively optimize one block of variables with all other blocks fixed. In each step, the objective function is decomposed into

$$\begin{aligned} & (\mathbf{w}_i \mathbf{z}_i^T - y_i \log(\mathbf{w}_i \mathbf{z}_i^T)) + \sum_{t \neq i}^N (\mathbf{w}_t \mathbf{z}_t^T - y_t \log(\mathbf{w}_t \mathbf{z}_t^T)) \\ & + \eta \sum_j S_{ij} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2 + \eta \sum_{t \neq i, j} S_{tj} \|\mathbf{w}_t - \mathbf{w}_j\|_2^2. \end{aligned}$$

The minimization of the objective function in each step is a constrained regression problem:

$$\begin{aligned} \min_{\mathbf{w}_i} \quad & (\mathbf{w}_i \mathbf{z}_i^T - y_i \log(\mathbf{w}_i \mathbf{z}_i^T)) + \eta \sum_j S_{ij} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2 \\ \text{s.t.} \quad & \mathbf{w}_i \mathbf{z}_i^T \geq 0. \end{aligned} \quad (6)$$

This sub-problem is a convex optimization problem. However, it has no closed-form solution. We apply the projection gradient method to solve it iteratively. In each iteration, we first apply the Newton-Raphson method on the objective function and then project the obtained solution to the feasible set. For the Newton-Raphson gradient step, the gradient and the Hessian of the objective (6) can be given as:

$$\begin{aligned} \mathbf{g}_i &= \left(1 - \frac{y_i}{\mathbf{w}_i \mathbf{z}_i^T} \right) \mathbf{z}_i + 2\eta \sum_j S_{ij} (\mathbf{w}_i - \mathbf{w}_j) \\ \mathbf{H}_i &= \frac{y_i}{(\mathbf{w}_i \mathbf{z}_i^T)^2} (\mathbf{z}_i^T \mathbf{z}_i) + 2\eta \sum_j S_{ij} \end{aligned} \quad (7)$$

Thus, the update rule for each \mathbf{w}_i is

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \mathbf{H}_i^{-1} \mathbf{g}_i.$$

After this update, we project the obtained weight to the feasible set by

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \frac{\mathbf{w}_i \mathbf{z}_i^T}{\mathbf{z}_i \mathbf{z}_i^T} \mathbf{z}_i.$$

Algorithm 2 Dynamic Poisson Autoregressive Model with Exogenous Variables (DPARX)

input data source \mathbf{X} , historical target \mathbf{y} .
1: Build the samples \mathbf{Z} , initial weight $\mathbf{W}^{(0)}$
2: **repeat**
3: **for** $i = 1, \dots, N$ **do**
4: Solve sub-problem (6) by Newton-Raphson method
5: **repeat**
6: Calculate the gradient \mathbf{g}_i and Hessian matrix \mathbf{H}_i
7: $\mathbf{w}_i \leftarrow \mathbf{w}_i - \mathbf{H}_i^{-1} \mathbf{g}_i$
8: Project the solution to the feasible set
 $\mathbf{w}_i \leftarrow \mathbf{w}_i - \frac{\mathbf{w}_i \mathbf{z}_i^T}{\mathbf{z}_i \mathbf{z}_i^T} \mathbf{z}_i^T$
9: **until** Terminated
10: **end for**
11: **until** Terminated
output Regression weight \mathbf{W} .

2.6 Multi-step Forecasting

Reports of historical ILI counts are in general lagged by several weeks and revised for several weeks after they are first reported. Thus, in the practical case, we usually need to forecast the target several steps ahead. In s -step forward forecasting, we are only given the target signal till s weeks ago. For example, we have the training data set $\{\mathbf{x}_i\}_1^{N+s}$, $\{y_i\}_1^N$, and we are required to predict the target signal at time $T+s$. In this case, we directly model the s -step forward forecast problem as regression problem as:

$$y_t = \mathbf{w}_t \mathbf{z}_t^T + \varepsilon_t$$

where, $\mathbf{z}_t = [\mathbf{x}_{t-d}, \dots, \mathbf{x}_{t-d-b}, y_{t-s}, \dots, y_{t-s-p+1}, 1]$

We can still apply the dynamic Poisson ARX model and the corresponding learning procedure to the general s -step ahead forecast. *In the forecasting phase, we apply the most recent model for predicting the target at the upcoming time point.*

3. EXPERIMENTS

In our experiments, we compare our proposed method with several state-of-the-art ILI case count forecasting methods on a series of datasets for multi-step short-term forecasting. The following algorithms are used in the evaluation: ARX, matrix factorization with nearest neighbor regression (MFN) [5], segmentation of ARX model (SARX) [15], our dynamic ARX model (DARX), and finally the dynamic Poisson ARX model (DPARX).

3.1 Datasets

In this section, we give a summary of the data sources used. Table 1 lists these sources and their characteristics. In general for all the sources we used data for the period Jan 1, 2010 to Oct 8, 2014 as available latest on Jan 5, 2015.

