Using approximate Bayesian computation and machine learning model selection techniques to understand the impact of climate on seasonal influenza-like illness in Australia

Jessica Penfold

March 27, 2019

Thesis submitted for the degree of

Master of Philosophy

in

Applied Mathematics

at The University of Adelaide

Faculty of Engineering, Computer and Mathematical Sciences

School of Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

Contents

List of Tables

List of Figures

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed: .... Date: 28/03/2019

Acknowledgements

I would first like to thank my supervisors Professor Joshua Ross, Dr Lewis Mitchell, and Dr Robert Cope for their support, patience and knowledge over the last two years.

Secondly, I would like to thank the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers for the development and networking opportunities they have provided me. I would also like to thank the Australian Sentinel Practices Research Network for their high-quality influenza-like illness notification data.

Finally, I must express my very sincere gratitude to my family, friends, and partner for providing me with support and encouragement throughout my years of study. This thesis would not have been possible without them. Thank you.

Abstract

Influenza-like illness exhibits a strong seasonal cycle in temperate climates, with a peak of varying intensity appearing each winter. However, the driving force of this seasonal cycle remains poorly understood. We develop stochastic epidemic models and a model selection framework to understand influenza-like illness seasonality, with the basic reproduction number R_0 being dependent on climate. We test four Functional Forms of transmissibility based on three different climate variables and select the best Functional Form for climate-dependent transmissibility via modern Bayesian machine learning model selection methods. By analysing a unique dataset comprising ten years of General Practitioner-reported influenza-like illness surveillance data from Adelaide, Brisbane, Perth and Sydney, Australia, we explore the relationship between influenza-like illness transmission and weather across Mediterranean and subtropical climate zones. We find that absolute humidity has the strongest impact on seasonal influenza-like illness, with two different Functional Forms both based on absolute humidity best describing influenza-like illness in Mediterranean and subtropical climates. Finally, we consider the problem of forecasting the timing of peak influenza-like illness using ensemble modelling techniques. We employ two score metrics and four techniques for calculating ensemble weights in a prototype ensemble forecasting framework. By implementing this method to predict the peak week of influenza-like illness in 2014 in each of the four different locations, we find that forecasting peak week from the start of the year is a challenging exercise providing mixed conclusions on the best training methods, with two approaches – traditional and prototype – producing comparable results. We find again that absolute humidity appears to be a strong factor in the seasonality of influenza-like illness, and find that random forests are a useful tool in informing ensemble forecast weights.

Chapter 1

Introduction

1.1 Influenza and influenza-like illness

Influenza is a virus that causes an acute infectious respiratory disease, commonly known as 'the flu'. It circulates globally in seasonal epidemics and has a significant economic impact, with a reduction in worker productivity and an increase in medical costs during influenza epidemics [3]. Annual influenza epidemics occur invariably in all major cities in Australia [4] during the winter season in temperate [5] climates, and during the wet season in tropical climates. However, the exact cause of this seasonality seen in influenza epidemics is not fully understood. This thesis will explore potential climate drivers of influenza in order to increase our understanding of factors contributing to the patterns of seasonal influenza epidemics.

Seasonal influenza epidemics place a significant burden on healthcare providers and hospital wards. In particular, certain groups of people such as the elderly, the very young, and people with chronic health conditions like diabetes and AIDS are at risk of serious complications requiring hospitalisation, including pneumonia or death. Approximately 3,500 individuals die each year in Australia due to influenza and complications caused by influenza [6]. In 2014, approximately 68,000 confirmed cases of influenza were reported in Australia with more than 18,000 hospitalisations occurring as a result. The ability to predict the behaviour of an annual influenza epidemic would enable healthcare workers to more effectively assign resources, and potentially provide early warnings for abnormally severe epidemic seasons.

As well as the burden of seasonal influenza epidemics on health care systems around

the world, there are also irregular influenza *pandemics*. An influenza pandemic is an epidemic of influenza that spreads worldwide and affects a large proportion of the global population [7]. Whilst there have been only 9 influenza pandemics in the last 300 years, each caused very high levels of mortality and had a lasting worldwide impact [8]. As such, it is important to be able to forecast the course not only of annual seasonal epidemics, but also of influenza pandemics that strike on a global scale. To accurately forecast influenza epidemics and pandemics, further understanding of the driving forces and the underlying seasonal activity of influenza is required.

The symptoms of influenza are often similar to that of the common cold, but tend to be more severe and longer-lasting. Symptoms usually start 1–3 days after exposure to the influenza virus, with fever, dry cough, headache, exhaustion, sore throat and joint pains being the most common symptoms [9]. Most major symptoms usually resolve within a week, though exhaustion and coughing may persist for much longer. Antiviral medication is available for severe cases of influenza, but is most effective if taken shortly after the symptoms appear [10]. Annual vaccinations for influenza are also available, but studies show that only approximately 45% of Australians are vaccinated each year [7]. As well as this, vaccines do not necessarily grant total immunity and do not necessarily cover all strains of influenza seen that year [11], leaving a large portion of the population susceptible.

Influenza shares similar symptoms to a number of other respiratory viruses. Influenza-like illness (ILI) is a 'catch-all' phrase used to describe individuals presenting with a set of symptoms common to influenza and other similar viral illnesses including rhinovirus and human respiratory syncytial virus (HRSV) [12, 13]. Symptoms include fever, chills, fatigue, cough and body aches, and may vary in severity from mildly uncomfortable to life-threatening. ILI is very common, with adults suffering 2–4 colds per year on average [14]. Most cases of ILI are mild and self-limiting, with no long-term effects.

Whilst ILI is a term that includes influenza, it is not possible to conclusively distinguish influenza from other respiratory diseases without the use of laboratory testing. Laboratory testing for influenza can be performed with a variety of methods, including polymerase chain reaction (PCR). PCR detects the presence of viral ribonucleic acid (RNA) in swab samples. It can identify strains of influenza and other respiratory viruses, and is fast and accurate. However, PCR testing tends to be costly, reducing the number

of samples sent to be tested for the presence of influenza [15].

Due to the expense and time involved in testing samples from patients with ILI, only a proportion of patients diagnosed with ILI are tested for influenza. The exact number of tests performed is generally not known, as only the samples testing positive for influenza are routinely reported. This leads to few notifications of confirmed influenza, and so it is sometimes preferable to use ILI data in the place of true influenza data, as influenza and ILI generally follow similar seasonal patterns.

Being able to accurately forecast the timing and severity of seasonal ILI will allow health care providers to better allocate resources, and help public health officials plan vaccination schedules and campaigns. Improving the ability to prepare for, and manage, seasonal influenza and ILI outbreaks is an important factor in reducing the mortality and economic impact caused by these annual winter-time epidemics.

1.2 Seasonality in influenza-like illness

Figure 1.1 shows the total number of ILI cases reported in Australia to the Australian Sentinel Research Practises Network (ASPREN) [16] each week. It shows that ILI in Australia displays a very clear seasonal pattern, with large peaks occurring each winter and few cases occurring in summer.

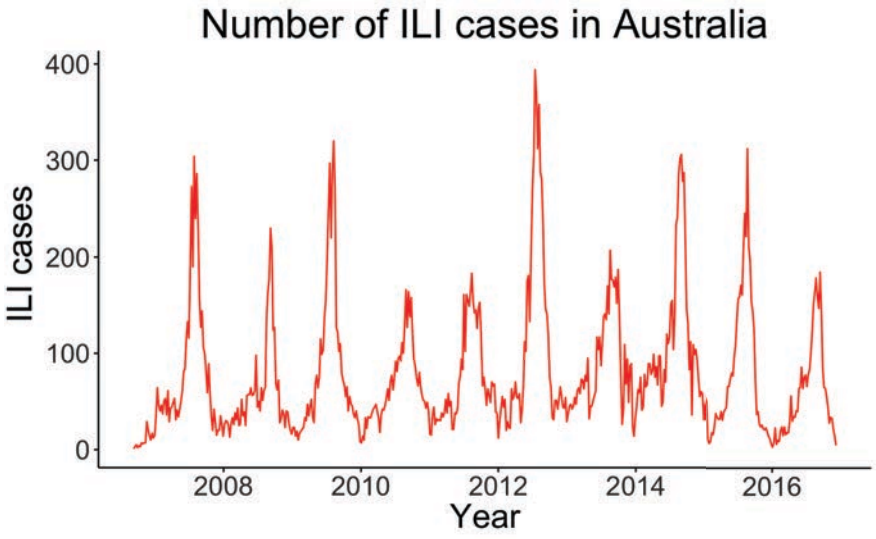

Figure 1.1: Total number of ILI cases reported weekly to ASPREN per week in Australia in the period 2006 – 2016, clearly showing defined epidemic peaks during winter months.

Different areas around the world see different seasonal patterns in ILI incidence. In

temperate areas, ILI epidemics tend to begin early winter, peak mid-winter and taper off as spring begins. In tropical areas, however, influenza and ILI can occur year-round, often with two or more peaks during the rainy season and a decrease in cases during the dry season. The exact cause of this seasonal behaviour is not fully understood, but is likely to be influenced by differences in large-scale and regional climate [17, 18, 19, 20]. Suggested factors include temperature, absolute humidity or relative humidity [21, 22, 23], which may affect the survival of viral particles between hosts or influence host behaviour. Other potential climate drivers could include rainfall or UV index [24, 25]. Influenza and ILI have more irregular annual periodicity than other seasonal infectious diseases [19], suggesting that there may be a complicated relationship between climate drivers, social factors, and viral strains.

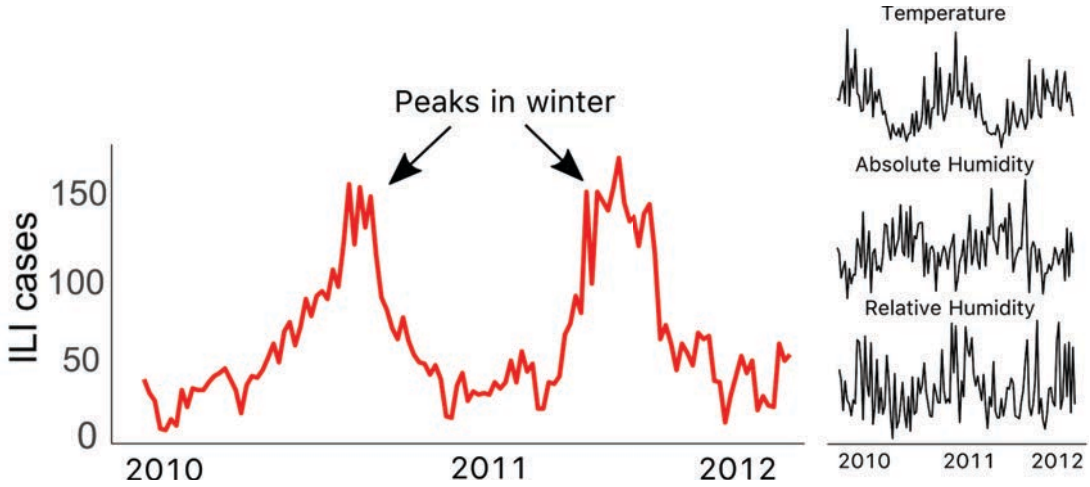

Figure 1.2: Influenza-like illness cases in Sydney over the period 2010-2012, with a comparison to the average daily temperature, absolute humidity and relative humidity over the same period.

Figure 1.2 shows a comparison between the number of ILI cases reported to ASPREN by participating doctors in the 2010 – 2012 period, and the temperature, absolute humidity and relative humidity in Sydney over that same time period. There appears to be a relationship between these climate variables and ILI, with the number of ILI cases generally increasing when temperature and relative humidity decrease [26], and absolute humidity increases. This thesis further explores the relationship between ILI, temperature, absolute humidity and relative humidity in four locations around Australia. We aim to determine a Functional Form governing the mechanistic relationship between climate and ILI, instead of a statistical association between variables which has been commonly

investigated to date.

In order to obtain a better understanding of ILI and how it spreads, we utilise mathematical models of ILI epidemics. The use of mathematical modelling for infectious diseases dates back to the 1800's, where it was used to model the effectiveness of smallpox inoculation [27]. Modern mathematical epidemiology research uses both deterministic and stochastic models to attempt to capture the details of infectious disease dynamics while remaining computationally feasible [28]. This thesis explores the use of a family of stochastic compartmental epidemic models to simulate ILI in four locations around Australia: Adelaide, Brisbane, Perth and Sydney.

Previous studies have explored the link between ILI, influenza, and climate variables; however, there is no consensus in the literature as to the exact driver of ILI seasonality. Experimental studies in a guinea pig model show evidence that relative humidity affects the transmission of influenza [22, 29], which may be explained by the effect of humidity on the between-host survival rates of influenza viral particles [30].

Similarly, studies from Israel have instead found a link between relative humidity and influenza in humans [31, 32]. These studies used a susceptible-infectious-recovered-susceptible (SIRS) compartmental model driven by both relative humidity and temperature, finding that the seasonality of ILI outbreaks can be explained by a combination of these climate drivers and antigenic drift leading to waning immunity.

However, recent mathematical studies based in the United States have found a link between absolute humidity and ILI [23, 33, 34]. These studies utilised a humidity-driven susceptible-infectious-recovered-susceptible (SIRS) compartmental model to explore the dynamics of ILI across multiple states in the U.S. They found that, compared to previous studies based on the same data [22, 29], absolute humidity has a stronger effect on ILI transmission than relative humidity.

A 2018 study found no significant difference in prediction when using temperature as a climate driver compared to specific humidity, a scaled form of absolute humidity [35]. A U.S. study of over 600 cities found that the key environmental driver was change in specific humidity, but that urbanisation also led to differences between cities not explained by climate variances [17]. A study in the Asia-Pacific region [36] found that particular climate drivers had different impact in different locations, with specific humidity positively associated with ILI in some places and negatively associated in others. Studies

on Australian data have also found a complicated relationship between climate and ILI, with relative humidity and temperature seemingly playing an important role [37, 38].

Other studies analysing the different methods currently used to model seasonal ILI [19, 39] found that forecasting ILI is considerably more challenging than for other infectious diseases. They found that ILI does not have the same seasonal predictability as other periodic infectious diseases (such as measles), and that the irregular periodicity of ILI epidemics presents a unique and difficult challenge for researchers looking to understand the drivers of ILI. Expanding on this, a study [40] found that different approaches and methods need to be explored, especially in long-term disease forecasting due to the predicted influence of climate change in the coming years.

The conflicting conclusions from current literature reinforce the need for further research in the area of ILI seasonality to increase understanding of the driving force behind seasonal ILI outbreaks. In this thesis, we use a stochastic compartmental epidemic model along with statistical and machine learning techniques to further explore the relationship between seasonal climate drivers and ILI in Australia. Faced with a range of different potential climate drivers, Functional Forms, and climate regions, we use Bayesian model selection techniques to find the best combination of factors explaining ILI seasonality.

1.3 Data

1.3.1 Influenza-like illness data

The Australian Sentinel Research Practises Network (ASPREN) [16] is a database of general practitioners (GPs), also known as 'family physicians' in the U.S., that provide weekly surveillance reports on ILI and influenza. These GPs are distributed across metropolitan and regional areas of Australia, with a target coverage of one GP per 200,000 people in metropolitan areas and one per 50,000 people in rural and regional areas. Participating GPs are also asked to take swab samples from 20 – 25% of ILI patients to be tested for influenza using PCR methods [41]. However, in reality not all GPs take swab samples, and so due to the sparsity of swab testing, we use influenza-like illness (ILI) as an approximation for influenza in this thesis. The incidence of ILI is highly correlated with the number of confirmed influenza cases [42] and so it is appropriate to use ILI as an approximate for influenza, rather than using only confirmed cases of influenza. ASPREN

and other influenza and ILI surveillance programs are used to understand and predict seasonal influenza, to aid public health officials in preparing for the yearly ILI epidemic. An Australian Influenza Surveillance Report (AISR)[1] is published fortnightly with data on the reported number of confirmed influenza cases.

1.3.2 Climate data

Within Australia, climate data is collected nationally by the Bureau of Meteorology[2] (BOM). There are weather stations throughout Australia, in both rural and metropolitan regions. For this analysis we select the weather station closest to the city centre for each of the cities studied: Adelaide Airport (Adelaide), Archerfield Airport (Brisbane), Perth Airport (Perth), and Observatory Hill (Sydney).

The BOM data provides us with readings of temperature (°C) and relative humidity (%) taken every three hours [43]. We also use absolute humidity in our comparisons, which we calculate from the temperature and relative humidity measurements. This calculation process is explained in Section 2.6. Absolute humidity measures the total amount of water vapour in the air in g/m^3, independent of the air temperature. However, as warmer air can hold more water than cold air, relative humidity is used to measure the total amount of water vapour in the air as a function of the theoretical maximum amount of water vapour the air could hold at that temperature. Relative humidity is also affected by air pressure; however, all four Australian cities studied are very close to sea level, and so we assume a constant atmospheric pressure, $P = 10.1$ N/cm^2, for ease of calculation.

1.4 Thesis outline

The aim of this thesis is to further understand which climate factors drive seasonal ILI outbreaks using new Bayesian machine learning model selection methods. We use the novel, high-quality influenza-like illness dataset provided by ASPREN in four locations around Australia: Adelaide, Brisbane, Perth and Sydney. In Chapter 2, we introduce the mathematical methods that we will be using.

In Chapter 3, we develop four Functional Forms of transmissibility based on the

[1] http://www.health.gov.au/flureport

[2] http://www.bom.gov.au/climate/data-services

basic reproduction number R_0 as a function of climate. These Functional Forms are dependent on three different climate variables – temperature, absolute humidity, and relative humidity. We then use the Functional Forms of R_0 within a stochastic SEIR-type epidemic model framework, where three new compartments are added to allow us to simulate an hierarchical observation process where not all infected individuals will be observed. This model is then used to simulate ILI epidemics over 10 years in each of the four locations.

Using these simulated ILI epidemics, we apply a Bayesian model selection method using Random Forests (RFs) to select which Functional Form of R_0 within the SEIR-type model best fits the known ILI data in each location. We then fit the top two Functional Forms in each location, as selected by the RF, to the ILI data using approximate Bayesian computation (ABC), allowing us to analyse the relationship between model fit and the RF model selection method.

In all locations, we find that absolute humidity is selected as the most important climate variable. However, the Functional Form selected varies depending on the climate zone. Fitting the top two Functional Forms in each location using ABC shows that the Functional Forms are able to fit well in all locations apart from Sydney, with the Functional Forms selected first by the RF showing more seasonality in their R_0 realisations than the Functional Forms selected second. In Sydney, the Functional Form selected by the RF method fits more closely to the data than the other fitted Functional Form, but neither Form is able to accurately fit the data.

We find that the RF model selection algorithm appears able to tell the Functional Forms apart with a high level of accuracy, and that the Functional Form selected most commonly in subtropical areas (Brisbane, Sydney) is different to the Functional Form most commonly selected in Mediterranean areas (Adelaide, Perth). However, absolute humidity is selected as the most 'important' climate variable in all four locations, adding further support to findings in the literature. We also find that the RF model selection method was accurate in selecting the Functional Form with the best fit in each location when comparing to the results of the ABC, suggesting that a RF is a reliable model selection tool for infectious disease modelling. These results, conclusions and discussions are summarised in Chapter 5.

In Chapter 4 we turn to the problem of forecasting ILI in Australian cities. We

consider a variety of methods to fit a weighted ensemble of Functional Forms to forecast the 2014 ILI season with a varying number of historical training years. We then analyse how the ensemble methods differ, and determine which ensemble of Functional Forms is best in each location.

We find in Chapter 4 that forecasting an ILI season from the start of the year is a challenging exercise, and that adding more years of training data does not necessarily lead to a more accurate forecast. There is no clear conclusion as to which training method is most appropriate for this forecasting process, but we find that random forests are again useful in this application and that Functional Forms using absolute humidity generally perform well compared to Functional Forms using other climate variables. We also find that the 'best' Functional Forms found in Chapter 3 are not necessarily useful for forecasting.

Chapter 2

Background

In this chapter, we introduce the methods and technical background used in this thesis. We introduce the methods used to develop our model for ILI, as well as the methods we use to develop our model selection process.

2.1 Epidemic models

In Chapters 3 and 4, we use a stochastic epidemic model (based on the mathematical models presented in this section) to simulate ILI datasets to use within a Bayesian model selection framework.

2.1.1 Basic reproductive number, R_0

The *basic reproductive number*, R_0, is the expected number of individuals infected by a primary infectious individual over the course of their infectious period, in an otherwise fully susceptible population. It is commonly used to measure the transmissibility of a disease [44]. This helps to determine if an infectious disease is likely to spread through a non-trivial proportion of the population, or if it is likely to die out reasonably quickly. It also helps to inform how rapidly a disease will spread.

In general, if

$$R_0 \leq 1$$

the disease will fail to invade, while if

$$R_0 > 1$$

the infection may invade, and may spread through the population, with each infectious individual infecting on average more than one other individual [45].

In reality, a population will rarely be composed of totally susceptible individuals. This leads to an *effective reproduction rate*, which is the expected number of secondary infections generated by one primary infection in a population accounting for both susceptible and non-susceptible individuals [46]. Non-susceptible individuals include both vaccinated and immune individuals.

R_0 is influenced by many different factors [47], and may not be the same in two different outbreaks of a disease. These factors include the duration of the infectious period, rate of contact between infectious and susceptible individuals in a population, and the probability of transmission occurring during contact between two individuals. The rate of contact and probability of transmission occurring during contact may be influenced by temperature or humidity, which can affect the survival of infectious viral particles as they are transferred between hosts.

2.1.2 Markov chains

In this section, we define both discrete and continuous-time Markov chains and introduce the types of Markov chains used in this thesis.

In the most basic terms, a Markov chain is a stochastic process which possesses the Markov property – meaning that, given the present, the future is independent of the past [48]. Mathematically, a collection of random variables $\{X_t\}_{t\in\mathcal{I}}$ defined on some countable state space $\mathcal{S}$, for some index set $\mathcal{I} \in \mathbb{R}^+$ is a Markov chain if it satisfies the property

$$P(X_t \in A|X_r, X_s) = P(X_t \in A|X_s),$$

for all r, s, $t \in \mathcal{I}$ with $r \leq s \leq t$ and $A \subseteq \mathcal{S}$. There are two types of Markov chains, called *discrete time* and *continuous time*. A Markov chain $\{X_t\}_{t\in\mathcal{I}}$ is a discrete time Markov chain (DTMC) if the index set $\mathcal{I}$ is countable, and a continuous time Markov chain (CTMC) if the index set $\mathcal{I}$ is uncountable.

In this thesis, the model we use is a DTMC. However, it is originally formulated as a CTMC, as disease transmission is a continuous-time process, and is then approximated in discrete time using 12 hour time-steps to facilitate more efficient simulation.

Within a DTMC, the probability of transitioning from some state i to state j in one

discrete time step is given by

$$P_{i,j} = P(X_{t+1} = j | X_t = i),$$

where

$$\sum_{j \in \mathcal{S}} P_{i,j} = 1.$$

Within a CTMC, on the other hand, transitions between states are described in terms of *infinitesimal transition rates*. Infinitesimal transition rates describe the rate at which transitions between states occur, instead of the probability of transitioning between states as in a DTMC. The rate of transition from some state i to another state j, with $i \neq j$, is given by

$$q_{i,j} = \lim_{h \to 0^+} \frac{P(X_{t+h} = j | X_t = i)}{h},$$

with

$$q_{i,i} = -\sum_{j \neq i} q_{i,j}.$$

We also define an absorbing state as one where, once the chain has entered the state, the probability or rate of leaving the state is zero. Absorbing states are seen in both DTMCs and CTMCs.

2.1.3 Stochastic compartmental epidemic models

Compartmental epidemiological models are used to simplify infectious disease modelling by dividing the population into compartments, where every individual in each compartment is assumed to have the same characteristics. They can be analysed deterministically, such as through ordinary differential equations (ODEs), or stochastically. Stochastic versions of compartmental epidemic models are more 'realistic' than their deterministic counterparts in the sense that they attempt to capture the randomness in true disease spread, but are more challenging to analyse. For suitably large populations, deterministic models are typically a highly accurate approximation to stochastic models [28]. However, capturing the stochastic variability in transmission is important for smaller populations where individual events have an impact on the overall epidemic outcome [49, 50]. In this thesis, we use a stochastic compartmental epidemic model.

In this section, we introduce an example stochastic compartmental epidemic model and describe it in terms of the underlying Markov chain.

SEIR model

One basic stochastic compartmental epidemic model is the 'SEIR' model [51]. This model describes a population of size N, where each individual is in one of four states at any time: *Susceptible, Exposed, Infected,* or *Recovered*. The *Exposed* state is also known as the latent state, where an individual has been exposed to the illness but is not yet infectious. Individuals can move between these states at certain rates. In an SEIR model, once an individual has entered the *Recovered* state they cannot leave. This effectively means that the individual has gained immunity from the disease. The standard SEIR model does not include population dynamics such as births or deaths, and is a *fully mixed model*, meaning that any individual in the population can interact with any other individual equally.

We define the parameter β as the effective transmission rate parameter, σ as the rate of exposed individuals becoming infectious, and the parameter γ as the per-capita rate of recovery. Then, the basic reproductive number is given by

$$R_0 = \frac{\beta}{\gamma}.$$

Let S be the number of susceptible individuals, E the number of exposed individuals, I be the number of infected individuals and R be the number of recovered individuals. Then, in a continuous-time Markov chain (CTMC) with state space

$$\mathcal{S} = \{(S, E, I, R) | S, E, I, R \in [0, N] \text{ such that } S + E + I + R = N\},$$

we have the following transition rates between the population compartments:

$$\begin{aligned} q_{(S,E,I,R),(S-1,E+1,I,R)} &= \frac{\beta SI}{N-1} \\ q_{(S,E,I,R),(S,E-1,I+1,R)} &= \sigma E, \\ q_{(S,E,I,R),(S,E,I-1,R+1)} &= \gamma I, \end{aligned} \tag{2.1}$$

where $q_{(S,E,I,R),(S-1,E+1,I,R)}$ is the rate at which susceptible individuals are exposed, $q_{(S,E,I,R),(S,E-1,I+1,R)}$ is the rate at which exposed individuals become infectious, and $q_{(S,E,I,R),(S,E,I-1,R+1)}$ is the rate at which infectious individuals recover from their illness. In this case, the recovered

state R is an absorbing state, so that once individuals have recovered from their illness they cannot be reinfected. No other state transitions are possible in this model. This model is summarised in Table 2.1.

State change	Event	Rate
$(S, E, I, R) \to (S-1, E+1, I, R)$	Exposure	$\frac{\beta SI}{N-1}$
$(S, E, I, R) \to (S, E-1, I+1, R)$	Infectiousness	σE
$(S, E, I, R) \to (S, E, I-1, R+1)$	Recovery	γI

Table 2.1: The state transition events that can occur in an SEIR model.

2.2 Approximate Bayesian computation

Approximate Bayesian computation (ABC) is a computational approach to Bayesian inference that does not require evaluation of the likelihood. For this reason, it is used when the likelihood function is difficult or infeasible to calculate. We use this ABC technique in Chapter 3, where it is used to fit selected Functional Forms within an ILI epidemic modelling framework.

Bayesian inference involves Bayes' Theorem,

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}, \tag{2.2}$$

which relates the probability of seeing a dataset D given a parameter θ – the likelihood – to the conditional probability of seeing θ given D – the posterior distribution; $p(\theta)$ denotes the prior distribution and $p(D)$ denotes the prior predictive probability of the data, which is also known as the normalising constant. The likelihood function $p(D|\theta)$, which expresses the probability of seeing the observed data from a specific model with parameter(s) θ, can be computationally expensive or infeasible to calculate in many applications. To circumvent this issue, we utilise a method called *approximate Bayesian computation* (ABC). In this thesis, the method used is the ABC rejection algorithm.

ABC methods share a common goal of approximating the likelihood function $p(D|\theta)$ [52]. In the ABC rejection algorithm, a large number of parameter values $\hat{\theta}$ are sampled from a prior distribution $p(\theta)$. Then, each sampled parameter point $\hat{\theta}_i$ is used to simulate

a dataset $\hat{D}_i$ under a chosen statistical model M specified by the parameter set $\hat{\theta}_i$. The simulated dataset $\hat{D}_i$ is then compared to the observed data D using a distance measure

$$\rho(\hat{D}_i, D).$$

The raw data, or a summary statistic, are compared with the true data D via a distance measure. The simulated dataset $\hat{D}_i$ is accepted if, for a chosen tolerance ϵ,

$$\rho(\hat{D}_i, D) \leq \epsilon.$$

Otherwise, the simulated dataset $\hat{D}_i$ is discarded and the process starts again. This continues until a pre-determined number of simulated datasets have been accepted. The parameter values $\hat{\theta}_i$ that generated the accepted simulations $\hat{D}_i$ then form the posterior distribution of θ. This process is described in Algorithm 1.

Input data D and prior distribution to draw $\hat{\theta}$ values;
Input the underlying model and choose a tolerance ϵ;
Choose a large j, the number of $\hat{\theta}$ to accept;
while # accepted $< j$ **do**
  Sample a parameter set $\hat{\theta}$;
  Simulate a data set $\hat{D}$ from the underlying model
  with parameters $\hat{\theta}$;
  Calculate distance $\rho(\hat{D}, D)$;
  if $\rho(\hat{D}, D) \leq \epsilon$ **then**
    Accept $\hat{\theta}$;
  else
    Discard $\hat{\theta}$;
  end
end
Output the accepted values of $\hat{\theta}$.

Algorithm 1: The ABC rejection algorithm.

Figure 2.1 shows the ABC rejection algorithm process, with a uniform prior being used to determine the posterior distribution of the parameter θ given a set of observational

data. Simulations that are 'close enough' to the observed data are accepted, and the parameters which generated those simulations form part of the posterior distribution.

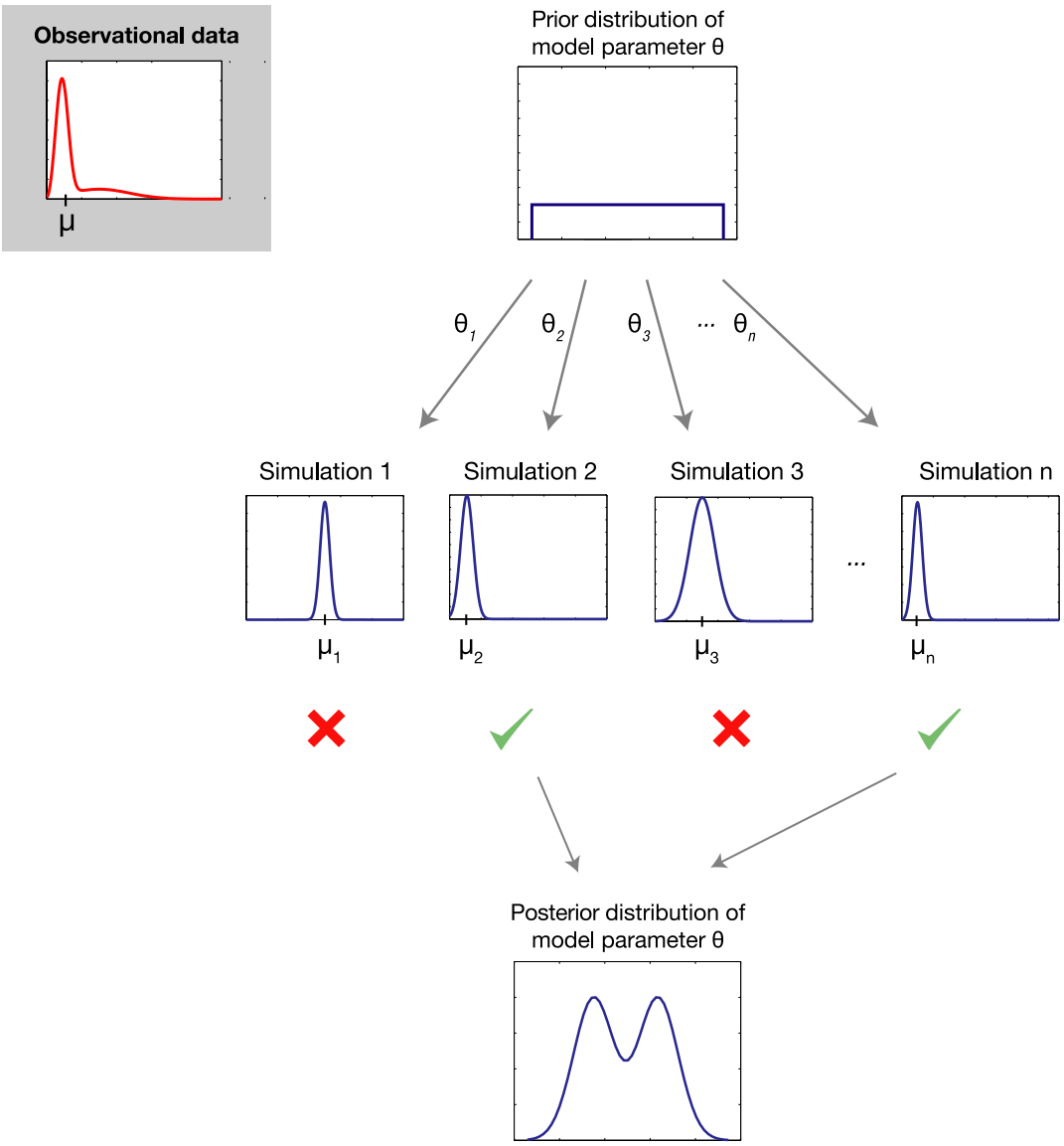

Figure 2.1: Credit: Sunnaker *et al.* [1]. A diagram showing an approximate Bayesian computation rejection algorithm.

The summary statistic that we use within the ABC rejection algorithm in this thesis is the *weighted root mean squared error* (WRMSE),

$$\rho(\hat{D}, D) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\sqrt{D_i + 1} \times (D_i - \hat{D}_i)^2}, \tag{2.3}$$

where D is the true ILI data, $\hat{D}$ is the simulated dataset, and N is the number of data points.

The WRMSE is a measure of accuracy similar to the *root mean squared error* (RMSE) [53], and is used in this context to determine the error between the true values and the predicted values with an increased emphasis on within-season observations rather than observations between seasons; the WRMSE metric gives larger observations a heavier weight than smaller observations.

2.3 Bayesian model selection

We use the Bayesian model selection techniques presented in this Section in Chapter 3 of this thesis, to select between Functional Forms of ILI transmission.

Model selection is the process of selecting the 'best' model out of a set of candidate models for a particular dataset. It is not only a problem of selecting which model best fits the given data, but also what metric or summary statistics are used to determine the 'best fit' [54]. There are a number of different approaches to model selection, from frequentist to Bayesian to machine learning techniques [1]. In this thesis, we focus on Bayesian model selection in conjunction with machine learning techniques.

Bayesian model selection involves comparing the likelihood of each candidate model to determine which model has the most support on the data, using one of a number of possible criteria [58]. The two most common criteria for model selection in a Bayesian framework are the *Bayes factor* (BF) and the *Bayesian information criterion* (BIC). The BF is the ratio of the likelihood probabilities of two competing candidate models, whereas BIC is a criterion that takes into account the likelihood probabilities, number of parameters, and the number of data points for a finite number of candidate models [55, 56]. In particular, BIC penalises models with a large number of parameters in order to reduce overfitting [57]. Once these criteria are used to select the 'best' model, the posterior distributions of model parameters are used to make inferences about the data.

