University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Zarebski, AE;Dawson, P;McCaw, JM;Moss, R

Title:
Model selection for seasonal influenza forecasting

Date:
2017-02-01

Citation:
Zarebski, A. E., Dawson, P., McCaw, J. M. & Moss, R. (2017). Model selection for seasonal influenza forecasting. Infectious Disease Modelling, 2 (1), pp.56-70. https://doi.org/10.1016/j.idm.2016.12.004.

Persistent Link:
https://hdl.handle.net/11343/198117

License:
CC BY-NC-ND

Infectious Disease Modelling 2 (2017) 56–70

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Model selection for seasonal influenza forecasting

Alexander E. Zarebski [a], Peter Dawson [b], James M. McCaw [a, c, d], Robert Moss [c, *]

[a] School of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia
[b] Land Personnel Protection Branch, Land Division, Defence Science and Technology Organisation, Melbourne, Australia
[c] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia
[d] 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.

http://dx.doi.org/10.1016/j.idm.2016.12.004
2468-0427/© 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/).

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, $\mathbf{X}(t) = [S(t), E(t), I(t)]^\top$. 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 $\mathbf{X}(0) = [N, 0, 0]^\top$. 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

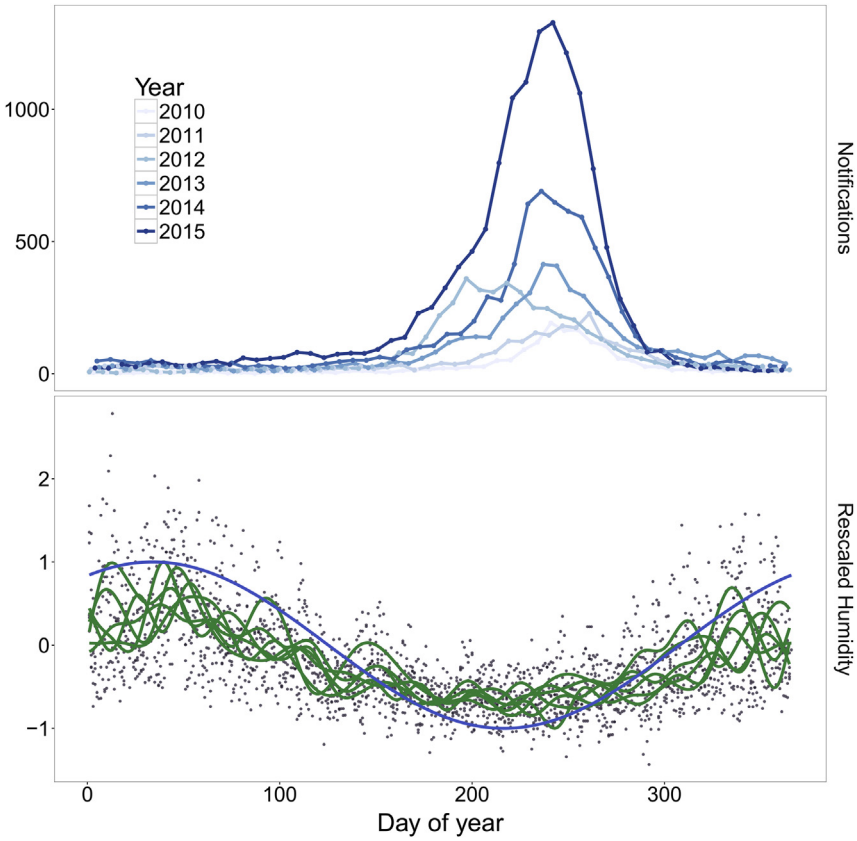


Fig. 1. (Top) Time series of the number of laboratory confirmed cases of influenza in Melbourne for 2010–15 aggregated by week. (Bottom) Scaled time series of the measurements of absolute humidity in Melbourne for 2010–15 in grey, with cubic spline smoothing in green and a sinusoidal approximation in blue. The minimum and maximum values over the whole six years were set to −1 and 1 respectively.

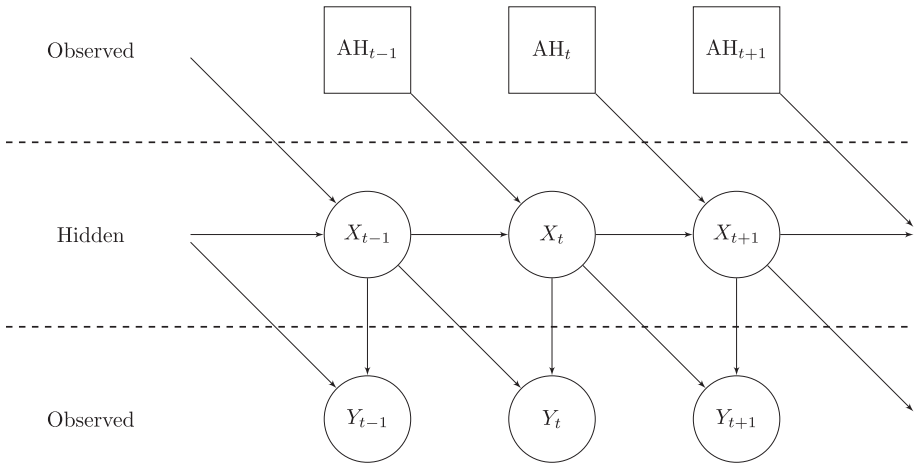


Fig. 2. Graphical representation of the hidden Markov model in which the hidden state, X_t, represents the state of the SEIR transmission model at time t and the observations, Y_t, the number of notifications over the week prior. The absolute humidity signal, AH_t, is assumed to be a deterministic function of time. The arrows indicate that: the evolution of the hidden state is dependent on its current state and the AH signal, and the observations are dependent on the current state of the hidden state and its state at the previous measurement.

reproductive number is lower to account for the increased transmission. This model supports a single outbreak assumed to start after a geometrically distributed number of days. During each day there is a fixed probability p_{exp} (set to $1/36$ (Moss et al., 2016a)) that a single individual will be exposed to the virus. This results in the state jumping to $[N-1, 1, 0]^\top$. After the initial exposure the states evolve by a system of stochastic differential equations (Allen et al., 2008a) of the form

$$d\mathbf{X}(t) = \underbrace{\mathbf{D}_1(\mathbf{X}(t))\ dt}_{\text{Drift}} + \varepsilon \underbrace{\mathbf{D}_2(\mathbf{X}(t))\ d\mathbf{W}(t)}_{\text{Diffusion}}. \tag{1}$$

The drift vector is given by

$$\mathbf{D}_1(\mathbf{X}) = \begin{bmatrix} -\beta\, N^{-1} SI \\ \beta\, N^{-1} SI - \sigma E \\ \sigma E - \gamma I \end{bmatrix}. \tag{2}$$

The average behaviour of the system over a small interval of time is the same as for the deterministic (ODE) SEIR model: $\beta I/N$ is the rate at which susceptible people are exposed to the virus, σ is the rate at which people become infectious after exposure, and γ is the rate at which people recover from being infectious. The diffusion matrix is given by

$$\mathbf{D}_2(\mathbf{X}) = \begin{bmatrix} -\sqrt{\beta\, N^{-1} SI} & 0 & 0 \\ \sqrt{\beta\, N^{-1} SI} & -\sqrt{\sigma E} & 0 \\ 0 & \sqrt{\sigma E} & -\sqrt{\gamma I} \end{bmatrix}. \tag{3}$$

Stochasticity is introduced into the transmission model by $d\mathbf{W}(t)$, a three-dimensional Weiner process. The diffusion matrix transforms this stochasticity so the variance is proportional to the mean behaviour (as specified by the drift vector). As in previous work (Moss et al., 2016a, 2016b; Skvortsov & Ristic, 2012) the parameter ε specifies this proportionality and is set to 0.025. This value was selected heuristically. A degree of stochasticity is required to avoid numerical issues (such as impoverishment) when using the particle filter, however too much reduces forecast skill. Experimentation with a larger value of ε did not qualitatively change the results but created a substantial bias towards underestimation as stochastic extinction is more likely with larger values of ε (data not shown).