3.1.1 Reference Source: PAHO/CDC

In this paper, we focus on United States (US) and 14 Latin American (LA) countries including Argentina (AR), Bolivia (BO), Chile (CL), Colombia (CO), Costa Rica (CR), Ecuador (EC), Guatemala (GT), Honduras (HN), Mexico (MX), Nicaragua (NI), Panama (PA), Peru (PE), Paraguay (PY) and, El Salvador (SV). For each country, we collected the weekly national counts of ILI infections as estimated

by the lab-based surveillance system present in the country (PAHO for LA and CDC for US). As discussed in [5], in general, these weekly estimates are lagged and regularly updated. However for this paper we focus on presenting a general purpose regression model allowing external variables and analyze the performance measures over the entire ILI season. As such, we considered the ILI count as available latest on Jan 5, 2015 and used a relatively stable portion of the ILI count spanning the period Jan 1, 2010 to Oct 8, 2014 for our analysis [5]. We considered the first 50 weeks as our fixed set and progressively used the remaining weeks to analyze the prediction accuracies.

3.1.2 Physical Indicators

Weather, specifically absolute humidity, has been found to be a strong indicator of ILI counts [18, 17, 19, 5]. In this paper, we used 6 hourly readings of 5 weather attributes including absolute humidity, relative humidity, precipitation, specific humidity, and temperature. These readings were accessed via the Global Data Assimilation System (GDAS) and downloaded in 1° lat/long resolution from <http://ladsweb.nascom.nasa.gov/>. Following [5], we used the weather readings for only 4 lat/long pairs around the surveillance centers for the country under investigation.

3.1.3 Social Indicators

Recent research [5, 9] has shown that there is significant signal hidden in online media, such as news, about prevalent ILI activity. We used Google Flu Trends (GFT), Google Search Trends (GST) and HealthMap (HM) as the selected social indicators sources.

GFT: Google Flu Trends (<http://google.org/flutrends>) is a tool based on [8] providing weekly estimates of ILI counts. These estimates are based on search query volumes as detected by Google. The system was recently updated [7] to make it more robust to sudden spikes in search activities. We used national level GFT data for the prespecified period. **GST:** As presented in [5], we used our in-house estimates of ILI related search activity by accessing Google trends (<https://google.com/trends>), which provides weekly estimates of search query volumes. We used the same 114 keywords as used in [5], constructed via a mix of pseudo-query expansion and correlation analysis. It is to be noted that these estimates are usually updated over time and we used a fixed download, similar to PAHO.

HealthMap: HealthMap (<http://healthmap.org>) is a global disease alert system capturing disease related data from over 50,000 electronic sources. We used flu-related news articles captured in the HealthMap system over the specified period and enriched the HTML content to create the weekly counts of 114 keywords for each country at a weekly level.

3.2 Evaluation Measures

Given the ground truth case counts $\{y_t\}$ and its prediction $\{\hat{y}_t\}$, the prediction error is measured in the form of

$$error = \frac{4}{N} \sum_{t=t_s}^{t_e} \frac{|y_t - \hat{y}_t|}{\max(y_t, \hat{y}_t, 10)},$$

where t_s and t_e indicate the starting and the ending time point for the predictions. N indicates the number of time points over the same time period (i.e. $N = t_e - t_s + 1$).

Table 1: Data source characteristics. Delayed refers to whether the data source is available for a given week in the same week or later. Revised refers to whether older values can get revised in future updates.

Characteristics	Num. Dimensions	Delayed?	Revised?	Temporal Resolution	Spatial Resolution
PAHO/CDC	1	Yes	Yes	Weekly	Country
Google Flue Trends (GFT)	1	No	Yes	Weekly	Country
Google Search Trends (GST)	114	No	Yes	Weekly	Country
Weather	5	No	No	6 hours \rightarrow Weekly	4 locations \rightarrow Country
HealthMap	114 \times 3	No	No	Daily \rightarrow Weekly	Country

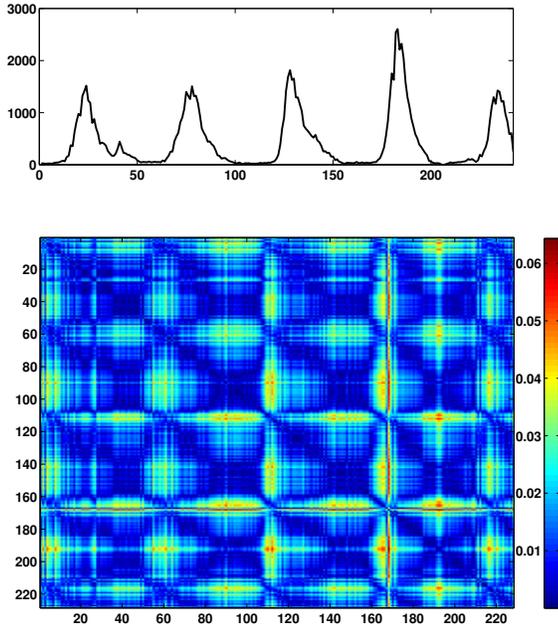


Figure 3: The distance matrix obtained from our learned DPARX model (bottom figure), associated with the ground truth ILI case count series (top figure) on the AR dataset. We can observe the strong seasonality automatically inferred in the matrix. Each element in the matrix is the Euclidean distance between a pair of the learned models at two corresponding time points after training. For the top figure, the x axis is the index of the weeks; the y axis is the number of ILI cases. For the bottom figure, both x and y axes are the index of the time points. Note that the starting time point (index 0) for the distance matrix is week 15 of the ILI case count series.