In this thesis, the likelihood is unable to be calculated directly and so ABC is used, as described in Section 2.2 [52, 59]. However, there are several limitations when using ABC for model selection directly. In order for the Bayesian model selection to be valid, the number of simulations n must be much larger than the number of parameters k [60]. This causes an issue in cases where there is a very large number of parameters, meaning that an extremely large number of simulations must be produced. This leads to issues

with computational time and expense. ABC methods also depend on the selection of summary statistics and initial prior distributions for each parameter in the candidate models. Incorrect calibration of these summary statistics or prior distributions can lead to incorrect or misleading results [61].

In 2015, a new technique for Bayesian model selection was proposed by Pudlo *et al.* [62]. This technique uses ABC in conjunction with random forests (RF), as described in Section 2.4, to circumvent these limitations. They propose using RF classification for model selection, instead of choosing particular summary statistics. This avoids the issue of needing to select an appropriate set of summary statistics. They further use a second RF within an ABC framework to estimate the posterior probability of the chosen model.

This new technique is particularly useful in situations where a sizeable number of model parameters require the use of an inconveniently large set of simulations. In the thesis, we utilise the first part of this technique by using RFs as a classification tool in model selection. We then use a traditional ABC framework as a comparison to analyse the performance of this new technique within a mathematical epidemiology application.

2.4 Random forests

This section introduces *classification and regression trees* (CARTs), and a method of utilising them called *random forests* (RFs). RFs are used in Chapters 3 and 4. In Chapter 3 they are used to select between multiple Functional Forms, and in Chapter 4 they are used to inform model weights in an ensemble forecasting framework.

CARTs, also known as decision trees, are a component of decision making algorithms in predictive modelling [63, 64], and are commonly used in machine learning applications [62, 65]. There are two types of CART: *classification decision trees*, and *regression decision trees*. Classification decision trees are used to classify data into one of a discrete number of outcome classes, whereas regression decision trees are used to predict a continuous outcome value. In this thesis we use classification decision trees to classify simulated datasets into one of a discrete number of candidate models.

A classification decision tree is a structure based on a tree graph, consisting of nodes and directed edges with no cycles, where each node has exactly one edge entering and two edges leaving. Each internal node represents a test on an attribute of the object being

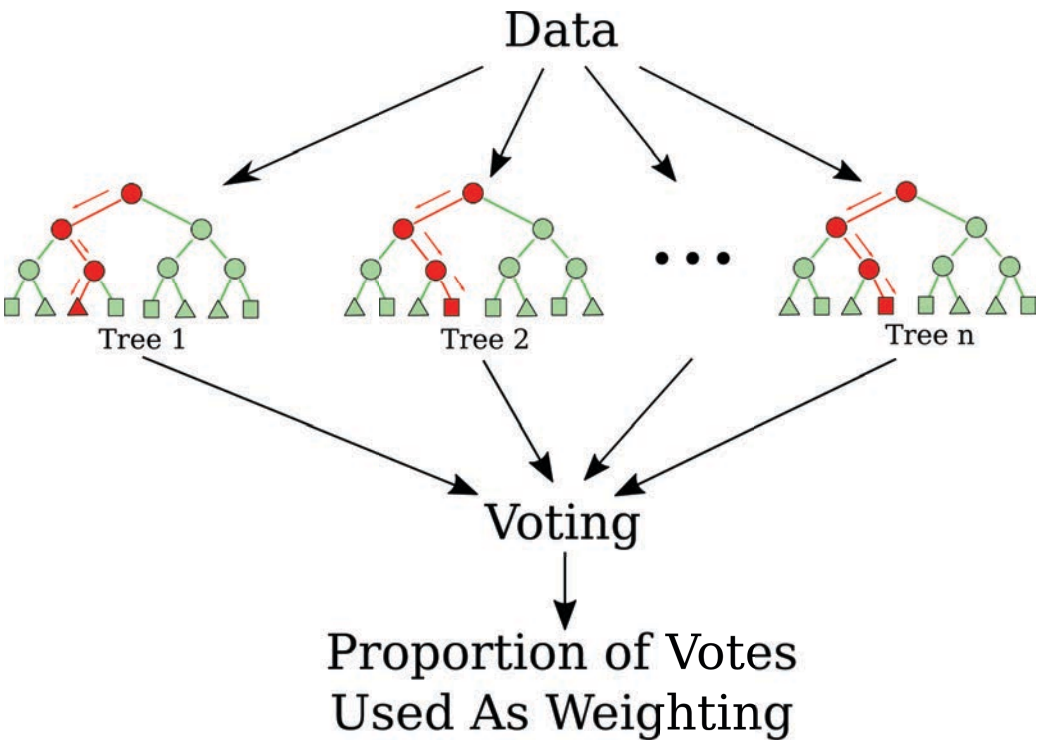

Figure 2.2: The process of using a random forest to classify data. In this diagram, the red circles represent the path the data takes as it is classified through each of the decision trees. The terminal node shape represents two different candidate models or classifications.

analysed, and each edge represents a decision, or the outcome of the test. The terminal nodes represent each of the possible classifications. The object is first tested at the top node, and then follows a path downwards along the edges until it reaches a terminal node, which represents the final classification given to the object. For example, if we are considering the outcome of a number of coin flips, we may look at the first coin flip in the initial decision node; if the coin is heads, we go left, if it is tails, we go right. We can continue this through all of the decision nodes to arrive at the classification, which is the number of heads and tails in a sequence of coin tosses.

RFs are a machine learning technique that use decision trees as basic building blocks [66]. RFs work by training an ensemble of decision trees on a set of data, and using the modal output of those trees to decide on the final classification. This is illustrated in Figure 2.2, which shows the progress of an object as it is classified by each of n decision trees. Each of these trees can cast a 'vote' as to which classification it belongs to, and the final classification of the object is the class with the majority of the votes.

The training algorithm for RFs involves the use of *bootstrap aggregating*, or 'bagging'. Bootstrap aggregating [67] involves taking a large number of random samples from the training dataset, with replacement, and training a decision tree on each sample. As well as this, the RF algorithm uses *feature bagging*, where at each proposed split (or node) in the decision tree, a random subset of the features is taken [68]. This means that the RF trains on a greater number of features than if feature bagging were not used, and so increases the accuracy of classification and reduces overfitting [69]. This is especially true in cases where the data has a large number of features to consider at each split.

Generate N simulations from m models, where each simulation is from a different set of parameters $\hat{\theta}_i$ sampled from a prior distribution;

for *s = 1 to N_{tree}* **do**

  select a sample from the simulations, with replacement;

  grow a randomized CART T_s;

end

Determine the predicted vote from each T_s for the real dataset;

Assign the model according to the majority vote among the CARTs.

Algorithm 2: Random forest algorithm for selecting R_0 Functional Form.

The RF algorithm is useful in data science because, provided there is a sufficient number of trees, it is unlikely to over-fit to the data. It is also straightforward and easy to use, with a number of software packages available. However, in applications with a very large amount of data, the RF algorithm can be slow and inefficient.

Classification random forests can be used for model selection, where the classes are candidate models and the decision nodes are variables within the data being classified. Algorithm 2 shows the algorithm used to train a RF for model selection. In this case, the training set consists of N simulations generated from m candidate models, where each simulation is labelled with the model from which it was generated. From this training

set, a number of random samples are drawn with replacement. The number of samples is equal to the number of decision trees in the final random forest, N_{tree}. A single decision tree is trained on each of the random samples. To use this RF to classify an object, each decision tree is used to individually classify the object as coming from one of m models. Then, the final classification is equal to the mode of the individual classifications.

Throughout the thesis, all random forests are computed using the `randomForest` package in the statistical computing environment R [66].

2.5 Kernel density estimation

Kernel density estimation is presented in this Section, and is used in Chapter 4 to estimate the posterior distribution of the 2014 ILI peak week from a variety of forecasting techniques.

Kernel density estimation is a non-parametric method to estimate the probability distribution of a random variable based on a sample of data. It is used to estimate a probability density function (PDF) for the data sample, called a *kernel density estimate* [70]. Kernel density estimates (KDEs) are similar in concept to histograms, in that they aim to estimate the probability of each point over the range of the data; however, histogram estimates rely heavily on the choice of bin width and can be unreliable due to this [71].

To form a KDE, a *kernel* is formed around each data point. A kernel is a PDF that is symmetrical around the centre point, in this case the data point. Kernels are most commonly formed using triangular, uniform or Gaussian distributions [72]. The kernel at each data point is identical, but kernels are closer to or further from each other depending on the distribution of the data.

The KDE is determined by the average sum of the kernels across the dataset. This is calculated using the equation

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$$

where n is the number of data points, h is the smoothing parameter, K is the chosen kernel PDF, and x_i is the i^{th} data point [73]. The smoothing parameter h determines the width of the kernel around each data point, where a larger h value results in a wider

distribution around each x_i. In general, a small h value should be chosen for large datasets and a large h value for small datasets [74].

The primary limitations of KDEs relate to the selection of the bandwidth and distribution limits. If a Gaussian kernel K is chosen, then the final KDE will have infinite support. This is an issue when the data points are from a positive-only measure, such as human height or number of individuals with an illness. This can be solved by more appropriate selection of the kernel PDF [75]. As well as this, the choice of bandwidth h plays a significant role in the shape of the final PDF. A bandwidth that is too large will produce an overly smoothed PDF, while an inappropriately small h will produce a very bumpy PDF over the same data. However, there are a number of techniques that can be utilised to select the optimal bandwidth. Many common statistical packages for implementing KDEs contain built-in bandwidth selection [73].

In this thesis, we utilise the `kde` function from the `ks` package in the statistical software R [76]. Within this package, we use the default bandwidth selector which is based on the methods presented by Sheather *et al.* [75]. The default kernel PDF is Gaussian, adjusted to ensure that the KDE has support only over the possible data values, i.e. weeks 1–52 of a year.

2.6 Relationship between types of humidity

In this thesis, we focus on Functional Forms of transmissibility based on climate variables including both absolute and relative humidity. The different types of humidity and the following calculations presented in this Section are used in both Chapters 3 and 4 of this thesis as part of the Functional Forms of ILI transmission.

Humidity is a quantity representing the amount of water vapour in the air. It is commonly described using one of three measures: *absolute, relative* or *specific* humidity. Absolute humidity (g/m^3) is the total amount of water vapour in the air, without considering the air temperature. Relative humidity (%) is the amount of water vapour in the air relative to the maximum possible amount of water vapour; this maximum amount depends on the temperature and pressure of the air [77]. Specific humidity is the ratio of the total water mass to the total air mass, and is a measure of absolute humidity.

The Australian Bureau of Meteorology (BOM) provides humidity measurements in

terms of relative humidity. Relative humidity depends on the absolute amount of water vapour in the air, the air temperature, and the air pressure. In Australia, the major cities are very close to sea-level and so we assume constant atmospheric air pressure of P = 10.1 N/cm^2. The BOM also provides measurements of temperature ($^\circ C$) taken at the same time and location as the relative humidity measurements. By using this, we can use the formula,

$$AH = \frac{6.1120 \times e^{\frac{17.670 \times T}{T+243.50}} \times RH \times 2.1674}{273.15 + T},$$

where AH is absolute humidity, RH is relative humidity, and T is temperature, to convert relative humidity measurements into absolute humidity [78]. This formula is accurate to 0.1% within the temperature range -30°C to +35°C. In Australia, temperatures do exceed 35°C during the summer. However, in this thesis we are primarily focussed on the climate during winter months where the temperature has never been recorded to be outside of this range in any of the cities under consideration: Adelaide, Brisbane, Perth and Sydney [43].

2.7 Conclusion

Now that we have established the necessary mathematical background and methods required for the remainder of this thesis, in the next Chapter we explore modelling ILI with a variety of Functional Forms for transmissibility. We then consider the problem of reliably selecting between these Functional Forms, and analyse our method by using ABC to compare the fit to true ILI data between these Functional Forms.

Chapter 3

Model selection via random forests

In this chapter, we explore the impact of climate variables on the incidence of seasonal influenza-like illness in Australia over the period 2006–2016. We use a modern Bayesian model selection technique to select between four seasonally-oscillating Functional Forms for R_0; three climate-dependent and one sinusoidal. We use a stochastic SEIR-type compartmental epidemic model (described in Section 3.3.2) with these four Functional Forms of R_0 to generate simulations of 10 years of influenza-like illness data for Adelaide, Brisbane, Perth and Sydney, sampling model parameters from specified prior distributions. Using these simulations to train a random forest model selection method, we determine which Functional Form and climate variable best fits observed historical influenza-like illness data in each location.

3.1 Introduction

Seasonal influenza epidemics exhibit significant inter-annual variation in the timing and severity of epidemic peaks, the true cause of which is not fully understood but can be at least partially explained by the seasonal climate cycle [5]. These seasonal epidemics place a significant burden on healthcare providers and hospital wards, with approximately 3,500 deaths each year in Australia due to influenza and complications caused by influenza [6]. The ability to predict the behaviour of a yearly influenza epidemic would enable healthcare workers to more effectively assign resources, as well to provide early warning for abnormally severe epidemic seasons.

Environmental factors can contribute to the transmission of ILI, and hence the value

of R_0, either by altering the effectiveness of disease transmission or the disease survival within and between hosts. Experimental studies have found complicated relationships between ILI and climate variables, most commonly temperature, relative humidity, and absolute humidity [22, 23, 31, 79]. Here we explore the impact of environmental drivers of R_0 over ten influenza-like illness seasons, comparing the results to a novel influenza-like illness surveillance data set across four different locations around Australia: Adelaide, Brisbane, Perth, and Sydney. We use modern Bayesian model selection to select between Functional Forms for R_0 within a stochastic compartmental SEIR-type model.

3.2 Data

We use a high-quality influenza-like illness dataset provided by the Australian Sentinel Research Practises Network (ASPREN) [16], as described in Section 1.3.1. We focus on four major cities in different regions of Australia, selecting only ILI reports from these metropolitan areas (Adelaide, Brisbane, Perth and Sydney).

Since Australia is geographically vast, rural areas within the same state may have significant variation in climate leading to issues regarding the daily weather in the locations where influenza-like illness was reported. We chose to focus on the metropolitan areas to avoid issues with variation in weather. Since the majority of notifications come from GPs based in metropolitan areas, we are still able to retain most of the data when focussing on these areas. We do not consider the gender or age group associated with each ILI notification as we assume homogeneous mixing within the population.

We use 20 years of historical climate data, from 1996–2016, obtained from the Australian Bureau of Meteorology (BOM) (introduced in Section 1.3.2). There are a large number of weather stations throughout Australia; we select the weather station closest to the city centre for each of the cities studied: Adelaide Airport (Adelaide), Archerfield Airport (Brisbane), Perth Airport (Perth), and Observatory Hill (Sydney).

Population dynamics are not included in our model to avoid increased complexity. We use a fixed population size based on the June 2016 census by the Australian Bureau of Statistics, with 1.2 million people in Adelaide, 2 million in Brisbane, 1.6 million in Perth and 5 million in Sydney.

3.3 Method

3.3.1 Functional Forms of R_0

We choose four candidate Functional Forms of R_0 to represent the seasonal change in the transmissibility of ILI. Three of these Functional Forms are dependent on climate data; the remaining form is independent of climate but displays seasonal variation throughout a year. The four Functional Forms are:

$$\text{Linear: } R_0(t) = \bar{R}_0(1 - \delta_i x_i(t) - \delta_3 x_3(t)) \qquad i = 1, 2, \tag{3.1}$$

$$\text{Exponential: } R_0(t) = R_0^{\min} + (R_0^{\max} - R_0^{\min})e^{a_i x_i(t)} \qquad i = 1, 2, 3, \tag{3.2}$$

$$\text{Step: } R_0(t) = \mathbb{1}_{\{x_i(t) \geq s\}} R_0^{\text{base}} + \mathbb{1}_{\{x_i(t) < s\}} R_0^{\text{elev}} \qquad i = 1, 2, 3, \tag{3.3}$$

$$\text{Sinusoidal: } R_0(t) = -A\cos\left(\left(\tfrac{2\pi}{730}\right)(t + v)\right) + \bar{R}_0, \tag{3.4}$$

where i indexes the climate variables (absolute humidity ($i = 1$), relative humidity ($i = 2$), and temperature ($i = 3$)); $x_i(t)$ represents the climate measurement of variable i at time t, each scaled to the interval $[-1, 1]$; the coefficient s represents the threshold at which R_0 switches from the base value, R_0^{base}, to the elevated value, R_0^{elev}; A is the amplitude of the sine function of R_0; and v is the number of time steps the sine function is shifted. Figure 3.1 shows an example of each type of Functional Form, showing a distinct difference in shape between the four.

The model parameters used in the Functional Forms, $\bar{R}_0$, δ_i, $R_0^{\min}$, $R_0^{\max}$, a_i, s, R_0^{base}, R_0^{elev}, A and v, are sampled from prior distributions for model selection, as described in Section 3.3.4.

Linear Functional Form

The linear Functional Form (3.1) is based on Yaari *et al.* [31] and Axelsen *et al.* [32]. Both used this Functional Form of R_0 within an SIRS model to explore seasonal influenza in Tel Aviv, Israel. This Functional Form uses a linear combination of temperature with either absolute humidity or relative humidity, each with some weighting parameter δ_i. There is also a parameter $\bar{R}_0$ which represents the value of R_0 in the absence of seasonality. The

values of these parameters are explained in Section 3.3.4.

Exponential Functional Form

The exponential Functional Form (3.2) is based on Shaman *et al.* [79]. From Shaman *et al.*, we take a generalised version of the R_0 Functional Form where we replace the -180 exponential coefficient with a variable, a. This is to allow more flexibility in the fit of the parameters of R_0 to the data. Shaman *et al.* use this Functional Form with specific humidity (a scaled form of absolute humidity) within a deterministic SIRS model framework to consider the impact of humidity in forecasting seasonal influenza in New York City. Within this Functional Form, the parameters are the minimum and maximum values of R_0, and the variable a which determines the impact of the climate variable on the value of R_0.

Step Functional Form

We introduce a step Functional Form (3.3) on the premise that there is a base value for R_0 when transmission rates are low, such as when temperatures are above some threshold s, as well as an elevated value of R_0 when transmission rates are higher, such as when temperatures drop below that threshold. When the value of the climate variable is higher than this threshold, the value of R_0 is at the base level, R_0^{base}, and when the climate value drops below the threshold the value of R_0 switches to the elevated value, R_0^{elev}. This Functional Form usually correlates with season, so that R_0 values are higher in winter and drop in summer. Figure 3.1 gives an example of the step Functional Form, showing the sharp switch between the base and elevated R_0 values. We do not require that the value of R_0^{base} be smaller than the value of R_0^{elev} when the values are sampled from their respective prior distributions, to ensure flexibility in shape within the Functional Form. This also takes into account that correlation between climate and transmissibility may be positive or negative, depending on the climate variable. In Section 3.3.4, we describe the validation method for generating simulations that prevents out-of-sync epidemic seasons from being produced.

Sinusoidal Functional Form

The sinusoidal Functional Form (3.4), based on a cosine function, is dependent only on time and not on any climate variable; this was included as a form of experimental control, to assess that the final selected model was informative. It has three coefficients: A, v and $\bar{R}_0$. The coefficient $\bar{R}_0$ is the mean R_0 value over a year. The coefficient A describes the amplitude of the sine function, while the coefficient v represents how much the function is shifted in terms of time steps. When v is zero, the Functional Form aligns so that the minimum R_0 value occurs at January 1st and the maximum value aligns with exactly six months later, the middle of the year. The stochastic epidemic model uses two time-steps per day, leading to 730 time steps per year; this is represented in the periodicity of the cosine function.

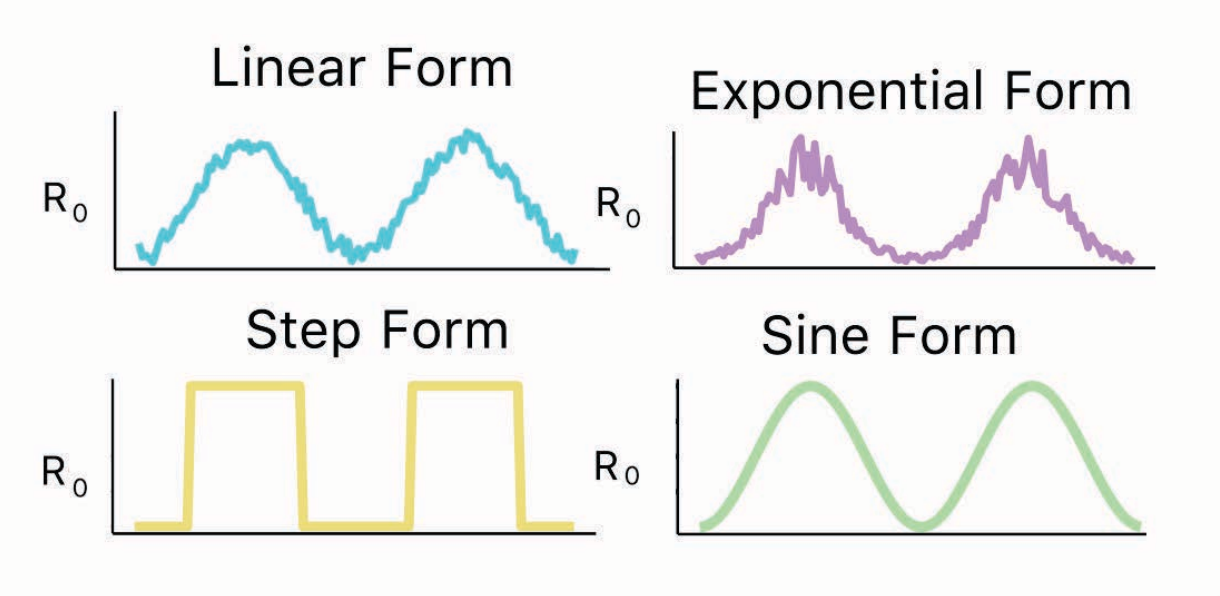

Figure 3.1: An illustration of each of the four types of Functional Forms over two years to compare shape.

3.3.2 The stochastic epidemic model

To model the influenza-like illness dynamics within each population over ten years, we used a stochastic compartmental epidemiological model. It is known that when people become infected with influenza or ILI, there is an exposed period prior to the stage where symptoms become apparent and the individual becomes infectious. As well as this, we know that not everyone who becomes infected will seek medical attention, and not all

those who seek medical attention will do so with an ASPREN participating GP. As such, we choose an SEIR-type model, with an observation process [35]. In this model, we include a waning immunity rate, η, which allows individuals to transition from the recovered state back to the susceptible state. The transitions between classes are stochastic. We use two consecutive infectious and observed states to produce an Erlang-2 infectious period.

Figure 3.2 describes the underlying ILI model being used. In this model, susceptible (S) individuals transition to an exposed (E) state when infected by an infectious individual; from the exposed state the individual moves to the first infectious state (I_1), and may choose to seek treatment from an ASPREN doctor. If they seek treatment then they transition to the first observed state (O_1), otherwise they transition into the second infected state (I_2). From here they can either be observed (O_2) or transition to the recovered state (R). From the first observed state (O_1) the individual transitions to the second observed state (O_2), and then to the recovered state (R). When in the recovered state (R), the individual cannot be reinfected. Over time the immunity wanes and the individual then transitions back to the susceptible state (S), allowing them to be reinfected. These transitions with the increments used are described in Table 3.1.

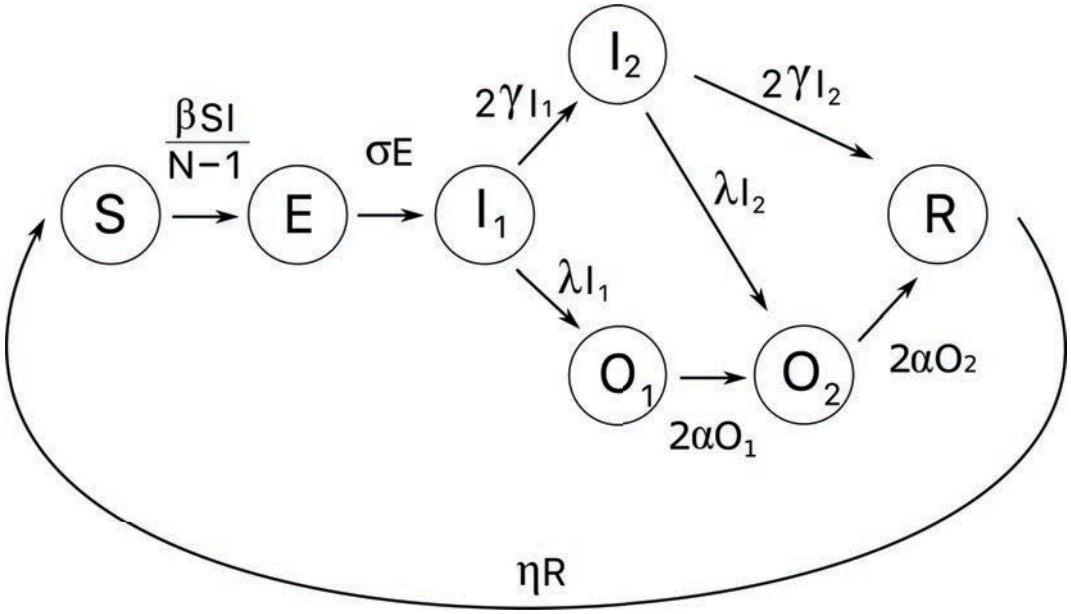

Figure 3.2: A diagram of the compartmental epidemic model.

The transition rates are based on quantities with physical meaning: P_{obs}, T_{inf}, T_{lat}, and T_{imm}. We utilise a hierarchical observation process for this model, where each infected individual has a certain probability P_{obs} of being observed by an ASPREN-reporting

Transitions
$S_{t+1} = S_t + \Delta R_t - \Delta S_t$
$E_{t+1} = E_t + \Delta S_t - \Delta E_t$
$I_{1,t+1} = I_{1,t} + \Delta E_t - \Delta I^I_{1,t} - \Delta I^O_{1,t}$
$I_{2,t+1} = I_{2,t} + \Delta I^I_{1,t} - \Delta I^R_{2,t} - \Delta I^O_{2,t}$
$O_{1,t+1} = O_{1,t} + \Delta I^O_{1,t} - \Delta O_{1,t}$
$O_{2,t+1} = O_{2,t} + \Delta I^O_{2,t} + \Delta O_{1,t} - \Delta O_{2,t}$
$R_{t+1} = R_t + \Delta I^R_{2,t} + \Delta O_{2,t} - \Delta R_t$
Increments
$\Delta S_t = \text{Binomial}(S_t, 1 - \exp(-(\beta \frac{(I_1+I_2+O_1+O_2)}{N-1} + \epsilon)))$
$\Delta E_t = \text{Binomial}(E_t, 1 - \exp(-\sigma))$
$(\Delta I^I_{1,t}, \Delta I^O_{1,t}) = \text{Multinomial}(I_{1,t}, \frac{2\gamma}{2\gamma+\lambda}(1 - \exp(-(2\gamma + \lambda))), \frac{\lambda}{2\gamma+\lambda}(1 - \exp(-(2\gamma + \lambda))))$
$(\Delta I^R_{2,t}, \Delta I^O_{2,t}) = \text{Multinomial}(I_{2,t}, \frac{2\gamma}{2\gamma+\lambda}(1 - \exp(-(2\gamma + \lambda))), \frac{\lambda}{2\gamma+\lambda}(1 - \exp(-(2\gamma + \lambda))))$
$\Delta O_{1,t} = \text{Binomial}(O_{1,t}, 1 - \exp(-2\alpha))$
$\Delta O_{2,t} = \text{Binomial}(O_{2,t}, 1 - \exp(-2\alpha))$
$\Delta R_t = \text{Binomial}(R_t, 1 - \exp(-\eta))$

Table 3.1: Transitions and corresponding stochastic increments used in the SEIIOOR model.

doctor. The probability of being observed is the same in each of the infected stages. The time that an individual spends infected is denoted T_{inf}, while the time an individual is immune after recovery from ILI is denoted T_{imm}. Finally, the latent period between when an individual is exposed and when they experience symptoms of infection is denoted T_{lat}. The relationship between the rates used in the model and these physical quantities is given in Table 3.2, using two time steps per day. Section 3.3.4 describes the prior distributions used to select the values of these physical quantities. These transition rates are derived in Appendix A.

Model parameter	Physical quantities
β	$\frac{R_0(2\gamma+\lambda)}{2}$
σ	$\frac{1}{2T_{\text{lat}}}$
γ	$\frac{\sqrt{1-P_{\text{obs}}}}{2T_{\text{inf}}}$
λ	$\frac{2-2\sqrt{1-P_{\text{obs}}}}{2T_{\text{inf}}}$
α	$3\gamma+\lambda$
η	$\frac{1}{2T_{\text{imm}}}$

Table 3.2: The transition rates given in terms of physical quantities.

3.3.3 Model selection by random forests

To perform Bayesian model selection we use random forests (RF), in the manner proposed by Pudlo *et al.* [62, 80]. The RF algorithm is an ensemble approach, where a 'forest' of binary decision trees is built to classify the data based on the output of the candidate models [54]. Decision trees, when used individually, have a tendency to overfit on the training data. By using an ensemble of decision trees, this reduces overfitting and variance by aggregating the information from multiple decision trees. RFs were introduced in Section 2.4.

At the core of the RF process is the classification and regression tree (CART) algorithm. We only consider classification for this application; in this case, the algorithm works by producing a decision tree which classifies a data entry – a simulation of 10 years of ILI notifications – as one of m candidate models. At each of the branches in this

decision tree, rules are allocated to determine which side of the branch the current entry falls on. The terminal nodes of the decision tree determine which candidate model the data entry is classified to.

The RF algorithm used in this study is presented in Section 2.4. We use $m = 9$ models for each city, where each Functional Form of R_0 with each climate variable is classed as a different model. We generate the N simulations from the SEIIOOR model described in Section 3.3.2, and the parameter sets $\hat{\theta}_i$ are sampled from priors as described in Section 3.3.4.

Using the stochastic epidemic model, we simulate 10,000 simulations of 10 years of ILI data for each of the Functional Forms of R_0 to use as RF training data, using prior distributions as described in Section 3.3.4. We repeat this for each of the four locations. When generating the simulations, we simulate 20 years, of which the first 10 years is discarded as burn-in. We use the 2016 Australian Bureau of Statistics (BOM) estimate for the populations in Adelaide, Brisbane, Sydney and Perth; and we assume that at the start of the simulation, 75% of the population is susceptible to ILI. The prior distributions of parameters and the conditions on the accepted simulations are described in Section 3.3.4.

We used two different approaches for RF model selection: one we call the 'tournament' approach, and the other the 'all-in' approach. The tournament approach first selects the different climate variables *within* each Functional Form, and then selects *between* the Functional Forms. For comparison, we also use the all-in approach which selects between all Functional Forms and all climate variables in a single step. Both of these approaches were implemented in R using the package `randomForest` [66]. We analyse the four cities separately, i.e. we do not consider correlations or interactions between cities.

The tournament approach was chosen as we felt that it was the more careful approach, providing further information about the relationship between climate variables and Functional Forms. By training a random forest on each Functional Form, we are able to gain insight into which climate variable within that Functional Form fits best, rather than in the all-in approach where we are only able to discern which combination of climate variable and Functional Form is most suitable. Comparison showed that the tournament approach produces very similar overall results to the all-in approach, whilst providing additional information through the individual stages.

Tournament approach

The tournament selection process is illustrated in Figure 3.3. In this approach, we use four different random forests in a tournament-style competition. We do this to consider the Functional Form and the associated climate variable separately. Having separate rounds for each Functional Form allows us to see the impact that a particular climate variable has within that Functional Form, and to select the climate variable which best represents influenza-like illness using that Functional Form. In the initial step, we select the best climate variable within each Functional Form by training a random forest on data from each climate variable in a Functional Form. The climate variable in each Functional Form which receives the highest proportion of votes based on the fit with ASPREN data is moved through to the final.

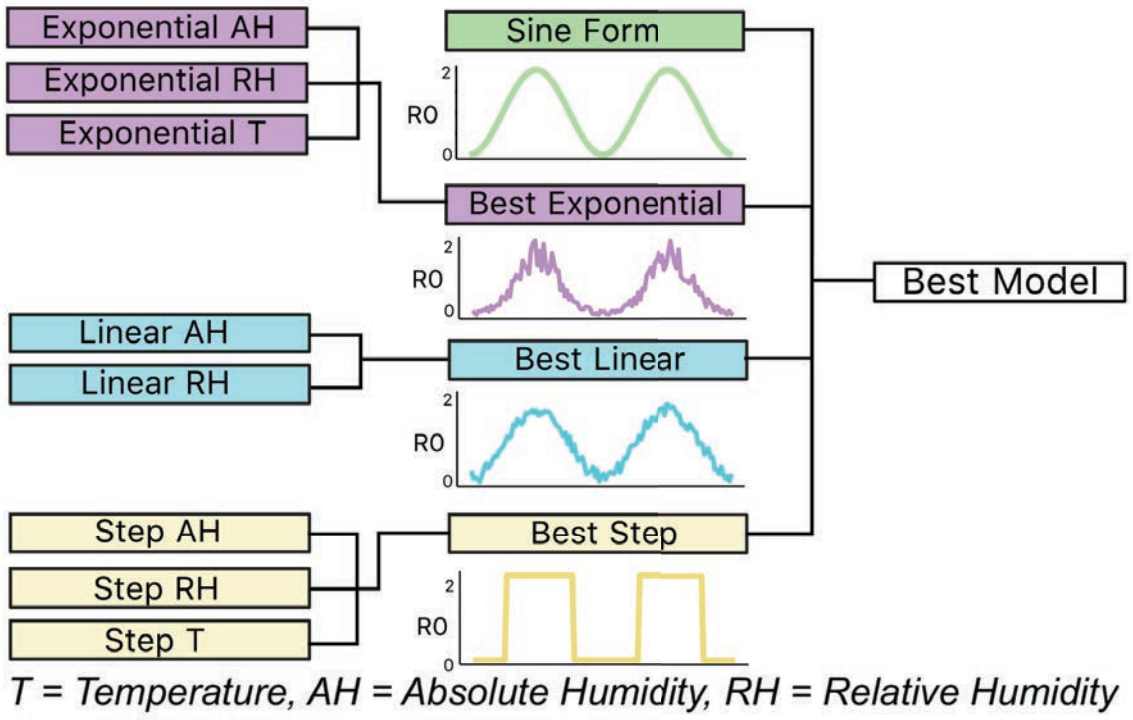

Figure 3.3: A depiction of the tournament-style random forest model selection process.

For the final round, we use a random forest to select between the three winning models from the penultimate round; the best climate variable from each of the three Functional Forms. As well as this, the sinusoidal Functional Form (3.4) is also included in the final round. For this final round, a random forest is again built using 10,000 simulations from each of the four candidate forms; the random forest then assigns votes to each of these forms based on the fit with ASPREN data. The form with the highest proportion of

votes is then chosen as the Functional Form best fitting the ILI data. For each random forest generated in this process, we use 500 decision trees.

All-in approach

For the all-in approach, we build a data set consisting of a training set of 10,000 simulations from each of the 9 Functional Forms; this results in a total data set of 90,000 simulations. The simulations are each labelled with which Functional Form they derive from. We then train a random forest on this large data set, using a forest of 500 decision trees. Once the forest has been trained, we use the ASPREN data from the city being considered and use the random forest to predict which model best fits this data. The output of this is a proportion of votes towards each model, revealing which model best fits to the data.