The model described above constitutes the null hypothesis, $\mathcal{H}_0$, for the transmission process, i.e., that transmissibility is constant and the rate of exposure varies linearly with the number of people who are infectious. Two ways of extending this model are considered: the first allows for the effect of absolute humidity (AH) on the transmissibility parameter, β, and is denoted by $\mathcal{H}_{\text{AH}}$ when using the spline smoothed AH measurements and $\mathcal{H}_{\text{sine}}$ when using a sinusoidal approximation of the AH. The second allows for the effect of inhomogeneous mixing in the population, and is denoted by $\mathcal{H}_{\text{mix}}$.

The effects of AH are introduced by allowing β to vary with a humidity signal (Shaman, Pitzer, Viboud, Grenfell, & Lipsitch, 2010; Yaari, Katriel, Huppert, Axelsen, & Stone, 2013). We consider two such signals; the smoothed measurements of AH and a sinusoidal approximation to these measurements. This produces a time dependent rate of infectious contact, β_t, that varies linearly with the AH signal so that

$$\beta_t = \beta_1(1 + \beta_2 \text{AH}(t)) \tag{4}$$

To allow for inhomogeneous mixing, the factor $N^{-1}S$ in Equations (2) and (3) becomes $(N^{-1}S)^{\eta}$ for $1 \leq \eta \leq 2$ (Roy & Pascual, 2006; Stroud et al., 2006). Since $0 < N^{-1}S < 1$ this has the effect of reducing the transmission rate. This modification encapsulates the idea that the number of contacts an individual has with distinct people in the population saturates as the size of the population grows, and so the probability of encountering a new susceptible individual is diminished.

2.2.2. Prior distributions for transmission model parameters

The average incubation and infectious periods, σ^{-1} and γ^{-1} respectively, have uniform priors, $\mathcal{U}(1/2, 3)$ (Beauchemin & Handel, 2011; Nicholson, Wood, & Zambon, 2003). The ratio β_t/γ is the time-dependent basic reproduction number, $\mathcal{R}_0$, which governs much of the behaviour of this model. Initial samples of β_1 (equivalent to β in the null model) are obtained by sampling the ratio β_1/γ (given γ) from $\mathcal{U}(1, 3/2)$ and solving for β_1. As a result β_1 initially takes values from $1/3$ to $3/2$, while this may seem restrictive, it only holds for the initial samples. Due to the use of regularisation in the particle filter (described in Section 3.2.1) the full range of values $\mathcal{R}_0$ can take is $1/3$ to $9/2$ (Keeling & Rohani, 2008).

When allowing for the effects of absolute humidity in $\mathcal{H}_{\text{AH}}$ and $\mathcal{H}_{\text{sine}}$, a time dependent β_t is used which can vary by as much as 20% of β_1. This is achieved by putting a uniform distribution, $\mathcal{U}(-1/5, 0)$, on β_2, (Equation (4)). In $\mathcal{H}_{\text{mix}}$ the parameter η is given a uniform prior distribution, $\mathcal{U}(1, 2)$, (Stroud et al., 2006). In the null model, $\mathcal{H}_0$, the mixing is homogeneous and absolute humidity plays no role so the parameter β_2 is set to 0 and η is set to 1.

2.2.3. Observation model

Incidence, as captured by epidemiological surveillance systems, is the rate at which people enter a diseased (symptomatic) state. For the data considered here, this is the number of notifications received by the VDHHS per week, which for week n we denote Y_n. Let W_n denote the day during the n^{th} week when Y_n is observed, i.e., Y_n is the number of laboratory-confirmations

received by the VDHHS in the period $(W_{n-1}, W_n]$. The observation model specifies the distribution of Y_n given the cumulative *model* incidence over $(W_{n-1}, W_n]$.

Model incidence is the rate at which people move from the exposed to the infectious state, i.e., at time t this is $\sigma E(t)$. Integrating the model incidence over $(W_{n-1}, W_n]$, gives the probability, $p_{\text{inf}}(n)$, a random individual becoming infectious during that period of time, hence

$$p_{\text{inf}}(n) = \frac{1}{N} \int_{(W_{n-1}, W_n)} \sigma E(t) dt.$$

Since people can only move through the compartments sequentially, this integral simplifies to the difference in the proportion of people who *could* become infectious at the start of the week and the proportion of people who could *still* become infectious at the end of the week. Therefore,

$$p_{\text{inf}}(n) = \frac{S(W_{n-1}) + E(W_{n-1}) - (S(W_n) + E(W_n))}{N}.$$

Model incidence only accounts for those who became infectious during the week, ignoring exposures that did not progress to the infectious stage. Those who were *only* exposed are counted in the week in which their incubation is completed and they become infectious. This reflects our assumption that people begin to experience symptoms, and subsequently appear in surveillance systems, at the same time they become infectious. The observation model assumes that of the people who become infectious, on average, a constant proportion, p_{obs}, will be observed. Therefore, the expected number of notifications over $(W_{n-1}, W_n]$ is

$$\mu_{\text{epi}}(n) = N p_{\text{obs}} p_{\text{inf}}(n).$$

While this explains notifications during the flu season, influenza notifications are received throughout the year. The additional notifications are attributed to a background signal, i.e., over the period $(W_{n-1}, W_n]$ there are, on average, $\mu_{\text{bg}}(n)$ background notifications. It is assumed that everyone who is not part of the model incidence has some fixed probability, p_{bg}, of generating such a notification. Therefore, the expected number of background notifications in week n is given by

$$\mu_{\text{bg}}(n) = N p_{\text{bg}} \Big(1 - p_{\text{inf}}(n)\Big).$$

The number of notifications, Y_n, given the expected number of notifications, $\mu_n = \mu_{\text{epi}}(n) + \mu_{\text{bg}}(n)$, is modelled as a negative binomial random variable (Lindén & Mäntyniemi, 2011; Thomas et al., 2015), i.e.,

$$Y_n \Big| \mu_n, k \sim \text{NB}\left(\mu_n, \mu_n + \frac{\mu_n^2}{k}\right), \tag{5}$$

where the variance depends upon the dispersion parameter, k. Therefore the observation model requires three parameters to be completely specified, p_{bg}, p_{obs} and k.

3. Methods

In Section 2.2.1 a null model, $\mathcal{H}_0$, for the transmission of influenza was defined. Alternatives were also given: humidity modulated transmission, $\mathcal{H}_{\text{AH}}$ and $\mathcal{H}_{\text{sine}}$, and transmission restrained by inhomogeneous mixing, $\mathcal{H}_{\text{mix}}$. By integrating the system over the interval between observations, the transmission model is treated as a discrete time Markov chain, enabling application of standard particle filtering techniques for hidden Markov models (Doucet, Freitas, & Gordon, 2001; Sanjeev Arulampalam, Maskell, Gordon, & Clapp, 2002). The observation process is defined by the observation model of Section 2.2.3. Together these constitute the hidden Markov model shown in Fig. 2. Parameter estimation is carried out in a sequential Bayesian framework in which a sequence of approximate posterior densities are computed. Each posterior gives the distribution of the transmission model state and its parameters, conditional on the observations $Y_{1:j}$ for $j = 1, \ldots, n$. A mathematical description of the problem is given in Section 3.1. The computational scheme used to generate these approximations is the particle filter (PF) (Doucet & Johansen, 2009; Doucet et al., 2001; Sanjeev Arulampalam et al., 2002). Section 3.2 describes the PF used, how forecasts are generated with it and the specifics of its implementation. Section 3.3 provides implementation details and Section 3.4 describes the estimation of the data likelihood and how this is used in model selection. Code used to generate all of the results presented in this manuscript are available online (http://dx.doi.org/10.4225/49/5851d9ea54c65).