In the following experiments, we present the prediction accuracy, which is $4 - \text{error}$, on different countries to show the performance of different prediction models. A score of 4 thus indicates a perfect forecast.

3.3 Model Similarity

First, we conduct experiments to investigate the model similarities posited by our proposed algorithm. In this experiment, we calculate the distance between all pairs of models learned by DPARX during a period of time on the AR dataset. We present the distance matrix associated with the ground truth ILI case count series in Figure 3. We see that the distance matrix has a strong seasonal pattern, which is consistent with the pattern of the ILI case count series. At the beginning of each flu season, the model is significantly

different from the rest of the models at other time points. This result demonstrates that ILI case counts have a strong periodic pattern and the dynamic modeling approach successfully captures this pattern. It also validates the necessity of conducting this level of modeling for flu forecasting.

In the next experiment, we run our proposed DPARX method on the US dataset under three different model similarity graphs including the fully connected graph, the 3-nearest neighbor graph and the seasonal 3-nearest neighbor graph. We then calculate the three corresponding distance matrices of the learned models, which are shown in Figure 4. The patterns in the three distance matrices are very similar. However, the distances between the pairs of models are smaller for the fully connected similarity graph. Without strong prior knowledge, the fully connected similarity graph is preferred, as during different seasons the target signal may still be very different. In the following experiments, we will use the fully connected similarity graph for the regularization term.

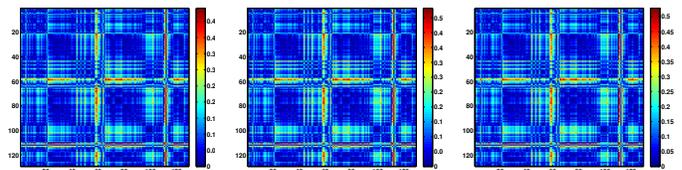


Figure 4: Model distance matrices for the US dataset. The three matrices are derived from the fully connected similarity graph, the 3-nearest neighbor similarity graph and the seasonal 3-nearest neighbor similarity graph, from left to right correspondingly.

3.4 Forecasting Results

In the ILI cast count forecast experiments, we use the data record from all 15 countries. All the case count data are associated with several data sources as listed in Table 1¹. We start with 50 given time points and test the prediction result on the remaining time points. We run all the competing methods in an online manner: the models are re-trained and updated after the arrival of values at every additional time point. For the DARX and DPARX models, we use the same parameter settings: $p = 1$, $b = 15$ for GFT and Weather data sources as these data sources have relatively small dimension; $p = 1$, $b = 4$ for GST and HealthMap data sources as these data sources have relatively high dimension. The ARX model does not provide numerical stable results for high dimensional data. Thus we present its results on GFT and Weather data sources with $p = 1$, $b = 15$. Likewise, the training of the SARX model is very time consuming, especially for high dimensional data. We thus only present

¹The GFT information is provided only for countries AR, BO, CL, MX, PE, PY and the US.

its results using the GFT data source with the same setting ($p = 1$ and $b = 15$). The remaining parameter in our model is the regularization parameter that controls the variation of the model. We fix it as $\eta = 1$ for the DARX model and $\eta = 5$ for the DPARX model during all experiments. For MFN algorithm, we follow the same procedure and parameter setting as in [5].

We present the results of short-term ILI case count forecasting for different countries with both 1-step forecast and multi-step forecasts with step sizes of 2, 3, and 4. The prediction accuracy on data sources GFT, Weather, GST, and HealthMap are presented in Tables 2, 3, 4, and 5, correspondingly.

The experiments show that our models yield better prediction accuracy, especially for multi-step forecasting. Multi-step forecast is a much harder task than 1-step forecast. The dynamic modeling of ARX provides more flexibility in handling the uncertainty associated with the target signal.

Table 2: Prediction accuracies for competing algorithms with different forecast steps over different countries using the GFT input source. GFT data is not available for other countries.