3.3.4 Selecting the priors

Within this Bayesian model selection framework, we take care with selecting the prior distributions. For the model selection method to accurately allocate the classification tree votes, there are several factors which must be kept consistent between the simulated influenza epidemics for the different Functional Forms. We ensure that model parameters are selected to represent physical quantities, and that the quantities are drawn from the same prior distribution across all of the models. We also perform conditioning to ensure that we compare between simulations that see an outbreak of ILI in each winter season over the ten years.

To ensure the parameters are capable of consistently producing ILI activity each year, we use a five-simulation validation process. Since we know that an epidemic invariably occurs each year, we only accept parameters which generate a simulation where there is a winter ILI outbreak in each of the 10 years. We define an ILI outbreak as at least one ILI infection occurring per day, where the peak ILI occurs between May and September. For a parameter set to be accepted, the first five simulations using that parameter set must have an outbreak each year. Otherwise, we discard the parameters. This allows us to keep only the parameter sets from the prior distributions which are most likely to generate an influenza outbreak each year. This means that, in effect, our Bayesian model selection operates on the prior conditioned on epidemic survival, rather than the initial

'naive' prior.

The value of the quantities used to determine rates in the SEIIOOR model are selected to have physical meaning, as given in Table 3.3. Each of these physical quantities has been assigned a prior distribution. At the start of each simulation, a value for each quantity is drawn from the prior distribution and kept constant throughout the simulation. The prior distributions for each quantity are listed in Table 3.3. Note that we select only a single value for each coefficient for a simulation; the value of the coefficients change between simulations but remain constant within any one simulation. The coefficients used for a simulation are referred to as the 'parameter set'.

Physical quantity	Prior distribution
Duration of infection, T_{inf} (days)	Uniform(0.5, 5)
Treatment seeking probability, P_{obs}	Uniform(0.00005,0.005)
Mean basic reproduction number, $\bar{R}_0$	Uniform(1,2)
Immune duration, T_{imm} (days)	Uniform(0.5,500)
Latent period, T_{lat} (days)	Uniform(0.5, 2)
Rate of imported infections, ϵ (infections/day)	Exponential(10^9)

Table 3.3: Prior distributions of the physical quantities used.

The prior distribution on the mean R_0 value, $\bar{R}_0$, is Uniform(1,2). This prior is kept consistent across the Functional Forms; however, each Functional Form has a different set of coefficients. The value of these coefficients determines $\bar{R}_0$. Determining the correct value of these coefficients to maintain a given $\bar{R}_0$ is non-trivial for the three climate-dependent Functional Forms, (3.1), (3.2) and (3.3). For the sine Functional Form (3.4), it is sufficient to simply set $\bar{R}_0$ equal to the mean R_0 value drawn from the prior, and sample a value of v to shift the Functional Form so that the maximum R_0 can happen any time between the 24^{th} and 34^{th} week, when ILI peaks historically occur.

To determine the correct value of these coefficients to maintain a given $\bar{R}_0$, we first set very wide limits on the value of each of the coefficients in the Functional Forms. We then sample 1,000,000 different combinations of the parameters for each Functional Form in a $100 \times 100 \times 100$ grid, equally-spaced over the parameter distributions given in Table 3.4. We then calculate the mean R_0 corresponding to each combination of parameters,

and accept all combinations resulting in mean R_0 within the interval [1,2]. Table 3.4 shows the prior distribution used for each variable in each Functional Form, where the distribution is the same between all four cities.

To create a reference table of coefficient values for each Functional Form, we uniformly draw 10,000 values of $\bar{R}_0$ from the prior in Table 3.3. For each of these $\bar{R}_0$ values, we select the combination of coefficients which produce a mean R_0 value closest to the $\bar{R}_0$ value. We repeat this process for each of the 10,000 values and each of the linear (3.1), exponential (3.2) and step (3.3) Functional Forms. This results in a reference table of 10,000 values of $\bar{R}_0$ and the corresponding coefficient combination for each of the Functional Forms, with the resulting prior remaining uniform in $\bar{R}_0$. To produce a simulation, when a value of $\bar{R}_0$ is drawn, the reference table is used to determine the combination of coefficients to use for the Functional Form of R_0 for that simulation.

Functional Form	Model parameter	Prior Distribution
Exponential	$R_0^{\min}$	Uniform(0,2)
	$R_0^{\max}$	Uniform(0,10)
	a	Log-Uniform(-5,0)
Linear	R_0^*	Uniform(0,12)
	δ_i, where $i = 1, 2$	Uniform(-1,1)
	δ_3	Uniform(0,2)
Step	s	Uniform(0,1)
	R_0^{elev}	Uniform(0,12)
	R_0^{base}	Uniform(0,12)
Sine	v (weeks)	Uniform(24,34)
	A	Uniform(0,$\bar{R}_0$)

Table 3.4: Prior distribution for all variables used in the Functional Forms of R_0, where $i = 1$ corresponds to absolute humidity, $i = 2$ corresponds to relative humidity, and $i = 3$ corresponds to temperature. The distributions are identical between the four cities.

3.3.5 Approximate Bayesian computation

To better understand the output of the Bayesian model selection process, we use approximate Bayesian computation (ABC) [1] to fit the best two Functional Forms to the ILI data in each of the four locations. We computed a weighted error score for each simulation used in the RF training. The score function

$$S := \sqrt{\frac{1}{\#\text{weeks}} \sum_{i=1}^{\#\text{weeks}} \sqrt{D_i + 1} \times (D_i - \hat{D}_i)^2} \tag{3.5}$$

is the weighted root mean squared error (WRMSE) introduced in Section 2.2, where D_i is the observed number of cases in the i^{th} week of ASPREN ILI data and $\hat{D}_i$ is the number of observed cases in the i^{th} week of the simulated ILI data.

We use this score function in an ABC framework to fit the Functional Form chosen by the RF process to our ASPREN data. To define a threshold that is consistent for each Functional Form and city, we calculate the score for each of the 20,000 simulations from the top two Functional Forms used in the RF process. We then choose the threshold to include the best 5% of these scores. In this case, the same threshold is used across both Functional Forms in each city, but scores are independent between cities. We then generate simulations from that Functional Form and accept those which meet the previously described acceptance criteria, and whose score value is better than the 5% threshold. We continue to generate these simulations until we have 1,000 simulations meeting these conditions. The parameter values which generated these 1,000 simulations then form the posterior distributions of the parameters [52].

3.4 Results

3.4.1 Random forest results

The results of the RF tournament showed that overall, absolute humidity was the best climate variable, selected in all cities across both climates. In Sydney and Brisbane, both subtropical climates, the step Functional Form, with absolute humidity ($i = 1$) as the climate variable, best represented the pattern of influenza-like illness. On the other hand, Adelaide and Perth, both Mediterranean climates, were best represented by the exponential Functional Form, also with absolute humidity ($i = 1$). The all-in RF

approach gave the same results, with minimal differences in the proportion of votes; see Tables 3.5 – 3.12 for further details. Figure 3.4 shows the proportion of votes for each Functional Form in each location, with a clear difference between the Functional Form preferences in Adelaide and Perth, compared with Brisbane and Sydney. Also in Figure 3.4, a climate map of Australia shows the proportion of votes given to each Functional Form in the final round of the RF process in each location, showing the Mediterranean climate of Adelaide and Perth presenting different results to the subtropical climate in Brisbane and Sydney. The rankings of Functional Forms in each stage of the tournament were consistent within each climate but different between climates.

We also analyse the RF results to determine how accurately the process can differentiate between the Functional Forms and climate variables. Tables 3.5 - 3.12 show the confusion matrices for each RF used. The confusion matrix is a table used to assess the performance of the RF, where the body of each table shows the number of simulations classified correctly and incorrectly, and the overall classification error for each Functional Form and RF. We find that each round of the tournament RF process has less than 10% misclassification error in each city, suggesting that the RF is proficient at distinguishing between the different Functional Forms and climate variables. We also find that the different Functional Forms had different rates of misclassification within cities, showing that some Functional Forms are more similar between climate variables than other Forms. For example, in row one of Table 3.5a, 9,994 simulations generated using the Linear Functional Form with climate variable absolute humidity were classified correctly, while 4 were erroneously classified as being from the Linear Functional Form with relative humidity, and 2 from the Linear Functional Form with temperature. This leads to an error for the Linear Functional Form with absolute humidity of 0.0006, or 0.06%. From the same table, the overall misclassification error for the Linear Functional Form RF round in Adelaide is 1.46%. From this data, we determine that the RF is able to accurately distinguish between the Functional Forms and climate variables. The most common misclassification across all cities is within the step Functional Form, between those using relative humidity and temperature. In subtropical locations, there is also notable misclassification of the sinusoidal Functional Form, most commonly misclassified as step Functional Form using relative humidity or temperature.

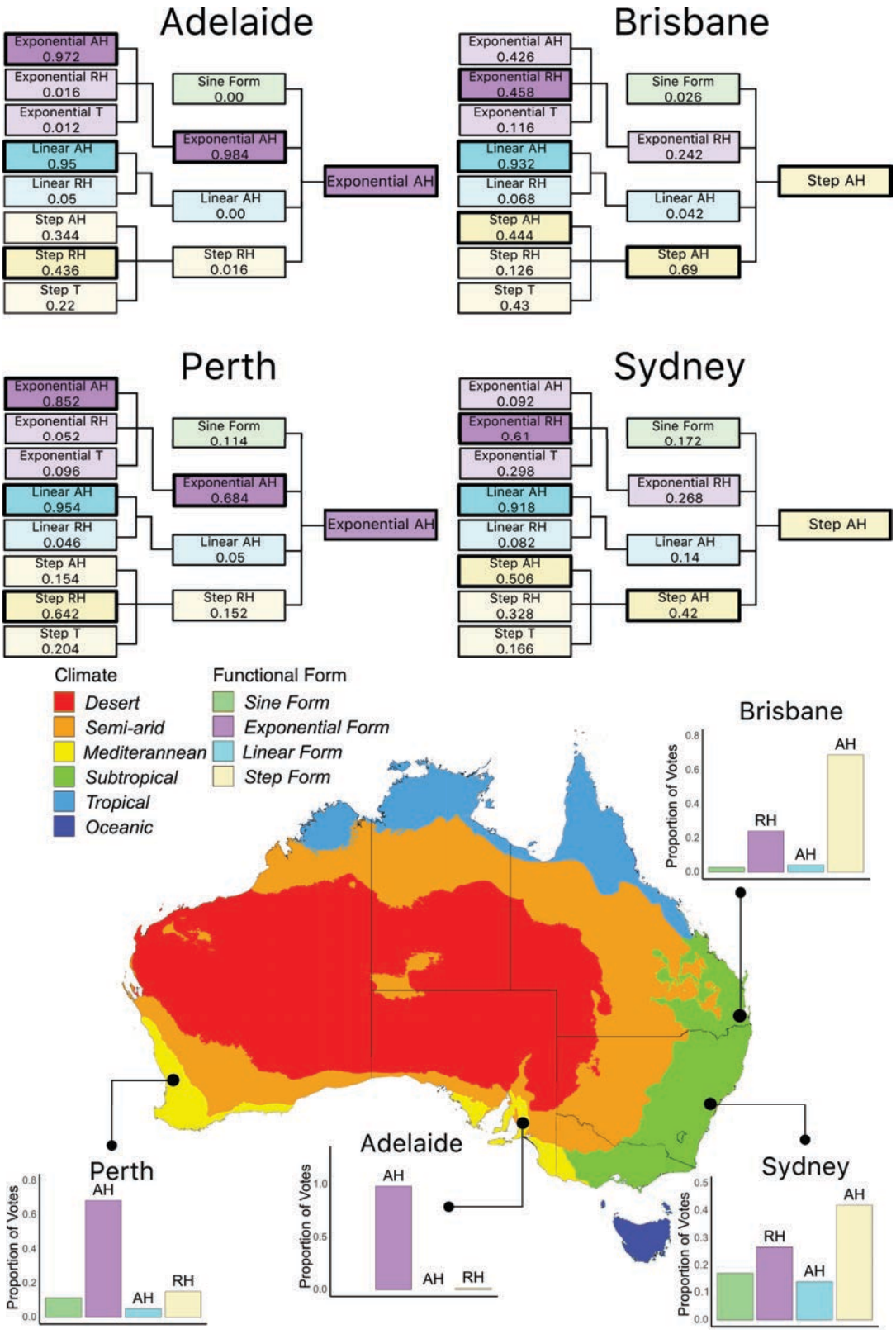

Figure 3.4: The resulting votes for each Functional Form in Adelaide, Brisbane, Perth and Sydney. The corresponding climate variable used in the final stage of the tournament model selection process is shown above each Functional Form in the histogram, with a modified Köppen climate map demonstrating the different climate zones around Australia [2].

The overall misclassification error is the out-of-bag (OOB) error. This is determined once the RF has been fully trained on the data. Out-of-bag samples are drawn from the set of training simulations and classified by the RF. The OOB error is the percentage of these samples which are incorrectly classified [66]. This method is explained in Section 2.4.

We can see that across all cities, the step Round of the RF tournament had a higher overall misclassification error than the other rounds, with the linear round producing the lowest error. This suggests that the different climate variables within the step Functional form may produce simulated epidemics with similar behaviour, leading to difficultly in correctly classifying each simulation.

The misclassification error is higher in Adelaide and Perth compared to Brisbane and Sydney for the all-in RF process; but this difference is not consistent in the tournament RF process, where the misclassification error is similar between all cities except for in the final round, where the error in Adelaide is higher than the other cities. As expected, the all-in process had a higher misclassification error than any of the individual rounds of the tournament process, due to the all-in RF process comparing all nine Functional Forms at once. However, the all-in process produces similar results to the tournament process.

Table 3.5: Confusion matrices for each round of the random forest tournament process in Adelaide.

(a) Exponential Round

true model \ predicted model	AH	RH	T	Error
AH	9994	4	2	0.0006
RH	12	9715	273	0.0285
T	20	127	9853	0.0147

Overall OOB Error: 1.46%

(b) Step Round

true model \ predicted model	AH	RH	T	Error
AH	9158	518	324	0.0842
RH	360	9448	192	0.0552
T	719	673	8608	0.1392

Overall OOB Error: 9.26%

(c) Linear Round

true model \ predicted model	AH	RH	Error
AH	9982	18	0.0018
RH	61	9939	0.0061

Overall OOB Error: 0.4%

(d) Final Round

true model \ predicted model	Exp AH	Lin AH	Step RH	Sine	Error
Exp AH	9997	0	3	0	0.0003
Lin AH	6	9646	289	59	0.0354
Step RH	24	137	9818	21	0.0182
Sine	17	166	260	9557	0.0443

Overall OOB Error: 2.46%

Table 3.6: Confusion matrix for the all-in RF in Adelaide.

true model \ predicted model	Exp AH	Exp RH	Exp T	Lin AH	Lin RH	Step AH	StepRH	Step T	Sine	Error
Exp AH	9995	0	2	0	0	0	0	3	0	0.0005
Exp RH	9	9244	256	14	1	14	179	199	84	0.0756
Exp T	17	157	9323	20	5	5	188	215	70	0.0677
Lin AH	5	52	698	8917	8	84	88	115	33	0.1083
Lin RH	1	133	31	12	9483	2	118	121	99	0.0517
Step AH	17	33	431	157	11	9112	110	89	40	0.0888
Step RH	29	359	284	26	22	5	7709	1085	481	0.2291
Step T	20	363	335	19	19	7	775	8233	229	0.1767
Sine	11	368	242	48	10	17	1203	1181	6920	0.3080
Overall OOB Error:										12.29%

Table 3.7: Confusion matrices for each round of the random forest tournament process in Brisbane.

(a) Exponential Round

true model \ predicted model	AH	RH	T	Error
AH	9891	65	44	0.0109
RH	33	9934	33	0.0066
T	25	117	9858	0.0142

Overall OOB Error: 1.06%

(b) Step Round

true model \ predicted model	AH	RH	T	Error
AH	9414	333	253	0.0586
RH	458	9054	488	0.0946
T	440	374	9816	0.0814

Overall OOB Error: 7.82%

(c) Linear Round

true model \ predicted model	AH	RH	Error
AH	9942	58	0.0058
RH	148	9852	0.0148

Overall OOB Error: 1.03%

(d) Final Round

true model \ predicted model	Exp RH	Lin AH	Step AH	Sine	Error
Exp RH	9450	18	531	1	0.0550
Lin AH	106	9350	413	131	0.0650
Step AH	11	25	9962	2	0.0038
Sine	99	101	225	9575	0.0425

Overall OOB Error: 4.16%

Table 3.8: Confusion matrix for the all-in RF in Brisbane.

true model \ predicted model	**Exp AH**	**Exp RH**	**Exp T**	**Lin AH**	**Lin RH**	**Step AH**	**StepRH**	**Step T**	**Sine**	**Error**
Exp AH	9837	25	35	19	0	6	36	12	30	0.0163
Exp RH	39	9271	17	19	0	1	299	153	201	0.0729
Exp T	56	14	9394	3	0	10	210	146	167	0.0606
Lin AH	88	59	365	8925	7	138	183	134	101	0.1075
Lin RH	1	67	0	31	9423	11	160	152	155	0.0577
Step AH	63	60	53	77	4	9492	103	72	76	0.0508
Step RH	8	94	27	5	0	2	8939	440	485	0.1061
Step T	8	152	51	8	5	4	914	8087	771	0.1913
Sine	11	147	47	13	8	15	1040	701	8081	0.1982
								Overall OOB Error:		9.57%

Table 3.9: Confusion matrices for each round of the random forest tournament process in Perth.

(a) Exponential Round

true model \ predicted model	AH	RH	T	Error
AH	9957	1	42	0.0043
RH	34	9647	319	0.0353
T	40	74	9886	0.0114

Overall OOB Error: 1.7%

(b) Step Round

true model \ predicted model	AH	RH	T	Error
AH	8666	682	652	0.1334
RH	117	9610	213	0.0390
T	597	568	8835	0.1165

Overall OOB Error: 9.63%

(c) Linear Round

true model \ predicted model	AH	RH	Error
AH	9976	24	0.0024
RH	95	9905	0.0095

Overall OOB Error: 0.6%

(d) Final Round

true model \ predicted model	Exp AH	Lin AH	Step RH	Sine	Error
Exp AH	9929	0	71	0	0.0071
Lin AH	19	9482	294	205	0.0518
Step RH	2	6	9991	1	0.0009
Sine	8	116	369	9507	0.0493

Overall OOB Error: 2.73%

Table 3.10: Confusion matrix for the all-in RF in Perth.

true model \ predicted model	**Exp AH**	**Exp RH**	**Exp T**	**Lin AH**	**Lin RH**	**Step AH**	**StepRH**	**Step T**	**Sine**	**Error**
Exp AH	9969	0	4	0	0	0	1	25	1	0.0031
Exp RH	5	9033	238	18	2	29	144	400	131	0.0967
Exp T	30	108	9164	12	4	6	113	442	121	0.0836
Lin AH	14	18	498	9013	13	220	50	124	50	0.0987
Lin RH	0	82	36	10	9584	1	87	119	81	0.0416
Step AH	3	18	444	146	2	9022	93	213	59	0.0978
Step RH	1	208	103	16	11	15	7321	1100	1225	0.2679
Step T	0	48	193	4	9	1	316	9113	316	0.0887
Sine	0	200	135	13	15	8	970	1110	7549	0.2451
Overall OOB Error:										11.37%

Table 3.11: Confusion matrices for each round of the random forest tournament process in Sydney.

(a) Exponential Round

true model \ predicted model	AH	RH	T	Error
AH	9830	35	135	0.0170
RH	0	9999	1	0.0001
T	22	168	9810	0.0190

Overall OOB Error: 1.2%

(b) Step Round

true model \ predicted model	AH	RH	T	Error
AH	9029	346	245	0.0614
RH	445	8822	393	0.0867
T	783	669	8108	0.1519

Overall OOB Error: 9.99%

(c) Linear Round

true model \ predicted model	AH	RH	Error
AH	9992	8	0.0008
RH	105	9895	0.0105

Overall OOB Error: 0.56%

(d) Final Round

true model \ predicted model	Exp RH	Lin AH	Step AH	Sine	Error
Exp RH	9902	0	98	0	0.0098
Lin AH	21	9547	401	31	0.0453
Step AH	5	15	9976	4	0.0025
Sine	29	170	296	9505	0.0495

Overall OOB Error: 2.7%

Table 3.12: Confusion matrix for the all-in RF in Sydney.

true model \ predicted model	**Exp AH**	**Exp RH**	**Exp T**	**Lin AH**	**Lin RH**	**Step AH**	**StepRH**	**Step T**	**Sine**	**Error**
Exp AH	9756	25	102	37	1	10	42	16	11	0.0244
Exp RH	0	9923	0	0	0	0	46	13	18	0.0077
Exp T	28	27	9037	48	1	11	454	208	186	0.0963
Lin AH	136	26	571	8815	5	29	230	93	95	0.1185
Lin RH	3	19	6	9	9644	2	154	80	83	0.0356
Step AH	99	19	233	144	10	9186	127	90	92	0.0814
Step RH	0	12	1	3	1	0	9273	459	251	0.0727
Step T	0	27	7	0	9	0	1004	8509	444	0.1491
Sine	2	25	33	7	14	1	931	888	8099	0.1901

Overall OOB Error: 8.62%

By analysing the WRMSE scores for the simulations used for each Functional Form, we are able to gain further insight and additional context to the RF model selection process. The scores assist with interpreting the RF results, as due to the 'black-box' nature of RFs it is difficult to understand the individual decisions behind this model selection process. Tables 3.13 - 3.16 show a summary of the WRMSE score distribution over the 10,000 simulations used for each Functional Form in each city. Highlighted in bold are the Functional Forms selected by the RF process. In three of the four cities, the RF process appears to be selecting the Functional Form with the most positive skew, suggesting that the 'best' scoring realisations of each Functional Form may have a significant impact on the RF selection process. In Adelaide (Table 3.13) the exponential Functional Form with absolute humidity is selected as best, and this is the Functional Form with the lowest minimum, median, mean and maximum score. We observe similar results in Brisbane, with the lowest median and mean occurring in the step Functional Form with absolute humidity. In Perth, the exponential Functional Form with absolute humidity has the lowest minimum, mean and maximum values but not the lowest median value. In contrast, in Sydney the step Functional Form with absolute humidity does not appear to have the lowest score in any of these summary statistics.

Table 3.13: Adelaide score statistics. Highlighted in bold are the Functional Forms selected by the RF process.

Functional Form	Min.	Median	Mean	Max.
Exponential AH	**18.51**	**26.63**	**29.08**	**1338.30**
Exponential RH	19.07	32.36	61.73	2398.01
Exponential T	18.89	30.51	59.98	3511.34
Linear AH	19.51	33.14	80.30	3944.41
Linear RH	22.60	41.93	106.75	4473.30
Step AH	22.57	29.79	72.61	2665.14
Step RH	22.60	28.95	68.33	3362.39
Step T	22.73	30.09	78.56	4120.53
Sine	19.23	36.65	88.78	2540.08

Table 3.14: Brisbane score statistics. Highlighted in bold are the Functional Forms selected by the RF process.

Functional Form	Min.	Median	Mean	Max.
Exponential AH	18.68	30.60	59.47	2346.06
Exponential RH	17.49	26.23	55.54	2408.21
Exponential T	17.72	27.55	57.82	2388.89
Linear AH	17.79	33.84	75.18	2040.55
Linear RH	20.79	34.28	87.15	3468.22
Step AH	**20.59**	**25.48**	**54.51**	**2083.06**
Step RH	20.53	25.94	60.53	3452.29
Step T	20.63	26.18	62.52	4106.06
Sine	15.48	37.59	86.80	2594.66

Table 3.15: Perth score statistics. Highlighted in bold are the Functional Forms selected by the RF process.

Functional Form	Min.	Median	Mean	Max.
Exponential AH	**57.54**	**98.16**	**100.16**	**1028.21**
Exponential RH	69.66	107.65	127.17	2535.72
Exponential T	59.95	102.70	118.73	2463.24
Linear AH	58.84	105.39	141.48	3427.41
Linear RH	78.56	104.28	164.70	3848.54
Step AH	78.10	96.02	134.31	3858.03
Step RH	77.56	100.54	120.77	3539.42
Step T	78.50	97.10	133.10	3763.61
Sine	58.95	104.74	140.85	3175.08

Table 3.16: Sydney score statistics. Highlighted in bold are the Functional Forms selected by the RF process.

Functional Form	Min.	Median	Mean	Max.
Exponential AH	68.92	117.82	139.87	3875.03
Exponential RH	75.83	120.73	130.88	2586.47
Exponential T	67.82	119.15	133.91	2165.85
Linear AH	70.65	121.56	145.60	2971.36
Linear RH	77.2	112.3	166.5	4226.5
Step AH	**77.26**	**120.90**	**137.50**	**4462.83**
Step RH	76.99	116.62	139.80	5112.55
Step T	76.84	111.41	146.00	4372.59
Sine	68.41	122.38	157.30	2840.50

3.4.2 ABC results

Figures 3.5 – 3.16 describe the results from the ABC section of this chapter. They show model fits and R_0 values over time for each of the four cities, as well as pair plots showing the distribution of parameters between Functional Forms.

Adelaide

In Adelaide, the RF model selection process selected the exponential Functional Form using absolute humidity as the most suitable, followed by the step Functional Form with relative humidity. Figure 3.5 shows a fan chart of the 1,000 accepted simulations from each Functional Form in Adelaide, with the true ILI data shown for reference. Both Functional Forms appear to fit well at this 5% threshold, however neither Functional Form accurately fit the extremely large ILI peaks seen in 2009 and 2015. However, in general both Functional Forms appear to accurately estimate the size and timing of ILI in most of the years.

To further explore why the random forest selected the exponential Functional Form using absolute humidity, we construct a fan chart with the realisations of R_0 that generated each of the 1,000 accepted simulations, shown in Figure 3.6. There is a clearly defined seasonal trend in R_0, which matches the seasonal behaviour of ILI. The difference between the two Functional Forms is more noticeable here, with the R_0 from the exponential Functional Form generally following a more distinct seasonal pattern than that of the step Functional Form. The exponential Functional Form also does not reach the very large values of R_0 that are seen in the outlying values of the step Functional Form. The step Functional Form overall takes a greater range of values of R_0 than the exponential Functional Form. We can see that the R_0 values in the step Functional Form range between 0–5 throughout the years, whilst the R_0 for the exponential Functional Form ranges between 0.5–3. Individual realisations of R_0 for the ten best scoring simulations of each Functional Form can be found in Appendix B, Figure B.1, showing the large amount of diversity in the R_0 realisations.

Figure 3.7 shows a pair plot of the posterior model parameter values from the accepted simulations from the two Functional Forms. We can see in this plot that generally, the distributions of the parameters are similar between the Functional Forms. The exception to this is the value of mean R_0, $\bar{R}_0$. We can see from the plots in the left-most column that the values of mean R_0 generally are higher for the exponential Functional Form when compared with the step Functional Form. This may contribute in some way to the random forest selecting the exponential Functional Form.

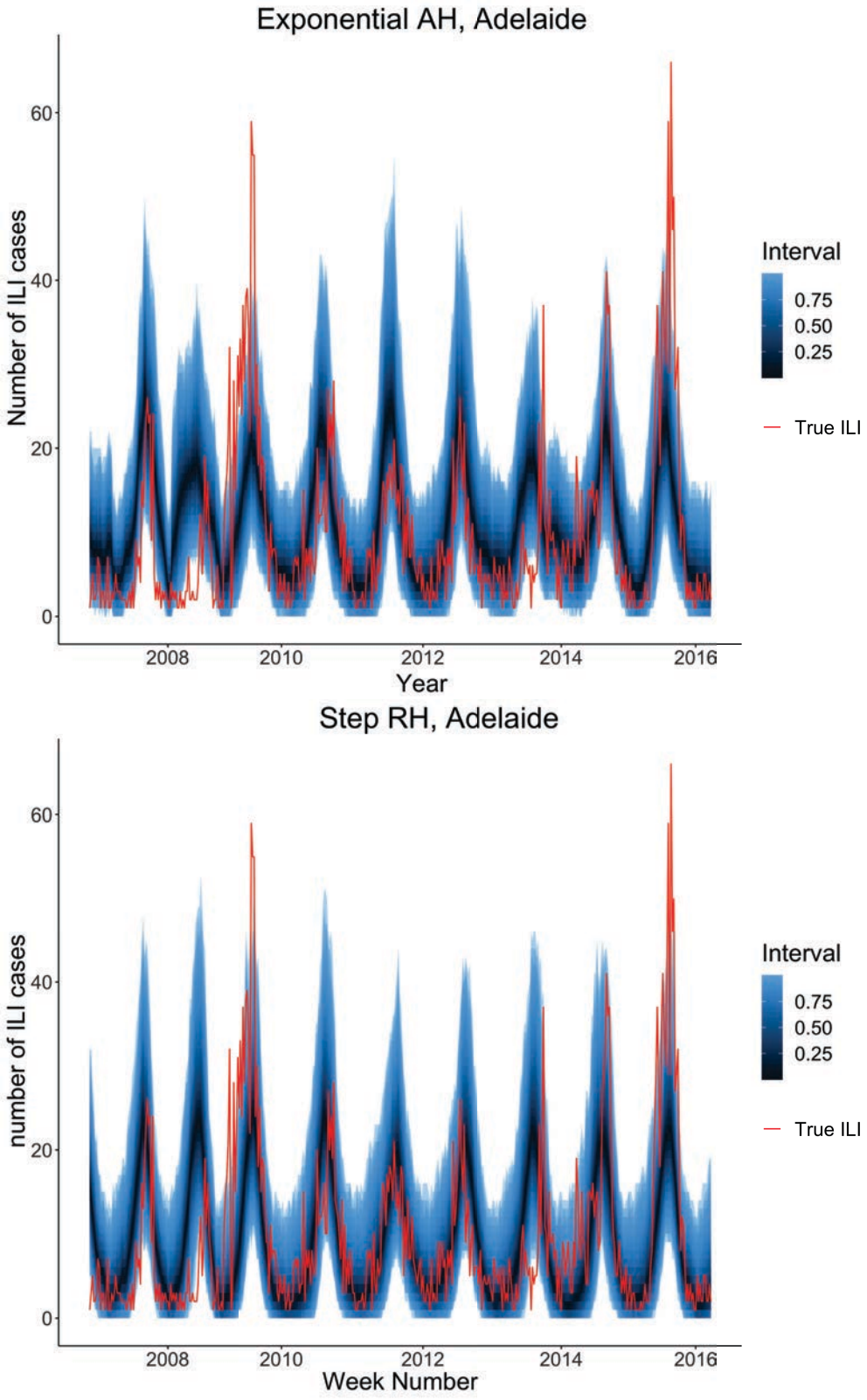

Figure 3.5: Fan charts showing the 1,000 simulated ILI datasets accepted by the ABC process over the years 2006–2016 for the top two Functional Forms as selected by the RF process in Adelaide, with the interval showing quartile ranges. The true ILI data is overlaid for comparison.

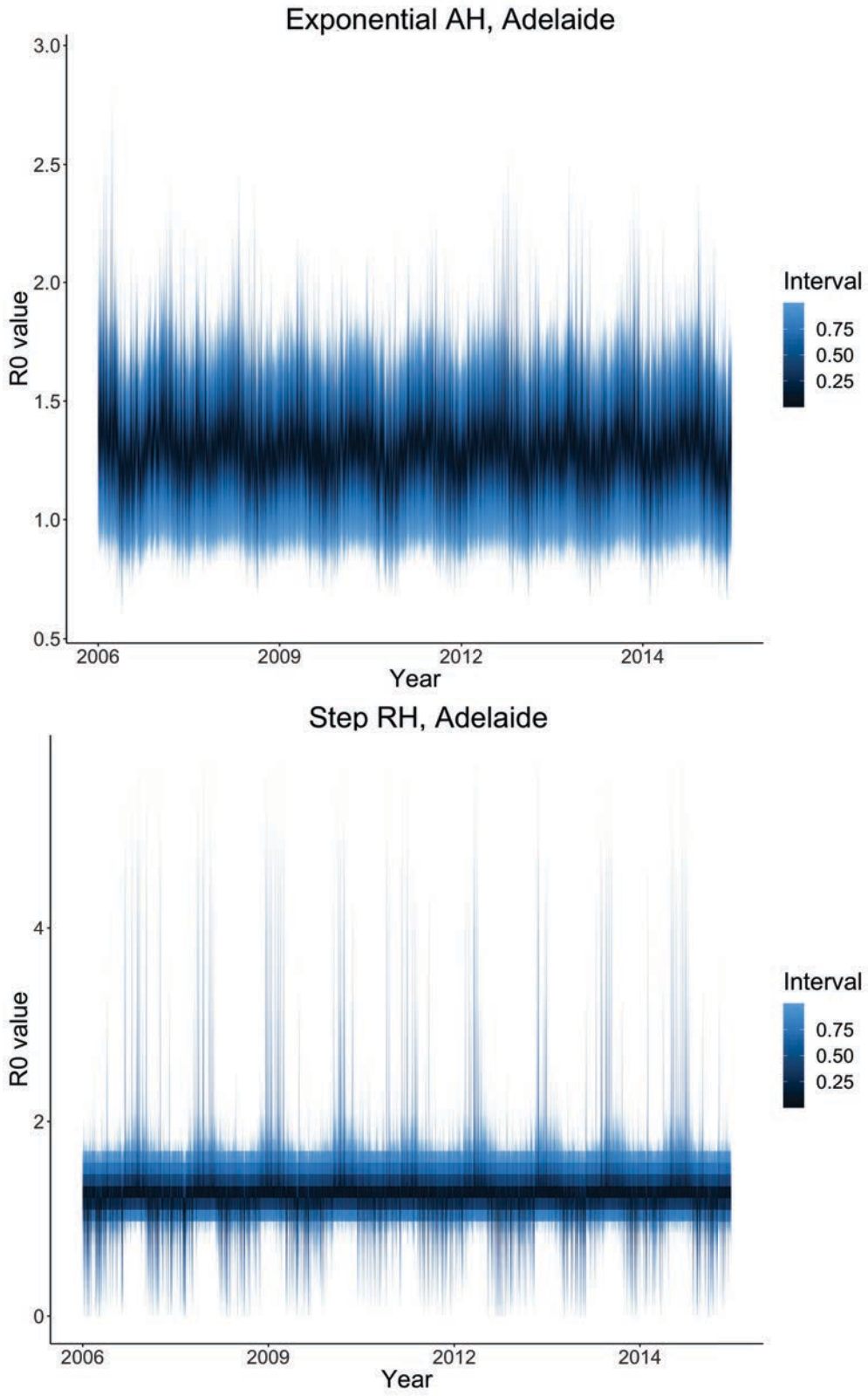

Figure 3.6: Fan charts showing the realisations of R_0 that generated the accepted simulations for the two selected Functional Forms in Adelaide, with the interval showing quartile ranges.