3.1. Problem statement

Except where specified otherwise, we assume the parameters of the observation model, p_{bg}, p_{obs} and k, are known. Let $Y_{1:T}$ denote the first T observations $Y_1, \ldots, Y_T$. The filtering problem involves computing the distribution of the hidden state, $\mathbf{X}(W_T) = [S(W_T), E(W_T), I(W_T)]^\top$, and its parameters, $\mathbf{\Theta} = [\beta, \sigma, \gamma]^\top$, conditional upon the available data at day W_T: $p(\mathbf{X}(W_T), \mathbf{\Theta} | Y_{1:T}, \mathcal{H}_X)$, which is referred to as the "filtering density". The inclusion of $\mathcal{H}_X$ in the notation makes explicit the assumption of a particular hypothesis for the transmission model. The *forecasting problem* requires computing the *predictive* distribution of the hidden state for the next H weeks, which is

$$p(\mathbf{X}(W_{T+i}) : i = 1, \ldots, H | Y_{1:T}, \mathcal{H}_X). \tag{6}$$

The predictive distribution of the hidden state can then be used to compute the predictive distribution of future observations, which is

$$p\left(Y_{(T+1):(T+H)} \middle| Y_{1:T}, \mathcal{H}_X\right). \tag{7}$$

Two model selection problems are considered: selecting the model which provides the best explanation of all the data, and the model which provides the best predictions of unobserved data given limited information from the start of the epidemic. The measure of "best" being the *a priori* probability of the data given the model, $\mathbb{P}(Y_{1:T} | \mathcal{H}_X)$ in the former and $\mathbb{P}(Y_{(T+1):(T+H)} \middle| Y_{1:T}, \mathcal{H}_X)$ in the latter.

3.2. Particle filter

3.2.1. Filtering

A bootstrap particle filter (PF) (Doucet & Johansen, 2009; Doucet et al., 2001; Sanjeev Arulampalam et al., 2002) was used to generate approximate samples ("particles") from the filtering density. During the filtering process re-sampling was used to avoid degeneracy, a numerical issue which occurs when the majority of the probability mass accumulates in a small subset of the particles. The particles were re-sampled if their effective number dropped below 25% of their total number using deterministic re-sampling (Douc & Cappé, 2005; Kitagawa, 1996).

Re-sampling causes particles with large weights to be duplicated, and those with low weights to be removed from the sample. Therefore, each re-sampling event reduces the number of distinct particles. Since the PF simultaneously estimates the hidden state, $\mathbf{X}(t)$, and the parameters, $\mathbf{\Theta}$, it is important that there is sufficient diversity among the particles to properly explore parameter space. Post-regularisation is used to maintain particle diversity, as described in (Musso, Oudjane, & Gland, 2001, pp. 247–271) using a Gaussian kernel. This involves randomly perturbing the particles in a systematic way during re-sampling to ensure they are distinct, while reducing the loss of information this causes.

After conditioning on the first T observations the PF consists of a weighted set of particles, $\mathcal{C}(W_T) = \{(\mathrm{P}^{(i)}, w^i) : \ i = 1, \ldots, M\}$. Together the particles and their weights define a discrete distribution which approximates the filtering density at day W_T. The particles, $\mathrm{P}^{(i)} = (\mathbf{X}^{(i)}(t), \mathbf{\Theta}^{(i)})$, consists of approximate samples from the filtering density over state and parameter space. These approximate samples make up the support points of the distribution, and the associated weights, w^i, their probabilities. The number of particles used for both filtering and forecasting was 7500.

3.2.2. Forecasting

Forecasting involves computing the distribution of future observations, i.e., approximating the distribution in Expression (7). This is done in two steps. First the filtering density is used to estimate the future hidden states using the set of particles and weights in $\mathcal{C}(W_T)$ to approximate the density in Expression (6). For each particle, $\mathrm{P}^{(i)}$, a trajectory $\widehat{\mathbf{X}}^{(i)}(\widehat{t})$ for $\widehat{t} \geq t$ is sampled (conditioning on $\widehat{\mathbf{X}}^{(i)}(t) = \mathbf{X}^{(i)}(t)$ and the parameters $\mathbf{\Theta}^{(i)}$). The ensemble of trajectories generated, the $\widehat{\mathbf{X}}^{(i)}$ and their corresponding weights w^i, are then treated as a weighted sample from the density in Expression (6).

The second step involves integrating over the hidden state to obtain the predictive distribution for the observations (Expression (7)). Using the discrete approximation from the first step the integral becomes a sum, and the distribution of the future observations can be expressed as a mixture of negative binomials (Equation (5)). The joint distribution of future observations is summarised by central credible intervals (CIs), i.e., a set the observations will fall with some specified probability. For example, the 100α% CI for a random variable Y (which takes values in $\mathbb{N}$) is the set of integers $[y_{\min}, y_{\max}] \subset \mathbb{N}$ such that $\mathbb{P}(y_{\min} \leq Y \leq y_{\max}) \geq \alpha$. Therefore the values of $y_{\min}$ and $y_{\max}$ are given by

$$y_{\min} = \max_{y \in \mathbb{N}} \left\{ y : \mathbb{P}(Y \geq y) \geq \alpha + \frac{1-\alpha}{2} \right\} \tag{8}$$

and

$$y_{\max} = \min_{y \in \mathbb{N}} \left\{ y : \mathbb{P}(Y \leq y) \geq \alpha + \frac{1-\alpha}{2} \right\}. \tag{9}$$

Since $y_{\min}$ and $y_{\max}$ are computed independently, this is not necessarily the smallest interval into which Y will fall with probability α, however it is simpler to calculate and the difference between the two methods is negligible. The computational cost of computing $y_{\min}$ and $y_{\max}$ can be significantly reduced by approximating each component in the mixture with a Gaussian distribution with the same mean and variance. An iterative method (Newton-Raphson) was used to compute the relevant percentiles in Equations (8) and (9).

3.3. Simulation parameters

The parameterisation of the observation model is treated separately to the estimation of the parameters in the transmission model. To assess the optimal performance of the transmission models a range of parameterisations of the observation model were tried and best parameters selected. This process of optimising the observation model, the method used to select the simulations dates, and the integration method for the transmission model are described in the following sections.

3.3.1. Simulation dates

Given the presence of the background signal in the notifications, there is no definitive method for identifying the start and end dates of a flu season. This raises the question of when one should start to generate forecasts. For each of the epidemic time series, 2010–15, we started the filtering process on the 1st of May. The end date of the filtering is defined to be the date of the first observation when the cumulative case count is at least 95% of the total cases for that year. Forecasts are generated for each of the 8 weeks prior to and including the week in which notifications peaked. The different forecasting dates are used investigate how the performance changes over the course of an epidemic. An example of the key dates in the filter/forecast process is given in Fig. 3. A full listing of the simulation dates is provided in Table 1.

3.3.2. Integration procedure

The system of SDEs describing the transmission model is integrated using the Euler-Maruyama method (Higham, 2001) with a time step of $\Delta t = 0.2$, i.e., the stochastic version of forward Euler. To ensure that during each day an epidemic is seeded with probability p_{exp} there is a Bernoulli trial at each time step with probability $1 - (1 - p_{\text{exp}})^{\Delta t}$ of seeding an epidemic. Once an epidemic has been seeded its evolution is governed by Equation (1). If a step results in any of the state variables leaving the

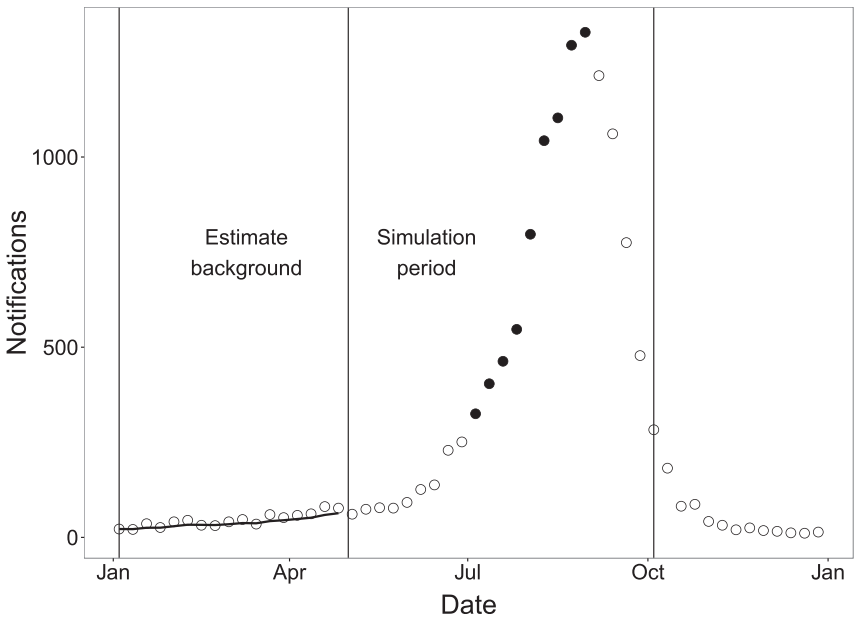


Fig. 3. Simulation periods for 2015. The first portion of the data (circles) is used to estimate the background notification rate via the exponentially weighted moving average (solid line). The second portion of the data is the target of the filtering and forecasting. The solid circles indicate the dates at which a forecast was generated.