Step	Method	AR	BO	CL	MX	PE	PY	US
1	ARX	2.85	2.63	3.18	2.61	2.51	2.82	3.71
	MFN	2.33	2.41	2.34	2.69	2.48	2.54	3.73
	SARX	3.02	2.42	3.11	2.90	2.81	2.69	3.67
	DARX	3.05	2.74	3.12	2.78	2.50	2.65	3.71
	DPARX	3.13	2.82	3.18	2.97	2.64	2.81	3.72
2	ARX	2.38	2.22	2.83	1.88	1.90	2.57	3.47
	MFN	2.12	2.00	2.13	2.33	2.21	2.19	3.63
	SARX	2.75	2.03	2.76	2.64	2.43	2.43	3.64
	DARX	2.94	2.68	3.02	2.58	2.38	2.58	3.60
	DPARX	2.86	2.70	2.89	2.64	2.52	2.65	3.61
3	ARX	2.11	1.86	2.61	1.28	1.44	2.31	3.19
	MFN	1.99	1.87	2.11	2.14	2.10	2.09	3.33
	SARX	2.33	1.61	2.46	2.42	2.16	2.23	3.40
	DARX	2.66	2.36	2.77	2.37	2.26	2.46	3.41
	DPARX	2.58	2.53	2.56	2.45	2.37	2.52	3.42
4	ARX	1.84	1.61	2.39	0.88	1.12	2.22	2.92
	MFN	1.85	1.83	2.00	2.05	2.01	1.94	3.15
	SARX	2.12	1.41	2.30	2.22	2.02	2.09	3.30
	DARX	2.34	2.21	2.52	1.98	2.19	2.22	3.18
	DPARX	2.29	2.35	2.32	2.26	2.29	2.40	3.20

3.5 Seasonal Analysis

Thus far we have primarily focused on the near-term forecasting capability of the system. In addition to these near-term forecasts, epidemiologists are often interested in predictions about seasonal characteristics. Typically, for a complete ILI season, the seasonal quantities of interest are ‘start week’, ‘peak week’, ‘end week’, ‘peak size’ and ‘season size’. We present interpretations of these quantities and our methods for calculating the same² as follows:

Start week: Within a particular ILI year (may not be calendar year, e.g., in the USA, the ILI year spans from Epi Week 40 to Epi Week 39 [1]), ‘start week’ is the week from

²Our definitions here are motivated by how CDC calculates the same for the United States.

which ILI is said to be in season. We define start week for a ILI year to be the first week where the ILI count for 3 consecutive past weeks (including itself) is greater than a pre-defined threshold.

Peak week: Within a particular ILI year, the peak week is the week for which the ILI count is highest for that ILI year.

Peak Size: Peak Size is the ILI count observed on the peak week.

End week: Within a particular ILI year, the end week is the first week after the peak week such that ILI counts for 3 consecutive past weeks (including itself) is lower than a pre-defined threshold. End week signifies the end of the ILI season and is thus of interest to epidemiologists.

Season Size: Season size is used as a proxy for the size of the epidemic. It is calculated by summing up the total ILI count from the start to the end week.

Typically, the relevant thresholds are estimated by surveillance agencies based on expert knowledge. However, from our experience we have found that using 40% quantile for a year gives a reasonable threshold.

In this paper we have not trained the models to predict the aforementioned metrics. However, we can construct ILI prediction curves for each ‘step-ahead’, i.e., 1-step ILI prediction curve, 2-step ILI prediction curve and so on. From these prediction curves we can then calculate the season-characteristics and compare them against those calculated from the observed PAHO (or CDC) ILI counts.

We compare the predicted and observed seasonal characteristics, for the last ILI year in our set for each country, in Table 6³. As Table 6 shows, the proposed algorithms work well for a number of countries. In general DPARX performs better in terms of the overall prediction characteristics. This is consistent with our results for near-term forecasts. For seasonal characteristics, Weather and GFT seem to be the most important sources for prediction. We also present the predicted and real curves for Mexico for the ILI season 2013 in Figure 5 based on 1-step ahead predictions. Excepting GST and HealthMap data for some of the state-of-the-arts, all the curves match up closely to the observed ILI curve.

4. CONCLUSION

This paper concerns a practical short-term ILI case count forecasting problem based on multiple digital data sources. One of the main contributions of the proposed model is that the underlying autoregressive model is allowed to change over time. In order to control the variation of the model, we build a model similarity graph to indicate the relationship between each pair of models at two different time points and embed the prior knowledge as the structure of the graph. The experiments demonstrate that our proposed algorithm provides consistently better forecasting results than state-of-the-art time series models used for short-term ILI case count forecasting. We also observed that the dynamic model successfully captures the seasonal pattern of flu activity. In our future work, we plan to extend our proposed model to the multi-source learning case, and learn the prediction model with different data sources simultaneously. We expect the forecasting performance to be significantly improved by

³We only show the most important algorithms over the most relevant sources. For full table see <http://www.yelab.net/publications/ILI-KDD15>.