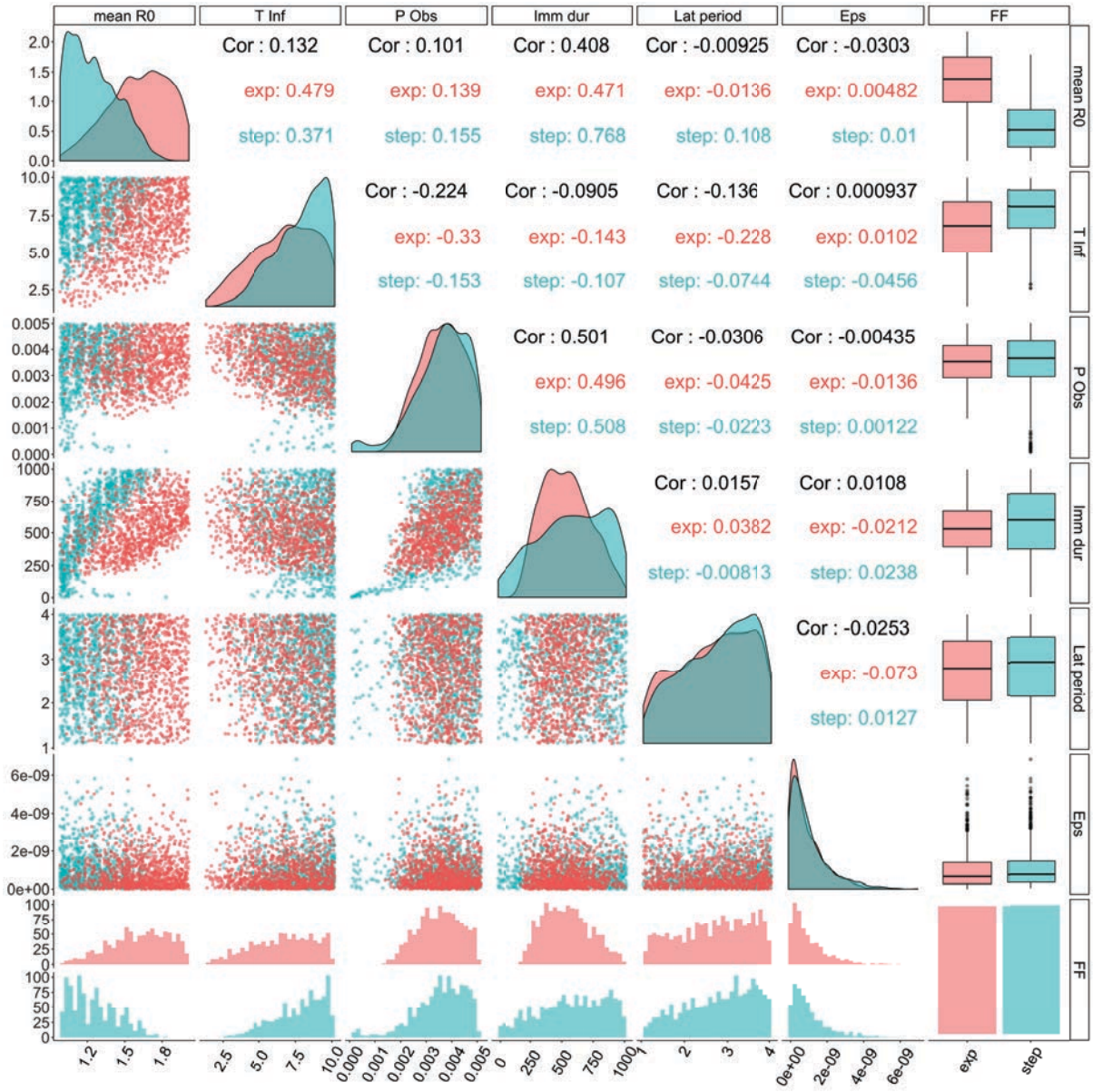

Figure 3.7: A pair plot comparing the posterior model parameters for accepted simulations from both Functional Forms in Adelaide. The parameter 'mean R0' is the mean value of R_0 over all seasons, $\bar{R}_0$; 'T inf' is the time infectious; 'P obs' is the probability of observation by an ASPREN doctor; 'Imm dur' is the duration of immunity after recovery from ILI; 'Lat period' is the latent period between exposure and infectiousness; 'Eps' is the number of external ILI cases introduced into the population; and 'FF' is the Functional Form, where 'exp' is exponential using absolute humidity and 'step' is step using relative humidity. The bottom right plot shows an equal number of particles from each Functional Form were used.

Brisbane

In Brisbane, the step Functional Form using absolute humidity was selected by the RF model selection process, followed by the exponential Functional Form using relative humidity. In Figure 3.8 we present a fan chart of the 1,000 accepted simulations from the two best Functional Forms. While the fit is not as good as that seen in Adelaide, the two Functional Forms are both able to fit the true ILI data quite accurately. Unlike Adelaide, Brisbane can have multiple ILI outbreaks per year, such as in 2010. As we saw in the Adelaide results, neither of these Functional Forms are able to correctly fit to the very large peak; in Brisbane, this very large ILI peak occurs in 2012. We can also see from these plots that simulations from both Functional Forms appear to have two peaks of ILI in some years, most notably in 2009, 2011 and 2012. This trend is more pronounced in the simulations from the step Functional Form, and is not seen as distinctly in the true ILI data.

Figure 3.9 presents a fan chart of the realisations of R_0 that generated the 1,000 accepted simulations from each Functional Form. The R_0 realisations are very different between the two Functional Forms. When compared to Figure 3.6, we can see that the realisations of R_0 are substantially more different between Functional Forms in Brisbane than in Adelaide. In Brisbane, the exponential Functional Form appears quite flat, with very large outliers. The seasonality of R_0 is visible but not well defined. The step Functional Form, however, shows the seasonality of ILI transmission much more noticeably. The realisations of R_0 generally follow a different pattern between Functional Forms, with the exponential Functional Form generally showing large upward peaks from a baseline while the step Functional Form showing deep troughs below the baseline. Individual realisations of R_0 for the ten best scoring simulations of each Functional Form can be found in Appendix B, Figure B.2.

Figure 3.10 shows a posterior pair plot for the parameter values from the simulations from both Functional Forms. In this Figure, we can see some noticeable differences between the Functional Forms. Like in Adelaide, the mean R_0 values tend to be lower in the step Functional Form and higher in the exponential Functional Form. Unlike in Adelaide however, this pattern is also seen in the values of P_{obs} and immDur - the probability of observation by ASPREN doctors and the duration of immunity after recovery, respectively.

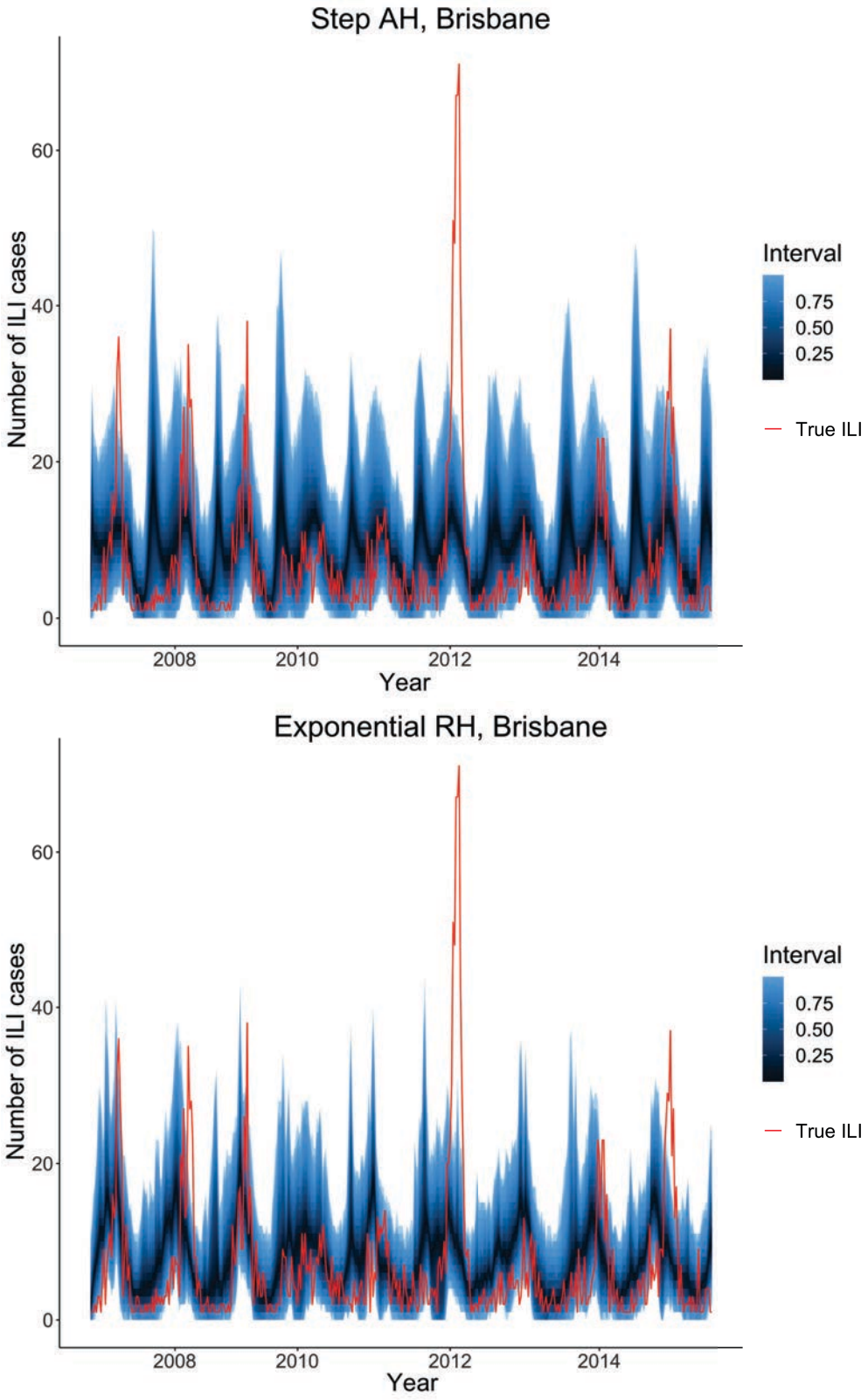

Figure 3.8: Fan charts showing the 1,000 simulated ILI datasets accepted by the ABC process over the years 2006–2016 for the top two Functional Forms as selected by the RF process in Brisbane, with the interval showing quartile ranges. The true ILI data is overlaid for comparison.

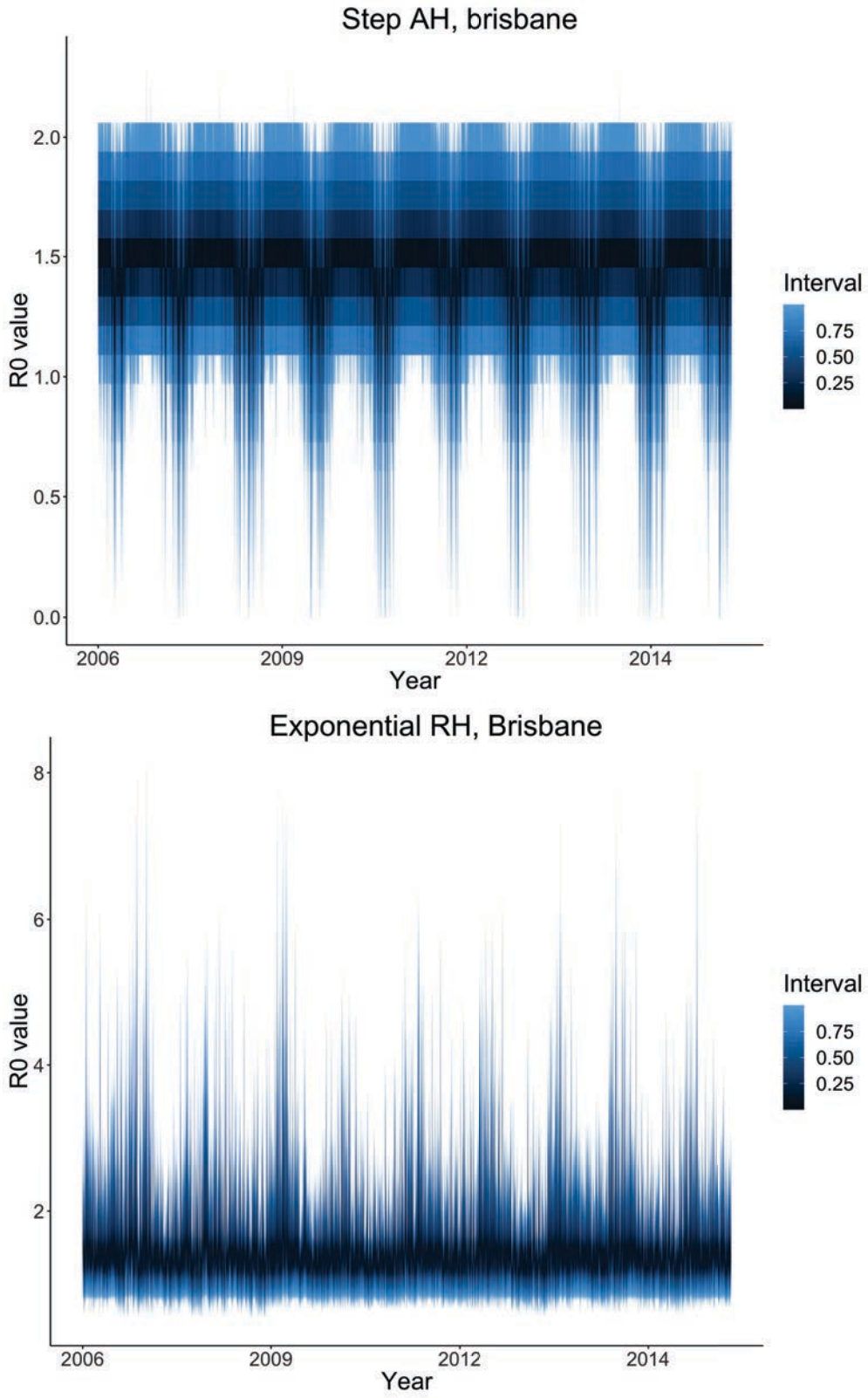

Figure 3.9: Fan charts showing the realisations of R_0 that generated the accepted simulations for the two selected Functional Forms in Brisbane, with the interval showing quartile ranges.

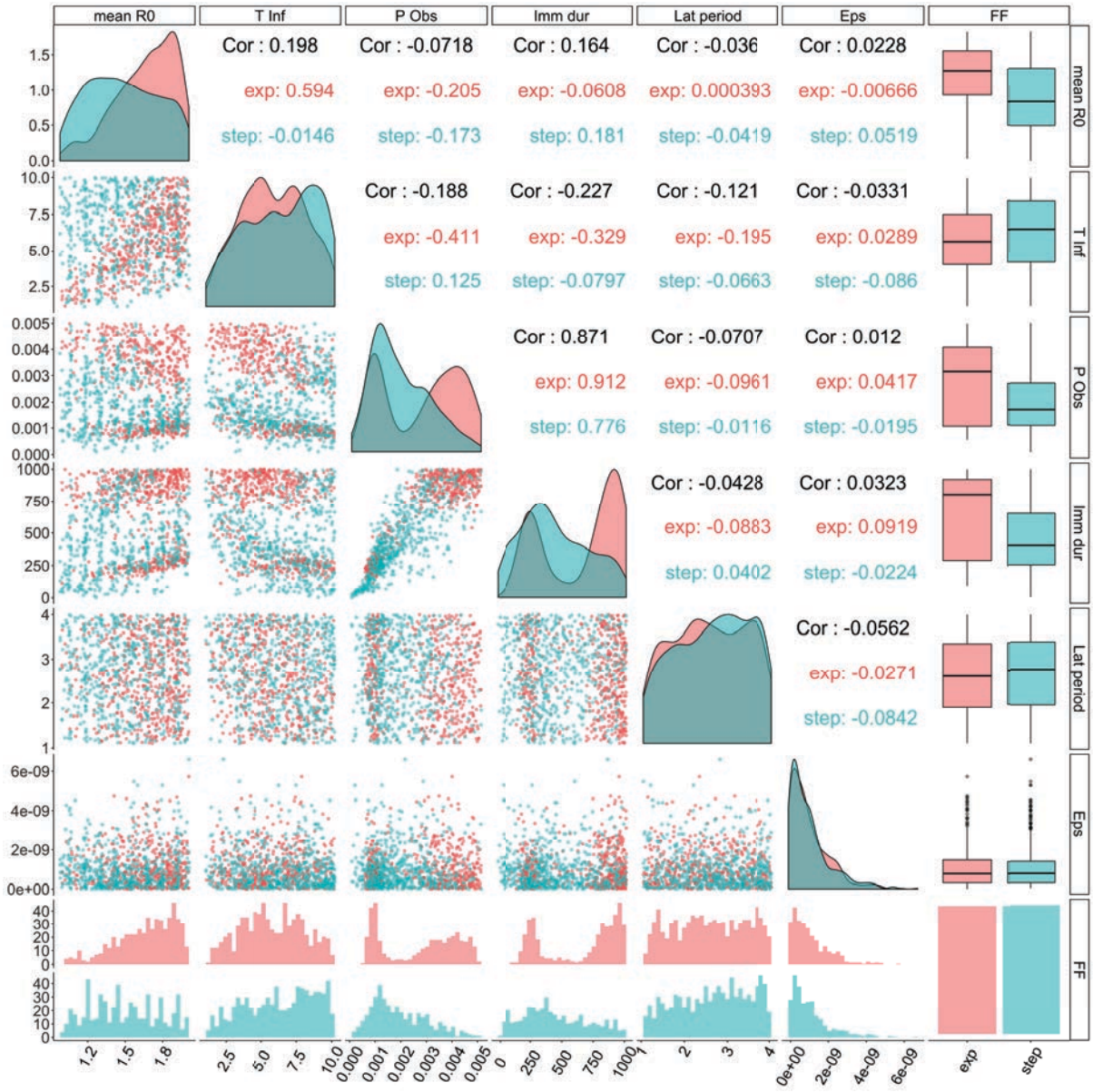

Figure 3.10: A pair plot comparing the posterior model parameters for accepted simulations from both Functional Forms in Brisbane. The parameter 'mean R0' is the mean value of R_0 over all seasons, $\bar{R}_0$; 'T inf' is the time infectious; 'P obs' is the probability of observation by an ASPREN doctor; 'Imm dur' is the duration of immunity after recovery from ILI; 'Lat period' is the latent period between exposure and infectiousness; 'Eps' is the number of external ILI cases introduced into the population; and 'FF' is the Functional Form, where 'exp' is exponential using relative humidity and 'step' is step using absolute humidity. The bottom right plot shows an equal number of particles from each Functional Form were used.

Perth

In Perth, the RF model selection process selected the exponential Functional Form with absolute humidity as the most suitable, followed by the step Functional Form with relative humidity. Figure 3.11 shows a fan chart of the 1,000 accepted simulations from each Functional Form in Perth, with the true ILI data shown for reference. We can see from this Figure that the results of the ABC fitting are quite good. The exponential Functional Form appears to fit the true ILI data slightly better than the step Functional Form. However, as with previous cities, the simulations from both Functional Forms struggle to accurately fit the very large peaks seen in 2007, 2009 and 2012 in Perth.

We constructed a fan chart with the realisations of R_0 that generated each of the 1,000 simulations, shown in Figure 3.12. There is a notable difference between the Functional Forms when considering the R_0 realisations. The exponential Functional Form appears to follow the same patterns as seasonal ILI, with a distinct peak in each year. The step Functional Form, on the other hand, appears to remain flat for most of the year, exhibiting large peaks during the ILI season and dropping back to the baseline level quite quickly. An interesting feature of the exponential Functional Form is a very large peak in R_0 during the 2010 ILI season. This large peak is not seen in the resulting simulations or the true ILI data. Individual realisations of R_0 for the ten best scoring simulations of each Functional Form can be found in Appendix B, Figure B.3.

Figure 3.13 shows a pair plot of the model parameter values from each of the accepted simulations from the two Functional Forms. In general, the parameter values are distributed very similarly between Functional Forms. This is different to what has been seen in previous cities, where the mean R_0 distribution usually varied between Functional Forms.

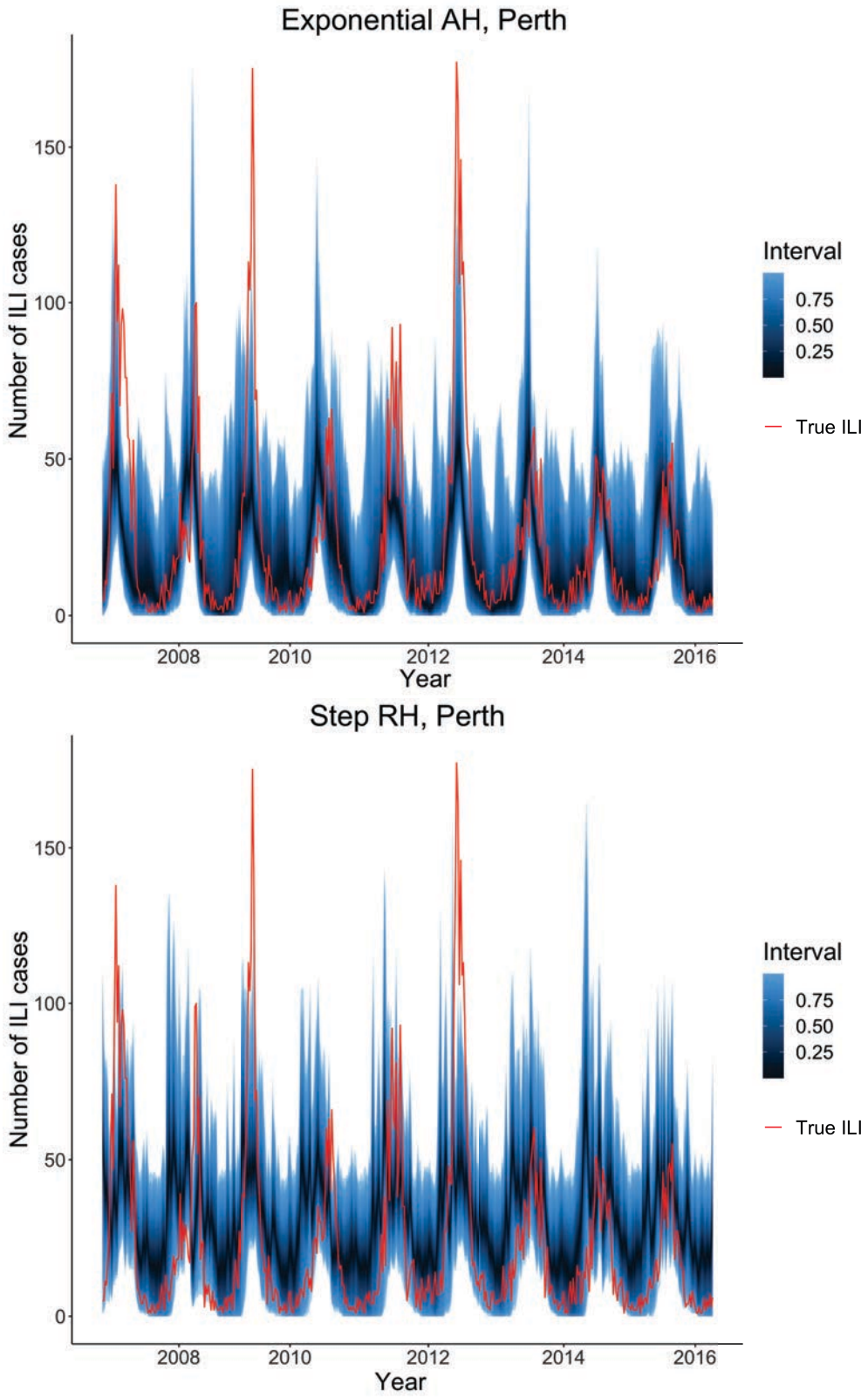

Figure 3.11: Fan charts showing the 1,000 simulated ILI datasets accepted by the ABC process over the years 2006–2016 for the top two Functional Forms as selected by the RF process in Perth, with the interval showing quartile ranges. The true ILI data is overlaid for comparison.

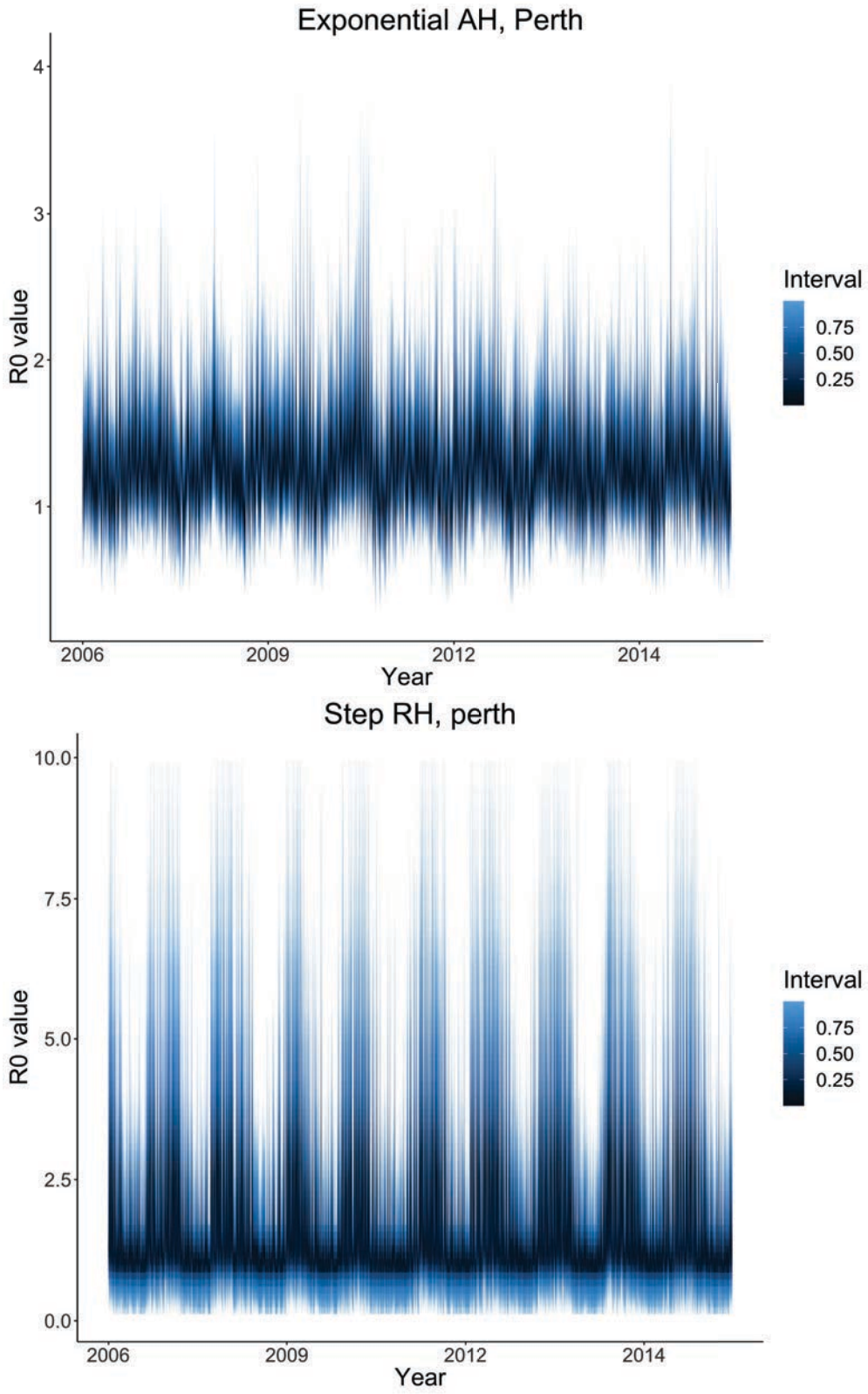

Figure 3.12: Fan charts showing the realisations of R_0 that generated the accepted simulations for the two selected Functional Forms in Perth, with the interval showing quartile ranges.

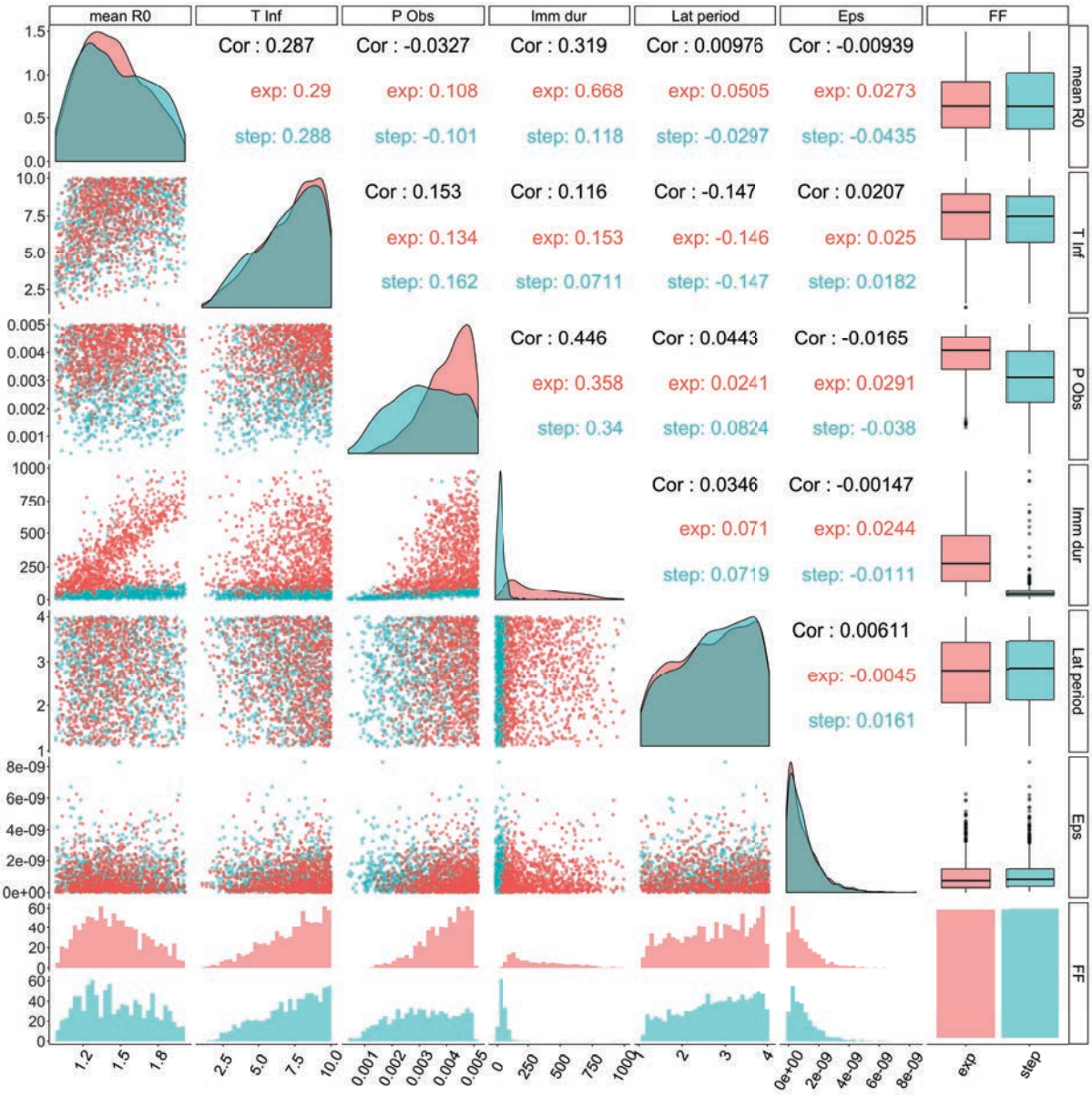

Figure 3.13: A pair plot comparing the posterior model parameters for accepted simulations from both Functional Forms in Perth. The parameter 'mean R0' is the mean value of R_0 over all seasons, $\bar{R}_0$; 'T inf' is the time infectious; 'P obs' is the probability of observation by an ASPREN doctor; 'Imm dur' is the duration of immunity after recovery from ILI; 'Lat period' is the latent period between exposure and infectiousness; 'Eps' is the number of external ILI cases introduced into the population; and 'FF' is the Functional Form, where 'exp' is exponential using absolute humidity and 'step' is step using relative humidity. The bottom right plot shows an equal number of particles from each Functional Form were used.

Sydney

In Sydney, the RF model selection process selected the step Functional Form with absolute humidity, followed by the exponential Functional Form with relative humidity. In Figure 3.14 we present a fan chart of these 1,000 accepted simulations from each of the two selected Functional Forms. We can see that, in comparison to the previous cities, the fit of the ABC simulations is not very accurate. In particular, in the exponential Functional Form, the model does not appear to fit to a defined peak in each year. In the preferenced step Functional Form, simulations do form a single peak during the ILI seasons and track the true ILI data, but with some bias.

We can also see when looking at Figure 3.14 that the true ILI data in Sydney shows more between-season ILI activity than in the other cities. There are many more fluctuations throughout each year in Sydney than other locations, as well as three distinctly large peaks occurring in the 2013–2015 seasons with an unusually large number of between-season cases.

Figure 3.15 shows a fan chart of the realisations of R_0 that generated the 1,000 accepted simulations from each Functional Form. The R_0 realisations are different between Functional Forms, but not as distinctly different as in previous cities. Individual realisations of R_0 for the ten best scoring simulations of each Functional Form can be found in Appendix B, Figure B.4.

Figure 3.16 shows a pair plot for the parameter values from the simulations from both Functional Forms. In this Figure, we can see that the distributions of model parameters are generally quite similar between Functional Forms. There is again a difference in the mean R_0 distribution, with the step Functional Form generally producing lower values and the exponential Functional Form generally producing higher values. However, this difference in mean R_0 distribution is not as distinct as it is in Adelaide and Brisbane. There is also an unusual distribution of the observation probability parameter ('P obs') in the exponential Functional Form, with two distinct peaks.