Table 1
The dates of the first observation of the year, the date the filtering process was started, the date of the week with the most notifications and the date at which the filtering process ended. The filtering process was terminated once **95**% of the year's cases had been observed. See Fig. 3 for an example of the division of the time series into a period for background estimation, and the filtering/forecasting period.

First observation	Start date	Peak date	End date	Background estimate
2010-01-03	2010-05-01	2010-08-29	2010-12-05	6
2011-01-02	2011-05-01	2011-09-18	2011-11-06	17
2012-01-01	2012-05-01	2012-07-15	2012-11-04	17
2013-01-06	2013-05-01	2013-08-25	2013-12-01	14
2014-01-05	2014-05-01	2014-08-24	2014-10-26	28
2015-01-04	2015-05-01	2015-08-30	2015-10-04	64

range $[0, N]$ then they are clipped; any subsequent change in population size is corrected by adjusting the size of the recovered compartment.

3.3.3. Observation parameters

The observation model described in Section 2.2.3 has three parameters: p_{bg}, p_{obs} and k. The performance of the PF appears to be relatively insensitive to the dispersion parameter, k, and previous work suggests a value of 100 is appropriate (Moss et al., 2015, 2016b). Subsequently k is fixed at 100.

An estimate of the background notification rate can be obtained from the elements of the time series prior to the start date. This rate specifies the background notification probability, p_{bg}. Let $B_{1:M}$ be the number of notifications in the M weeks prior to the start of the simulation period. An exponentially weighted moving average of the signal, at the i^{th} week, A_i, is given by

$$A_1 = B_1 \quad \text{and} \quad A_n = \lambda B_n + (1-\lambda)A_{n-1}$$

with $\lambda = 0.25$ (Hunter, 1986). It is assumed that in the pre-simulation period the epidemic has not begun and so every notification is part of the background signal. This leads to the following running estimate of the background probability: $p_{\text{bg}} = [A_M]/N$, where A_M has been rounded to the nearest integer. To assess the sensitivity of the models to this estimate, several perturbations of the probability where used: $p_{\text{bg}} + j\Delta/N$ for $j = 0, 1, 2$ with $\Delta = 5$. The weighting constant, λ, was selected following the recommendations of Hunter (Hunter, 1986). However, its influence on p_{bg} is small relative to the perturbations, hence its selection is not expected to affect the fit.

Previous work suggests the performance of the PF is most sensitive to the observation probability, p_{obs}, (Moss et al., 2015). A range of plausible values can be obtained from an order of magnitude estimation. The population of Melbourne is in the millions, and each year there are thousands of notifications. Estimates of the annual attack rate of seasonal influenza range from $5-10$% in adults and $20-30$% in children (World Health Organization, 2014). This leads to an estimated observation probability of $\approx 10^{-2}$. Subsequently, values in $[10^{-3}, 5\times 10^{-2}]$ where considered. The PF appears to be most sensitive to p_{obs} at the lower end of the range so the set of values considered was uniformly spaced in the logarithm.

3.4. Model selection

In Section 2.2.1 alternative transmission models were defined: the null model, $\mathcal{H}_0$, the model accounting for absolute humidity, $\mathcal{H}_{\text{AH}}$ and the model accounting for inhomogeneous mixing, $\mathcal{H}_{\text{mix}}$. This paper aims to solve two model selection problems: the first is to determine the best fitting model, the second to determine the model which performs best at prediction. The former is about understanding the importance of climate and contact structure on transmission, the latter purely pragmatic. Bayes factors (BF) are used for the model selection in both cases; the estimation procedure for the BF and its application to each problem are described below.

3.5. Likelihood estimation

The key quantity for the model selection is the likelihood of the data, $Y_{1:T}$. The PF can estimate this quantity recursively. First observe that the likelihood can be factorised as

$$\mathbb{P}(Y_{1:T}) = \mathbb{P}(Y_1)\prod_{i=2}^{T}\mathbb{P}\left(Y_i\middle|Y_{1:(i-1)}\right). \tag{10}$$

While not explicit, in the equation above and for the rest of the derivation it is assumed that we are conditioning on a given hypothesis, $\mathcal{H}$. Each factor in the product can be expressed as an integral by conditioning on the relevant hidden state, therefore

$$\mathbb{P}\left(Y_i\middle|Y_{1:(i-1)}\right) = \int \mathbb{P}\left(Y_i\middle|x_{(i-1):i}\right)p\left(x_{(i-1):i}\middle|Y_{1:(i-1)}\right)dx_{(i-1):i}.$$

The PF provides an estimate of the integral with

$$\widehat{\mathbb{P}}\left(Y_i \middle| Y_{1:(i-1)}\right) = M^{-1} \sum_{\ell=1}^{M} \mathbb{P}\left(Y_i \middle| \widehat{X}^{(\ell)}_{(i-1):i}\right),$$

where $\widehat{X}^{(\ell)}_{(i-1):i}$ is the ℓ^{th} sample from the density $p(x_{(i-1):i} \middle| Y_{1:(i-1)})$. Since the distribution of the observation given the hidden state is known, each term of the sum can be computed. The PF does this by using stratified re-sampling to obtain a set of $\widehat{X}^{(\ell)}_{(i-1):i}$. This ensures that we have a uniformly weighted set of samples and reduces the variance in the estimator (Kitagawa, 1996; Ross, 1990). An estimator, $\widehat{\mathbb{P}}(Y_{(T+1):(T+H)} \middle| Y_{1:T})$, is constructed in the same way, however the trajectory of each sample is extended for H weeks.

3.5.1. Fitting

Bayes factors were estimated to determine which of $\mathcal{H}_0$, $\mathcal{H}_{\text{AH}}$, or $\mathcal{H}_{\text{mix}}$ provides the best fit. The Bayes factor is given by

$$B_{m0} = \frac{\mathbb{P}(Y_{0:T}|\mathcal{H}_m)}{\mathbb{P}(Y_{0:T}|\mathcal{H}_0)}, \tag{11}$$

for $\mathcal{H}_m \in \{\mathcal{H}_{\text{AH}}, \mathcal{H}_{\text{mix}}\}$. The probabilities in the numerator and denominator of Equation (11) are estimated using the PF as described above.

3.5.2. Forecasting

As with the model selection for fit, the forecasting ability of the PF using the various transmission models was assessed via Bayes factors, which in this situation is given by

$$\tilde{B}_{m0}(T) = \frac{\mathbb{P}\left(Y_{(T+1):(T+H)} \middle| Y_{0:T}, \mathcal{H}_m\right)}{\mathbb{P}\left(Y_{(T+1):(T+H)} \middle| Y_{0:T}, \mathcal{H}_0\right)}. \tag{12}$$

To estimate $\mathbb{P}(Y_{(T+1):(T+H)} \middle| Y_{0:T}, \mathcal{H}_m)$ the particle approximation for the joint density of $\mathbf{X}(W_T), \boldsymbol{\Theta} | Y_{0:T}, \mathcal{H}_m$ was integrated until W_{T+H}. Then the probability of the observations $Y_{(T+1):(T+H)}$ was estimated by summing the conditional probabilities of the observations for each of the trajectories in the ensemble.