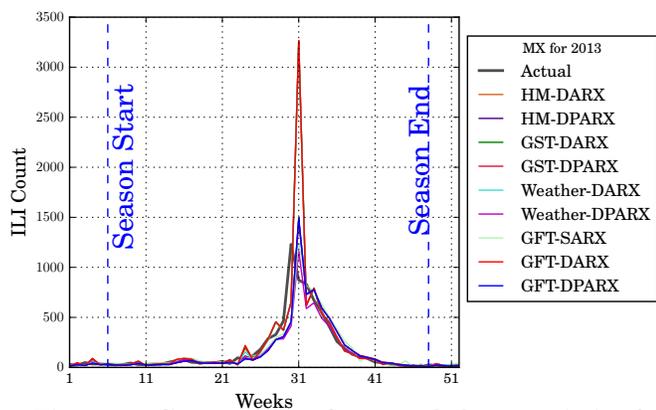


Figure 5: Comparison of seasonal characteristics for Mexico using different algorithms for one-step ahead prediction. Blue vertical dashed lines indicate the actual start and end of the season. ILI season considered: 2013.

properly fusing the data sources and the model. At the end, we would like to mention that though our paper focuses on predicting ILI case counts, the proposed dynamic ARX model is a general time series modeling technique and it is broadly applicable to a wider range of time series prediction problems.

Acknowledgments

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC000337. The US Government is authorized to reproduce and distribute reprints of this work for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusion contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the US Government.

5. REFERENCES

- [1] U.S. Flu Forecasting 2014 - SciCast. <https://scicast.org/flu>. Last Accessed: 2015-02-20.
- [2] A. Apolloni, V. A. Kumar, M. V. Marathe, and S. Swarup. Computational epidemiology in a connected world. *Computer*, 42(12):0083–86, 2009.
- [3] M. T. Bahadori, Y. Liu, and E. P. Xing. Fast structure learning in generalized stochastic processes with latent factors. In *Proceedings of KDD '13*, 2013.
- [4] K. R. Bisset, J. Chen, X. Feng, V. Kumar, and M. V. Marathe. Epifast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In *Proceedings of the ICS '09*, 2009.
- [5] P. Chakraborty, P. Khadivi, B. Lewis, A. Mahendiran, J. Chen, P. Butler, E. O. Nsoesie, S. R. Mekaru, J. S. Brownstein, M. V. Marathe, and N. Ramakrishnan. Forecasting a moving target: Ensemble models for ILI case count predictions. In *Proceedings of SDM '14*, 2014.
- [6] Y. Chen, D. Pavlov, and J. F. Canny. Large-scale behavioral targeting. In *Proceedings of KDD '09*, 2009.
- [7] P. Copeland, R. Romano, T. Zhang, G. Hecht, D. Zigmond, and C. Stefansen. Google disease trends: an update. *Nature*, 457:1012–1014, 2013.
- [8] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2008.
- [9] K. S. Hickmann, G. Fairchild, R. Priedhorsky, N. Generous, J. M. Hyman, A. Deshpande, and S. Y. Del Valle. Forecasting the 2013–2014 influenza season using wikipedia. *arXiv preprint arXiv:1410.7716*, 2014.
- [10] K. Lee, A. Agrawal, and A. Choudhary. Real-time disease surveillance using twitter data: demonstration on flu and cancer. In *Proceedings of the KDD '13*, 2013.
- [11] Y. Liu, M. T. Bahadori, and H. Li. Sparse-gev: Sparse latent space model for multivariate extreme value time series modelling. In *Proceedings of ICML '12*, 2012.
- [12] M. Marathe and A. K. S. Vullikanti. Computational epidemiology. *Communications of the ACM*, 56(7):88–96, 2013.
- [13] E. Nsoesie, M. Marathe, and J. Brownstein. Forecasting peaks of seasonal influenza epidemics. *PLoS currents*, 5, 2013.
- [14] E. O. Nsoesie, D. L. Buckeridge, and J. S. Brownstein. Who's not coming to dinner? evaluating trends in online restaurant reservations for outbreak surveillance. *Online Journal of Public Health Informatics*, 5(1), 2013.
- [15] H. Ohlsson, L. Ljung, and S. Boyd. Segmentation of arx-models using sum-of-norms regularization. *Automatica*, 46:1107–1111, 2010.
- [16] M. J. Paul, M. Dredze, and D. Broniatowski. Twitter improves influenza forecasting. *PLoS Currents*, 6, 2014.
- [17] J. Shaman, E. Goldstein, and M. Lipsitch. Absolute humidity and pandemic versus epidemic influenza. *American journal of epidemiology*, 173(2):127–135, 2010.
- [18] J. Shaman and A. Karspeck. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50):20425–20430, 2012.
- [19] J. Shaman, V. E. Pitzer, C. Viboud, B. T. Grenfell, and M. Lipsitch. Absolute humidity and the seasonal onset of influenza in the continental united states. *PLoS biology*, 8(2):e1000316, 2010.
- [20] J. D. Tamerius, J. Shaman, W. J. Alonso, K. Bloom-Feshbach, C. K. Uejio, A. Comrie, and C. Viboud. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathog*, 9(3):68–72, 2013.
- [21] M. Tizzoni, P. Bajardi, C. Poletto, J. J. Ramasco, D. Balcan, B. Gonçalves, N. Perra, V. Colizza, and A. Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 a/h1N1pdm. *BMC medicine*, 10(1):165, 2012.
- [22] W. Yang, A. Karspeck, and J. Shaman. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. *PLoS computational biology*, 10(4):e1003583, 2014.

Table 3: Prediction accuracies for competing algorithms with different forecast steps over different countries using the weather data source.