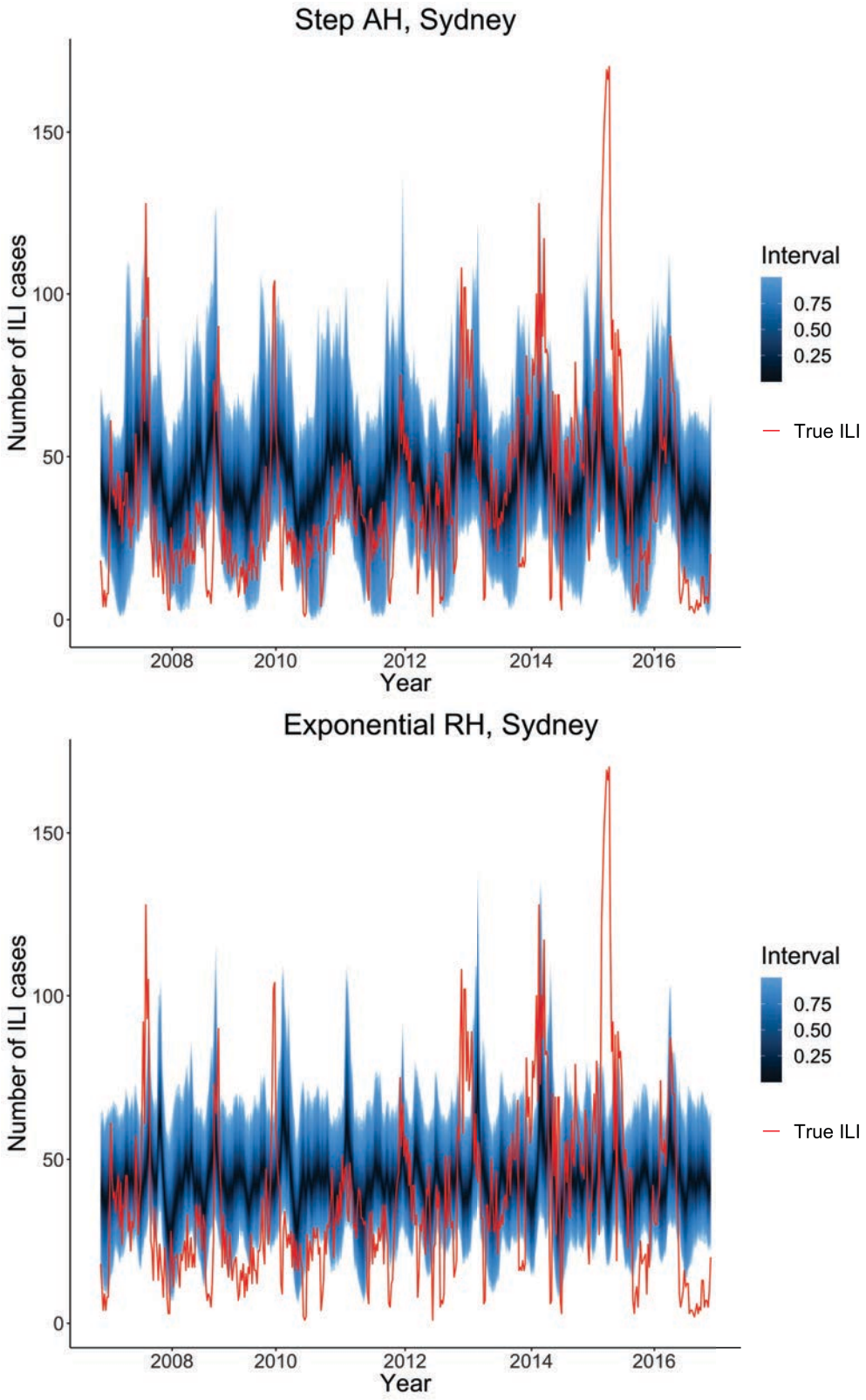

Figure 3.14: Fan charts showing the 1,000 simulated ILI datasets accepted by the ABC process over the years 2006–2016 for the top two Functional Forms as selected by the RF process in Sydney, with the interval showing quartile ranges. The true ILI data is overlaid for comparison.

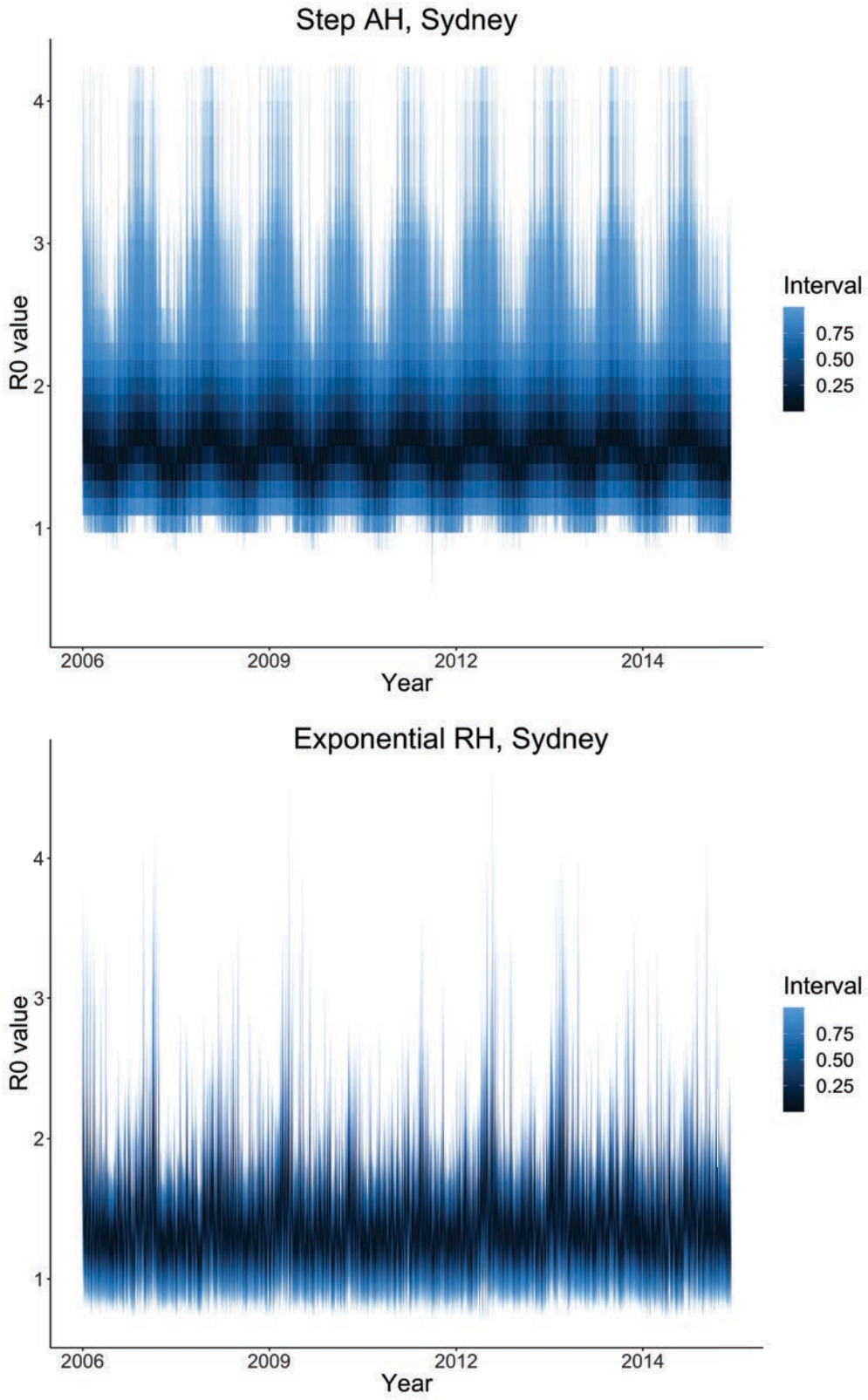

Figure 3.15: Fan charts showing the realisations of R_0 that generated the accepted simulations for the two selected Functional Forms in Sydney, with the interval showing quartile ranges.

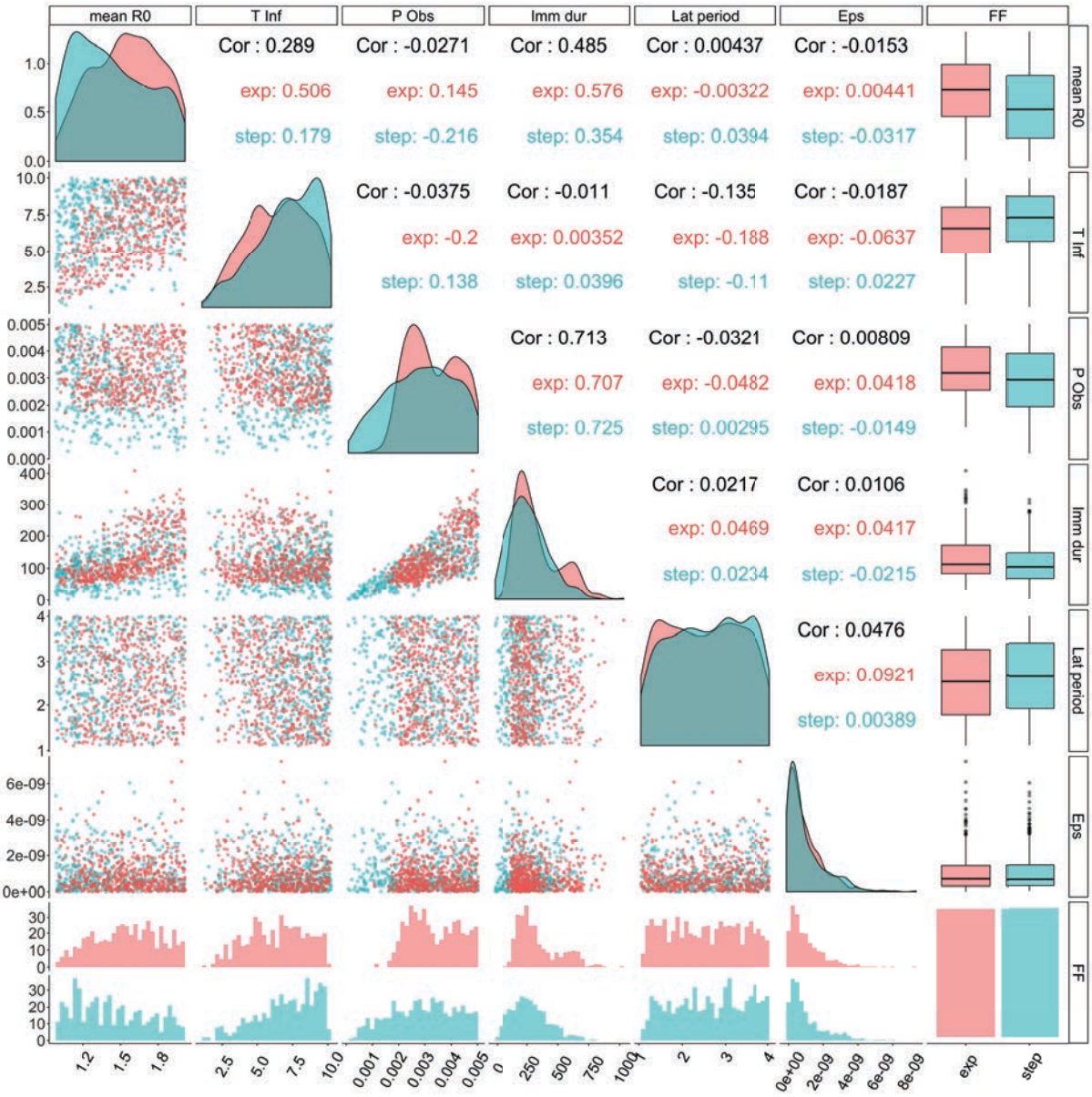

Figure 3.16: A pair plot comparing the posterior model parameters for accepted simulations from both Functional Forms in Sydney. The parameter 'mean R0' is the mean value of R_0 over all seasons, $\bar{R}_0$; 'T inf' is the time infectious; 'P obs' is the probability of observation by an ASPREN doctor; 'Imm dur' is the duration of immunity after recovery from ILI; 'Lat period' is the latent period between exposure and infectiousness; 'Eps' is the number of external ILI cases introduced into the population; and 'FF' is the Functional Form, where 'exp' is exponential using relative humidity and 'step' is step using absolute humidity. The bottom right plot shows an equal number of particles from each Functional Form were used.

3.5 Discussion

3.5.1 Summary

The exact reason for seasonal ILI outbreaks is not fully understood, with experimental models often showing conflicting evidence as to which climate variables may be driving ILI – suggesting a complicated relationship between ILI virus transmission and climate variables [21, 26, 35, 79]. In this chapter, we used a novel, high-quality ILI dataset in four locations around Australia coupled with a modern Bayesian model selection technique to gain further insight into the climate drivers behind ILI, and found absolute humidity to be the most important factor in the development and prediction of winter-time ILI outbreaks.

We developed four Functional Forms of transmissibility, based on the basic reproduction number R_0 as a function of climate. These Functional Forms are dependent on three different climate variables – temperature, absolute humidity, and relative humidity. We then used the Functional Forms of R_0 within a stochastic SEIR-type epidemic model framework, where three new compartments were added to allow us to simulate an hierarchical observation process where not all infected individuals will be observed. These Functional Forms were then used within a Bayesian model selection framework, utilising random forests (RFs) to decide between the models in each location.

The results of this process showed that the exponential Functional Form with absolute humidity best describes ILI in Mediterranean regions (Adelaide, Perth), while the step Functional Form with absolute humidity best fits observed ILI patterns in subtropical areas (Brisbane, Sydney). This suggests a complicated relationship between ILI and climate, where absolute humidity is an important but not sole influencing factor in the behaviour of seasonal ILI in Australia. We found that there is limited difference between the all-in and tournament RF model selection methods, suggesting that it does not matter which method is used. However, the all-in process reduced the computational time by approximately half.

In the Mediterranean regions where the exponential Functional Form was selected as the best fit, the step Functional Form was selected as the second best fit; and vice versa in subtropical regions where the step Functional Form was selected as the best fit.

By considering the out-of-bag (OOB) and misclassification error for each RF round,

we saw that the RF process was able to reliably identify the underlying Functional Form given simulations from multiple Functional Forms. Each round of the tournament RF process had an overall OOB error of less than 10%, while the all-in rounds all had OOB errors less than 13%. However, the selection between climate variables within Functional Forms had varying levels of error, with the step Functional Form being the most similar between climate variables and thus the hardest to correctly classify.

We then used ABC to fit simulations from the top two Functional Forms chosen by the RF process in each location. We determined the WRMSE score of each of the 10,000 training simulations from each Functional Form and each location and determined the best 5% score threshold. Using this, we generated simulations from the two chosen Functional Forms until we had a set of 1,000 simulations from each Functional Form with WRMSE score below that threshold.

The results of this showed that in Adelaide, Brisbane and Perth, the simulations generated using ABC generally fit the true ILI data very well, but failed to accurately model unusually large ILI peaks. In Sydney, neither Functional Form modelled the true ILI data as accurately as in the other cities. However, when considering the realisations of R_0 that generated the simulations, we saw in all cities that the Functional Form selected by the random forest had R_0 realisations that generally showed the most distinct seasonality similar in shape to the true ILI data.

3.5.2 Discussion of findings

Our results are in line with other experimental studies. Shaman *et al.* [79] found that absolute humidity was the strongest driving factor in seasonal influenza, and the exponential Functional Form, which was selected in subtropical climates, was first proposed by these authors. However, other research indicates that relative humidity or temperature also strongly influence ILI patterns [22]. Further, there is limited literature regarding the step Functional Form with which to compare results, suggesting that more research in this area is needed. In particular, further experiments to understand the effect of humidity and temperature on influenza and ILI transmission are needed to fully understand the cause of ILI seasonality and how it changes by climate and location.

Within the RF model selection method, we chose prior distributions for each of the parameters involved. These prior distributions were chosen in line with physical values

where possible. In cases where the parameters did not represent physical values, the prior distributions were chosen to be very wide, and uniform, to allow a wide range of values. However, different prior distributions could be chosen and may have lead to a different outcome in the RF model selection method. For example, the range of values of R_0 for ILI, in the absence of seasonality, is debated in current literature and so other ranges are possible [35].

We used the results of a RF to select the Functional Form with the highest proportion of votes. However, this means that some information is discarded by not considering the proportion of votes towards other Functional Forms. In Chapter 4, we will explore ensemble methods to maximise the information from the RF that is used in our analysis. As well as this, before simulations are used to train a RF, they undergo a validation process. A parameter set is accepted if the first five simulations from a parameter set show a persisting epidemic across all 10 years, with peaks occurring between May and September (the winter months in Australia) each year. This effectively conditions on the parameter sets producing 'realistic' epidemics. There are other conditions that could be placed on the simulations to ensure that the parameter sets are more consistent with what we see in the data; however, over-conditioning can lead to over-fitting during the ABC stage and so we chose to limit the conditions to only those two.

Interpreting the results from RF model selection method is challenging. There is limited information about the decisions and tested variables leading to the development of each individual tree in the RF, leading to difficulties in understanding why one model was selected over another in each location. To aid in interpretation of the RF method results, we considered the WRMSE scores and ABC fitting of the best two Functional Forms in each location.

In terms of understanding why the RF method selected the particular Functional Forms, we consider the shape of the R_0 realisations that generated the simulations accepted by the ABC method. In doing this, there is a noticeable pattern where the Functional Form with R_0 realisations that have the strongest seasonal shape are the Functional Forms selected by the RF process. This may contribute to why those Functional Forms were selected.

Further, we can see that in general the mean R_0 values in the simulations accepted by the ABC method were higher for the exponential Functional Form and lower for the step

Functional Form. The exponential Functional Form was selected by the RF method in the two Mediterranean locations (Adelaide and Perth), while the step Functional Form was selected by the RF method in subtropical locations (Brisbane and Sydney). This suggests that there may be some underlying relationship between the general climate of an area and the transmissibility of ILI in the area.

Figures B.5 – B.7 in Appendix B show a comparison of temperature, absolute humidity and relative humidity between the four locations over the years 2011–2013. Whilst the climate variables follow a mostly similar pattern between locations, there are some differences. In particular, the absolute humidity seen in Brisbane and Sydney is more varied than that seen in Adelaide and Perth. This may help to explain the different Functional Forms selected in Adelaide and Perth (exponential Functional Form) compared to Brisbane and Sydney (step Functional Form).

3.5.3 Assumptions and limitations

In future, it would be advantageous to perform this ABC fitting for all nine of the Functional Forms in each city. It would also be more informative to use a greater number of simulations within the ABC framework. However, due to computational and time constraints this was outside the scope of the thesis.

Whilst the RF method appears to accurately select between Functional Forms and climate variables, the choice of prior distributions for model parameters play a significant role in the model selection process. It is therefore important that the prior distributions are selected to have physical meaning, to ensure that the model selection process is accurate to true ILI transmission. The assumptions we place on the prior distributions are a limitation on the model, as inappropriate prior distributions may affect the accuracy of results. As well as this, the ASPREN data provides only a subset of actual ILI cases, and so this provides only an estimate of the number of ILI cases in the real population. This means that the accuracy of the RF model selection process depends on the accuracy of the data. In some areas of Australia, such as Darwin or Hobart, there is very limited ASPREN notification data and so it is not feasible to perform this method to analyse the behaviour of ILI. As such, the results could be improved by finding a more complete source of data.

The primary limitations in this model is the fact that we consider ILI as one illness

with a single value of $\bar{R}_0$; in reality, ILI is a mixture of a large number of illnesses with each having distinct characteristics. However, it is difficult to determine between the different ILI-contributing illnesses in a clinical setting, and so ILI is used due to the lack of data on each individual virus. We also do not consider the evolution of viral strains between years, which can cause a significant difference in the value of $\bar{R}_0$ each year [31].

Other limitations in the SEIR-type model that we use include that we do not consider population dynamics, we assume that a fixed proportion of the population is susceptible, and we utilise a fully-mixed model. We also do not use consider an age structure, which may be significant due to the differences in immunity and susceptibility between different age groups. Further, we assume a static population size in each city and consider each city independently, while in reality there is significant travel and migration between cities.

A limitation of the RF model selection process is that it can be a 'black box' – in other words, we see only the input and the output of the RF, but we are not able to observe the variables and decisions used to generate each tree [81]. This means that it can be difficult to analyse why the particular model is selected as best, and can mean that issues and undesirable effects may go unnoticed. However, utilising ABC with the same model and prior distributions and comparing the results allows us to further understand whether the RF model is effective at selecting the 'best' Functional Form. Analysing variable importance in the RF may also help to improve understanding and interpretation of the model selection results. As well as this, the true relationship between climate and ILI is likely to be far more complicated than any of our individual Functional Forms, meaning that the RF model selection process is being performed without a true model as a candidate.

Finally, a limitation within the ABC model fitting is the visualisation of the results. In Section 3.4.2, fan charts are shown to give an idea of the average behaviour of the accepted ILI datasets and realisations of R_0. However, as we see in the plots in Appendix B (Figures B.1 – B.4), the individual realisations of R_0 are very diverse. This means that it is challenging to accurately interpret and visualise the ABC results.

3.5.4 Future work

Possible further extensions to this model include adding an age structure, a household structure, or including human movement within the population. This would allow us

to better understand how ILI spreads in human populations that are not fully mixed. Adding in other possible factors such as school holiday dates or antigenic drift may also refine results and increase understanding of the exact drivers of ILI. However, further experimental models play a significant part of this understanding, as they allow us to study how the viral capsule and resulting transmission is affected by differing temperature and humidity which can then be applied to population-level studies.

The goal of understanding the impact of climate variables on annual ILI epidemics is to be able to forecast the start and peak of the ILI season, allowing healthcare workers and the public health system to effectively manage resources. In the following chapter, we explore using these results within an ensemble framework to forecast ILI, and further understand which climate variables may be affecting the spread of seasonal ILI.

Chapter 4

Ensemble forecasting

In this chapter, we consider the problem of forecasting the peak timing of ILI incidence in Australia using between one and seven years of historical data. We explore the use of ensemble forecasting with the nine Functional Forms introduced in Chapter 3 to predict the peak week of ILI in 2014 in each of the four Australian cities: Adelaide, Brisbane, Perth and Sydney. We compare two different score metrics to select top posterior forecasts from each of the nine Functional Forms, and four different techniques for calculating ensemble weights. Finally, we compare the forecast skill of the resulting ensemble forecasts to determine whether one method is most effective at forecasting the peak timing of ILI.

4.1 Introduction

The effective management of seasonal influenza and ILI requires careful resource allocation by health care services, and to effectively manage resources it is important to know when the ILI season is likely to reach a peak [82]. However, the problem of forecasting ILI is challenging due to complex interactions between the large number of factors that affect disease transmissibility. As well as this, the range of diseases comprising ILI, the seasonal and geographic variation in transmissibility, and the irregularity of seasonal ILI epidemic timing make forecasting ILI a formidable undertaking [19, 39].

Due to the complexity of seasonal ILI, it is unlikely that there exists any one model to fully explain the behaviour and characteristics of ILI across multiple seasons. Instead, some models will accurately capture distinct features and fail to capture others. By combining multiple models via an ensemble, we may increase the likelihood of accurately

representing multiple features of ILI behaviour [83, 84].

In this chapter, we present the framework for ILI forecasting using ensemble forecasting techniques along with approximate Bayesian computation and machine learning model selection. This work builds on the research presented in Chapter 3, with a focus on forecasting future ILI seasons using historical ILI data. We present an example of this prototype forecasting technique using 10,000 simulated ILI datasets from candidate models, and forecast the full 2014 ILI season from the end of 2013. Performing ILI forecasts 6 months in advance is of course impractical. We apply our proposed method in this context for consistency with the previous chapter, and for simplicity of implementation. We expect the method to be able to capture general features of seasonal ILI such as the overall shape of the season, but should not expect highly accurate forecasts. Nonetheless, as we will show, the method performs well as a proof-of-concept.

There are many aspects of seasonal ILI epidemics that can affect the allocation of health care resources in public health systems, including the number of people infected, the peak of the epidemic, and the severity of particular strains of ILI. Here, we focus on forecasting the peak week of the epidemic – the week with the greatest number of individuals suffering from ILI symptoms. We use a variety of ensemble methods to forecast the timing of the 2014 peak of seasonal ILI, in Adelaide, Brisbane, Perth and Sydney. We compare two different score metrics for constructing component forecasts, and four techniques for generating the model weights.

4.2 Method summary

Recall that in Chapter 3 we introduced nine Functional Forms of transmissibility,

$$\text{Linear: } R_0(t) = \bar{R}_0(1 - \delta_i x_i(t) - \delta_3 x_3(t)) \qquad i = 1, 2, \tag{4.1}$$

$$\text{Exponential: } R_0(t) = R_0^{\min} + (R_0^{\max} - R_0^{\min})e^{a_i x_i(t)} \qquad i = 1, 2, 3, \tag{4.2}$$

$$\text{Step: } R_0(t) = \mathbb{1}_{\{x_i(t) \geq s\}} R_0^{\text{base}} + \mathbb{1}_{\{x_i(t) < s\}} R_0^{\text{elev}} \qquad i = 1, 2, 3, \tag{4.3}$$

$$\text{Sinusoidal: } R_0(t) = -A\cos((\tfrac{2\pi}{730})(t + v)) + \bar{R}_0, \tag{4.4}$$

where i indexes the climate variables (absolute humidity ($i = 1$), relative humidity ($i = 2$), and temperature ($i = 3$)), and $x_i(t)$ represents the climate measurement of variable i at

time t, each scaled to the interval $[-1, 1]$. Section 3.3.1 describes the Functional Forms in detail.

These Functional Forms each approximate disease transmission in a different way, and are used in an SEIR-type stochastic compartmental model described in Section 3.3.2. In Chapter 3 we used a random forest model selection process to determine which Functional Form best represented ILI transmission in the four different locations around Australia. We found that no single Functional Form was best everywhere; in Adelaide and Perth the exponential Functional Form using absolute humidity was the selected model, whilst ILI in Brisbane and Sydney was best described by the step Functional Form, also using absolute humidity. This difference suggests that while some Functional Forms are good, none of them are an ideal representation of ILI transmission.

To improve the approximation of ILI in a forecasting framework, we consider *ensembles* of forecasts from these Functional Forms. Ensemble modelling works by combining forecasts from multiple different models to improve predictions. Since there is no single correct model for describing ILI, it is often advantageous to combine forecasts, because different models are potentially predictive in different ways. Ensembles of forecasts also allows for uncertainty in model choice [85], and as such they are often used in meteorological applications for forecasting weather. In this chapter, we develop ensembles using a weighted sum of the nine Functional Form forecasts, comparing four different methods of selecting the ensemble weights.

We choose to test our approach to forecasting the peak week of ILI using between 1 and 7 years of training data, to understand how forecast skill may be affected by the amount of available historical ILI data. With this goal, we perform our forecast with multiple sets of training years: 2013 alone, 2012–2013, 2011–2013, 2010–2013, 2009–2013, 2008–2013, and 2007–2013, all used to predict the peak week of ILI in 2014. We choose to predict the peak week in 2014 as this is the latest year for which we have full ILI notification data in all cities, allowing us to use the maximum possible number of training years.

For each Functional Form and range of training data, we construct a posterior forecast using each of the ABC score metrics described in Sections 4.3.1 and 4.3.2. In practise, this forecast consisted of 100 simulations. We call this set of 100 simulations the 'forecast set'. From the simulations, we can then construct a probabilistic forecast of peak week using Kernel Density Estimates (KDEs) as described in Section 2.5, to estimate the

distribution of peak weeks from the individual simulations. Then we use four different methods, described in Section 4.5, to determine the model weightings used to form the ensembles. The ensemble forecast is then the weighted sum of the forecasts from each of the nine Functional Forms, where the model weightings sum to 1.

This process is repeated for each set of training years, for each score metric, and each method of selecting model weightings. By generating a forecast for the peak week in 2014 from each of these methods, we can compare the skill and determine which method is most appropriate for this application.

This ensemble forecasting method is summarised in Algorithm 3, where we utilise $m = 9$ Functional Forms, $1 \leq t \leq 7$ training years, and select our threshold ϵ to be the 99^{th} percentile of simulation scores. We always test our predictions on the year 2014 to maintain consistency between cities.

Generate N simulations from m models, over t training years;
for *s = 1 to t* **do**
  calculate score $\mathcal{D}$ for each simulation of each model;
end
Sum the scores for each simulation over the t training years;
Select the simulations with the lowest scores up to a chosen threshold ϵ to form set of forecasting simulations S^m;
Calculate the Kernel Density Estimate of peak week for the following year for each model m using the set of simulations S^m;
Determine the ensemble model weightings;
Calculate the weighted sum of m Kernel Density Estimates using these model weightings, to form the ensemble forecast of the peak week of the following year;
Use this ensemble KDE to predict the peak week of the following year.

Algorithm 3: KDE-based ensemble forecasting method.

4.3 Training methods

To forecast future ILI epidemics, we use an ABC-style forecasting framework, producing a forecast consisting of posterior simulations, from each of 9 different Functional Forms. From each Functional Form, we start with 10,000 simulated 10-year ILI epidemics from the prior distributions described in Section 3.3.4. These simulations are not fitted in any way and so may vary wildly from the true ILI observations. To obtain the set of posterior simulations, we choose those simulations that are most similar to the true ILI observations. We do this by calculating a score for each simulation, and accepting only simulations with a score better than the threshold ϵ over the set of training years.

Specifically, for each simulation, we calculate a score $\mathcal{D}_t$ for every training year t by two different methods, detailed below. Then simulations are accepted if the sum of these annual scores is below the threshold ϵ, i.e:

$$\sum_{t=1}^{N} \mathcal{D}_t \leq \epsilon,$$

where N is the number of years used for training. The resulting set of posterior simulations are then simulated forward to forecast the following year. In our case, we choose our threshold ϵ to keep the best-fitting 1% of simulations, or the 99^{th} percentile, with each Functional Form having a different threshold.

There are a number of possible score metrics that could be chosen. In this chapter, we consider two different score metrics - one training only on the peak week of historical ILI seasons (Section 4.3.1), and another which uses a weighted root mean squared error (WRMSE) to take into account all ILI observations in the training years (Section 4.3.2).

4.3.1 Peak week

The first score metric that we consider is dependent only on the timing of historical ILI peaks. We consider this score metric as it is the same metric that we are trying to predict through our forecasts. In this score metric, $\mathcal{D}_t$ is given by the difference between the week number of the peak week of ILI in the simulated data and in the true data. The weeks in a year are numbered from 1–52, where week 1 is the week starting on the 1^{st} Monday of the year. We use the function

$$\mathcal{D}_t = |i_{\text{true}} - i_{\text{candidate}}|,$$

where i_{true} is the number of the peak week of ILI in the true data for training year t, and $i_{\text{candidate}}$ is the number of the peak ILI week in the simulated data over the same year t. The absolute value of this is taken as we consider only the difference in peak week in the simulated ILI and the true ILI data, not whether the simulated ILI peak occurred before or after the true peak.

4.3.2 Weighted root mean squared error

Our second approach is to instead fit to the full season of ILI data. We choose the weighted root mean squared error (WRMSE), introduced in Section 2.2:

$$\mathcal{D}_t = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \sqrt{\text{true}_i + 1} \times (\text{true}_i - \text{candidate}_i)^2}.$$

We calculate the score $\mathcal{D}_t$ over each of the training years t individually, where $n = 52$ weeks, and for each week i, true_i is the number of infectious individuals recorded in the true ILI data, and candidate_i is the number of infectious individuals in the simulated data in that week.

Using this WRMSE score metric has the advantage that the score can be updated each week; this allows the forecast to take into account new observations during the season to improve the forecast as the year progresses. By using only peak week, it is not possible to update progressively as the entire year of observations is needed to determine the peak week timing. However, in this chapter we focus solely on producing peak week forecasts at the beginning of the year.

4.4 Developing forecasts and assessing skill

The ensemble model is a weighted sum of the probabilistic forecasts from the nine Functional Forms, calculated under the chosen training method, and with model weights determined using the chosen weighting method. We determine the probabilistic forecast for the 2014 peak week using each Functional Form individually and calculate the weighted sum of those forecasts. This process is illustrated in Figure 4.1, showing an example

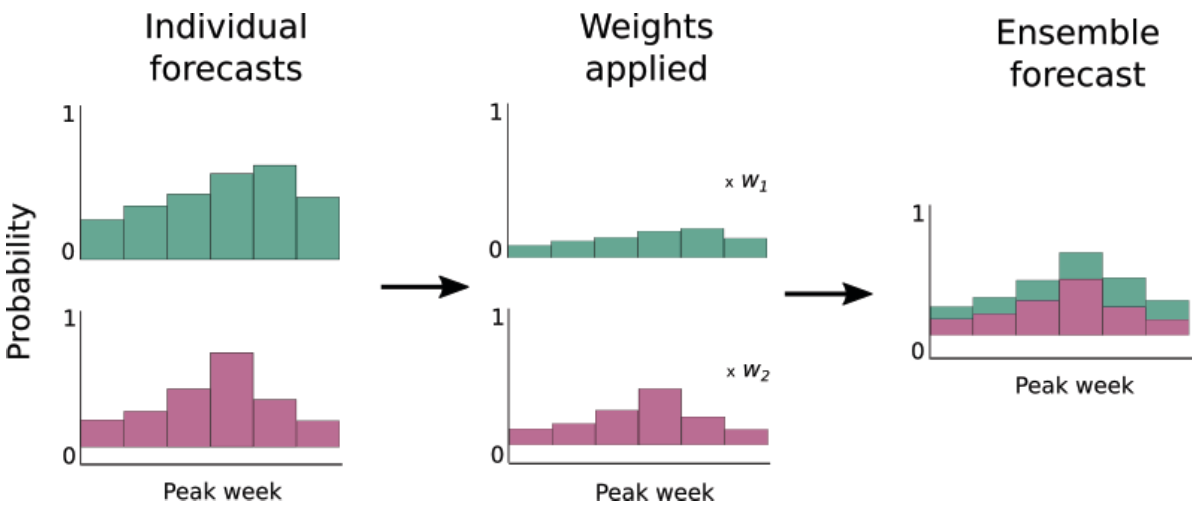

Figure 4.1: A conceptual diagram showing the process of creating an ensemble forecast from individual model forecasts.

of two individual model forecasts being combined into a single probabilistic ensemble forecast with weights w_1 and w_2.

To produce an estimate of the distribution of the predictions of the 2014 peak ILI week from each Functional Form, we use Kernel Density Estimation (KDEs), as described in Section 2.5. Each Functional Form has a forecasting set that has been determined using the training data. From these forecasting sets, we determine the peak week of ILI in 2014 in each of the 100 simulations, and construct a KDE. We can then directly calculate the weighted sum of these forecasts, to produce a single probabilistic forecast that is the weighted ensemble of the forecasts of the nine Functional Forms.

To assess the performance of each ensemble technique, we consider the *forecast skill* [86]. Forecast skill is a common measure of the accuracy of a forecast. In this case, we consider the forecast skill to be the probability that the ensemble method predicts that the 2014 peak ILI week correctly. For example, in Figure 4.2, if the true peak week was week 36, then the model would have a forecast skill of 0.11. It is also possible to reduce the rigidity of the forecast skill assessment by considering peak weeks within ± 1 week of true peak week to be 'correct' [33].

4.5 Ensemble weighting methods

To develop weighted ensemble forecasts, we need to calculate the weighting given to each model. There are many possible methods of calculating model weights. Here, we

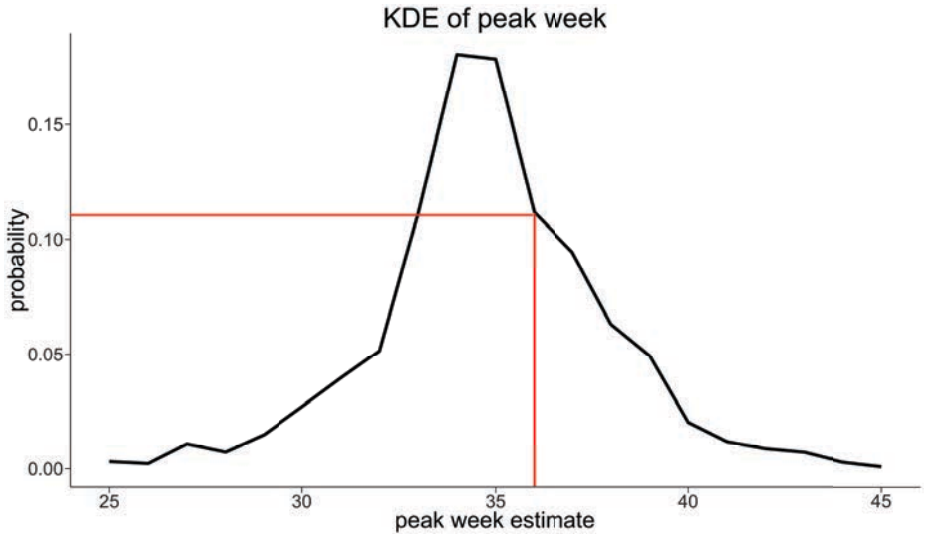

Figure 4.2: An example probabilistic forecast determined using a Kernel Density Estimate (KDE) for the peak week of ILI in 2014, showing the probability of the peak week occurring in each week between weeks 25 and 45. The red lines show the calculated skill, assuming that the true peak week occurred in week 36.

explore three different methods of calculating the model weights within our ensemble as well as considering a single Functional Form in each location as selected in Chapter 3. Specifically, we select the Functional Form that was voted as best by the random forest in each location: exponential (Equation 4.2) using absolute humidity in Adelaide and Perth, and step (Equation 4.3) using absolute humidity in Brisbane and Sydney. We do this to assess whether the forecasting is improved by using ensemble techniques rather than a single model. We calculate forecast skill (described in Section 4.4) for each ensemble as well as the best Functional Form by location.