The aggregate Bayes factors for the alternative models considering all the epidemics is calculated by assuming that each epidemic is independent and has its own parameters. As in the case of individual epidemics we assume that the optimal parameters of the observation model are known. Let $Y^{(y)}_{A:B}$ denote the notifications time series from year y, then the probability in the numerator of Equation (11) becomes

$$\prod_{y=2010}^{2015} \mathbb{P}\left(Y^{(y)}_{1:H_y} \middle| \mathcal{H}_m\right)$$

and the probability in the numerator or Equation (12) becomes

$$\prod_{y=2010}^{2015} \mathbb{P}\left(Y^{(y)}_{(T_y+1):(T_y+H_y)} \middle| Y_{0:T_y}, \mathcal{H}_m\right)$$

where the T_y and H_y are now year dependent. The former due to the selection of the forecasting dates which relies on aligning the epidemics by peak week, and the latter because of the method used to determine the end of the forecasting period.

4. Results

This paper considers two model selection problems, the first is to determine which of the transmission model hypotheses provides the best *fit* to the notification data (Section 4.1) and the second is to determine which produces the best *forecast* based on limited data from the start of an epidemic. The results for 2015 are given in Section 4.2 and the aggregate results over 2010–15 in Section 4.3.

4.1. Filtering results

Fig. 4 shows an example of the evolution of the filtering density for the null and sinusoidally forced models over the 2015 epidemic (See Supplementary Material Text 1 Fig. S1-1–S1-6 for the results from 2010 to 15). The whole notification time series for 2015 is shown in the figure, although the filtering density was only estimated for the observations falling in the

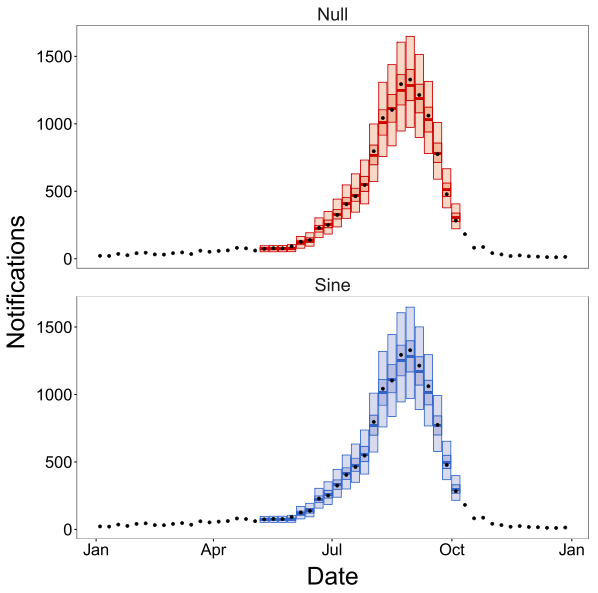


Fig. 4. The 50% and 95% credible interval for the observations under the filtering distribution for the null and sinusoidally forced models for the 2015 notification data. These running summaries of the observation distribution demonstrate both the null and sinusoidally forced models have near identical ability to assimilate new data, i.e., they have equal now-casting capabilities.

simulation period as described in Section 3.3.1. The number of notifications received for each of the weeks are represented by solid points. The box plots summarise the approximate distribution for each of the observations under the filtering distribution by showing the 50% and 95% credible intervals. As such, they act as a running fit of the model to the data and represent the most up-to-date posterior at any point in the simulation period. Supplementary Material Text 2 Fig. S2-1–S2-24 present the posterior samples for the transmission model parameters.

Using the method described in Section 3.5 the probability of the time series from the years 2010–15 arising from each of the transmission models can be estimated. The ratio of these probabilities are the Bayes factors, and the logarithm of these ratios are shown for each of the alternative models in Fig. 5. This figure shows that the transmission model that uses smoothed measurements of AH gives the best explanation of the data, however the improvement over the null model is not statistically significant. Allowing the transmissibility to vary sinusoidally, or allowing for inhomogeneous mixing, reduces the explanatory ability of the model. Again for the sinusoidal transmission hypothesis the difference is not statistically significant, however there is strong evidence that inhomogeneous mixing provides a weaker fit.

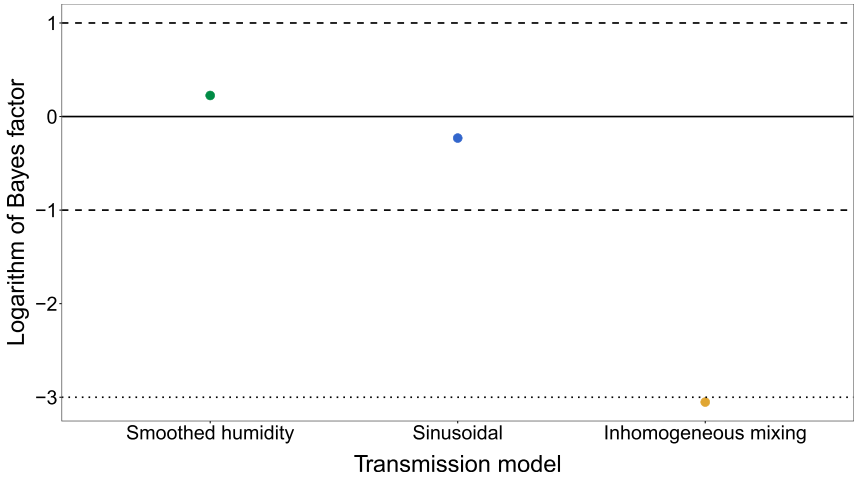


Fig. 5. The logarithm of the aggregate Bayes factor (across all the epidemics 2010–15) for each of the alternative transmission models. The solid horizontal line indicates parity with the null model, anything above this line is an improvement in model fit over the null, and below the fit is weaker. The dashed horizontal lines indicate the significance threshold.

4.2. Retrospective forecast for 2015

Much of the motivation for this work is in generating forecasts based on limited data from the start of an epidemic. Fig. 6 demonstrates some of these forecasts for the 2015 epidemic. It shows a sequence of forecasts produced by the null and sinusoidal models using increasing amounts of data from the start of the epidemic. The sinusoidal model is presented as it demonstrates a realistic forecasting tool, i.e., without the requirement of a long term, detailed forecast of AH. The full set of forecasts for all the models and years are available in Supplementary Material Text 1 Fig. S1-7–S1-30. As in Fig. 4 the number of notifications are shown as points. The solid points at the start of the epidemic were used to fit the model, which was then used to predict the values of the subsequent hollow points. This time the box plots represent a summary of the posterior distribution of the "future" observations. The rectangles collectively form the 50% and 95% credible interval (CI) of the forecast, i.e., the regions in which all future notifications are expected to fall with probabilities 0.5 and 0.95 respectively.

The first column of Fig. 6, ("Sinusoidal") contains the forecasts generated when the transmissability varies sinusoidally over the year, the second column, ("Null") those from the null model. For each row, the models were fit using all the data in the simulation period up until the date shown on the right. For example the first row contains the forecast generated using all the observations available on the 19th of July. For each row the model is fit to an additional element of the time series. As more data is used to train the PF the CIs should converge as the particles concentrate in regions of high posterior likelihood, resulting in narrower rectangles.

The estimated Bayes factors for the sinusoidally forced model's forecasts are also shown. The Bayes factor is largest in the second row, showing the improvement of the sinusoidal model over the null was the greatest for the forecasts generated on the 26th of July (for the weeks shown). The next section describes how the models performed when judged in this way considering all of the epidemics 2010–15.

