Imperial College London

Imperial College London

Life Sciences

The Role of Temperature-Dependent Human Behaviour and Virus Stability on Respiratory Disease Dynamics

Author:

Ruth Keane

Supervisors:

Dr Samraat Pawar

Dr Michael Tristem

Word Count: 5888

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Science at Imperial College London

Formatted in the style of Ecology

Submitted for the MSc in Computational Methods in Ecology and Evolution

27 August 2020

Abstract

Respiratory viruses can show strong seasonality and are a substantial burden on health services. Understanding their seasonality could help to reduce their impact. Previous models of disease have not explicitly modelled contact rate and the decay rate of the virus simultaneously. I suggest a model for temperature dependent transmissbility based on both viral survival and contact rate to be used in an SEIR model. Using this model and estimated average weekly temperatures, I predict the number of infectious individuals each day for 77 countries (for 10 years, then averaged over one year). As contact rate patterns with climate are not fully understood, I vary the climate at which contact rate is assumed to be highest, resulting in repeats at different levels of mismatch, where mismatch measures the difference between the climates where contact rate and virus survival are highest. To assess which level of mismatch was best, I compare the correlation between influenza data and model results (averaged for one year) for each each mismatch for each country. In temperate regions, the best mismatch result was consistent with highest contact rate and virus survival occurring at the same temperature for influenza. In the tropics the mismatch was much more varied and there was a high variability between countries. A mismatch of 0.5 led to lower mean R_0 values and a mismatch of 1 led to lower maximum R_0 values. When I applied this model to COVID-19 (using the best mismatch from influenza results), temperate regions had a consistent winter peak. Shifting the mismatch value decreased the mean and maximum R_0. This is new model which attempts to provide a mechanistic model for respiratory virus seasonality which could be used to improve current epidemiological methods. More research into the climate-dependence of contact rate is very important in understand how human behaviour might effect seasonal forcing.

Contents

List of Figures

List of Tables

1 Introduction

Respiratory viruses put a huge burden on the world at both endemic and epidemic levels. Influenzas are a problem both as seasonal influenza and pandemic influenzas such as H1N1 (World Health Organization (WHO) 2018b; World Health Organization (WHO) 2018a). Most recently COVID-19, a respiratory virus caused by SARS-CoV-2 is having significant impacts across the world, first seen in December 2019 and declared a pandemic on the 11th March 2020. (World Health Organization (WHO) 2020b; World Health Organization (WHO) 2020a). Previous coronovirus epidemics such as SARS have prompted discussions about the importance of understanding the nature of seasonality infectious disease but much is still unknown (Dowell and Ho 2004). The future of COVID-19 is uncertain but some predict that repeated wintertime outbreaks are likely, at least in the short term (Kissler et al. 2020). Understanding the causes for the dynamics of these diseases is helpful for planning the timing and types of interventions that should take place. This can be especially important for seasonal pathogens because periods with fewer cases allow time for planning in advance of the next outbreak.

Differences in climate have been frequently linked to disease, both temporally and spatially. In terms of differences over the course of a year, influenza is perhaps the most well known seasonal respiratory virus but others include respiratory syncytial virus (RSV) (J. W. Tang et al. 2010), mild betacoronoviruses (Kissler et al. 2020) pneumococcus and rubella (Dowell and Ho 2004). These tend to peak in the winters of temperate regions. Climate may also mediate difference between severity of outbreaks in different years. In terms of differences in space, in general the tropics and temperate regions have different disease patterns. Marked yearly winter peaks may be more common in temperate regions but in the tropics, disease may peak in the rainy season, have two peaks or remain high all year round (J. D. Tamerius et al. 2013).

For the most part, work to understand the impact of climate is statistical work that links climate with disease including time series analysis. While not being mechanistic these can be very valuable in understanding the factors that lead to seasonality and heterogeneity of disease. A lot of phenomenological work has explored whether COVID-19 is linked to environmental variables. These tend to indicate that lower temperatures and lower relative humidity are associated with more disease but results do vary based on the area and the study methods *(add references/specific examples). (add phenomenological flu)* Attempts have been made to demonstrate causality between climate and influenza. Deyle et al. (2016) used convergent cross mapping to find that absolute humidity mediated by temperature was the most likely cause but whether this truly demonstrates causality is in question because the same approach can be used to demonstrate that influenza causes the climatic conditions, which is obviously not the case

(Baskerville and Cobey 2017; Sugihara, Deyle, and Ye 2017) . As O. N. Bjørnstad and Viboud (2016) point out, combination of such work with mechanistic models is important to enable prediction of future disease.

