



## Model selection for seasonal influenza forecasting



Alexander E. Zarebski <sup>a</sup>, Peter Dawson <sup>b</sup>, James M. McCaw <sup>a, c, d</sup>, Robert Moss <sup>c, \*</sup>

<sup>a</sup> School of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia

<sup>b</sup> Land Personnel Protection Branch, Land Division, Defence Science and Technology Organisation, Melbourne, Australia

<sup>c</sup> Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia

<sup>d</sup> Modelling & Simulation, Murdoch Childrens Research Institute, Royal Childrens Hospital, Melbourne, Australia

### ARTICLE INFO

#### Article history:

Received 31 October 2016

Received in revised form 16 December 2016

Accepted 16 December 2016

Available online 10 January 2017

### ABSTRACT

Epidemics of seasonal influenza inflict a huge burden in temperate climes such as Melbourne (Australia) where there is also significant variability in their timing and magnitude. Particle filters combined with mechanistic transmission models for the spread of influenza have emerged as a popular method for forecasting the progression of these epidemics. Despite extensive research it is still unclear what the optimal models are for forecasting influenza, and how one even measures forecast performance.

In this paper, we present a likelihood-based method, akin to Bayes factors, for model selection when the aim is to select for predictive skill. Here, “predictive skill” is measured by the probability of the data *after* the forecasting date, conditional on the data from *before* the forecasting date. Using this method we choose an optimal model of influenza transmission to forecast the number of laboratory-confirmed cases of influenza in Melbourne in each of the 2010–15 epidemics. The basic transmission model considered has the susceptible-exposed-infectious-recovered structure with extensions allowing for the effects of absolute humidity and inhomogeneous mixing in the population. While neither of the extensions provides a significant improvement in fit to the data they do differ in terms of their predictive skill. Both measurements of absolute humidity and a sinusoidal approximation of those measurements are observed to increase the predictive skill of the forecasts, while allowing for inhomogeneous mixing reduces the skill.

We discuss how our work could be integrated into a forecasting system and how the model selection method could be used to evaluate forecasts when comparing to multiple surveillance systems providing disparate views of influenza activity.

© 2017 KeAi Communications Co., Ltd. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Influenza causes regular but unpredictable seasonal epidemics in temperate climes. Due to the difficulties of large scale data collection there is increasing interest in “now-casting” the state of influenza to improve situational awareness (Ginsberg et al., 2009; Lazer, Kennedy, King, & Vespignani, 2014). Experimental evidence suggests a decrease in absolute humidity increases the influenza virus’ ability to transmit between hosts (Shaman & Kohn, 2009), potentially driving the distinctive seasonality of influenza epidemics in temperate climes. There is also much interest in understanding the impact of contact

\* Corresponding author.

networks on disease transmission, and how these effect the dynamics at a population level. Many infectious disease models assume the population mixes homogeneously (Allen, Brauer, van den Driessche & Wu, 2008; Stroud et al., 2006). An alternative approach is to restrain the rate of transmission in the model to account for inhomogeneity in the mixing of real populations (Chowell, Sattenspiel, Bansal, & Viboud, 2016; Ristic, Skvortsov, & Morelande, 2009; Roy & Pascual, 2006).

This paper investigates how allowing for the effects of absolute humidity and inhomogeneous mixing in the transmission process can improve our ability to: explain the observed influenza activity (now-casting) and predict future incidence (forecasting). Building on previous work with mechanistic models (Moss et al., 2015, 2016a, 2016b), a particle filter is used to predict the number of lab-confirmed cases of influenza observed by the Victorian Department of Health and Human Services (Australia), and to determine which model is most suitable for now-casting and forecasting. A Bayesian approach is used which can be applied more generally to the problem of selecting a model with the most predictive skill, where by “predictive skill”, we refer to the likelihood of “future” data (to be forecast), conditional on the data already observed.