Step	Method	AR	BO	CL	CO	CR	EC	GT	HN	MX	NI	PA	PE	PY	SV	US
1	ARX	2.94	2.51	3.10	2.90	2.21	2.81	2.83	2.96	2.25	2.18	2.78	2.51	2.84	2.83	3.51
	MFN	2.99	3.01	2.88	2.53	2.78	2.81	2.77	2.83	2.61	2.70	2.56	2.82	2.66	2.79	3.81
	DARX	3.09	2.84	3.17	2.84	2.57	2.94	2.83	2.89	2.91	2.77	2.72	2.67	2.79	2.72	3.71
	DPARX	2.98	2.84	3.07	3.01	2.70	2.97	2.87	2.93	2.84	2.86	2.82	2.78	2.86	2.77	3.72
2	ARX	2.56	2.05	2.63	2.71	1.61	2.56	2.63	2.76	1.15	1.36	2.56	2.05	2.62	2.64	3.21
	MFN	2.86	2.89	2.81	2.49	2.71	2.67	2.72	2.41	2.55	2.31	2.50	2.59	2.71	2.30	3.75
	DARX	2.98	2.69	3.00	2.69	2.63	2.79	2.72	2.81	2.66	2.28	2.55	2.49	2.68	2.66	3.60
	DPARX	2.67	2.73	2.86	2.83	2.66	2.79	2.78	2.78	2.62	2.49	2.71	2.63	2.64	2.68	3.61
3	ARX	2.25	1.65	2.21	2.50	1.06	2.30	2.39	2.59	0.60	0.94	2.42	1.72	2.39	2.46	2.92
	MFN	2.49	2.38	2.41	2.33	2.45	2.31	2.32	2.10	2.21	2.11	2.19	2.22	2.40	2.08	3.64
	DARX	2.68	2.32	2.68	2.57	2.52	2.72	2.50	2.65	2.47	2.00	2.52	2.32	2.54	2.53	3.41
	DPARX	2.33	2.44	2.63	2.70	2.58	2.66	2.59	2.61	2.36	2.31	2.75	2.44	2.51	2.55	3.42
4	ARX	1.98	1.37	1.73	2.31	0.72	2.07	2.22	2.41	0.39	0.83	2.21	1.46	2.21	2.30	2.56
	MFN	2.10	2.13	2.15	2.04	2.25	2.11	2.22	1.94	1.99	1.87	2.01	1.86	2.10	1.77	3.54
	DARX	2.42	2.12	2.39	2.49	2.34	2.52	2.42	2.51	2.17	1.74	2.38	2.27	2.30	2.42	3.18
	DPARX	2.10	2.23	2.32	2.64	2.38	2.52	2.55	2.45	2.06	2.15	2.72	2.38	2.27	2.53	3.20

Table 4: Prediction accuracies for competing algorithms with different forecast steps over different countries using the GST data source.

Step	Dataset	AR	BO	CL	CO	CR	EC	GT	HN	MX	NI	PA	PE	PY	SV
1	MFN	2.61	2.44	2.55	2.22	2.61	2.52	2.31	2.62	2.48	2.61	2.31	2.23	2.53	2.13
	DARX	2.99	2.65	3.09	2.74	2.41	2.86	2.72	2.83	2.82	2.84	2.59	2.56	2.75	2.63
	DPARX	3.07	2.74	3.15	2.85	2.72	2.80	2.51	2.80	2.96	2.77	2.59	2.66	2.82	2.61
2	MFN	2.50	2.33	2.31	2.10	2.44	2.29	2.11	2.43	2.37	2.39	2.20	2.01	2.27	2.00
	DARX	2.83	2.54	2.94	2.57	2.53	2.69	2.58	2.72	2.59	2.40	2.35	2.40	2.54	2.51
	DPARX	2.78	2.59	2.86	2.67	2.63	2.67	2.35	2.71	2.60	2.48	2.43	2.53	2.57	2.59
3	MFN	2.33	2.10	2.16	1.99	2.21	2.03	1.99	2.14	2.20	2.14	2.02	1.91	2.13	1.92
	DARX	2.51	2.07	2.69	2.45	2.36	2.47	2.41	2.54	2.34	2.06	2.48	2.10	2.49	2.44
	DPARX	2.46	2.41	2.53	2.56	2.48	2.51	2.26	2.58	2.38	2.30	2.41	2.34	2.49	2.51
4	MFN	1.99	2.00	2.01	1.82	1.97	1.88	1.92	1.93	1.81	1.77	1.79	1.70	1.82	1.71
	DARX	2.16	1.91	2.36	2.24	2.20	2.17	2.28	2.40	1.80	1.86	2.40	2.06	2.23	2.36
	DPARX	2.17	2.21	2.29	2.46	2.35	2.33	2.14	2.46	2.10	2.13	2.33	2.21	2.30	2.44

Table 5: Prediction accuracies for competing algorithms with different forecast steps over different countries using the HealthMap data source.