The causes for seasonality have been discussed in lots of detail but can broadly be split into three potential reasons: human behaviour, human immunity and viral survival (J. Tamerius et al. 2011). Viral survival during transmission is essential for successful transmission. Numerous studies have explored the effect of climate variables of survival of viruses in a lab setting and they find that for lipid enveloped viruses, lower temperatures and lower relative humidities have better virus survival. (Julian W. Tang 2009; J. Tamerius et al. 2011). The effect of temperature on influenza is reviewed by Irwin et al. (2011) who found that across air, water, feces and surfaces, temperature is a significant predictor of half life. Additionally, it is generally accepted that virus decay is faster at higher humidities but there is some confusion about whether relative humidity or absolute humidity is a better measure (Shaman and Kohn 2009; Marr et al. 2019). Relative humidity is the ratio of the concentration of water vapour in the air to the saturation concentration whereas absolute humidity describes the amount of water vapour in a volume of air (Marr et al. 2019). Absolute humidity is sometimes preferred because it may be a better predictor than temperature or relative humidity alone (Shaman and Kohn 2009). However, at higher temperatures, more water can be carried in the air (the absolute humidity can be higher) so the effect of higher absolute humidities may be confounded by the effects of higher temperatures, making it a potentially less mechanistic measure. Coronoviruses are also lipid-enveloped viruses so may react similarly to influenza in response to climate. There may be fewer studies but higher temperatures and higher relative humidities have been linked with faster breakdown on surfaces (Biryukov et al. 2020), in aerosols (Doremalen, Bushmaker, and Munster 2013) and in liquids (Chin et al. 2020). In terms of transmission, experimental evidence has found transmission of influenza between guinea pigs in controlled environments is higher in colder, less humid environments (Lowen, Mubareka, Steel, et al. 2007). As with viral stability, some suggest that absolute humidity may be more important than relative humidity or even temperature, for example the reanalysis of Lowen, Mubareka, Steel, et al. (2007) by Shaman and Kohn (2009).

Human behaviour may change with climate due to seasonality in school terms, in particular a long summer holiday. As some diseases mostly effect children a strong effect of school terms is expected. Work by Fine and Clarkson (1982) and Finkenstädt and Bryan T. Grenfell (2000) suggests that measles seasonality may be mediated by school terms. School terms are is also likely to play a role in influenza (J. Tamerius et al. 2011; Cauchemez, Valleron, et al. 2008) The weather is also likely to impact human contact rates or type of contacts. Graham and McCurdy (2004) found that in cooler and rainier conditions people

spent more time indoors. However there is yet to be work exploring how contact rate changes with time or attempts to separate the effects of climate on the virus from the effect of climate on human behaviours.

Seasonality has frequently been added to SEIR models by modifying the transmissability of the virus. Frequently this has been by assuming a sinusoidal shape of transmissibility with time. This has occurred since work on measles (Bolker and B. T. Grenfell 1993) and more recently (Neher et al. 2020) used it to explore the potential impact of seasonal forcing on COVID-19. Other work has allowed transmissability to vary linearly with temperature, humidity or both (Shi et al. 2020; Postnikov 2016) which will result in a similar shape if the climate variables change sinusoidally with time. One issue with these models is that parameterisation is necessary for both the intercept and the effect of climate or time on the transmission rate which has to be found from data. Brenner, Marwan, and Hoffmann (2017) overcomes this issue by using experimental data from Lowen, Mubareka, Steel, et al. (2007) reanalysed by Shaman and Kohn (2009) to find the vapor pressure, a measure of absolute humidity. A lookup table is used to estimate the transmissability for a given day (here the transmission rate is a percentage) given the data. This approach is limited by the data being from guinea pigs, which do not experience the same influenza symptoms as humans (Lowen, Mubareka, Tumpey, et al. 2006; Julian W. Tang 2009). Changes to contact rate and diet with climate were not included in the experiment but may effect transmissibility in the real world. These models are very valuable in improving our understanding of disease dynamics but are limited by not including the multiple components that may cause seasonality.

The majority of the potential causes for seasonality are mediated by climate factors. Although it is frequently possible to find evidence that certain traits are effected by climate-dependent traits, it can be difficult to determine whether this is the cause of seasonality so evidence linking climate and disease have not been able to clarify which factor is the cause. A combination of factors may well be contributing (Lofgren et al. 2007). Dushoff et al. (2004) suggest that only small fluctuations in transmission with climate are required for seasonal disease because the similarity between the period of climatic changes and the intrinsic period of the model (from non-climate dependent factors) allows seasonal disease peaks to occur due to dynamic resonance. Altizer et al. (2006) demonstrates how the amplitude of the seasonal forcing may also impact the patterns of disease. Due to this, understanding how the different factors may interact is really important. For example if one factor leads to better conditions for the virus in cold temperatures, and another factor leads to better conditions for the virus in warm temperatures, the amplitude of the seasonal forcing will be much lower than if they peak at the same value. Understanding the effect of individual components may be useful to understand how interventions can help. For example social behaviour can be changed but climate cannot, which could be used to manipulate seasonal forcing.

Additionally, this knowledge could help us to understand how climate change may impact disease.