Section 2 contains a description of the materials used in this analysis: the data for influenza activity and absolute humidity, the basic “transmission” model describing how influenza spreads in the community and its alternatives, and the “observation” model which connects the time series of notifications to the transmission model. Section 3 contains a description of the statistical techniques used to fit and interrogate the model along with the statistical basis for the model selection.

Section 4 contains the results of these analyses and in Section 5 we discuss the implications of this work for forecasting seasonal epidemics and how this methodology can be used for model selection when working with multiple surveillance systems providing disparate views of influenza activity.

## 2. Materials

### 2.1. Data

Influenza is a nationally notifiable disease in Australia, subsequently the Victorian Department of Health and Human Services (VDHHS) receives a notification for each specimen which tests positive for influenza in Victoria, Australia (Lambert et al., 2010). These notifications form the time series investigated in this paper. While the VDHHS is notified of positive tests, there are no data for negative tests. As a result it is difficult to distinguish between high levels of influenza activity and high ascertainment. The VDHHS captures only a small fraction of the total incidence of influenza in Victoria, the peak of the “burden of illness pyramid” (O’Brien et al., 2010; Wheeler et al., 1999). Despite these limitations, previous work ((Thomas, McCaw, Kelly, Grant, & McVernon, 2015) and (Moss et al., 2016b)) suggests that — of the data pertaining to influenza-like illness (ILI) and influenza activity generated by systems surveying this population — the VDHHS data are the least variable and most amenable to prediction. Subsequently, these data are thought to provide the best possible source of information surrounding the underlying dynamics.

While available at a daily resolution, the influenza notifications were aggregated by week to smooth the signal. Time series of relative humidity and temperature in Melbourne were obtained from the (Australian) Bureau of Meteorology (measurements taken every 3 hours). The absolute humidity (AH) was calculated from these and the results averaged over each day. These averages were then smoothed using a cubic spline (default `smooth.spline` in R) and scaled so the minimum and maximum values (over the whole 6 years) were  $-1$  and  $1$  respectively. Fig. 1 displays the AH and notification time series for each of the years considered in this study.

### 2.2. Model

The VDHHS notification time series has previously been modelled (Moss et al., 2015, 2016b) as the observations from a hidden Markov model (HMM) as represented in Fig. 2. The hidden Markov chain, and the model for the observations are described below. The hidden “transmission model” describes how influenza spreads in the community. This is represented by the middle layer of Fig. 2 and is described in Section 2.2.1. Section 2.2.2 describes the priors for the transmission models. The “observation model” links transmission to the data collected by the VDHHS; it is represented by the bottom layer of Fig. 2 and is described in Section 2.2.3. Absolute humidity is included in the top layer of the figure with each of the  $AH_t$  indicating the time series of AH between the observations  $Y_{t-1}$  and  $Y_t$ .

#### 2.2.1. Transmission model

The transmission model describes the spread of influenza in the population. The model is a susceptible-exposed-infectious-recovered (SEIR) type compartmental model (Anderson & May, 1992; Keeling & Rohani, 2008), where the state at time  $t$  (measured in days) is the number of people in each of the compartments,  $\mathbf{X}(t) = [S(t), E(t), I(t)]^T$ . The evolution of the state vector is governed by a system of stochastic differential equations described below. A closed population of  $N = 4,108,541$  is used; this figure was derived from population statistics for metropolitan Melbourne (Department of Health & Human Services, 2013). Since the population is closed, the number of people in the “recovered” compartment can be obtained from the conservation law,  $R = N - (S + E + I)$ .

Initially everyone in the population is assumed to be susceptible to the virus, hence the initial condition for the state vector is  $\mathbf{X}(0) = [N, 0, 0]^T$ . In real populations there will be people who are immune to the circulating strains (John et al., 2009; McCaw et al., 2009), this is not problematic; the particle filter will converge in regions of parameter space where the basic

Download English Version:

<https://daneshyari.com/en/article/5662904>

Download Persian Version:

<https://daneshyari.com/article/5662904>

[Daneshyari.com](https://daneshyari.com)