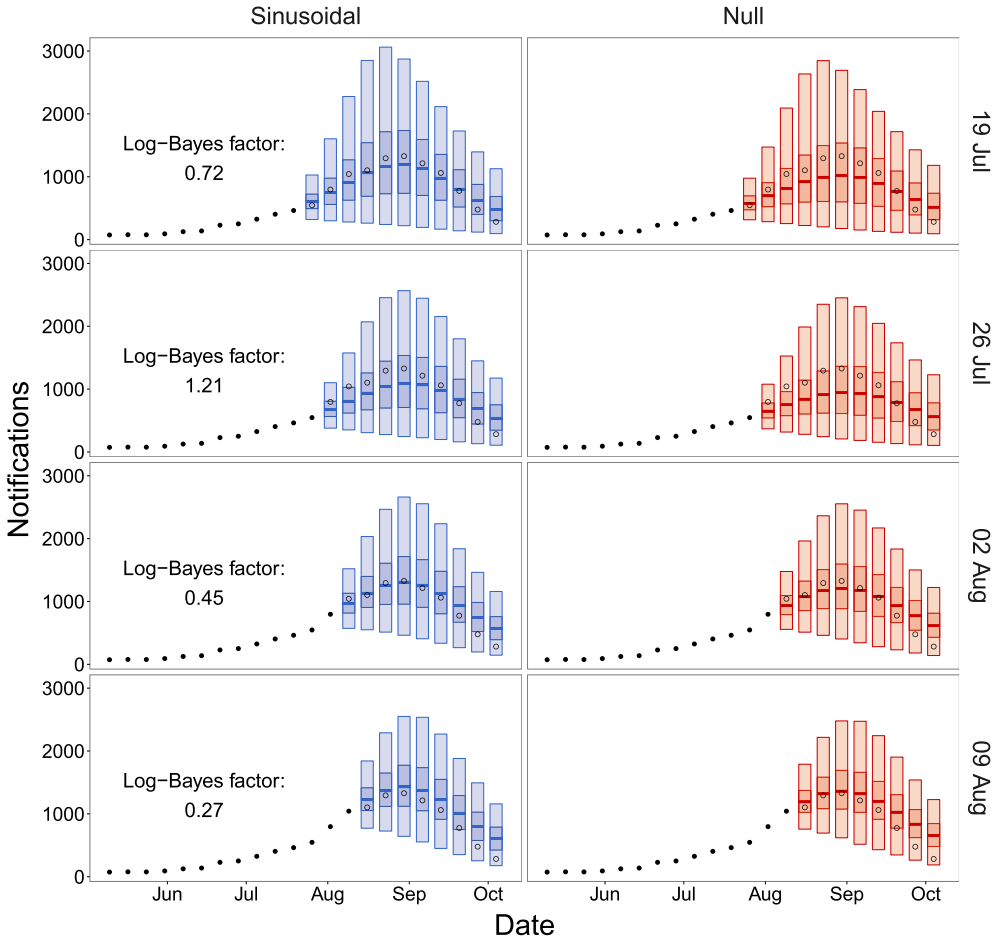


Fig. 6. Comparison of the forecasts from the null and sinusoidally forced transmission models using increasing amounts of data from the 2015 epidemic. The solid points represent "observed" data used to fit the model and the hollow points represent the "future" data, the target of the forecast. The logarithms of the Bayes factors reported describe the improvement in forecast performance by the sinusoidally forced model over the null for each of the forecasts generated.

4.3. Forecasting performance 2010–15

Fig. 7 shows the Bayes factors of the forecasts (across all the years) when aligning the epidemics by their peak week. For example, the values on the blue line were obtained by fixing the forecasting date (for all the years) relative to the peak week and computing the Bayes factor for the sinusoidally forced transmission model. Doing so gives one of the values on the curve, the full set of values coming from varying the forecasting date from 8 to 0 weeks prior to the peak week. The coloured curves represent the Bayes factors for the different alternative transmission models. This figure demonstrates that, given the correct observation model, there is strong evidence that the spline-smoothed absolute humidity model provides better forecasts than the null model over the 8 weeks leading up to the peak week. The improvement over the null model by the spline-smoothed model is largest when forecasting approximately one month prior to the peak. The sinusoidal model also outperforms the null model when generating forecasts over an interval of approximately a month, a month prior to the peak week. However, when generating forecasts more than 6 weeks prior to the peak or within a week of it, the performance of the sinusoidal model is weaker than that of the null model. As with the model fit, the forecasting performance of the inhomogeneous mixing model is poor for the majority of the season, although it does improve around the time of the peak.

An alternative method for investigating the differences between the forecasts is to consider the errors. Fig. 8 shows the average error in the forecast median as a function of the number of observed positive tests for all of the observations across the years 2010–2015. A point at (x, y) indicates that when forecasting an observation of x cases the average error in the prediction was y, so negative and positive values of y indicate underestimation and overestimation respectively. Since forecasts are generated at multiple weeks we average this error over the different forecasting dates. The solid lines show a LOESS smoothing of the scatter plot for each of the models to highlight the general trend. These show that for small numbers of positive tests the forecasts are reasonably unbiased, however all the models tend to underestimate larger observations.

5. Discussion

5.1. Principal findings

Accounting for the effects of absolute humidity (AH) does not significantly alter model fit and allowing for inhomogeneous mixing leads to worse fits. However, with a well parameterised observation model, allowing for the effects of AH improves forecasts of seasonal influenza. Moreover, even with a simple approximation of the AH (i.e., a sine wave) forecast performance is still improved. While the model accounting for inhomogeneous mixing leads to poor forecasts in the ascent phase of the epidemic once the peak has been reached this model appears to produce better forecasts of the descent phase. These results for model fit, and predictive skill relate to the aggregate performance of the models over all the epidemics from 2010 to 2015.

5.2. Study strengths and weaknesses

5.2.1. Strengths

While it is interesting to see that accounting for AH can improve forecast performance, a substitute signal (sine wave) has also been shown to improve forecast performance. This is an important observation as even the best predictor may be useless if we cannot obtain/predict it reliably.

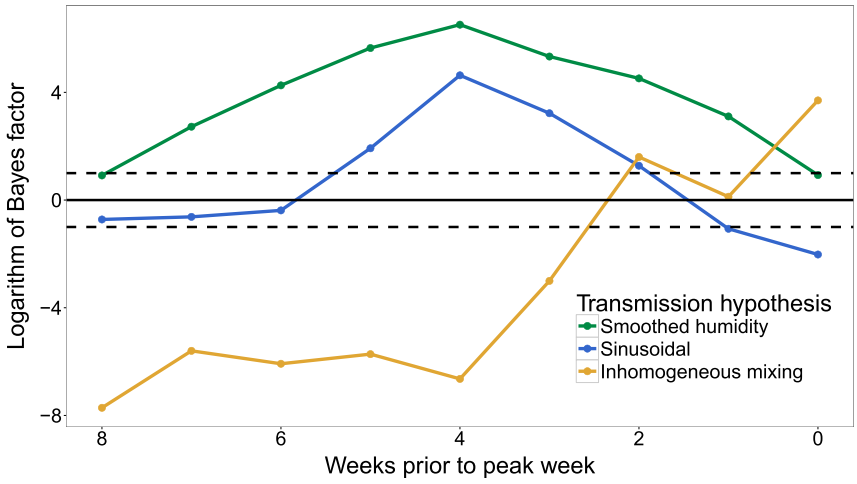


Fig. 7. The logarithm of the aggregate forecast Bayes factor (across all the epidemics 2010–15) for each of the alternative transmission models. The solid horizontal line indicates parity with the null model, anything above this line is an improvement in predictive skill. The dashed horizontal lines indicate the significance threshold.