Transmission is likely to be the major component effected by climate as human behaviour, virus properties and even some components of human physiology can effect it. Previous work where transmission varies with time tends to assume a sinusoidal shape without including mechanistic reasoning for this. The main aim of this project is to explore the effect of temperature on respiratory disease by making a modified SEIR model to mechanistically model some of the possible causes for temperature dependence of disease. This will be important in understanding the dynamics of COVID-19 if it were to become endemic. The aim is to explore the effects of viral decay rate and human contact rate. Contact rate can encompass the different aspects of human behaviour such a social behaviour and school terms. Virus decay determines how long a virus can survive in the environment. If the virus cannot survive for long it is less likely to be transmitted. This offers a mechanistic basis for differences in transmissibility due to climatic conditions. An important consideration is how human behaviour changes with temperature. Previous work (Mossong et al. 2008; Hoang et al. 2019) has explored distribution of contact rates in detail. However there is a lack of high quality data exploring temporal or climatic variation in contact rates. For this reason this work will explore different scenarios for the temperature at which contact rate is highest, described as the "mismatch" between the temperatures at which the virus survival and human contact rate are the highest. Temperature is focussed on as the climate variable of interest here because it has the most clear links with both behaviour and viral survival. The main aims of this project are:

- To model the potential effect of temperature dependence of human contacts and virus survival.
- To find which level of mismatch is most likely and what may determine this.
- Understand how mismatch effects severity of disease outbreaks.
- Understand how COVID-19 may behave as an endemic disease.

This will be achieved by:

- Making an SEIR model with a climate dependent transmissible term containing both virus survival and contact rates as climate-dependent terms.
- Parameterising this for influenza and simulating the model for many different countries with different levels of mismatch (different temperatures at which contact rate peaks).
- Use this data to determine what level of mismatch is the most likely for each country by comparing it to long term Influenza data
- Explore R_0 at different levels of mismatch.
- Using the best mismatch level for each country, estimate the seasonality of COVID-19 and the potential impact of shifting contact rates.

128

2 Methods

2.1 Model

The basis of the model is a simple SEIRS model. This is based on models in O. Bjørnstad (2018) and Keeling and Rohani (2007). SEIR models are a type of compartmental model which compartments for susceptible, exposed (but not yet infected), infected and recovered individuals. SEIRS models are modifications of SEIR models where individuals can lose immunity. They are described by a series of differential equations, where each differential equation describes the change in the number of individuals in that compartment.

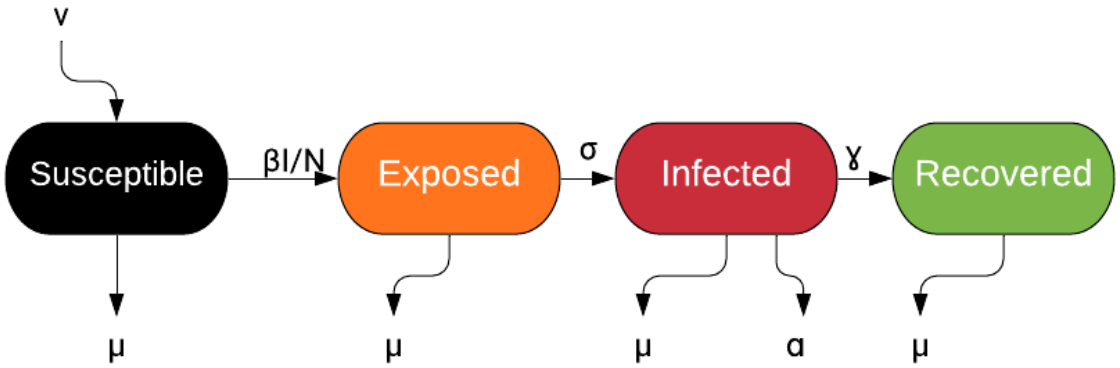


Figure 1: Visual Representation of SEIR model

$$\frac{dS}{dt} = \nu N - \frac{\beta IS}{N} - \mu S + fR \tag{1a}$$

$$\frac{dE}{dt} = \frac{\beta IS}{N} - (\sigma + \mu)E \tag{1b}$$

$$\frac{dI}{dt} = \sigma E - (\alpha + \mu + \gamma)I \tag{1c}$$

$$\frac{dR}{dt} = \gamma I - \mu R - fR \tag{1d}$$

α, the rate of disease induced mortality, is difficult to parameterise so was replaced by the case fatality rate p as demonstrated by Keeling and Rohani (2007), slightly changing the equation for the change in numbers of infectious individuals (2). It is worth noting that disease induced mortality was included

(unusual for a typical influenza model) due to the relatively high death rate of COVID-19.

$$\alpha = \frac{p}{1-p}(\mu + \gamma) \tag{2a}$$

$$\frac{dI}{dt} = \sigma E - (\mu + \gamma)\left(\frac{1}{1-p}\right) I \tag{2b}$$

Parameters are described in Table 1.

Parameter	Meaning
S	Number of Susceptible Individuals
E	Number of Exposed Individuals
I	Number of Infected Individuals
R	Number of Recovered Individuals
N	Total Number of Individuals
β	transmissibility (number of infected individuals per infected individual per day)
f	rate of loss of immunity
ν	natural per capita birth rate
σ	rate of movement from E to I (reciprocal of latent period)
μ	natural per capita death rate
γ	recovery rate
α	disease induced mortality rate
p	case fatality rate

Table 1: Basic Parameter meanings for SEIR model. Parameters are described fully (with units and sources) in Table 2 of Supplementary information