Step	Dataset	AR	BO	CL	CO	CR	EC	GT	HN	MX	NI	PA	PE	PY	SV	US
1	MFN	2.81	3.13	2.63	2.58	2.91	2.77	2.63	2.73	2.50	2.61	2.54	2.69	2.51	2.61	3.78
	DARX	3.00	2.69	3.11	2.79	2.44	2.89	2.75	2.91	2.85	2.86	2.60	2.65	2.75	2.64	3.71
	DPARX	3.07	2.74	3.15	2.84	2.69	2.83	2.58	2.82	2.95	2.79	2.59	2.70	2.83	2.62	3.72
2	MFN	2.71	2.91	2.30	2.21	2.77	2.49	2.40	2.38	2.44	2.36	2.15	2.33	2.22	2.33	3.64
	DARX	2.86	2.60	3.01	2.62	2.54	2.74	2.64	2.77	2.66	2.47	2.37	2.47	2.53	2.58	3.60
	DPARX	2.78	2.60	2.88	2.67	2.62	2.71	2.44	2.72	2.60	2.50	2.45	2.58	2.58	2.60	3.61
3	MFN	2.44	2.30	2.42	2.07	2.31	2.14	2.28	2.01	2.19	2.12	1.99	2.00	1.97	1.95	3.35
	DARX	2.58	2.18	2.78	2.49	2.35	2.63	2.51	2.62	2.48	2.15	2.49	2.33	2.48	2.51	3.41
	DPARX	2.46	2.42	2.55	2.56	2.47	2.58	2.36	2.59	2.38	2.31	2.45	2.37	2.49	2.50	3.42
4	MFN	1.93	1.99	2.20	1.88	2.00	1.95	2.15	1.95	1.89	1.85	1.72	1.78	1.91	1.81	3.13
	DARX	2.28	2.02	2.46	2.39	2.19	2.37	2.39	2.45	2.22	1.97	2.45	2.26	2.20	2.42	3.18
	DPARX	2.17	2.21	2.30	2.44	2.34	2.42	2.25	2.47	2.12	2.14	2.37	2.25	2.30	2.47	3.21

Table 6: Comparison of seasonal characteristics extracted from predictions. For week matches such as start week, peak week and end week, we apply a lenient interpretation wherein any deviation less than 2 is considered to be true. Similarly, for value scores, an accuracy score of at least 3.0 (out of 4.0) was used to denote a correct prediction. Bold highlighted rows indicate forecasts where all the metrics were matched correctly.