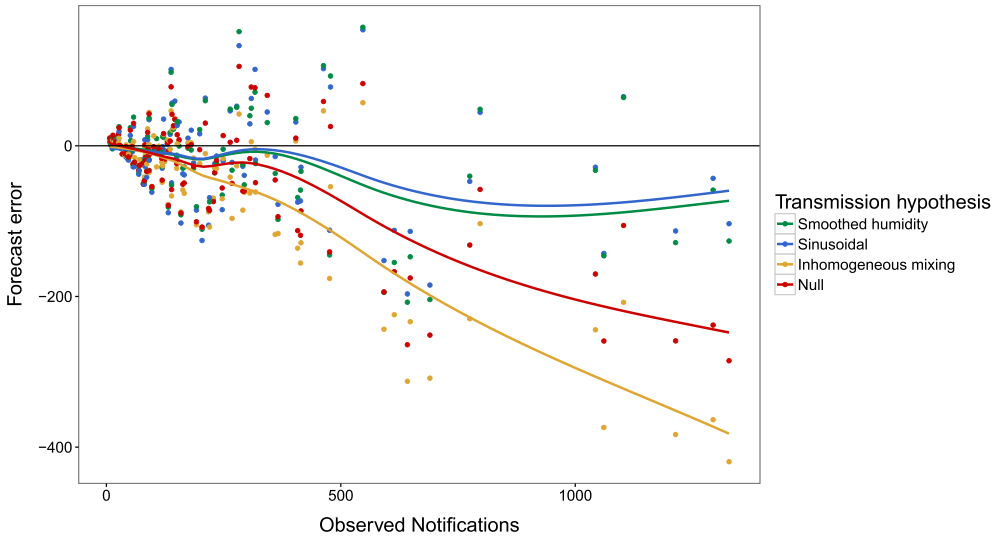


Fig. 8. Forecast error plotted against the size of the observation being forecast. Each point represents the error in attempts to forecast a single observation (averaged over the forecasts made at different points in the season). A point at (x, y) indicates that when forecasting an observation of x cases the average error in the prediction was y, so negative and positive values of y indicate underestimation and overestimation respectively. The colour of each point indicates which model was used to generate the forecast. The solid lines represents a LOESS smoothing of the data.

Attempts to forecast epidemics are often judged by their ability to predict one-dimensional summary statistics such as the timing of the peak of the epidemic (Chretien et al., 2014; Nsoesie, Brownstein, Ramakrishnan, & Marathe, 2014), in this paper a likelihood based measure is used. This measure is built upon the probabilistic model of the data and therefore avoids the need for further decision making in what to optimise for i.e., ability to predict peak time, magnitude or final size. Such an approach has several desirable properties, such as providing a natural way to assess forecast performance when there are multiple data streams. For instance, when using data from two surveillance systems showing the epidemic peaking at different times, it is unclear how to judge the accuracy of the forecast's peak time prediction. If forecast performance is assessed by the same metric as model fit, this method could be used to estimating forecast performance based on current fit to the data (i.e., does the model which best fits the most recent observation also best predict the next one?). Furthermore, the use of Bayes factors implicitly accounts for model complexity through the integration over the whole parameter space.

The PF lends itself naturally to estimation of credible intervals (CIs) of the parameters. In previous analyses of these data, forecasts consisted of the median estimate of the expected number of observations and only the CIs of this estimate. This analysis improves upon this by providing a summary of the predictive distribution of the observations. An iterative method has been presented to efficiently estimate the relevant quartiles. While the difference is negligible when case counts are high, when they are low, such as at the beginning and end of an epidemic, there is a substantial difference. Showing the uncertainty in the observation model gives a more realistic representation of the forecasts and should improve communication of these results.

5.2.2. Weaknesses

The greatest weakness of this work is the assumption of a known observation model. While estimation of these parameters was not a goal of this paper it should be kept in mind that the results presented here do require some knowledge of these parameters. Previous attempts at live forecasting in Melbourne, Australia (Moss et al., 2015) have highlighted that the observation probability, p_{obs}, in particular plays an important role in forecast performance. This presents a challenge since it appears that for Melbourne's influenza surveillance systems, this parameter changes from year to year (Moss et al., 2016b). Moreover, an important assumption of this work is that the observation probability is constant *within* a season. Without such an assumption, or a plausible alternative, it would be difficult to distinguish changes in transmission from changes in observation.

Another weakness of this study is the quantity of data. Each of the six years presents only a single time series, making if difficult to draw strong conclusions about either the transmission or observation processes. Due to this dearth of data the Bayesian framework is particularly attractive as it provides a more satisfactory quantification of the uncertainty.

When fitting nested models an effect size analysis can be informative. While the model selection revealed allowing for AH effects (i.e., allowing $\beta_2 > 0$) improved forecast performance, no analysis of the size of β_2 was performed. In part this is because of the sensitivity of the model to this parameter, which makes such an investigation difficult.

The transmission model assumes there is only a single circulating influenza strain and that initially everyone in the population is susceptible. While this is clearly false it is a useful assumption and with the existing data it would be challenging to parameterise a multi-strain model.

5.3. Comparison with other studies

5.3.1. Transmission model

Modelling the rate of transmission as dependent on AH has been done in several studies. The precise way in which the dependence is expressed varies though, with some models (such as the one presented here) using linear dependence (Bock Axelsen et al., 2014), while others using exponential dependence (Shaman & Kohn, 2009). To analyse this data set it has been assumed that the rate of infection should increase as AH decreases. This is because Melbourne has a temperate climate. In tropical climates the interaction between humidity and influenza transmission appears to be more complex and such an assumption may not be valid.

The method used to account for inhomogeneous mixing is also not the only one which appears in the literature (Chowell et al., 2016). However the majority of these approaches are similar at the core: the transmission rate should be retarded by saturation of local contacts, and this idea is realised in the functional form used here.

In the models considered in this paper it is only possible for individuals to progress through the compartments once. As a result an individual cannot be infected twice and since the population is closed the null model only supports a single epidemic. Extensions to this model allow for a loss of immunity resulting in transition from the recovered compartment back into the susceptible compartment and the incorporation of births and deaths which achieves similar changes in the dynamics. However, over the time scales considered here accounting for births and deaths will have a negligible effect.

5.3.2. Computational method

In this paper a particle filter (PF) has been used to fit and forecast the notification time series. The PF is an attractive method for this sort of analysis, and for practical forecasting work for several reasons. It is flexible in allowing for an arbitrary observation model, and maintains the non-linearity of the transmission model, whereas the Kalman filter and variants thereof usually require some form of linearisation and normality assumptions. However, this flexibility comes at the cost of potential numerical issues that arise from the use of a finite sample of particles, such as particle impoverishment and degeneracy. We have used re-sampling and regularisation to address these issues, however alternatives exist (Yang & Shaman, 2014). Alternative methods which do not require modification of the model include: particle MCMC (Doucet et al., 2001) and iterated filtering (Ionides et al., 2015). The former is significantly more computationally expensive and for the later it is less clear how one would incorporate the uncertainty in the parameter estimates into forecast generation.

5.4. Further work

The metric for forecast performance presented here provides a natural way to assess the performance of forecasts when comparing them to the data from multiple surveillance systems. This will assist in our future efforts to assimilate data from multiple sources, each with their own biasses. Another approach to improve predictive skill is to incorporate more prior knowledge into the forecasts, a task to which the Bayesian framework is well suited. Specifically, by constructing an informative prior based on previous epidemics it is reasonable to expect that initial uncertainty in the forecast can be reduced.

In addition to the main challenges listed above there are a number of attractive changes which could be made to the model. For instance, the background signal is a phenomenological modification and ignores the infectious potential of these individuals, this could be corrected for by forcing a proportion of the non-incident population into the infectious class. The analysis in (Mercer, Glass, & Becker, 2011) suggests this may improve the initial estimation of the reproduction number, potentially improving forecasts early in the season. Moreover, by allowing the dispersion parameter k to vary in Equation (5) it is possible to have quite a flexible mean-variance relationship in the observation model.

5.5. Meaning and implications

We have demonstrated that our existing forecast technology (Moss et al., 2015, 2016a, 2016b) can be improved by allowing for the effects of absolute humidity (AH) in the transmission model. While the true values of AH provide the largest improvement, even a simple approximation to the AH data (e.g. sinusoidal) is sufficient to improve predictive skill. The greatest improvement using a sine wave is seen when forecasting 5 to 2 weeks prior to the peak week. After the peak it is still possible to improve upon the null model by accounting for inhomogeneous mixing. However, there is little to gain in terms of now-casting by modifying the transmission model.

Methodologically this paper presents an objective function for forecast optimisation and an iterative scheme for approximating the credible intervals of the forecast. The former allows for a model selection based on a more comprehensive comparison of forecast and data, and provides a sensible way to optimise the forecasting tool to multiple data streams. The latter, by more naturally describing forecast uncertainty, improves the communication of the results.

Acknowledgements

This work was funded by the DSTG project “Bioterrorism Preparedness Strategic Research Initiative 07/301”. James M. McCaw is supported by an ARC Future Fellowship (FT110100250). We thank Nicola Stephens, Lucinda Franklin and Trevor Lauer (VDHHS) for providing access to, and interpretation of, Victorian influenza surveillance data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.idm.2016.12.004.