The transmissability (β) is a climate-dependent term. This model was obtained by modifying the model in Valle, Hyman, and Chitnis (2013) to add climate dependence. Transmissability is split into the number of contacts per unit time (c_r) and a probability of transmission given a contact between a susceptible and infected individual. Within this probability there is a climate-dependent term and a scaling term (ϵ). The climate-dependent term is based on the expected amount of virus in the environment given a specific duration of contact compared the the amount of virus that would be in the environment after the time needed for infection to occur (3). It is based on the assumption that the change in the amount of virus in the environment at time t is $a - bV$ where a is a constant shedding rate, b is the rate of decay of the virus and V is the amount of virus. By integrating this, we can find the the amount of virus at a time equals to the average duration of contact ($V = \frac{a}{b}(1 - e^{-bd})$). This was divided by itself plus ah, which represents how much virus would be shed in the time required for successful infection to occur (h is the expected contact time needed for infection). This cancels to become the fraction in equation 3. This

can represent the amount of virus in the environment at a given time considering the rate of decay of the virus as a probability of infection.

The scaling term covers the multitude of other reasons why β is not the same as c_r such as the fact that risk from contacts vary depending on reasons other than duration and climate, such as type of contact.

β = number contacts per unit time $\times$ probability of disease transmission per contact

$$\beta = c_r \times \epsilon \times \frac{\frac{1}{b}(1 - e^{-bd})}{\frac{1}{b}(1 - e^{-bd}) + h} \tag{3}$$

Climate dependence is included in the model through variation in contact rate (c_r) and decay rate (b) of the virus in the environment. In early models, the average duration of contact (d) was also climate-dependent however the effect of it was minimal.

Contact rate as a function of climate was assumed to be based on a normal distribution scaled up so the maximum is the maximum contact rate (4). This was calculated for each country such that each country was assumed to have the same maximum contact rate. This was done because this work focuses on seasonal differences rather than differences between countries. The climate where this peak occurs (i.e. the mean of the normal distribution) was not known so 5 values at regular intervals along the climate range were modelled. This was converted to the levels of "mismatch" where mismatch is the difference between the minimum climate (climate at which survival of the of the virus is highest or viral decay is lowest) and the climate at which contact peaks, divided by the maximum possible difference for that climate range (see Figure 2a for visualisation of this). A mismatch of 0 means the the contact rate peaks at the lower range of the climate variable (which is also where the virus survival is highest and virus decay is lowest). The value of s, the standard deviation of the normal distribution, is determined such that 95% of the contact rate of the area under the curve occurs between the minimum and maximum climate. This is based on the fact that $Z = \frac{x-\mu}{s}$ and the z-score required to include 95% is 1.96, where x is the raw score for that z-score. In this case, x is the upper or lower climate bound (max $\_C$ or min $\_C$). Rearranging this gives $s = \frac{x-\mu}{1.96}$. The choice of lower versus upper bound is determined by whether the climate of the peak contact rate ($C\_\max\_c_r$) happens closer to the upper or lower bound of the climate. When the maximum contact rate happens closer to the lower bound of climate, the maximum is used (4b), otherwise the miminum is used (4c).

$$c_r(C) = \max\_c_r \times e^{\frac{-(C - C\_\max\_c_r)^2}{2s^2}} \tag{4a}$$

$$s = \frac{(\max\_C - C\_\max\_c_r)}{1.96} \tag{4b}$$

$$s = \frac{-(\min\_C - C\_\max\_c_r)}{1.96} \tag{4c}$$

Experimental data were obtained for viability of influenza over time in different climatic conditions from Harper (1961). For temperature this was at 7.5, 22.5 and 32 degrees celcius at up to seven timepoints (unless virus had decayed past a detectable level by that point). Where the given temperature or humidity were a range of values (e.g. 21-23 degrees),the mean of this range was used. Exponential decay models (5) were fitted to each set of time series of viabilities using nlsLM in R. b was bounded between 0 and 5000 per day; $v0$ was bounded between 0 and 100%. The Clausius–Clapeyron relation was used to find the absolute humidities for each combination of temperature and relative humidities (as described by Shaman and Kohn (2009)) The decay rates of the virus with temperature, humidity and absolute humidity were fitted to exponential models (6) using nlsLM using the climate values and the b values found by the earlier model. This obtained an estimate for the rate of decay at time 0 (b_0) and the rate of growth of the rate of decay (g) with the climate variable. This resulted in equations which could find b for different climates (for temperature, relative humidity and absolute humidity seperately). Experimental data for SARS-COV-2 and climate were also found (Biryukov et al. 2020). This data contained the half life for eight different combinations of temperature and relative humidity conditions. This was converted into b by rearranging Equation 5 and setting t to the half life and v to equal $\frac{v_0}{2}$. nlsLM was used to fit Equation 6 to the climate and b values.

$$v = v_0 e^{-bt} \tag{5}$$

$$b = b_0 e^{gT} \tag{6}$$

2.2 Paramaterisation

A literature search was used to estimate parameter values for COVID-19 and Influenza (Table 3 in Supplementary Information).