4.5.1 Equal weighting

To act as a baseline against which to compare other ensemble weighting techniques, we consider an ensemble with equal weightings between models. This allows us to determine if the ensemble weighting techniques are improving the forecast skill.

4.5.2 Optimised weighting

Another method to determine the weights of the ensemble is to optimise model weightings over the training years. This maximises forecast skill of the ensemble model over the

historical ILI training years, with the aim of producing the best forecast. This method is equivalent to calculating the maximum likelihood over the ensemble weights.

This method uses the probabilistic forecast for each Functional Form, determined using one of the two training methods described in Section 4.3. We then consider each of the training years individually, using this forecast. First, we determine the true peak week of ILI in each of the training years. Then, we calculate $p_j^{(t)}$ as the proportion of simulations in the forecasting set for model j and year t that peak in exactly the correct year. This means that $p_j^{(t)}$ is the proportion of the top 1% of simulations from model j that hit exactly the correct peak week in year t.

We aim to determine weightings on the Functional Forms that maximises the forecast skill of the ensemble over the training data. We define $\mathbf{w}$ as the vector of weightings on the Functional Forms, and maximise the forecast skill over the training years with the objective equation,

$$\max_{\mathbf{w}} \sum_t \ln\left(\sum_j w_j p_j^{(t)}\right),$$

subject to

$$0 \leq w_j \leq 1$$

and

$$\sum_j w_j = 1,$$

where t is the year and j is the model, and w_j is the weighting given to model j. The sum $\sum_j w_j p_j^{(t)}$ is the ensemble probability of the correct peak week. The initial weightings $\mathbf{w}$ are taken to be equal across all Functional Forms.

We evaluate the solution to this objective equation using the R function `optim`, from the R package `stats` [87], to produce the vector of optimal weights $\mathbf{w}$, for each combination of ILI training years and individual forecast fitting method.

4.5.3 Random forest weighting

We propose a method to choose ensemble weights based on the output of a RF. We utilise a RF framework based on Chapter 3 of this thesis, where random forests are used to select

between the 9 Functional Forms given in Equations (4.1)–(4.4). Using the RF framework provides an alternative method to the traditional optimisation, and may potentially be able to provide accurate solutions using less data than other methods [88].

Using this method, a random forest is trained on the 10,000 simulated ILI datasets from each of the Functional Forms across the training years. The random forests produce an output where a proportion of decision tree votes are assigned to each of the Functional Forms, so that the nine proportions sum to 1. We can therefore use these vote proportions directly as model weightings.

4.6 Results

Figures 4.3–4.10 show the results of the ensemble forecasting of peak week using the different methods of training and determining weights in each city. Figures 4.3, 4.5, 4.7 and 4.9 show results trained using peak week, while Figures 4.4, 4.6, 4.8 and 4.10 show WRMSE. In each figure, Plot (A) shows the training data, Plot (B) shows the forecast from the optimised and random forest ensembles, Plot (C) shows the forecast of each individual Functional Form, and Plot (D) shows the evolution of forecast skill over time for each weighting method

Adelaide

The results in Adelaide are given in Figures 4.3 and 4.4. From the Plot (D) in each of these Figures, we can see that the random forest method consistently produces the highest skill when compared to the optimisation method. However, the skill of the exponential Functional Form using absolute humidity is slightly higher than the random forest, suggesting that in Adelaide (in 2014), an ensemble of Functional Forms does not necessarily improve forecast skill over having only the Functional Form voted best by the random forest; this is likely because the best Functional Form produces a very accurate forecast in this case, as seen in both figures' Plot (C). We can see from the Plots (B) in both of these figures that the forecast generally improves once three or more years of training data are added, with the plots becoming more tightly centred around the true peak week, for both weighting methods. Plots (A) show that the 2014 ILI season appears to peak at a relatively similar time to the previous years of ILI, which contributes to the forecast skill improving as more training years are added.

Considering Plot (D) in Figure 4.3, we can see that initially adding more training years improves the forecast skill when using the peak week training method. However, when more than three training years are considered the forecast does not improve further. For the random forest and the exponential Functional Form using absolute humidity, forecast skill decreases slightly as more training years are added. The optimisation forecast method shows a very inconsistent forecast skill, improving with some added years and worsening with others. Plot (D) in Figure 4.4 shows no major changes in forecast skill when more training years are added using the WRMSE training method.

Plots (C) show the peak week prediction of each of the individual Functional Forms. The linear Functional Form using relative humidity shows a distinct spikiness; this is due to a small number of weeks being forecast as peak week with a high level of certainty across the 100 simulations. We can see in both Figures that the exponential Functional Form using absolute humidity has the highest value at the correct peak week, further reinforcing this Functional Form as being the most appropriate for Adelaide. When comparing between the peak week and WRMSE methods in Plots (D), we can see that the peak week method (Figure 4.3) results in higher skill scores than the WRMSE method (Figure 4.4). This is intuitive, as we are training only on the feature that we wish to predict.

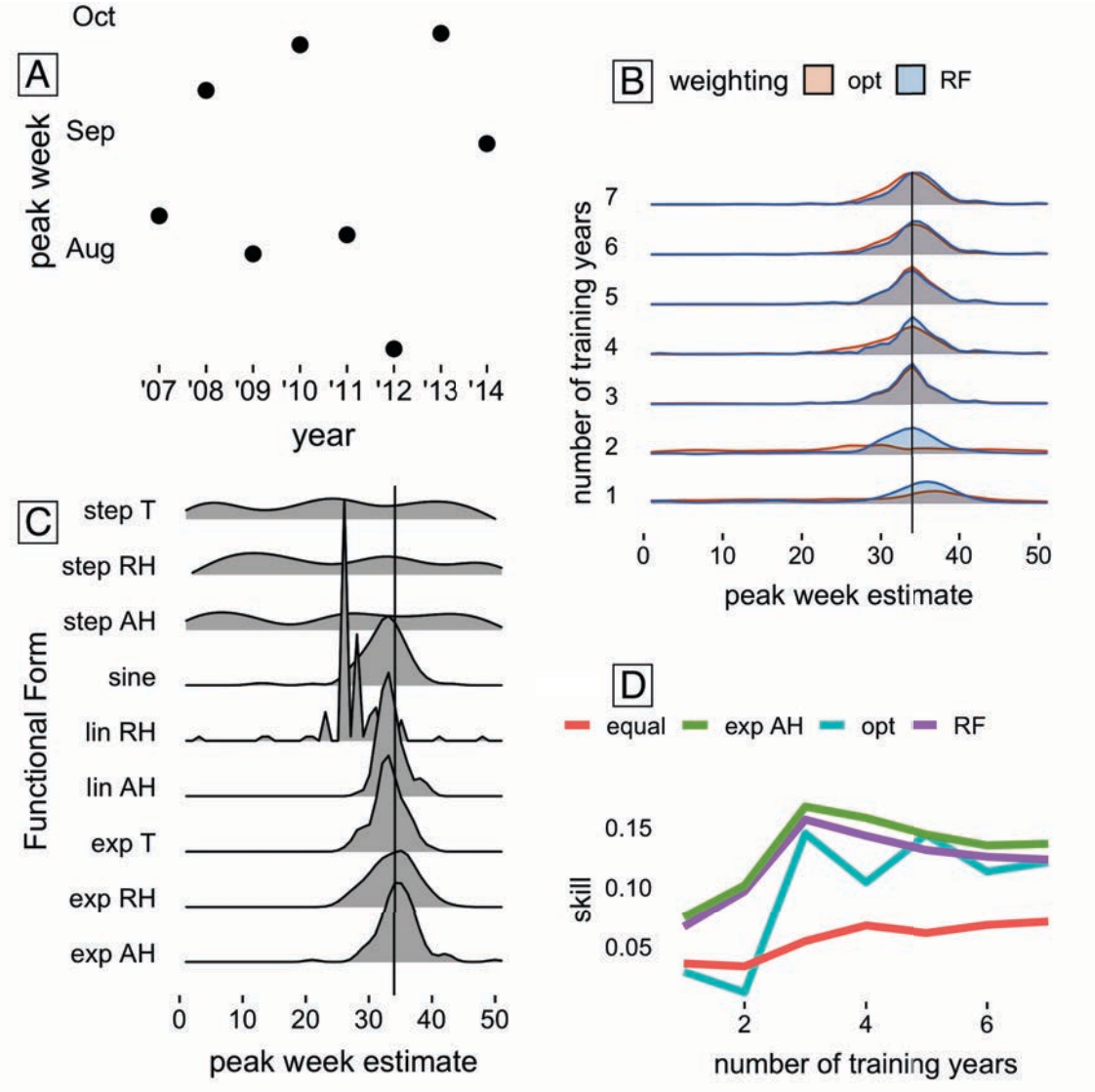

Figure 4.3: The probabilistic forecast results in Adelaide, where the forecasting set is trained using the peak week method. Plot (A) shows the peak week of each year from 2007–2014. Plot (B) compares the forecast of ensembles developed using the optimisation and random forest methods for model weighting, and shows how these results evolve as more years of training data are included. Plot (C) compares the forecast of each individual Functional Form, with seven years of training data. Finally, Plot (D) compares the forecast skill of the four different weighting methods: optimisation, random forests, equal weighting and exponential Functional Form using absolute humidity. The vertical line in Plots (B) and (D) represents the true ILI peak week in 2014.

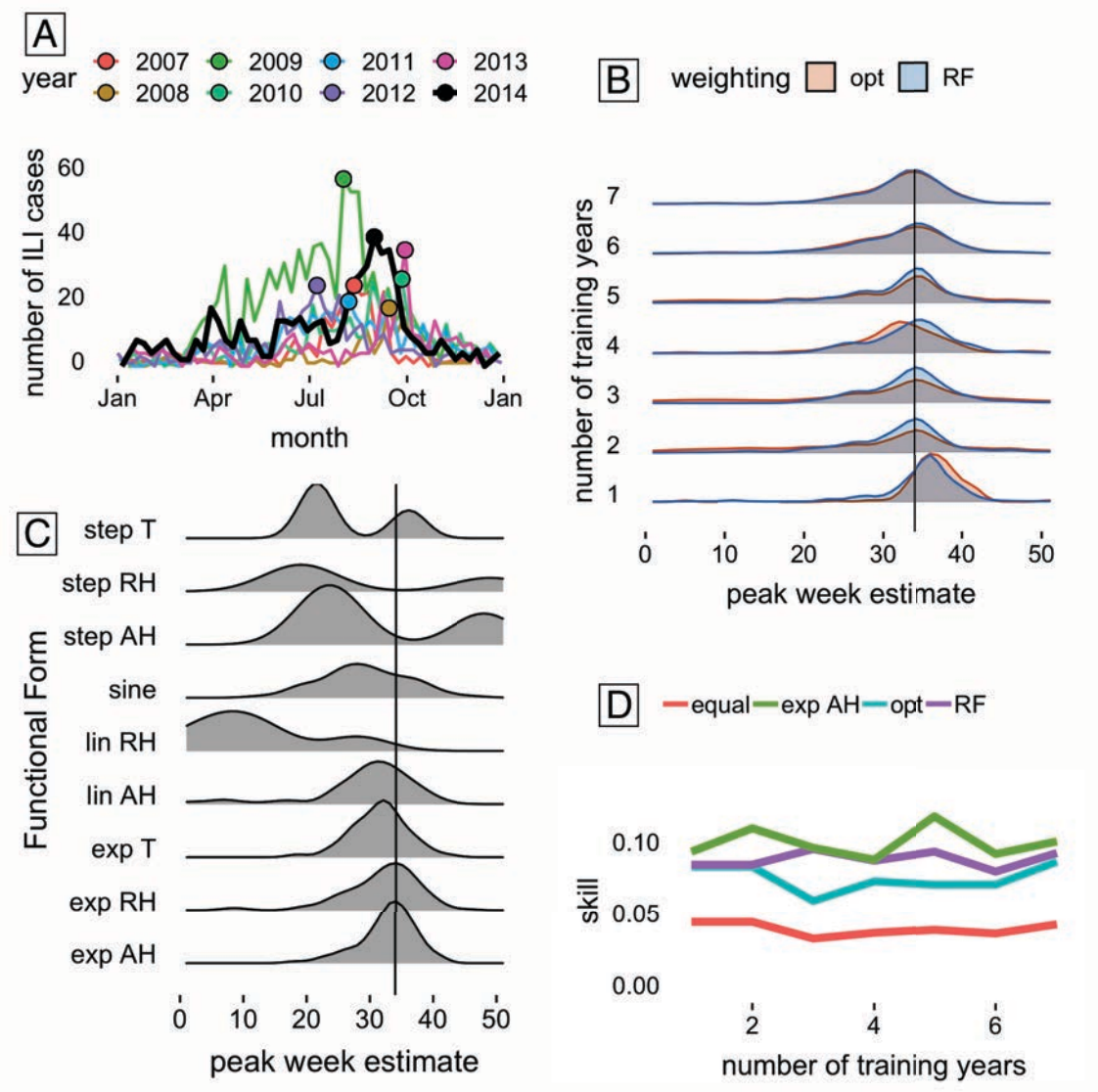

Figure 4.4: The probabilistic forecast results in Adelaide, where the forecasting set is trained using the WRMSE method. Plot (A) shows the true ILI observations for each year from 2007–2014. Plot (B) compares the forecast of ensembles developed using the optimisation and random forest methods for model weighting, and shows how these results evolve as more years of training data are included. Plot (C) compares the forecast of each individual Functional Form, with seven years of training data. Finally, Plot (D) compares the forecast skill of the four different weighting methods: optimisation, random forests, equal weighting and exponential Functional Form using absolute humidity. The vertical line in Plots (B) and (D) represents the true ILI peak week in 2014.

Brisbane

Figures 4.5 and 4.6 present the results from the ensemble forecasting in Brisbane. Unlike in Adelaide, in Brisbane Plots (D) show that, with both training methods, the optimisation weighting tends to perform better than the random forest weighting. The general trends in skill are the same for both training methods: optimisation weighting is the best, followed by random forest weighting, then equal weights, then the step Functional Form.

Plot (D) in Figures 4.5 and 4.6 show quite similar results. The forecast skill for the random forest method, equal weights, and the step Functional Form using absolute humidity generally stays fairly constant regardless of how many training years are added. The optimisation method, however, at first increases forecast skill as more training years are added. When more than three training years are added, the forecast skill of the optimisation method decreases. This is likely due to different combinations of Functional Forms being selected by the optimisation method. In particular, Plot (D) in Figure 4.6 shows a very distinct drop in forecast skill when the fourth training year is added. This inconsistency in forecast skill over the training years is likely because the peak week of 2014 is unusually early, as seen in Plot (A) of both figures.

The random forest weighting technique, combined with the peak week training method, generally outperforms both the individual Functional Form and the equal weighting. The step Functional Form is outperformed by all of the other methods, as shown in Plot (D) of both figures. In Plot (B) in both figures, we can see that the forecast does not generally improve as more training data is added, with the optimisation forecast predicting a later peak week and the RF forecast predicting an earlier peak week compared to the true ILI peak. In Plot (C) in both figures, the KDE of the step Functional Form does not have a single strong peak, for any climate variable. This means that the simulations that fit well to previous years of ILI do not predict the same peak week, suggesting that a different training method may be more effective or that the model is overfitting to the earlier training years. Finally, when comparing between the peak week and WRMSE training methods using Plot (D) from both figures, we can see that using the peak week method consistently produces a forecast with higher skill.

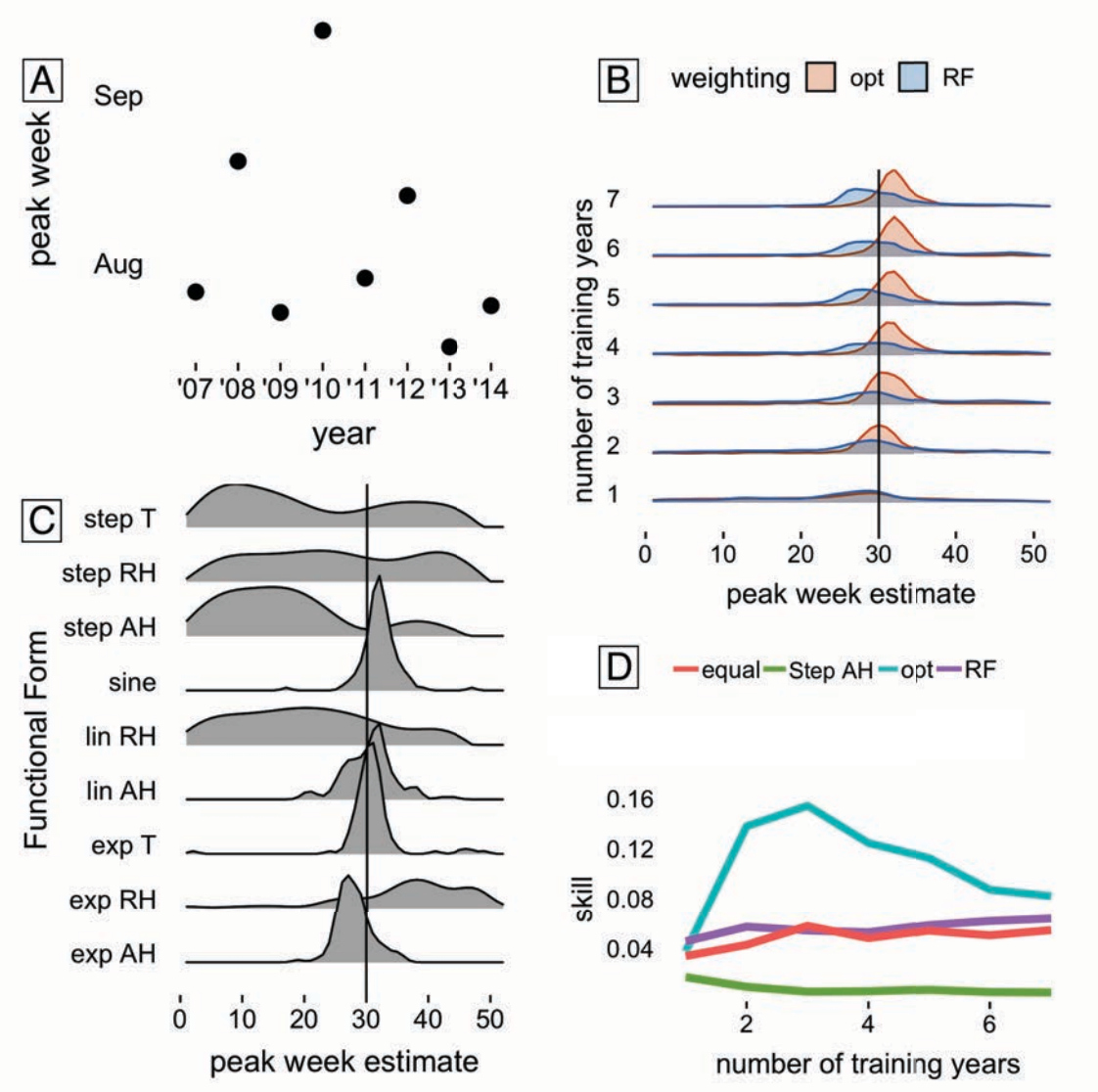

Figure 4.5: The probabilistic forecast results in Brisbane, where the forecasting set is trained using the peak week method. Plot (A) shows the peak week of each year from 2007–2014. Plot (B) compares the forecast of ensembles developed using the optimisation and random forest methods for model weighting, and shows how these results evolve as more years of training data are included. Plot (C) compares the forecast of each individual Functional Form, with seven years of training data. Finally, Plot (D) compares the forecast skill of the four different weighting methods: optimisation, random forests, equal weighting and step Functional Form using absolute humidity. The vertical line in Plots (B) and (D) represents the true ILI peak week in 2014.

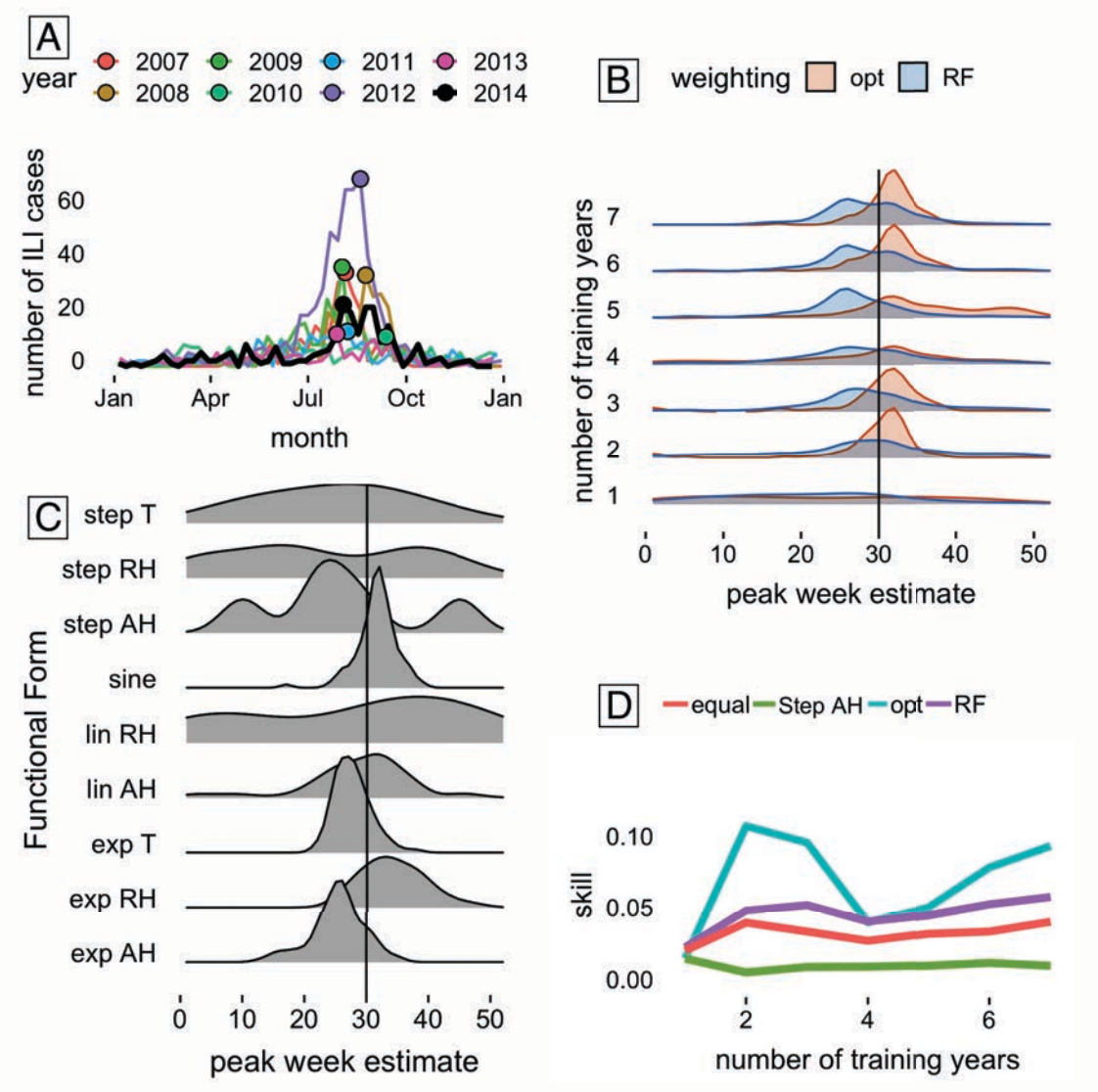

Figure 4.6: The probabilistic forecast results in Brisbane, where the forecasting set is trained using the WRMSE method. Plot (A) shows the true ILI observations for each year from 2007–2014. Plot (B) compares the forecast of ensembles developed using the optimisation and random forest methods for model weighting, and shows how these results evolve as more years of training data are included. Plot (C) compares the forecast of each individual Functional Form, with seven years of training data. Finally, Plot (D) compares the forecast skill of the four different weighting methods: optimisation, random forests, equal weighting and step Functional Form using absolute humidity. The vertical line in Plots (B) and (D) represents the true ILI peak week in 2014.

Perth

The results in Perth are given in Figures 4.7 and 4.8. From Plot (D) in Figure 4.7, we can see that the exponential Functional Form using absolute humidity has the highest forecast skill when using the peak week, closely followed by the optimised method. However, there is not a large difference between methods, and generally all methods had a low forecast skill regardless of the number of training years. When using the peak week method, the random forest method performs nearly as well as the optimisation method and the exponential Functional Form using absolute humidity; however, when using the WRMSE method it does not perform as well. The equal weighting method performed similarly to the other three weighting methods regardless of the training method used.

Plot (D) in Figures 4.7 and 4.8 shows the forecast skill decreasing as more training years are included. We can see in Figure 4.7 that in general, as more training years are added, the forecast skill decreases quite significantly. This is likely due to the unusually early timing of the peak week in 2014, as shown in Plot (A). When the second training year (2012) is added, the forecast improves slightly. Plot (A) shows that the 2012 peak week occurred at a similar time to 2014, but that the other years generally peaked much later. This helps to explain why the forecast skill drops significantly when the next training years are added. Plot (D) in Figure 4.8 shows a slightly different pattern, with the forecast skill changing a lot with the addition of more training years but with no visible long-term trends.

In Plot (B) in both figures, we can see that the forecast appears to improve slightly as more training years are added, but that it consistently predicts a later peak week than the true peak using the WRMSE method. This can be explained by looking at Plot (A) in Figure 4.7, which shows the peak week in 2014 occurring earlier than in most of the previous years. In Plot (C) in Figure 4.7, we can see that the exponential Functional Form using absolute humidity appears to perform better than the other Functional Forms when using the peak week training method. However, this is not true when using the WRMSE training method as seen in Plot (C) of Figure 4.8, where the exponential Functional Form using absolute humidity shows a distinct spikiness, due to certain weeks being predicted as peak week across the 100 simulations with a high level of certainty. Finally, when comparing the peak week and WRMSE training methods, the WRMSE method appears to outperform the peak week method, producing predictions with higher skill than the

peak week method.

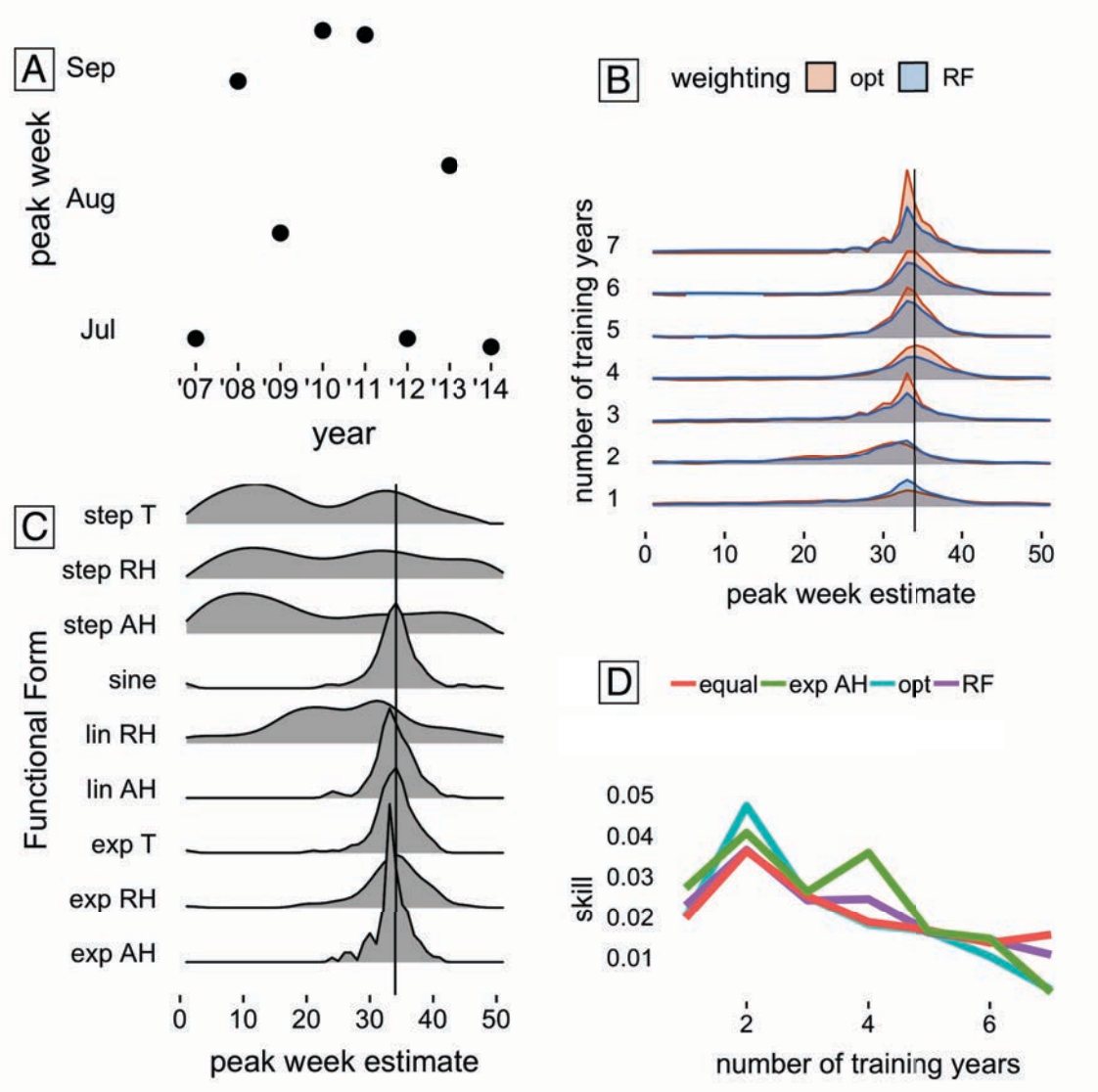

Figure 4.7: The probabilistic forecast results in Perth, where the forecasting set is trained using the peak week method. Plot (A) shows the peak week of each year from 2007–2014. Plot (B) compares the forecast of ensembles developed using the optimisation and random forest methods for model weighting, and shows how these results evolve as more years of training data are included. Plot (C) compares the forecast of each individual Functional Form, with seven years of training data. Finally, Plot (D) compares the forecast skill of the four different weighting methods: optimisation, random forests, equal weighting and exponential Functional Form using absolute humidity. The vertical line in Plots (B) and (D) represents the true ILI peak week in 2014.

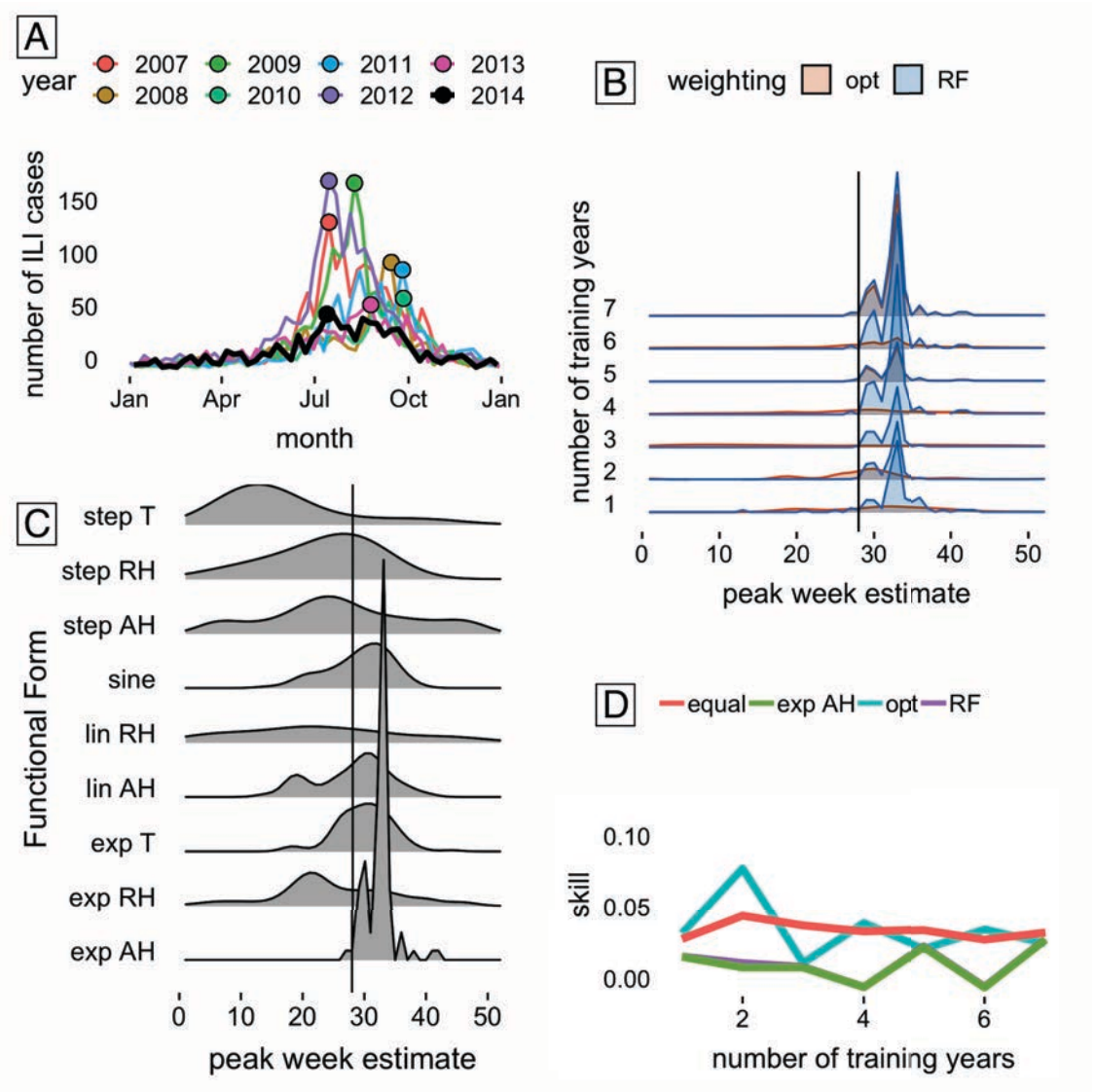

Figure 4.8: The probabilistic forecast results in Perth, where the forecasting set is trained using the WRMSE method. Plot (A) shows the true ILI observations for each year from 2007–2014. Plot (B) compares the forecast of ensembles developed using the optimisation and random forest methods for model weighting, and shows how these results evolve as more years of training data are included. Plot (C) compares the forecast of each individual Functional Form, with seven years of training data. Finally, Plot (D) compares the forecast skill of the four different weighting methods: optimisation, random forests, equal weighting and exponential Functional Form using absolute humidity. The vertical line in Plots (B) and (D) represents the true ILI peak week in 2014.

Sydney

Figures 4.9 and 4.10 show the forecasting analysis results in Sydney. From Plot (A) in both figures, we can see that the peak week in 2014 occurs later than in any previous training year; this explains the results in Plot (B) of both figures, which shows that both the random forest and optimisation weighting methods predict the peak week to fall earlier than the true peak week. This forecast does not improve as more training years are added, suggesting that the 2014 peak week occurs unusually late.