References

Allen, Edward J., Allen, Linda J. S., Arciniega, Armando, & Greenwood, Priscilla E. (2008). Construction of equivalent stochastic differential equation models. *Stochastic Analysis and Applications, 26*(2), 274–297.

Allen, Linda J. S., Brauer, Fred, van den Driessche, Pauline, & Wu, Jianhong (2008). *Mathematical epidemiology*. Springer.

Anderson, Roy M., & Robert, May M. (1992). *Infectious diseases of Humans: Dynamics and control, 28*. Wiley Online Library.

Beauchemin, Catherine A. A., & Handel, Andreas (2011). A review of mathematical models of influenza A infections within a host or cell culture: Lessons learned and challenges ahead. *BMC Public Health, 11*(1), 1.

Bock Axelsen, Jacob, Yaari, Rami, Grenfell, Bryan T., & Stone, Lewi (2014). Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. *Proceedings of the National Academy of Sciences, 111*(26), 9538–9542.

Chowell, Gerardo, Sattenspiel, Lisa, Bansal, Shweta, & Viboud, Cécile (2016). Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*.

Chretien, Jean-Paul, George, Dylan, Shaman, Jeffrey, Chitale, Rohit A., & Ellis McKenzie, F. (2014). Influenza forecasting in human populations: A scoping review. *PloS ONE, 9*(4), e94130.

Department of Health & Human Services. (2013). *2013 local government area profiles*. Technical report, Victorian Government (Accessed 3 September 2014).

Douc, Randal, & Cappé, Olivier (2005). Comparison of resampling schemes for particle filtering. In *ISPA 2005. Proceedings of the 4th International Symposium on image and signal processing and analysis, 2005* (pp. 64–69). IEEE.

Doucet, Arnaud, Freitas, Nando de, & Gordon, Neil (2001). *Sequential monte carlo methods in practice*. Springer Science & Business Media.

Doucet, Arnaud, & Johansen, Adam M. (2009). A tutorial on particle filtering and Smoothing: Fifteen years later. *Handbook of Nonlinear Filtering, 12*(656–704), 3.

Ginsberg, Jeremy, Mohebbi, Matthew H., Patel, Rajan S., Brammer, Lynnette, Smolinski, Mark S., & Brilliant, Larry (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012–1014.

Higham, Desmond J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review, 43*(3), 525–546.

Hunter, Stuart J. (1986). The exponentially weighted moving average. *Journal of Quality Technology, 18*(4), 203–210.

Ionides, Edward L., Nguyen, Dao, Atchadé, Yves, Stoev, Stilian, & King, Aaron A. (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences, 112*(3), 719–724.

Mathews, John D., Chesson, Joanne M., McCaw, James M., & McVernon, Jodie (2009). Understanding influenza transmission, immunity and pandemic threats. *Influenza and Other Respiratory Viruses, 3*(4), 143–149.

Keeling, Matt J., & Rohani, Pejman (2008). *Modeling infectious diseases in humans and animals*. Princeton University Press.

Kitagawa, Genshiro (1996). Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics, 5*(1), 1–25.

Lambert, Stephen B., Faux, Cassandra E., Grant, Kristina A., Williams, Simon H., Bletchly, Cheryl, Catton, Michael G., et al. (2010). Influenza surveillance in Australia: We need to do more than count. *Medical Journal of Australia, 193*(1), 43–45.

Lazer, David, Kennedy, Ryan, King, Gary, & Vespignani, Alessandro (2014). The parable of google Flu: Traps in big data analysis. *Science, 343*(14 March).

Lindén, Andreas, & Mäntyniemi, Samu (2011). Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology, 92*(7), 1414–1421.

McCaw, James M., McVernon, Jodie, McBryde, Emma S., & Mathews, John D. (2009). Influenza: Accounting for prior immunity. *Science, 325*(5944), 1071.

Mercer, Geoff N., Glass, Kathryn, & Becker, Niels G. (2011). Effective reproduction numbers are commonly overestimated early in a disease outbreak. *Statistics in Medicine, 30*(9), 984–994.

Moss, Robert, Fielding, James E., Franklin, Lucinda J., Kelly, Heath A., Stephens, Nicola, McVernon, Jodie, et al. Live forecasting of the 2015 Melbourne influenza season using lab-confirmed influenza cases. Under review.

Moss, Robert, Zarebski, Alexander, Dawson, Peter, & McCaw, James M. (2016). Forecasting influenza outbreak dynamics in Melbourne from Internet search query surveillance data. *Influenza and Other Respiratory Viruses*.

Moss, Robert, Zarebski, Alexander, Dawson, Peter, & McCaw, James M. (2016). Retrospective forecasting of the 2010–2014 melbourne influenza seasons using multiple surveillance systems. *Epidemiology & Infection*, 1–14.

Musso, Christian, Oudjane, Nadia, & Gland, Francois Le (2001). Improving regularised particle filters. In *Sequential monte carlo methods in practice*. Springer.

Nicholson, Karl G., Wood, John M., & Zambon, Maria (2003). *Influenza. Lancet, 362*, 1733–1745.

Nsoesie, Elaine O., Brownstein, John S., Ramakrishnan, Naren, & Marathe, Madhav V. (2014). A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses, 8*(3), 309–316.

Ristic, Branko, Skvortsov, Alex, & Morelande, Mark (2009). Predicting the progress and the peak of an epidemic. In *Acoustics, Speech and signal processing, 2009. ICASSP 2009. IEEE International conference on* (pp. 513–516). IEEE.

O'Brien, Sarah J., Rait, Greta, Hunter, Paul R., Gray, James J., Bolton, Frederick J., et al. (2010). Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious intestinal disease in the community (IID2 study). *BMC Medical Research Methodology, 10*(39).

Ross, Sheldon M. (1990). *A course in simulation*. Prentice Hall PTR.

Roy, Manojit, & Pascual, Mercedes (2006). On representing network heterogeneities in the incidence rate of simple epidemic models. *Ecological Complexity, 3*, 80–90.

Sanjeev Arulampalam, M., Maskell, Simon, Gordon, Neil, & Clapp, Tim (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing, 50*(2), 174–188.

Shaman, Jeffrey, & Kohn, Melvin (2009). Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences, 106*(9), 3243–3248.

Shaman, Jeffrey, Pitzer, Virginia E., Viboud, C. écile, Grenfell, Bryan T., & Lipsitch, Marc (2010). Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol, 8*(2), e1000316.

Skvortsov, Alex, & Ristic, Branko (2012). Monitoring and prediction of an epidemic outbreak using syndromic observations. *Mathematical Biosciences, 240*, 12–19.

Stroud, Phillip D., Sydoriak, Stephen J., Riese, Jane M., Smith, James P., Mniszewski, Susan M., & Romero, Phillip R. (2006). Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing. *Mathematical Biosciences, 203*, 301–318.

Thomas, Emma G., McCaw, James M., Kelly, Heath A., Grant, Kristina A., & McVernon, Jodie (2015). Quantifying differences in the epidemic curves from three influenza surveillance systems: A nonlinear regression analysis. *Epidemiology and Infection, 143*, 427–439.

Wheeler, Jeremy G., Sethi, Dinesh, Cowden, John M., Wall, Patrick G., Rodrigues, Laura C., Tompkins, David S., et al. (1999). Study of infectious intestinal disease in england: Rates in the community, presenting to general practice, and reported to national surveillance. *The BMJ, 318*, 1046–1050.

World Health Organization. (2014). *Influenza (seasonal) fact sheet N° 211*. http://www.who.int/mediacentre/factsheets/fs211/en/.

Yaari, Rami, Katriel, Guy, Huppert, Amit, Axelsen, Jacob, & Stone, Lewi (2013). Modelling seasonal influenza: The role of weather and punctuated antigenic drift. *Journal of The Royal Society Interface, 10*(84), 20130298.

Yang, Wan, & Shaman, Jeffrey (2014). *A simple modification for improving inference of non-linear dynamical systems*. arXiv preprint arXiv:1403.6804.