2.3 Data Sources and Sorting

Climate and flu data were obtained from Deyle et al. (2016). This contained relative humidity, absolute humidity, temperature and positive flu results per capita for 79 countries. One country was removed from this because the relative humidities were clearly incorrect. The time was given in the year, month of the year and day of the month (but values were only present up to every week). To simplify analysis this

was converted into week of the year. Only years with 52 or 53 weeks were included so 11090 rows were removed (about 3% of rows). This removed years where only part of the year was known because future averaging steps may have been biased by this. Then the average weekly flu and climate values were found for each country. For each country the minimum, maximum and week of the maximum climate variables were found. This was used to create as a sinusoidal function for the climate for each country (7). This was necessary because weekly time intervals were not sufficient for the integration in later steps.

$$C(t) = \frac{\max\_C - \min\_C}{2} \times \cos\left(\frac{2\pi}{365} \times (t - week\_\max\_C)\right) + \frac{\max\_C + \min\_C}{2} \tag{7}$$

Population data was found for these countries from The World Bank (2019). One country was missing population data so was removed from the analysis (French Guinea). Latitude and longitude data was obtained from Google (2019) for each of the remaining countries.

2.4 Analytical Sensitivity Analysis

An explicit equation for R_0 was found using an equation from Bjornstad, Finkenstadt, and Greenfell (2002) and modifying it with the previous equation for β (3) and converting from α to p (8). SymPy in Python (Meurer et al. 2017) was used to find the partial derivative of R_0 with respect to each parameter. Then this equation was divided by the R_0 equation to find an equation for the change in the parameter as a proportion of R_0 ($\frac{\frac{\partial R_0}{\partial P}}{R_0}$). For different parameter combinations for each parameter (P) the value of $\frac{\frac{\partial R_0}{\partial P}}{R_0}$ was found by substituting different parameter combinations into this equation. q and c_r were assumed to be constant. The temperature range for this was the mean coldest week and the mean hottest week. In addition the temperature where contact rate was highest was set to be the minimum of the temperature range (same value when maximum is used). Then the current temperature was set to the midpoint of the temperature range. Each parameter (including climate dependent parameters) were varied from 50% to 150% of the best estimate of the parameter, at 100 different regularly spaced points. $\frac{\frac{\partial R_0}{\partial P}}{R_0}$ was found for every parameter at each parameter combination (i.e the number of intervals timed by the number of parameters being varied). This was found as both a value and a rank of the relative importance. Over the different parameters, h, ϵ and μ were most frequently the most important parameters. For this reason, these were varied during the integration step. To further test robustness, the analytical sensitivity analysis was repeated with the temperature was set the the lower,middle and upper bound of the temperature range where c_r and q were not varied. The results at these three values were very similar. Only the relative importance of d changed. However as d was always one of the three least important parameters, this

difference could be ignored and a the temperature was kept at the midpoint of the range (but c_r and d were still varied as described above).

$$R_0 = \frac{\sigma}{\sigma + \mu} \times \frac{c_r \times \epsilon \times \frac{\frac{1}{b} \times (1 - e^{-bd})}{\frac{1}{b} \times (1 - e^{-bd}) + h}}{(\mu + \gamma)\frac{1}{1-p}} \tag{8}$$

2.5 Model Simulations

Integration was performed for each country for 5 different levels of mismatch. For each country, the climate range (used for finding the value of contact rate at different climates) was set to the lowest and highest mean weekly temperatures for that country and the population size of the model was set to the 2019 population of the country. To investigate different levels of mismatch, the integration was repeated with the maximum contact rate occurring at different climates, evenly spaced between the higher and lower climate values. ode in deSolve in R (with lsoda) Soetaert, Petzoldt, and Setzer (2010) was used to integrate the SEIR model over 11 years (intervals of one day) to estimate the number of susceptible, exposed, infected and recovered individuals at each time point. The model started with infected (0.1% and susceptible 99.9%) individuals. The population size from 2019 was used as the starting population for each country. The simulation began at the calendar start of the first year. The values for the first year were removed to allow the model to burn in and reduce the impact of the starting time and starting values. At each time point the value of the climate variable was found (7). From this the contact rate and decay rate were found (Equations 6 and 4). These were used to find the transmissibility (β), part of the SEIR model. R_0 based on these values was found (3).

This process was repeated for temperature, relative humidity and absolute humidity.

2.5.1 Sensitivity Analysis in Integration

Many of the parameters used were estimates so some of the parameters were varied and the integration repeated for different combinations. The parameters varied were the three most important parameters (h,μ and ϵ) in the analytical sensitivity analysis and f which could not be included in the analytical sensitivity analysis because it does not contribute to R_0. Latin hypercube sampling was used to find 160 combinations of these parameters. "lhs" in R was used to find a values between 0 and 1 given the number of repeats and number of parameters. Then "qunif" was used to find the value at which this quantile occurred from a uniform distribution between 0.5 and 1.5 times the parameter estimate. This resulted in a different parameter combination for each repeat.

Additionally the integration was repeated with different values for the standard deviation of c_r. It was multiplied by X values between 0.75 and 5 using just the parameter estimates.

2.6 Analysis of Integration Results and Data

The mean number of infected individuals and R_0 for each week of the year was found for each integration. The mean number of positive flu tests per capita each week was found from Deyle et al. (2016) for each country. The correlation (r) between the number of infected individuals and the number of positive flu tests per capita was found for each level of mismatch, combination of parameters and country to obtain an approximate value for the similarity between data and model.