Country, Step, Model	Start week			Peak Week			End Week			Peak Size			Season Size		
	act	pred	res	act	pred	res	act	pred	res	act	pred	res	act	pred	res
PE, 1, GFT-DPARX	3	3	True	21	22	True	36	37	True	308	259	True	2139	1712	True
PE, 1, Weather-DPARX	3	3	True	21	22	True	36	37	True	308	225	False	2139	1767	True
PE, 1, GFT-DARX	3	3	True	21	21	True	36	37	True	308	1028	False	2139	2922	False
PE, 1, Weather-DARX	3	3	True	21	21	True	36	37	True	308	733	False	2139	2579	True
PE, 2, GFT-DPARX	3	4	True	21	23	True	36	38	True	308	292	True	2139	1760	True
PE, 2, Weather-DPARX	3	4	True	21	23	True	36	38	True	308	229	False	2139	1744	True
PE, 2, GFT-DARX	3	4	True	21	23	True	36	36	True	308	1760	False	2139	4702	False
PE, 2, Weather-DARX	3	5	True	21	23	True	36	38	True	308	1228	False	2139	3411	False
PE, 3, GFT-DPARX	3	5	True	21	24	False	36	39	False	308	294	True	2139	1804	True
PE, 3, Weather-DPARX	3	5	True	21	24	False	36	39	False	308	229	False	2139	1731	True
PE, 3, GFT-DARX	3	5	True	21	24	False	36	37	True	308	2517	False	2139	7136	False
PE, 3, Weather-DARX	3	6	False	21	24	False	36	38	True	308	1276	False	2139	4101	False
PE, 4, GFT-DPARX	3	6	False	21	25	False	36	40	False	308	295	True	2139	1846	True
PE, 4, Weather-DPARX	3	6	False	21	25	False	36	39	False	308	227	False	2139	1721	True
PE, 4, GFT-DARX	3	6	False	21	25	False	36	37	True	308	3323	False	2139	9698	False
PE, 4, Weather-DARX	3	7	False	21	25	False	36	38	True	308	1136	False	2139	4658	False
BO, 1, GFT-DPARX	11	3	False	27	28	True	50	51	True	46	30	False	804	580	False
BO, 1, Weather-DPARX	11	13	True	27	28	True	50	51	True	46	44	True	804	752	True
BO, 1, GFT-DARX	11	3	False	27	15	False	50	21	False	46	57	True	804	217	False
BO, 1, Weather-DARX	11	3	False	27	28	True	50	51	True	46	51	True	804	886	True
BO, 2, GFT-DPARX	11	3	False	27	29	True	50	52	True	46	30	False	804	578	False
BO, 2, Weather-DPARX	11	3	False	27	29	True	50	52	True	46	45	True	804	841	True
BO, 2, GFT-DARX	11	3	False	27	29	True	50	52	True	46	84	False	804	1016	True
BO, 2, Weather-DARX	11	3	False	27	29	True	50	52	True	46	59	True	804	926	True
BO, 3, GFT-DPARX	11	3	False	27	30	False	50	52	True	46	29	False	804	573	False
BO, 3, Weather-DPARX	11	3	False	27	30	False	50	52	True	46	44	True	804	855	True
BO, 3, GFT-DARX	11	3	False	27	17	False	50	23	False	46	78	False	804	297	False
BO, 3, Weather-DARX	11	3	False	27	30	False	50	52	True	46	51	True	804	882	True
BO, 4, GFT-DPARX	11	3	False	27	31	False	50	52	True	46	29	False	804	565	False
BO, 4, Weather-DPARX	11	3	False	27	31	False	50	52	True	46	44	True	804	862	True
BO, 4, GFT-DARX	11	3	False	27	18	False	50	24	False	46	82	False	804	323	False
BO, 4, Weather-DARX	11	3	False	27	31	False	50	52	True	46	56	True	804	883	True
MX, 1, GFT-DPARX	5	6	True	30	31	True	47	48	True	1230	1494	True	7839	7445	True
MX, 1, Weather-DPARX	5	4	True	30	31	True	47	47	True	1230	1163	True	7839	6657	True
MX, 1, GFT-DARX	5	4	True	30	31	True	47	47	True	1230	3265	False	7839	9745	True
MX, 1, Weather-DARX	5	4	True	30	31	True	47	47	True	1230	3254	False	7839	9670	True
MX, 2, GFT-DPARX	5	7	True	30	32	True	47	49	True	1230	1562	True	7839	7627	True
MX, 2, Weather-DPARX	5	5	True	30	32	True	47	48	True	1230	1110	True	7839	6426	True
MX, 2, GFT-DARX	5	5	True	30	32	True	47	48	True	1230	4603	False	7839	12247	False
MX, 2, Weather-DARX	5	5	True	30	32	True	47	48	True	1230	4557	False	7839	12065	False
MX, 3, GFT-DPARX	5	8	False	30	33	False	47	50	False	1230	1579	True	7839	7804	True
MX, 3, Weather-DPARX	5	6	True	30	33	False	47	48	True	1230	1021	True	7839	6152	True
MX, 3, GFT-DARX	5	6	True	30	33	False	47	47	True	1230	5218	False	7839	15276	False
MX, 3, Weather-DARX	5	6	True	30	33	False	47	47	True	1230	5146	False	7839	14829	False
MX, 4, GFT-DPARX	5	9	False	30	34	False	47	51	False	1230	1621	True	7839	7966	True
MX, 4, Weather-DPARX	5	7	True	30	34	False	47	48	True	1230	908	False	7839	5856	False
MX, 4, GFT-DARX	5	5	True	30	34	False	47	46	True	1230	8195	False	7839	20954	False
MX, 4, Weather-DARX	5	7	True	30	34	False	47	46	True	1230	7869	False	7839	19719	False
CL, 1, GFT-DPARX	9	10	True	19	20	True	44	45	True	964	1013	True	12985	12587	True
CL, 1, Weather-DPARX	9	9	True	19	20	True	44	46	True	964	922	True	12985	12003	True
CL, 1, GFT-DARX	9	8	True	19	20	True	44	45	True	964	1215	True	12985	13699	True
CL, 1, Weather-DARX	9	8	True	19	20	True	44	46	True	964	1216	True	12985	13700	True
CL, 2, GFT-DPARX	9	11	True	19	21	True	44	46	True	964	1050	True	12985	12760	True
CL, 2, Weather-DPARX	9	10	True	19	21	True	44	43	True	964	929	True	12985	11682	True
CL, 2, GFT-DARX	9	9	True	19	21	True	44	44	True	964	1335	False	12985	14500	True
CL, 2, Weather-DARX	9	9	True	19	21	True	44	44	True	964	1338	False	12985	14440	True
CL, 3, GFT-DPARX	9	12	False	19	22	False	44	47	False	964	1093	True	12985	12904	True
CL, 3, Weather-DPARX	9	10	True	19	22	False	44	45	True	964	930	True	12985	11584	True
CL, 3, GFT-DARX	9	10	True	19	22	False	44	44	True	964	1552	False	12985	16003	True
CL, 3, Weather-DARX	9	10	True	19	22	False	44	44	True	964	1558	False	12985	15848	True
CL, 4, GFT-DPARX	9	13	False	19	23	False	44	48	False	964	1121	True	12985	13039	True
CL, 4, Weather-DPARX	9	11	True	19	23	False	44	49	False	964	902	True	12985	11348	True
CL, 4, GFT-DARX	9	10	True	19	23	False	44	47	False	964	1742	False	12985	18800	False
CL, 4, Weather-DARX	9	11	True	19	23	False	44	47	False	964	1752	False	12985	18068	False