Plot (D) in both of these figures shows that the random forest weighting performs better than the optimisation weighting, regardless of the training method used. The step Functional Form using absolute humidity also performs well, outperforming all other weighting methods when using the WRMSE training method. The pattern of forecast skill is distinct between the peak week and WRMSE methods as more years of training data are added. Plot (D) of Figure 4.9 shows that for the equal and RF weighting methods, the forecast skill is fairly consistent regardless of the number of training years. The exponential Functional Form with absolute humidity improves slightly as more years are added. The optimisation method, however, initially decreases as more years are added. However, once four training years are used, the forecast skill begins to improve. This is likely related to the timing of ILI peaks seen in Plot (A). In Plot (D) in Figure 4.10, we can see that when more than three training years are added using the WRMSE method, the forecast skill of all weighting techniques remains fairly consistent.

Plot (C) in both figures shows that the step Functional Form with absolute humidity is not accurate in predicting the peak week; this again suggests that a different training method should be implemented. Finally, when comparing performance of the peak week and WRMSE training methods, we see that the peak week training method produces considerably higher forecast skill than the WRMSE method.

Overall, we saw across all four cities that generally the forecast skill did not improve as more training years were added. We also found that in all locations apart from Perth, that the peak week training method produced a higher forecast skill than the WRMSE method. There were mixed conclusions about the best technique of selecting ensemble weights, with the random forest method outperforming the optimisation in some locations but not others.

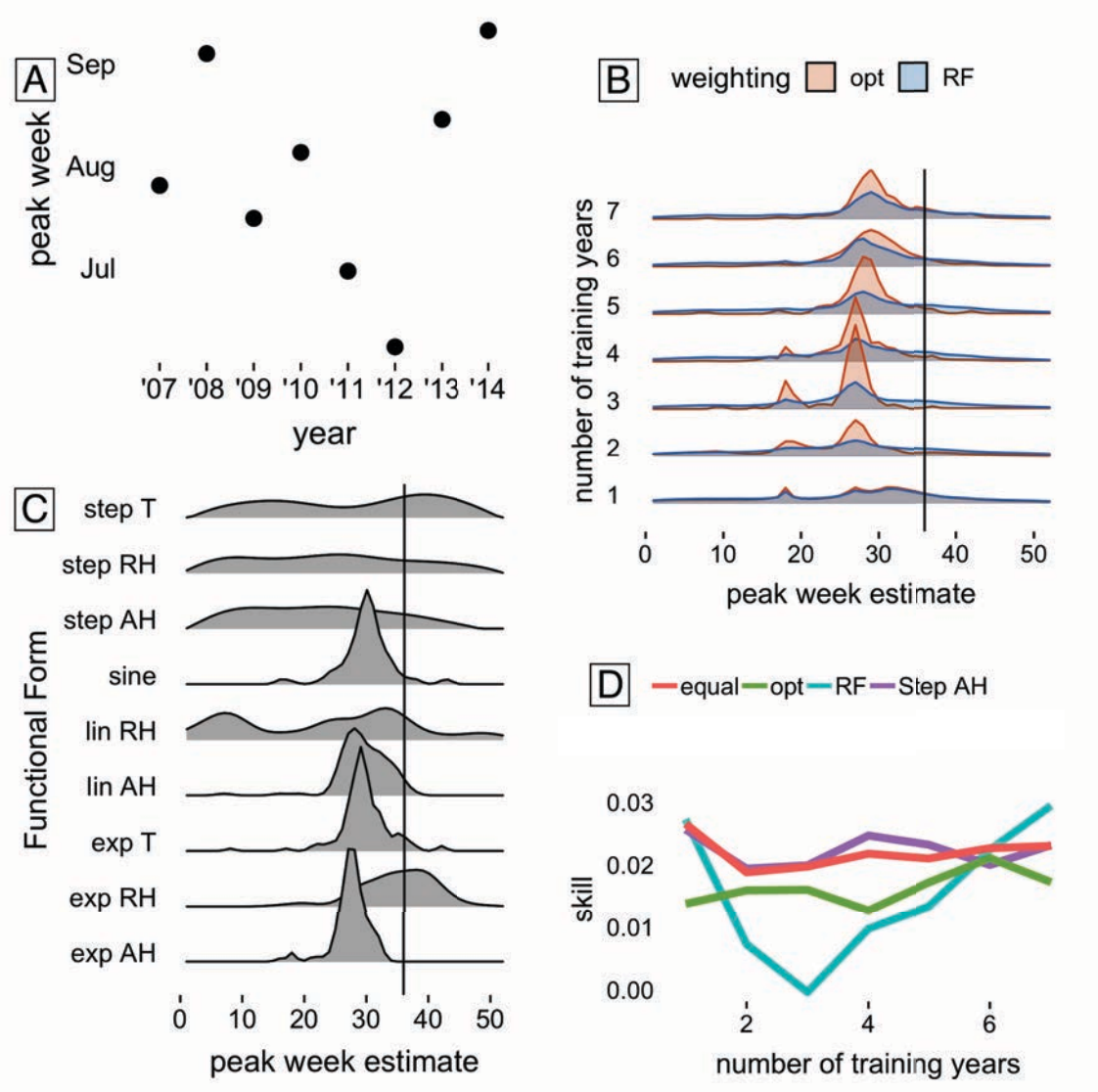

Figure 4.9: The probabilistic forecast results in Sydney, where the forecasting set is trained using the peak week method. Plot (A) shows the peak week of each year from 2007–2014. Plot (B) compares the forecast of ensembles developed using the optimisation and random forest methods for model weighting, and shows how these results evolve as more years of training data are included. Plot (C) compares the forecast of each individual Functional Form, with seven years of training data. Finally, Plot (D) compares the forecast skill of the four different weighting methods: optimisation, random forests, equal weighting and step Functional Form using absolute humidity. The vertical line in Plots (B) and (D) represents the true ILI peak week in 2014.

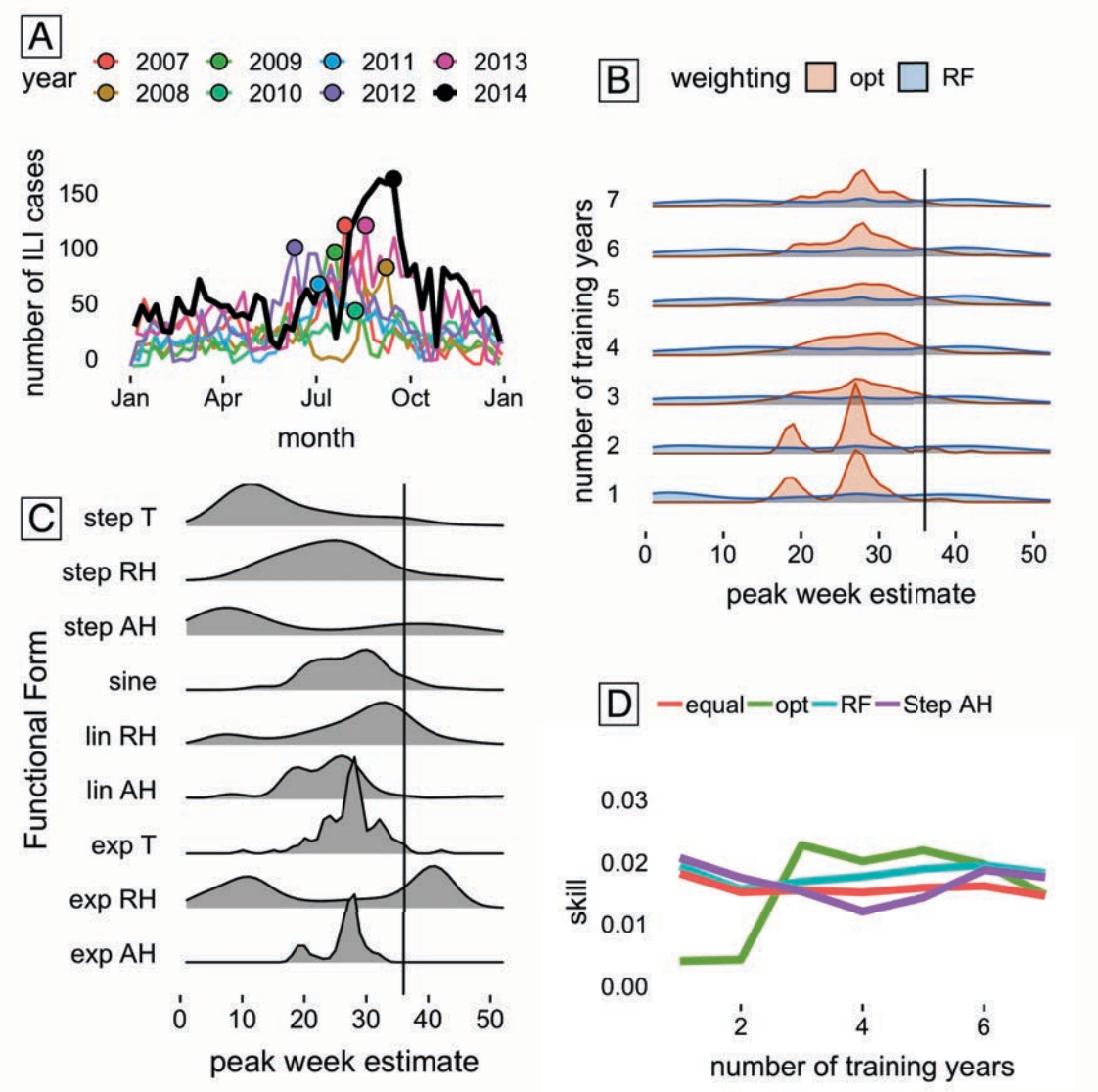

Figure 4.10: The probabilistic forecast results in Sydney, where the forecasting set is trained using the WRMSE method. Plot (A) shows the true ILI observations for each year from 2007–2014. Plot (B) compares the forecast of ensembles developed using the optimisation and random forest methods for model weighting, and shows how these results evolve as more years of training data are included. Plot (C) compares the forecast of each individual Functional Form, with seven years of training data. Finally, Plot (D) compares the forecast skill of the four different weighting methods: optimisation, random forests, equal weighting and step Functional Form using absolute humidity. The vertical line in Plots (B) and (D) represents the true ILI peak week in 2014.

4.7 Discussion

4.7.1 Summary

A goal of understanding the impact of climate variables on annual ILI epidemics is to be able to forecast the start and peak of the ILI season, allowing healthcare workers and the public health system to effectively manage resources. In this chapter, we have introduced a method to forecast the peak week of ILI infection using an ensemble of models based on the nine Functional Forms for climate-driven transmission introduced in Chapter 3. We have explored two different training methods, and four techniques for selecting ensemble model weightings. We have also used a number of years of historical ILI data to explore the effect of changing the amount of training data using these techniques.

We tested the forecasting skill for each ensemble, and analysed how the ensemble methods differ to determine which ensemble of Functional Forms is most suitable in each location.

We introduced two training methods to determine which simulations from each Functional Form to use in the forecast. One training method was based only on the timing of the peak week in previous years, while the other training method was the weighted root mean squared error (WRMSE) which uses more information from past ILI data.

We also considered four different methods to determine the weighting on each Functional Form within the ensemble: optimisation, random forest, equal weights, and best Functional Form. The optimisation method maximised the proportion of simulations that correctly predict the peak week in the training data. The random forest weighting used the same random forest technique as used in Chapter 3, and used the proportion of votes from the random forest as the weightings on the model. The equal weights method is simply equal weights given to each model, and used as a control. Finally, we also tested using only one Functional Form rather than an ensemble of all nine. The Functional Form was chosen as the one voted as best by the random forest in each location in Chapter 3: exponential Functional Form using absolute humidity in Adelaide and Perth, and step Functional Form using absolute humidity in Brisbane and Sydney.

We also used a varying number of training years to assess how the forecast skill changes as the amount of data used in training is increased. Finally, we assessed the performance of the ensembles by calculating the skill; that is, the probability that the

ensemble predicted the correct peak week of ILI in 2014.

4.7.2 Discussion of findings and limitations

It is important to note that this chapter is presented as a prototype of a new method for developing ILI forecasts. If this were to be done in practice, it would be advantageous to include many more simulations and to update the forecast weekly; unfortunately, this is outside the scope of this thesis due to computational and time constraints. In this chapter, we use the top 1% of simulations from the training set of 10,000 simulations. This means that each forecast is made up of only 100 simulations from each Functional Form. If the number of simulations were increased significantly, the forecast skill may improve.

As well as this, we forecast the peak week only from the start of 2014. In practice, we would update the forecast weekly as new ILI data is received. This should significantly improve the forecasts. As it is, by forecasting only from the start of the year we have no information about the ILI cases at the start of the season. This makes it very difficult to accurately predict the peak week, especially if the 2014 peak week occurs unusually early or late. We believe that the forecast would be greatly improved if the number of simulations was increased and the forecast updated weekly. It would also be more informative to consider the forecast for multiple different individual seasons. If the 2014 ILI peak week occurs at an unusual time in a location, then the results would be influenced according to that. If we instead consider the forecasts in multiple individual years, then we would gain more insight into the true behaviour and accuracy of the forecasts.

Surprisingly, we found overall that increasing the number of years of historical ILI training data is not necessarily advantageous. This is especially the case when the ILI season being predicted has an unusually early or late peak. In Adelaide, the forecast skill improved as more training years were added, but this was not true in the other locations being considered. This is likely because the 2014 peak ILI week in Adelaide occurred at a similar time to previous years, which again was not true in the other locations. We also see that the skill of the forecast depends on the specific training years being added; when a training year is added that has a very different peak to the 2014 peak, the forecast skill drops, and vice-versa. This suggests that testing the method on a single season is not ideal. Due to time constraints we were only able to test the method on one forecasting

year, and chose to predict the peak week in 2014 as it was the latest year for which we had full ILI notification data in all cities.

In Sydney and Brisbane, the step Functional Form using absolute humidity is selected as best by the random forest. However, we can see that the estimate of peak week in 2014 has no distinct peak in either of these locations; meaning that the simulations selected by both training methods appear to predict the peak to occur at any time between week 25 and 45. It is surprising that the step Functional Form is not good at forecasting, as Chapter 3 saw the step Functional Form fitting well to ILI in most locations following ABC fitting. This may be partially due to the step Functional Form being highly sensitive to parameter choice. It also suggests that neither the peak week nor the WRMSE training method is particularly appropriate. This is an area that may be improved on with further research.

As well as this, considering other scoring metrics alongside forecast skill may provide further insight into the performance of the ensemble techniques. We also saw no strong conclusion as to the best method of selecting weights, with the RF method outperforming the optimisation method in some locations but not others.

We saw that in Adelaide, Brisbane and Sydney, the peak week training method performed better than the WRMSE method; while the WRMSE method outperformed the peak week method only in Perth. This suggests that using the peak week is the most appropriate training method for this application, meaning that we train the ensembles only on the factor that we are trying to predict: peak week. However, this is only the case when predicting the ILI season from the start of the year. In situations where it is possible to include weekly surveillance data in the forecast, the WRMSE method would be most appropriate as it can be updated as more data is received. The peak week method can only be updated with a full year of ILI data.

Overall, we saw that there is no clear best weighting method, with the random forest performing better than the optimisation method in some locations but not in others. However, we see that in general the Functional Form selected as best by the random forest appears to fit well, reinforcing the idea that the random forest is a useful technique in epidemic model selection and could be used to inform ensemble weights. We also saw that in most cases, the ensemble forecast showed improved skill over using a single Functional Form.

We recommend that model ensembles be used for forecasting, with the RF technique for informing model weights over the optimisation technique. The RF ensembles showed more stability in forecast skill, while the optimisation ensembles varied considerably between years. However, further testing would provide greater insight to which ensemble technique is most effective.

Chapter 5

Conclusions and further research

5.1 Summary

Influenza-like illness (ILI) describes a syndrome characterised by a set of symptoms common to influenza and other similar viral illnesses. These symptoms are generally mild, but can occasionally lead to serious illness or death, especially in patients with compromised immunity, such as newborn babies or individuals suffering from chronic illnesses. In Mediterranean and subtropical climates, such as in Australia, ILI outbreaks occur each winter leading to significant pressure on healthcare systems. The exact reason for the seasonality of these outbreaks is not fully understood, with animal and human models showing conflicting evidence as to which climate variables may be driving ILI – suggesting a complicated relationship between ILI transmission and climate variables.

The aim of this thesis was to further understand what climate factors drive these seasonal ILI outbreaks by using new Bayesian machine learning model selection methods. We utilised a high-quality influenza-like illness dataset provided by the Australian Sentinel Research Practises Network (ASPREN), and considered four locations around Australia.

In Chapter 3, we developed four Functional Forms of transmissibility, based on the basic reproduction number R_0 as a function of climate. These Functional Forms are dependent on three different climate variables – temperature, absolute humidity, and relative humidity. We then used these Functional Forms of R_0 within a stochastic SEIR-type epidemic model framework, where three new compartments are added to allow us to simulate an hierarchical observation process where not all infected individuals will be

observed. This model was then used to simulate ILI epidemics over 10 years in each of the four locations.

Using these simulated ILI epidemics, we applied a random forest (RF) model selection method to select which Functional Form of R_0 within the SEIR-type model best fits the known ILI data in each location. We then fit the top two Functional Forms in each location, as selected by the RF, to the ILI data using approximate Bayesian computation (ABC), allowing us to analyse the relationship between model fit and the RF model selection method.

We found that the RF model selection framework was able to select Functional Forms in each location with a low out-of-bag error, indicating that the results were meaningful. In Adelaide and Perth, both Mediterranean climates, the exponential Functional Form using absolute humidity was selected first by the RF, followed by the step Functional Form using relative humidity. In Brisbane and Sydney, both subtropical climates, the step Functional Form using absolute humidity was selected first, followed by the exponential Functional Form using relative humidity. This indicates that while the overall climate plays a part in the behaviour of seasonal ILI, absolute humidity appears to be the most significant climate factor.

The Functional Forms selected first and second by the RF were then fitted to the true ILI data using our ABC framework. We found that in all locations apart from Sydney, both Functional Forms were able to accurately fit to the true data. In Sydney, the fit was not as accurate as the other locations. We also found, when looking at the realisations of R_0 that generated the accepted simulations in each location, that the Functional Forms selected as best by the RF method were the ones with R_0 realisations that most closely followed the pattern of the true ILI data.

Interpreting RF results in this type of framework is challenging. We chose to use ABC to fit the first and second best Functional Form in each city, as voted by the RF method, as a way of testing and validating our interpretation. We found that the Functional Form voted best in each location seemed to fit slightly better than the Functional Form voted second. This supports our interpretation of the RF model selection results. However, correct interpretation of the RF remains a challenging task.

In Chapter 4, we turned to the problem of forecasting seasonal ILI using historical data. We focussed on forecasting the timing of the peak week of ILI in 2014, using

a variety of training year sets. We considered a variety of methods to fit a weighted ensemble of forecasts from each of the Functional Forms, and tested the forecasting skill for each ensemble. We then analysed how the ensemble methods differ to determine which ensemble of Functional Forms is most suitable in each location and overall.

We used two different training methods to determine the set of simulations from each Functional Form to use in the forecast. One training method was based only on the timing of the peak week in previous years, while the other training method was the weighted root mean squared error (WRMSE).

As well as the two methods of selecting simulations, we also considered four methods of determining the weights on each Functional Form within the ensemble: equal weights, optimisation, random forest, and best Functional Form. The equal weights method assigns equal weights to each model, and used as a control. The optimisation method is a method traditionally used in ensemble modelling [83]. We used the optimisation method to maximise the proportion of simulations that correctly predict the peak week in the training data. The random forest weighting used the proportion of votes from the random forest as the weights on each Functional Form, following the same technique as used in Chapter 3. Finally, the best Functional Form method used a single Functional Form. The Functional Form was chosen as the one selected by the random forest in each location in Chapter 3: exponential Functional Form using absolute humidity in Adelaide and Perth, and step Functional Form using absolute humidity in Brisbane and Sydney.

We found that increasing the number of years of historical ILI training data used to train the forecast is not necessarily advantageous, especially in cases where the 2014 ILI season has an unusually early or late peak. We found that in Adelaide, the forecast skill improved as more training years were added. However, this was not consistent in the other locations considered. We also saw that the forecast skill depends on the specific training years being added; when a training year is added that has a very different peak to the 2014 peak, the forecast skill drops, and vice-versa. However, the task of forecasting an entire ILI season from the start of the year is challenging, and so the variability in forecast score is not necessarily a reflection of the method, but of the difficulty of the task itself.

In Adelaide, Brisbane and Sydney, using peak week to train the forecast outperformed forecasts trained using the WRMSE method; while in Perth, the WRMSE method out-

performed the peak week method. We also saw that the step Functional Form did not produce very high prediction confidence, with the estimated distribution of predicted peak week being much wider than that of the other Functional Forms. We also saw that there did not appear to be a clear best weighting method, with the random forest performing better than the optimisation method in some locations but not in others.

5.2 Discussion

The aim of this thesis was to increase our understanding of what climate factors drive these seasonal ILI outbreaks. We found that absolute humidity appears to be the most significant climate factor in driving seasonal ILI in Australia. In Chapter 3, the RF model selection process selected Functional Forms using absolute humidity in all cities. In Chapter 4, the forecasts using these selected absolute humidity-based Functional Forms performed well in all cities except Brisbane. This suggests that the results are robust, with absolute humidity playing an important role in seasonal ILI transmission.

The methods presented here are quite sensitive to the true ILI data. The ILI data in Sydney is much noisier than that in the other cities, with more between-season ILI activity. The forecast skill in Sydney is lower than the other cities, and the ABC fit of both selected Functional Forms in Sydney is not as accurate as in the other cities. This may be due to the noisiness of the ILI data in Sydney, suggesting that these results may be improved if some level of data smoothing were to be implemented. There is also uncertainty in interpreting the RF, raising the question of to what degree the results are influenced by the data itself rather than the underlying climate-dependency of ILI transmission.

We found throughout the thesis that the RF technique was informative for both model selection and ensemble modelling applications. We found in Chapter 3 that the RF model selection framework was able to select a Functional Form with a low out-of-bag error. In Chapter 4, we saw that in general, the RF method for informing ensemble model weights performed as well as, or better than, the traditional optimisation method. Further refining the techniques presented in Chapter 4 would give further insight into the difference in forecast skill between the RF and optimisation methods. As well as this, the black-box nature of RFs mean that results are challenging to interpret, leading to

further uncertainty. However, in general, random forests appear to be a useful technique for model classification in an infectious disease modelling framework.

Throughout this thesis, we focussed on nine climate-dependent Functional Forms for transmissibility. We found that the exponential and step Functional Forms with absolute humidity were selected by the RF method in Chapter 3. However, in Chapter 4, the step Functional Form was unable to predict a particular peak week of ILI with a high degree of certainty. This suggests that the step Functional Form is highly sensitive to parameter choice, while the exponential Functional Form may be more robust. This also calls into question the choice of score parameter. The WRMSE score weights within-season fit more strongly than between-season fit. However, the step Functional Form does not show distinct seasonality if unsuitable parameters are chosen; due to the nature of the WRMSE score function, these realisations may still receive a good score due to their fit within-season, despite not showing seasonality. Further refinement of the step Functional Form and consideration to parameter prior distribution choice and score function would likely increase the performance of the step Functional Form in the forecasting framework.

5.3 Further research and outlook

In Chapter 3, we use a modified SEIR-type stochastic compartmental epidemic model. Possible further extensions to this model include adding an age structure, household structure, or including human movement within the population. This would allow us to better understand how ILI spreads in human populations that are not fully-mixed. Adding in other possible factors such as school holiday dates or antigenic drift may also refine results and increase understanding of the exact drivers of ILI. However, further experimental models play a significant part of this understanding, as they allow us to study how the viral capsule and resulting transmission is affected by differing temperature and humidity which can then be applied to population-level studies. We could also consider other Functional Forms for ILI transmission, and other machine learning techniques alongside random forests.

For Chapter 4, there are many possible extensions that would be interesting to explore. These extensions include considering more ensemble techniques; different methods of training and model weighting; and considering different score metrics to gain further

insight into model fit. It would also be interesting to consider the challenge of forecasting other details of future ILI seasons, such as the size of the ILI peak as well or the length of the ILI season. Testing the ensemble forecasting technique over more ILI seasons, rather than only the 2014 season, may also provide further insight into the effectiveness of the method.

Another significant extension to Chapter 4 is to utilise this method in a framework where the ensemble forecast can be updated as weekly ILI notifications are added. This would increase the accuracy and usefulness of the technique. As well as this, it is important to implement this method with a large number of simulations. We used 10,000 simulations in this method, but ideally the use of 1,000,000 simulations should increase the accuracy and reliability of the method. It would also be informative to consider other techniques of choosing the set of simulations to be used in the forecast.

Another aspect that may be informative is to consider the geographic location of ILI notifications to post-code level rather than simply state level, which may provide further insight on the interactions between climate variables and ILI. It will also allow more information to be gathered about the socio-economic factors in each location, which is likely to be another factor relating to ILI transmission. If access to a large dataset of confirmed influenza cases were available, it would be of great interest to repeat many of these models and techniques for confirmed influenza. This would allow greater insight into the transmission and spread of true influenza.

Appendix A

In this Appendix, we present the derivation of the transition rates given in Table 3.2.

In deriving these equations, we ensure that the *mean* time to recovery is identical regardless of the path taken through the system shown in Figure 3.2.

First, we see that the time to recovery when *not* observed is

$$T_{\text{not observed}} = \frac{2}{2\gamma + \lambda},$$

and the time to recovery when observed at some stage is

$$T_{\text{observed}} = P(\text{observed first stage})T_{\text{observed first}} + P(\text{observed second stage})T_{\text{observed second}},$$

where $T_{\text{not observed}}$ denotes the time until recovery$T_{\text{observed first}}$ denotes the time to recovery when observed at the first stage, and $T_{\text{observed second}}$ denotes the time to recovery when observed at the second stage. Therefore, by substituting in the values from Figure 3.2, we get

$$T_{\text{observed}} = P(\text{observed first stage})\left(\frac{1}{2\gamma + \lambda} + \frac{2}{\alpha}\right) + P(\text{observed second stage})\left(\frac{2}{2\gamma + \lambda} + \frac{1}{\alpha}\right).$$

We can then consider the probabilities separately, with

$$P(\text{observed first stage}) = \frac{\frac{\lambda}{2\gamma+\lambda}}{\frac{\lambda(2\gamma+\lambda)+2\gamma\lambda}{(2\gamma+\lambda)^2}} = \frac{2\gamma + \lambda}{4\gamma + \lambda}$$

and

$$P(\text{observed second stage}) = \frac{\frac{2\gamma\lambda}{(2\gamma+\lambda)^2}}{\frac{\lambda(2\gamma+\lambda)+2\gamma\lambda}{(2\gamma+\lambda)^2}} = \frac{2\gamma}{4\gamma + \lambda}.$$

Substituting these values into the above equation shows that

$$T_{\text{observed}} = \frac{2}{2\gamma + \lambda} = T_{\text{not observed}},$$

which can then be rearranged to show that

$$\alpha = 3\gamma + \lambda,$$

as given in Table 3.2. The remaining values in Table 3.2 are simply derived from this and from the values in Figure 3.2, taking into account that there are two time-steps per day within this model.

Appendix B

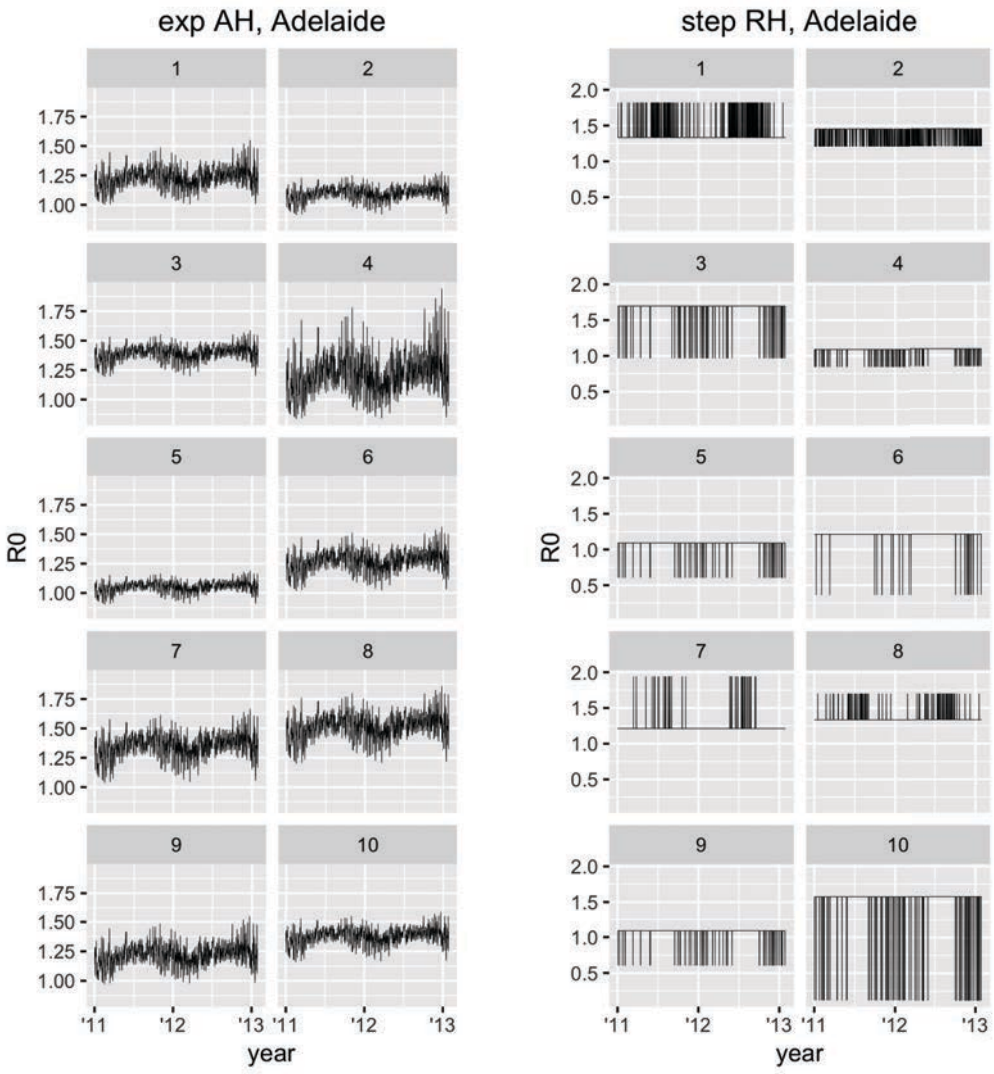

Figure B.1: Individual realisations of R_0 over the years 2011–2013 from the 10 top scoring simulations from each Functional Form in Adelaide. The two Functional Forms, exponential using absolute humidity and step using relative humidity, are the first and second best as voted by the random forest model selection in Chapter 3.

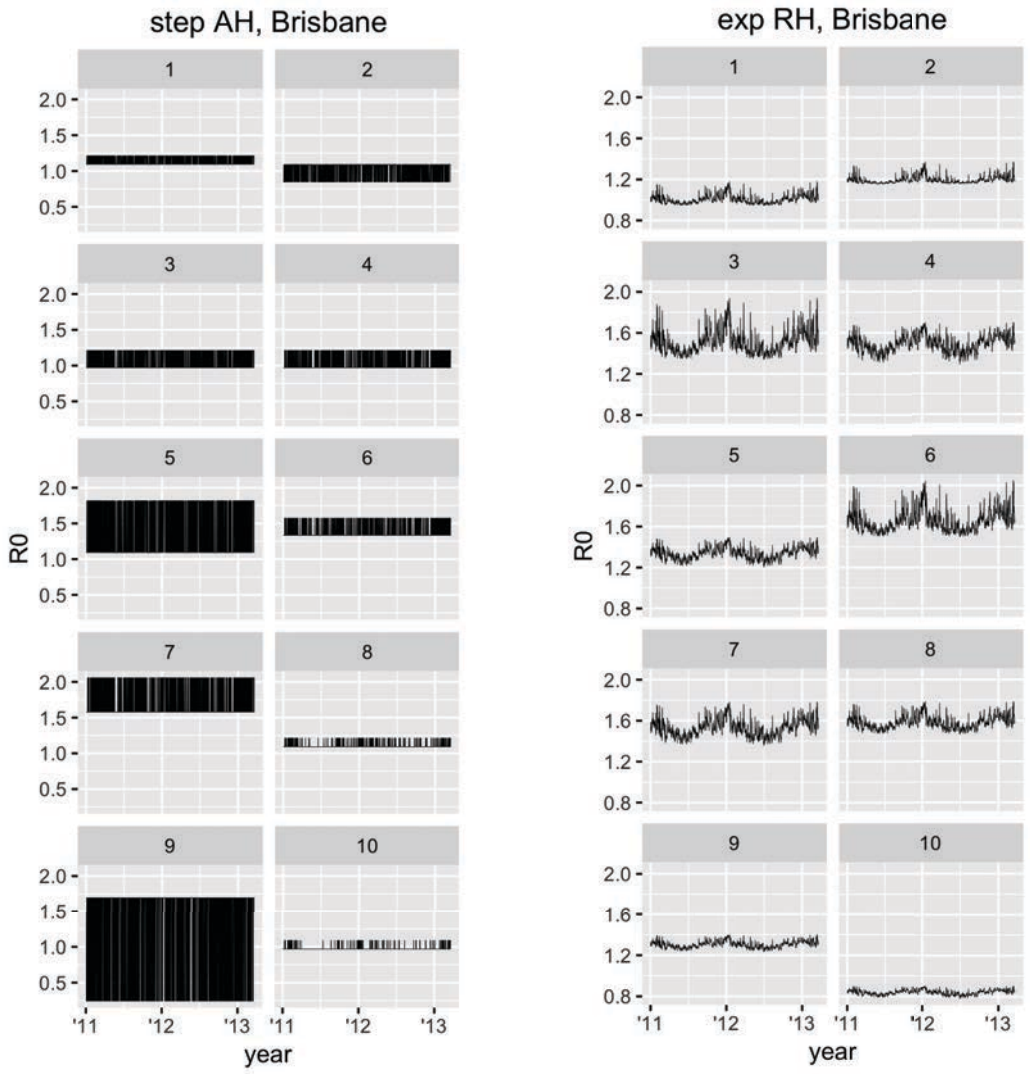

Figure B.2: Individual realisations of R_0 over the years 2011–2013 from the 10 top scoring simulations from each Functional Form in Brisbane. The two Functional Forms, step using absolute humidity and exponential using relative humidity, are the first and second best as voted by the random forest model selection in Chapter 3.