2.7 COVID-19

For each country, the best mismatch level for influenza was found from the mismatch at which the correlation was highest. The integration was reran using COVID-19 parameters (no sensitivity analysis) with the best mismatch level for each country for influenza used as the estimated mismatch level for COVID-19. This was then repeated with different mismatch shifts, where the mismatch was 0.25,0.5 or 0.75 more than predicted by the influenza data.

2.8 Statistical Analysis

Countries were separated by latitude into temperate and tropics based on whether the absolute value of their latitude was greater than 23.5. Data analysis was carried out in R and plotting was done using ggplot2 (Wickham 2016).

3 Results

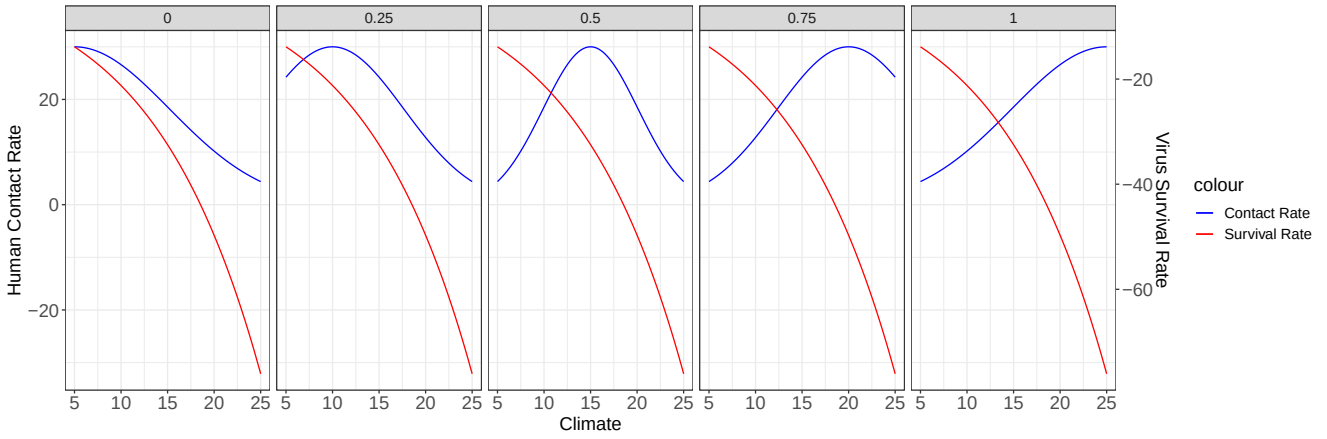
3.1 Models and Data

According to the data, the temperature range was lower in tropical regions (mean: 5.35 degrees, standard error: 0.64 degrees, n=29) than temperate regions (mean: 20.70 degrees, standard error: 1.12 degrees, n=48). The virus survival was based on these temperatures and the virus survival range was also lower in tropical regions (mean difference of 38.12/day between lowest and highest viral decay, standard error: 4.89/day, n=29) than temperate regions (mean difference of 52.88/day,standard error: 3.44/day, n=48).

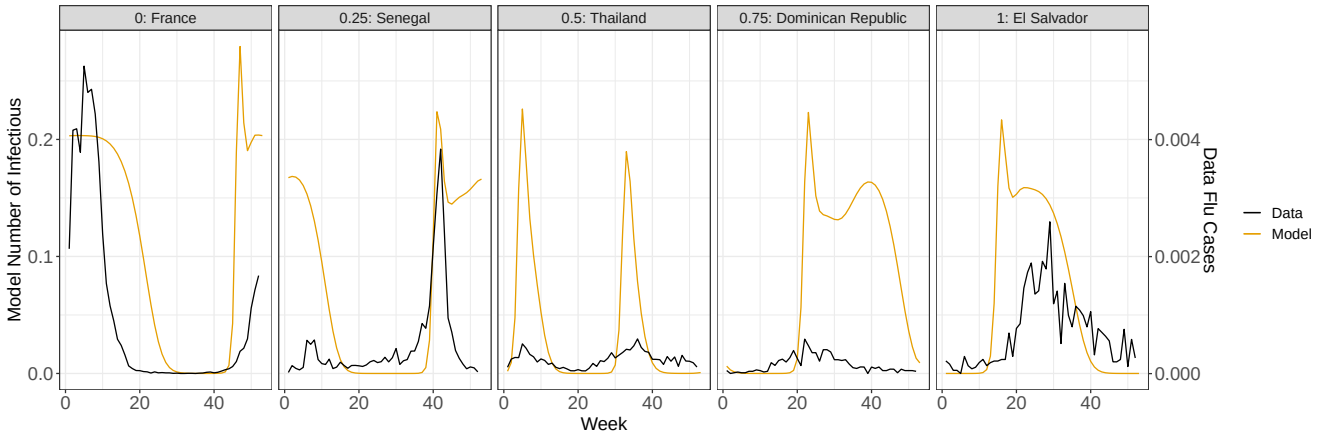
The mismatch represents the climate at which contact rate is highest (Figure 2a). Virus Survival rate is always highest when the climate variable is lowest. A mismatch of 0 means that the survival and the contact rate are highest at the same climate (so the contact rate peaks at low values). A mismatch of 1 means that the survival and the contact rate are highest at very different values (so the contact rate peaks at high values). The correlation varied between mismatches meaning that the climate at which contact rate peaked played a key role in determining how similar the model was to the data. The mismatch level with the highest correlation varied between countries implying that in different countries, contact rate has a different relationship with climate (Figure 2).

3.2 Mismatch and Temperature

In the temperate regions, all countries were most similar to models when the contact rate peaked at low temperatures (46 countries has a mismatch of 0, 2 countries had a mismatch of 0.25). In tropical regions, the temperature at which contact rate was most likely to peak varied between countries. In 11 countries, contact rate was highest at low temperatures (mismatch of 0 or 0.25), in 4 countries, contact rate was highest at intermediate temperatures (mismatch of 0.5) and in 13 countries, contact rate was highest in high temperatures (mismatch of 0.75 or 1). In temperate regions, the models were better (i.e. the correlation between the model and data was highest) at low mismatches and worst at high mismatches. In tropicsal regions there was not a consistent difference in correlation with different mismatch levels (Figure 3). Using absolute humidity as the climate variable had the same effect as temperature (Figure 8 in Supplementary information). Using relative humidity as the climate variable led to a weaker effect in the opposite direction and there was a less pronounced difference between temperate and tropical regions (8 in Supplementary information). Temperature was focused on for further analysis as it may be linked more to behaviour. Changing the standard deviation for the contact range does not change the overall relationships between region and mismatch but the effect becomes less strong (Supplementary Information



(a) Mismatch Visualisation



(b) Data and best model for selected countries

Figure 2: Visualisation for different levels of mismatch. Top panel is the model contact rates and survival rate with temperature for different levels of mismatch. Bottom panel is the number of infectious individuals according to the best model for that country and the mean number of positive flu tests per capita (black). Note: the model and results have different axes because the models and data show very different results, due to both the accuracy of the model and the underestimations of number of flu cases.

Figure 9).

3.3 Sensitivity Analysis

Analytical sensitivity analysis found that the most important parameters were h,μ and ϵ(Figure 4). Further, the results were robust to the variation in parameter values. Figure 3 shows a mean over 100 combinations of parameters. The results are qualitatively the same when just the estimates are used.

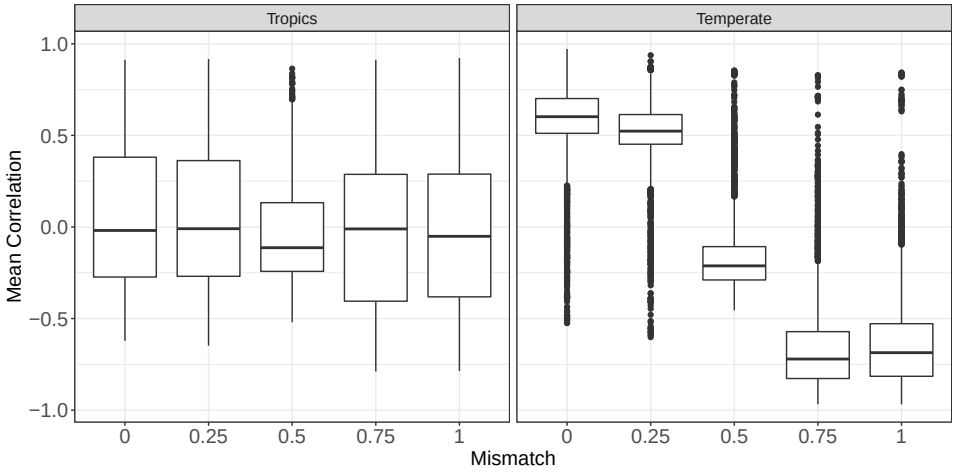


Figure 3: Box and whisker plot of the correlation between data and the model for 5 different mismatches split by whether the country is tropic or temperate. The data is the weekly mean per capita positive influenza tests and the model is the weekly mean proportion of infected individuals. A mismatch of 0 has a higher correlation. n=29 in tropics, 48 in temperate region