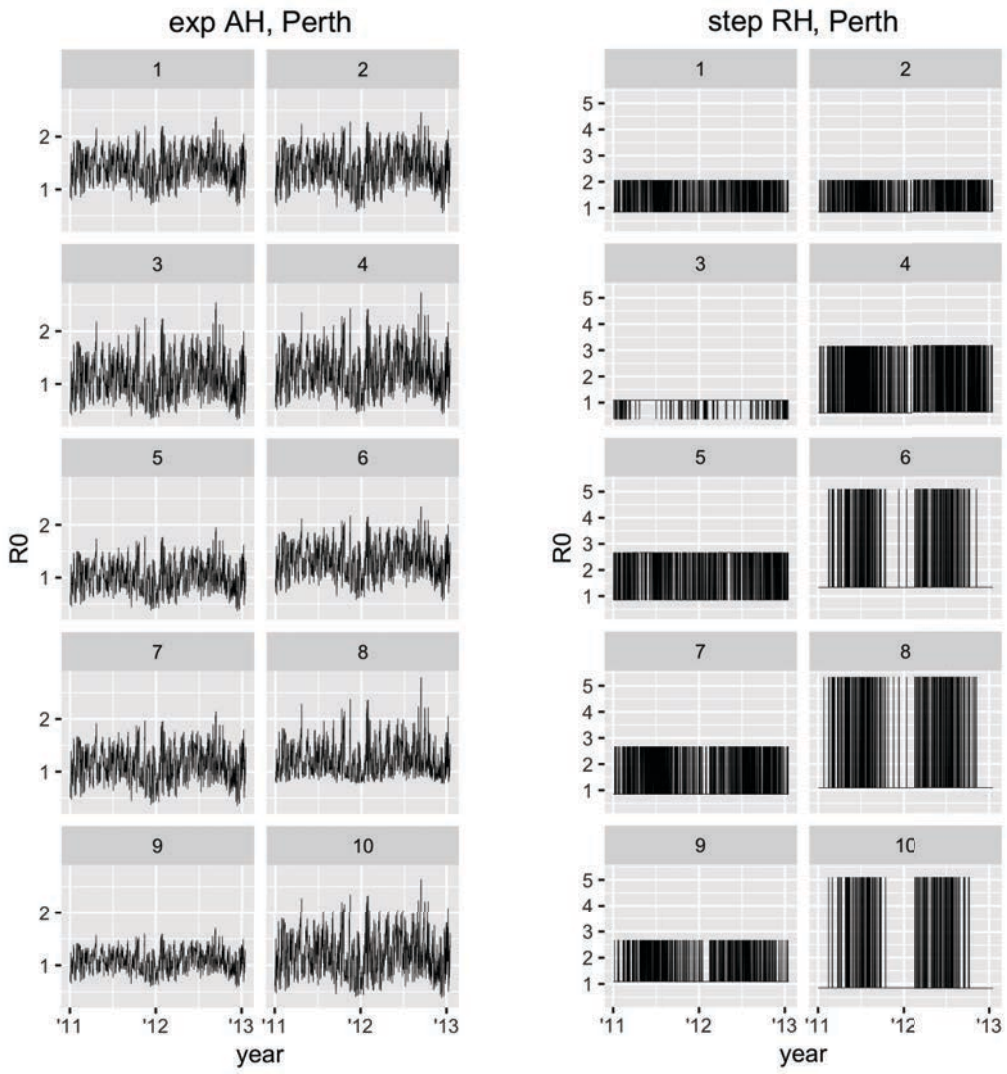

Figure B.3: Individual realisations of R_0 over the years 2011–2013 from the 10 top scoring simulations from each Functional Form in Perth. The two Functional Forms, exponential using absolute humidity and step using relative humidity, are the first and second best as voted by the random forest model selection in Chapter 3.

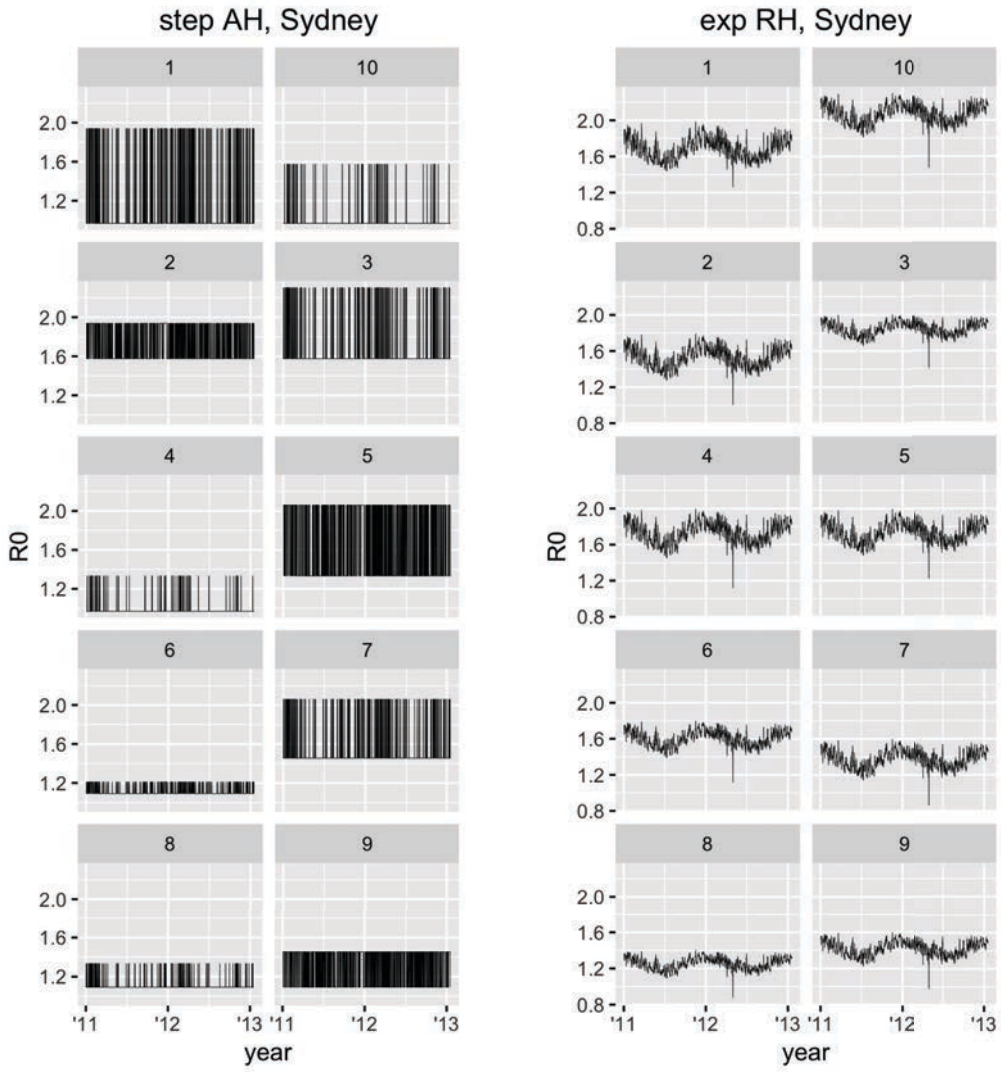

Figure B.4: Individual realisations of R_0 over the years 2011–2013 from the 10 top scoring simulations from each Functional Form in Sydney. The two Functional Forms, step using absolute humidity and exponential using relative humidity, are the first and second best as voted by the random forest model selection in Chapter 3.

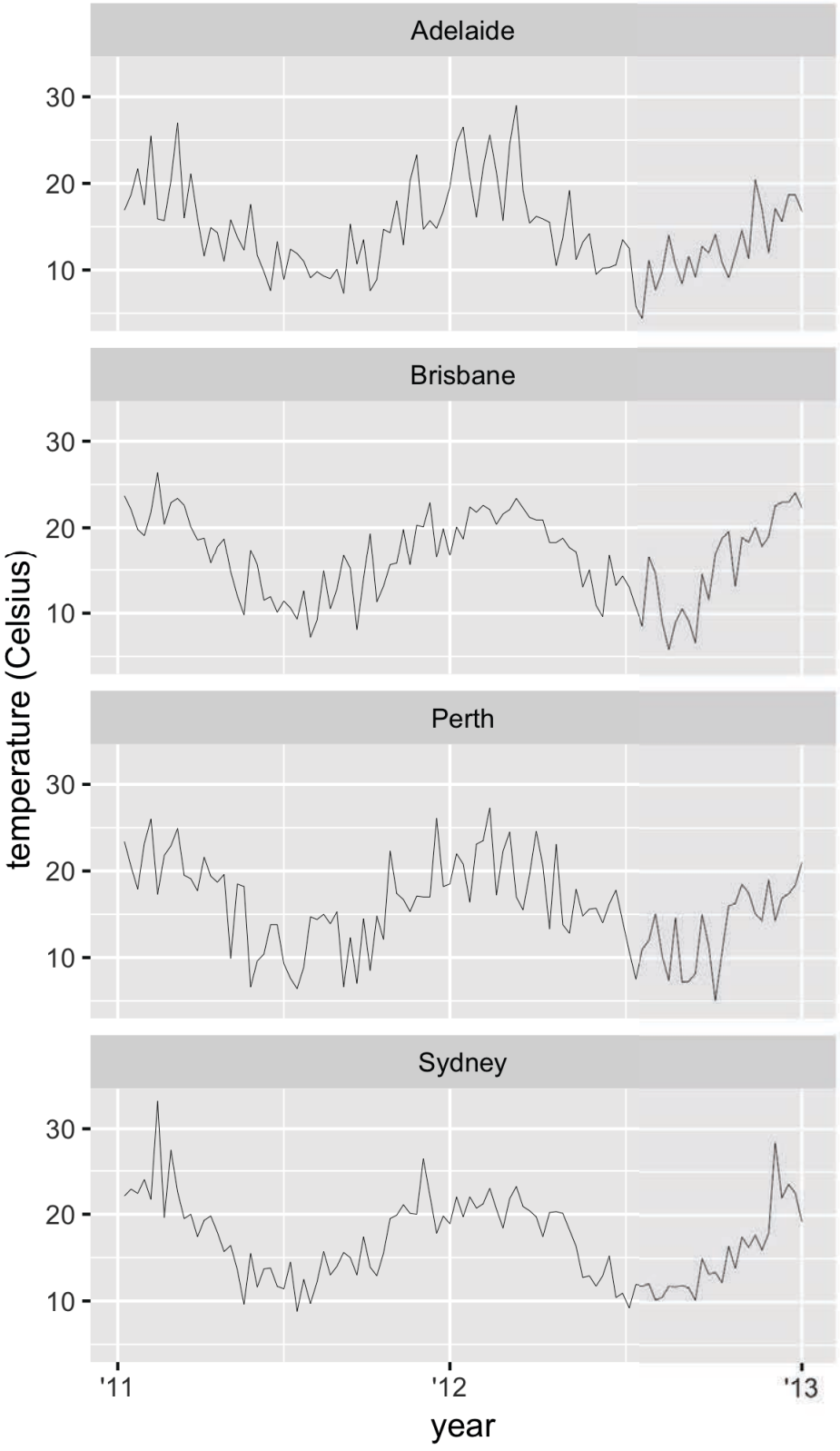

Figure B.5: A comparison of temperature between locations over the years 2011–2013 with weekly averages.

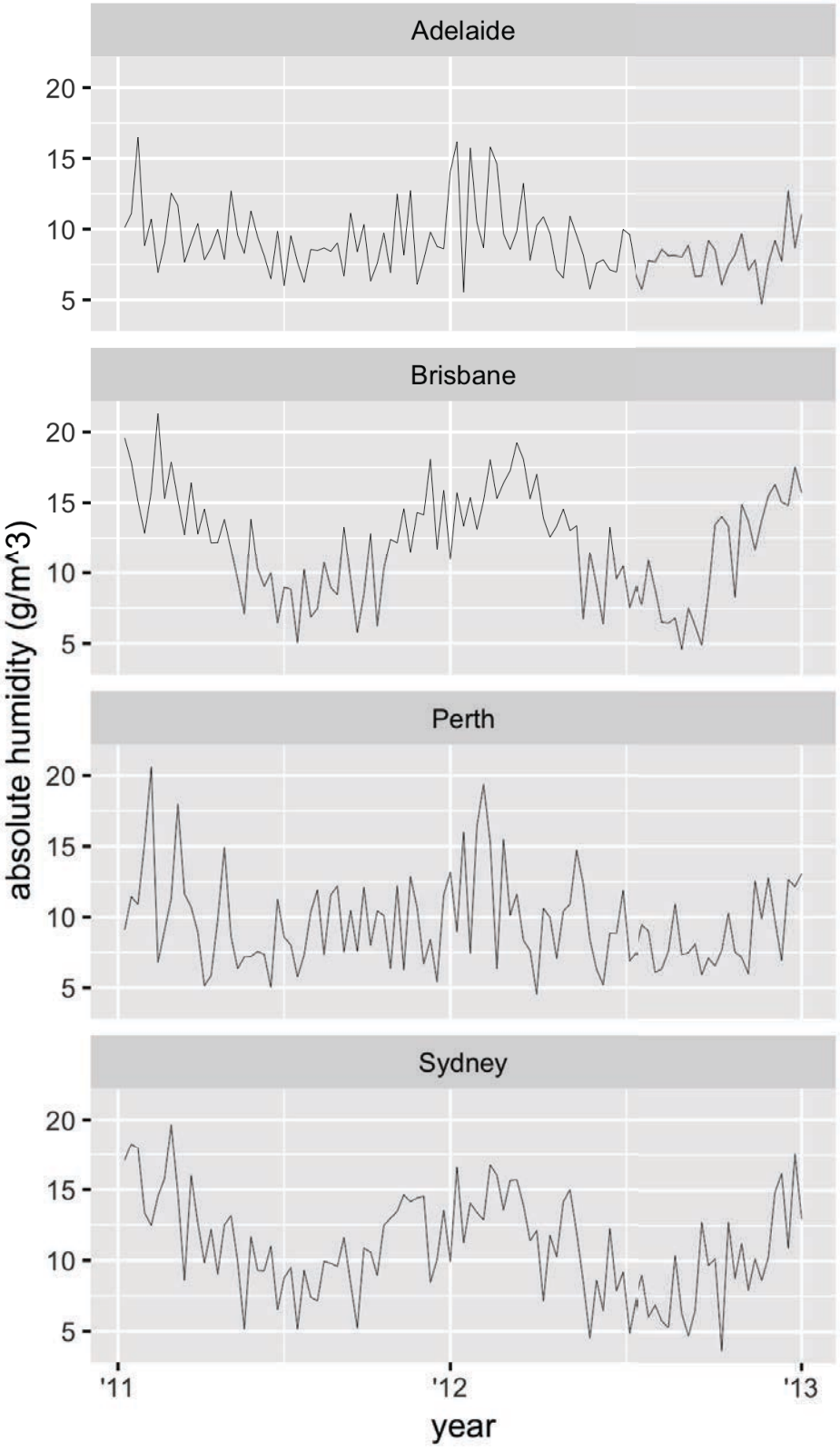

Figure B.6: A comparison of absolute humidity between locations over the years 2011–2013 with weekly averages.

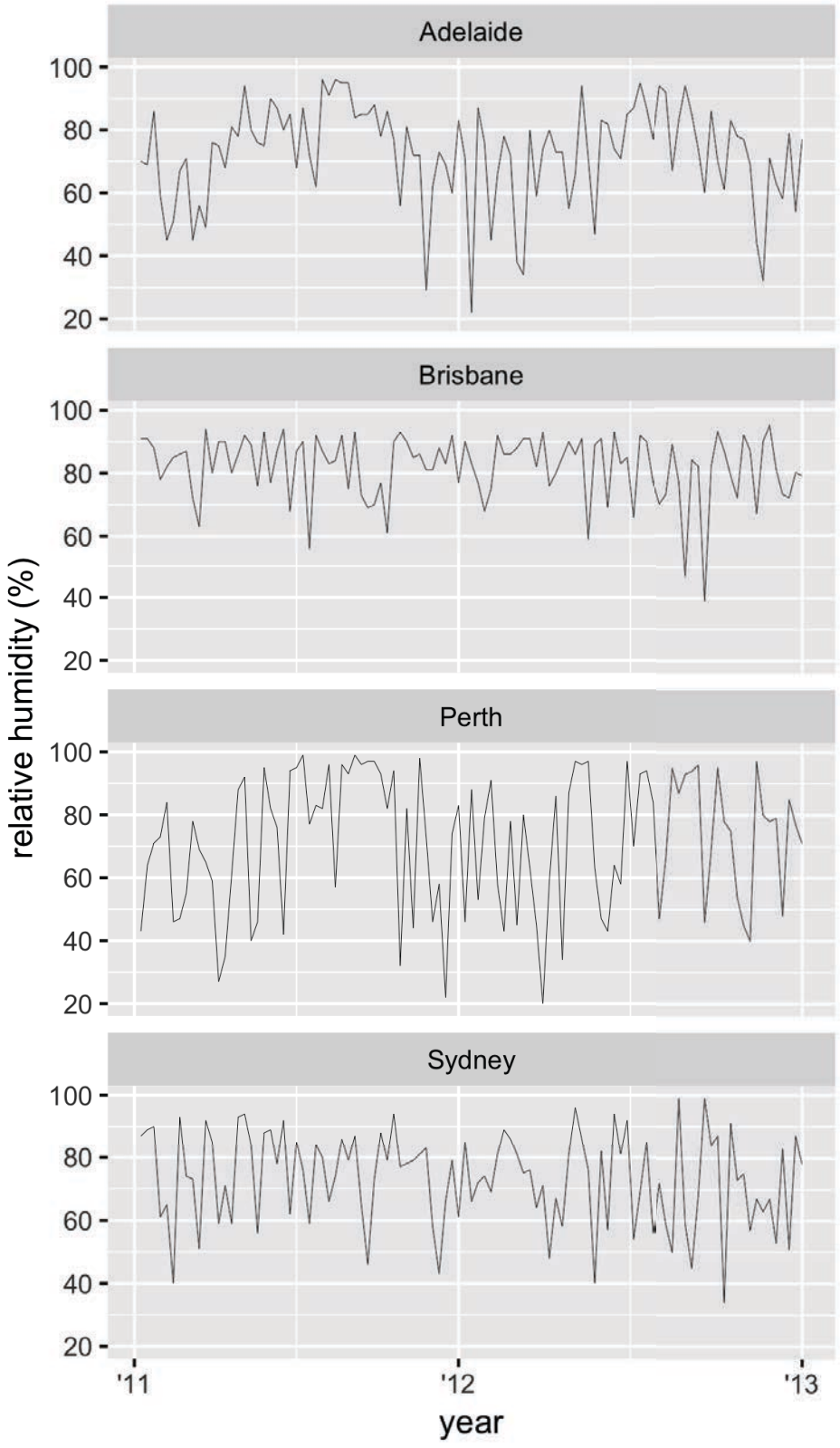

Figure B.7: A comparison of relative humidity between locations over the years 2011–2013 with weekly averages.

Bibliography

[1] M. Sunnåker, A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz, "Approximate Bayesian computation," *PLoS Computational Biology*, vol. 9, 2013.

[2] M. C. Peel, B. L. Finlayson, and T. A. McMahon, "Updated world map of the Köppen-Geiger climate classification," *Hydrology and Earth System Sciences*, vol. 11, pp. 1633–1644, 2007.

[3] A. T. Newall and P. A. Scuffham, "Influenza-related disease: The cost to the Australian healthcare system," *Vaccine*, vol. 26, pp. 6818–6823, 2008.

[4] J. W. Tang, F. Y. Lai, P. Nymadawa, Y.-M. Deng, M. Ratnamohan, M. Petric, T. P. Loh, N. W. Tee, D. E. Dwyer, I. G. Barr, and F. Y. Wong, "Comparison of the incidence of influenza in relation to climate factors during 2000-2007 in five countries," *Journal of Medical Virology*, vol. 82, no. 11, pp. 1958–1965, 2010.

[5] E. A. Baumgartner, C. N. Dao, S. Nasreen, M. U. Bhuiyan, S. Mah-E-Muneer, A. A. Mamun, M. A. Y. Sharker, R. U. Zaman, P.-Y. Cheng, A. I. Klimov, M.-A. Widdowson, T. M. Uyeki, S. P. Luby, A. Mounts, and J. Bresee, "Seasonality, timing, and climate drivers of influenza activity worldwide," *The Journal of Infectious Diseases*, vol. 206, pp. 838–846, 2012.

[6] The Australian Department of Health, "Australian influenza 2016: Laboratory confirmed influenza activity," pp. 1–10, 2016.

[7] H. Seale, A. E. Heywood, M.-L. McLaws, K. F. Ward, C. P. Lowbridge, D. Van, and C. R. MacIntyre, "Why do I need it? I am not at risk! Public perceptions towards the pandemic (H1N1) 2009 vaccine," *BMC Infectious Diseases*, vol. 10, pp. 10–99, 2010.

[8] E. M. Hill, M. J. Tildesley, and T. House, "Evidence for history-dependence of influenza pandemic emergence," *Scientific Reports*, vol. 7, article 43623, 2017.

[9] P. J. Kumar and M. L. Clark, *Kumar & Clark's clinical medicine*. Saunders Elsevier, 2009.

[10] R. Lehnert, M. Pletz, A. Reuss, and T. Schaberg, "Antiviral medications in seasonal and pandemic influenza," *Deutsches Aerzteblatt International*, vol. 113, pp. 799–807, 2016.

[11] S. G. Sullivan, M. B. Chilver, K. S. Carville, Y.-M. Deng, K. A. Grant, G. Higgins, N. Komadina, V. K. Leung, C. A. Minney-Smith, D. Teng, T. Tran, N. Stocks, and J. E. Fielding, "Low interim influenza vaccine effectiveness, Australia, 1 May to 24 September 2017," *Eurosurveillance*, vol. 22, pp. 22–43, 2017.

[12] M. R. Kasper, T. F. Wierzba, L. Sovann, P. J. Blair, and S. D. Putnam, "Evaluation of an influenza-like illness case definition in the diagnosis of influenza among patients with acute febrile illness in Cambodia," *BMC Infectious Diseases*, vol. 10, p. 320, 2010.

[13] J. Fitzner, S. Qasmieh, A. W. Mounts, B. Alexander, T. Besselaar, S. Briand, C. Brown, S. Clark, E. Dueger, D. Gross, S. Hauge, S. Hirve, P. Jorgensen, M. A. Katz, A. Mafi, M. Malik, M. McCarron, T. Meerhoff, Y. Mori, J. Mott, M. T. da Costa Olivera, J. R. Ortiz, R. Palekar, H. R. de Andrade, L. Soetens, A. A. Yahaya, W. Zhang, and K. Vandemaele, "Revision of clinical case definitions: influenza-like illness and severe acute respiratory infection," *Bulletin of the World Health Organization*, vol. 96, pp. 122–128, 2017.

[14] T. Heikkinen and A. Järvinen, "The common cold," *The Lancet*, vol. 361, pp. 51–59, 2003.

[15] L. A. Wallace, T. C. Collins, J. D. Douglas, S. McIntyre, J. Millar, and W. F. Carman, "Virological surveillance of influenza-like illness in the community using PCR and serology," *Journal of Clinical Virology*, vol. 31, pp. 40–45, 2004.

[16] S. G. Sullivan, M. B. Chilver, G. Higgins, A. C. Cheng, and N. P. Stocks, "Influenza

vaccine effectiveness in Australia: results from the Australian Sentinel Practices Research Network," *The Medical Journal of Australia*, vol. 201, pp. 109–111, 2014.

[17] B. D. Dalziel, S. Kissler, J. R. Gog, C. Viboud, O. N. Bjørnstad, C. J. E. Metcalf, and B. T. Grenfell, "Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities," *Science*, vol. 362, pp. 75–79, 2018.

[18] M. Lipsitch and C. Viboud, "Influenza seasonality: Lifting the fog," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 3645–3646, 2009.

[19] O. N. Bjørnstad and C. Viboud, "Timing and periodicity of influenza epidemics," *Proceedings of the National Academy of Sciences*, vol. 113, pp. 12899–12901, 2016.

[20] C. Viboud, K. Pakdaman, P. Boelle, M. Wilson, M. Myers, A. Valleron, and A. Flahault, "Association of influenza epidemics with global climate variability," *European Journal of Epidemiology*, vol. 19, pp. 1055–1095, 2004.

[21] 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 Pathogens*, vol. 9, pp. 9–11, 2013.

[22] A. C. Lowen, S. Mubareka, J. Steel, and P. Palese, "Influenza virus transmission is dependent on relative humidity and temperature," *PLoS Pathogens*, vol. 3, p. e151, 2007.

[23] J. Shaman and M. Kohn, "Absolute humidity modulates influenza survival, transmission, and seasonality," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 3243–3248, 2009.

[24] A. Mahamat, P. Dussart, A. Bouix, L. Carvalho, F. Eltges, S. Matheus, M. Miller, P. Quenel, and C. Viboud, "Climatic drivers of seasonal influenza epidemics in french guiana, 2006–2010," *Journal of Infection*, vol. 67, pp. 141–147, 2013.

[25] D. Slusky and R. J. Zeckhauser, "Sunlight and protection against influenza," *National Bureau of Economic Research, Cambridge, MA*, 2018.

[26] K. Jaakkola, A. Saukkoriipi, J. Jokelainen, R. Juvonen, J. Kauppila, O. Vainio, R. Ziegler, E. Ronkko, J. Jaakkla, and T. Ikaheimo, "Decline in temperature and

humidity increases the occurrence of influenza in cold climate," *American Journal of Epidemiology*, 2007.

[27] H. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, pp. 599–653, 2000.

[28] M. Roberts, V. Andreasen, A. Lloyd, and L. Pellis, "Nine challenges for deterministic epidemic models," *Epidemics*, vol. 10, pp. 49 – 53, 2015.

[29] A. C. Lowen and J. Steel, "Roles of humidity and temperature in shaping influenza seasonality," *Journal of Virology*, vol. 88, pp. 7692–7695, 2014.

[30] M. Ud-dean, "Structural explanation for the effect of humidity on persistence of airborne virus: Seasonality of influenza," *Journal of Theoretical Biology*, vol. 264, pp. 822–829, 2010.

[31] R. Yaari, G. Katriel, A. Huppert, J. B. Axelsen, and L. Stone, "Modelling seasonal influenza : the role of weather and punctuated antigenic drift," *Journal of the Royal Society*, vol. 10, no. 84, p. 20130298, 2013.

[32] J. Axelsen, R. Yaari, B. T. Grenfell, and L. Stone, "Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 9538–9542, 2014.

[33] J. Shaman and A. Karspeck, "Forecasting seasonal outbreaks of influenza," *Proceedings of the National Academy of Sciences*, vol. 109, no. 50, pp. 20425–20430, 2012.

[34] 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*, vol. 8, 2010.

[35] R. C. Cope, J. V. Ross, M. Chilver, N. P. Stocks, and L. Mitchell, "Characterising seasonal influenza epidemiology using primary care surveillance data," *PLoS Computational Biology*, vol. 14, p. e1006377, 2018.

[36] T. Kamigaki, L. Chaw, A. G. Tan, R. Tamaki, P. P. Alday, J. B. Javier, R. M. Olveda, H. Oshitani, and V. L. Tallo, "Seasonality of influenza and respiratory syn-

cytial viruses and the effect of climate factors in subtropical and tropical asia using influenza-like illness surveillance data, 2010 – 2012," *PLoS ONE*, vol. 11, 2016.

[37] Z. P. Ross, N. Komadina, Y.-M. Deng, N. Spirason, H. A. Kelly, S. G. Sullivan, I. G. Barr, and E. C. Holmes, "Inter-seasonal influenza is characterized by extended virus transmission and persistence," *PLoS Pathogens*, vol. 11, 2015.

[38] Y. Zhang, H. Bambrick, K. Mengersen, S. Tong, and W. Hu, "Using Google Trends and ambient temperature to predict seasonal influenza outbreaks," *Environment International*, vol. 117, pp. 284–291, 2018.

[39] P. Chakraborty, B. Lewis, S. Eubank, J. S. Brownstein, M. Marathe, and N. Ramakrishnan, "What to know before forecasting the flu," *PLoS Computational Biology*, vol. 14, 2018.

[40] C. J. E. Metcalf, K. S. Walter, A. Wesolowski, C. O. Buckee, E. Shevliakova, A. J. Tatem, W. R. Boos, D. M. Weinberger, and V. E. Pitzer, "Identifying climate drivers of infectious disease dynamics: recent advances and challenges ahead," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 284, 2017.

[41] R. C. Cope, J. V. Ross, M. Chilver, N. P. Stocks, and L. Mitchell, "Connecting surveillance and population-level influenza incidence," 2018, www.biorxiv.org/content/early/2018/09/28/427708.

[42] M. Peppa, W. John Edmunds, and S. Funk, "Disease severity determines health-seeking behaviour amongst individuals with influenza-like illness in an internet-based cohort," *BMC Infectious Diseases*, vol. 17, p. 238, 2017.

[43] Bureau of Meteorology, "Climate data online," *Australian Bureau of Meteorology*, 2013, www.bom.gov.au.

[44] K. Dietz, "The estimation of the basic reproduction number for infectious diseases," *Statistical Methods in Medical Research*, vol. 2, pp. 23–41, 1993.

[45] Z. Ma, Y. Zhou, and J. Wu, *Modeling and Dynamics of Infectious Diseases*. World Scientific, Co-published with Higher Education Press, Park City, 2009.

[46] B. J. Cowling, M. S. Lau, L.-M. Ho, S.-K. Chuang, T. Tsang, S.-H. Liu, P.-Y. Leung,

S.-V. Lo, and E. H. Lau, "The effective reproduction number of pandemic influenza: prospective estimation," *Epidemiology*, vol. 21, p. 842, 2010.

[47] N. Bacaër and R. Ouifki, "Growth rate and basic reproduction number for population models with a simple periodic factor," *Mathematical Biosciences*, vol. 210, pp. 647–658, 2007.

[48] E. Behrends, *Introduction to Markov chains*, vol. 228. Springer, 2000.

[49] N. P. Rebuli, N. G. Bean, and J. V. Ross, "Hybrid Markov chain models of S–I–R disease dynamics," *Journal of Mathematical Biology*, vol. 75, no. 3, pp. 521–541, 2016.

[50] T. G. Kurtz, *Approximation of Population Processes*. Society for Industrial and Applied Mathematics, Madison, WI, 1981.

[51] L. J. Allen, F. Brauer, P. Van den Driessche, and J. Wu, *Mathematical Epidemiology*, vol. 1945. Springer, 2008.

[52] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," *Journal of the Royal Society Interface*, pp. 187–202, 2009.

[53] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? –arguments against avoiding RMSE in the literature," *Geoscientific model development*, vol. 7, pp. 1247–1250, 2014.

[54] W. Zucchini, "An introduction to model selection," *Journal of Mathematical Psychology*, vol. 61, pp. 41–61, 2000.

[55] K. Csilléry, M. G. Blum, O. E. Gaggiotti, and O. François, "Approximate Bayesian computation (ABC) in practice," *Trends in Ecology & Evolution*, vol. 25, no. 7, pp. 410–418, 2010.

[56] J.-M. Marin, N. S. Pillai, C. P. Robert, and J. Rousseau, "Relevant statistics for Bayesian model choice," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, no. 5, pp. 833–859, 2013.

[57] A. A. Neath and J. E. Cavanaugh, "The Bayesian information criterion: background, derivation, and applications," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 4, no. 2, pp. 199–203, 2011.

[58] G. Claeskens and N. L. Hjort, *Model Selection and Model Averaging*. Cambridge University Press, Cambridge, 2008.

[59] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder, "Approximate Bayesian computational methods," *Statistics and Computing*, vol. 22, no. 6, pp. 1167–1180, 2011.

[60] C. C. Drovandi, A. N. Pettitt, and R. A. McCutchan, "Exact and approximate bayesian inference for low integer-valued time series models with intractable likelihoods," *Bayesian Analysis*, vol. 11, no. 2, pp. 325–352, 2016.

[61] C. C. Drovandi and A. N. Pettitt, "Estimation of parameters for macroparasite population evolution using approximate bayesian computation," *Biometrics*, vol. 67, no. 1, pp. 225–233, 2010.

[62] P. Pudlo, J. M. Marin, A. Estoup, J. M. Cornuet, M. Gautier, and C. P. Robert, "Reliable ABC model choice via random forests," *Bioinformatics*, vol. 32, pp. 859–866, 2015.

[63] R. Kaundal, A. S. Kapoor, and G. P. Raghava, "Machine learning techniques in disease forecasting: a case study on rice blast prediction," *BMC Bioinformatics*, vol. 7, no. 1, p. 485, 2006.

[64] G. Pokharel and R. Deardon, "Supervised learning and prediction of spatial epidemics," *Spatial and Spatio-temporal Epidemiology*, vol. 11, pp. 59–77, 2014.

[65] E. O. Nsoesie, R. Beckman, M. Marathe, and B. Lewis, "Prediction of an epidemic curve: A supervised classification approach," *Statistical Communications in Infectious Diseases*, vol. 3, 2011.

[66] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news*, vol. 2, pp. 18–22, 2002.

[67] D. H. Wolpert and W. G. Macready, "An efficient method to estimate bagging's generalization error," *Machine Learning*, vol. 35, pp. 41–55, 1999.

[68] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[69] C. Strobl, J. Malley, and G. Tutz, "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests.," *Psychological Methods*, vol. 14, p. 323, 2009.

[70] M. P. Wand and M. C. Jones, *Kernel Smoothing*. Springer US, 1995.

[71] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, pp. 605–610, 1979.

[72] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory of Probability & Its Applications*, vol. 14, no. 1, pp. 153–158, 1969.

[73] W. N. Venables, "An introduction to statistical inference and its applications with R by Trosset, M. W.," *Biometrics*, vol. 66, no. 2, pp. 656–657, 2010.

[74] C. P. Robert, B. S. Everitt, and T. Hothorn, "A handbook of statistical analyses using R, second edition," *International Statistical Review*, vol. 79, no. 2, pp. 276–277, 2011.

[75] S. Sheather and C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 53, pp. 683–690, 1991.

[76] T. Duong, *ks: Kernel Smoothing*, 2018. R package version 1.11.3.

[77] D. Green and R. Perry, *Perry's Chemical Engineers' Handbook, Eighth Edition*. McGraw Hill professional, McGraw-Hill Education, 2007.

[78] P. Mander, "How to convert relative humidity to absolute humidity," 2012, https://carnotcycle.wordpress.com/2012/08/04/how-to-convert-relative-humidity-to-absolute-humidity.

[79] J. Shaman, C. Y. Jeon, E. Giovannucci, and M. Lipsitch, "Shortcomings of vitamin D-based model simulations of seasonal influenza," *PLoS ONE*, vol. 6, pp. 1–7, 2011.

[80] J.-M. Marin, P. Pudlo, and A. Estoup, "Likelihood-free model choice," *Handbook of ABC*, pp. 1–21, 2015.

[81] A. Palczewska, J. Palczewski, R. M. Robinson, and D. Neagu, "Interpreting random forest classification models using a feature contribution method," in *Integration of Reusable Systems*, pp. 193–218, Springer International Publishing, 2014.

[82] M. Myers, D. Rogers, J. Cox, A. Flahault, and S. Hay, "Forecasting disease risk for increased epidemic preparedness in public health," in *Remote Sensing and Geographical Information Systems in Epidemiology*, pp. 309–330, Elsevier, 2000.

[83] E. L. Ray and N. G. Reich, "Prediction of infectious disease epidemics via weighted density ensembles," *PLoS Computational Biology*, vol. 14, no. 2, p. e1005910, 2018.

[84] P. Smyth and D. Wolpert *Machine Learning*, vol. 36, no. 1/2, pp. 59–83, 1999.

[85] J. Slingo and T. Palmer, "Uncertainty in weather and climate prediction," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 369, no. 1956, pp. 4751–4767, 2011.

[86] R. Benedetti, "Scoring rules for forecast verification," *Monthly Weather Review*, vol. 138, no. 1, pp. 203–211, 2010.

[87] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[88] R. Caruana, A. Niculescu-Mizil, G. Crew, and A. Ksikes, "Ensemble selection from libraries of models," in *Twenty-first international conference on Machine learning - ICML '04*, ACM Press, 2004.