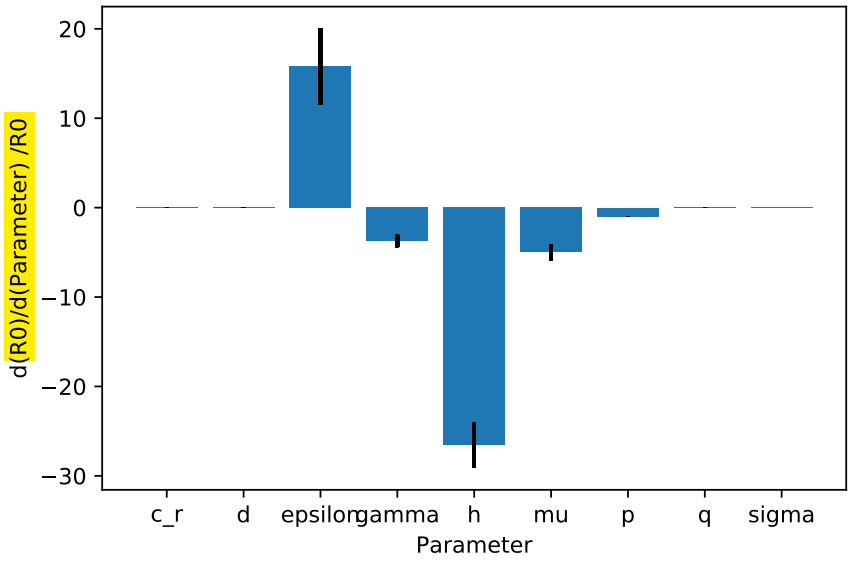
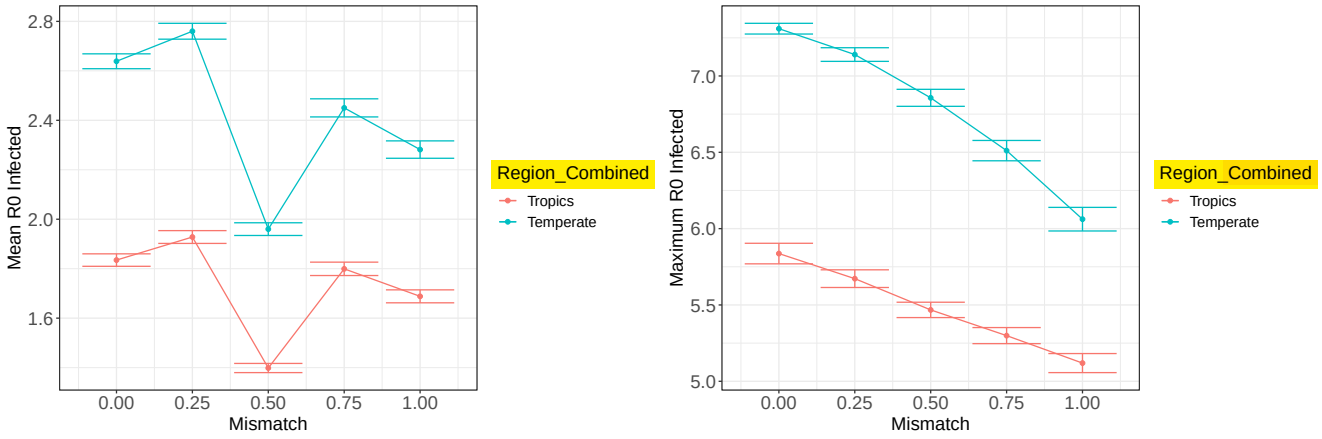


Figure 4: Results of Sensitivity analysis for each paramater in equation 8. Each parameter was varied between 0.5 and 1.5 times the estimated value with 100 intervals and the change in R_0 partial differential of each parameter divided by the R_0 at that point was calculated for each parameter. The mean value for each parameter was plotted with the standard deviation as error bars

3.4 Mismatch and Disease Prevalence

In general, tropical countries had lower disease for all levels of mismatch. Models where the contact rate was highest at intermediate temperatures had the lowest mean R_0 and models where the contact rate was

highest at high temperatures had the lowest maximum R_0 (Figure 5).



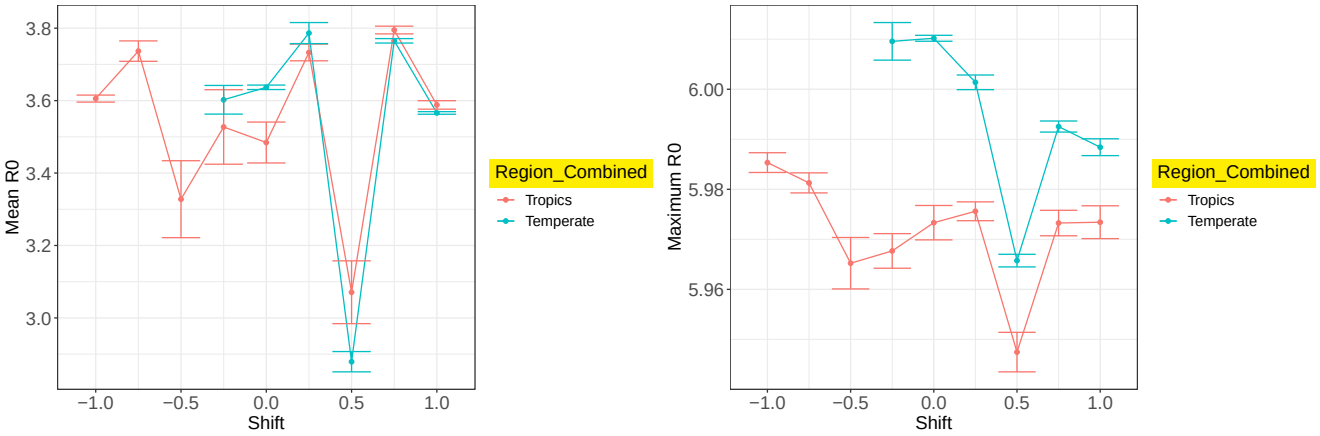
(a) Mean R_0

(b) Maximum R_0

Figure 5: Yearly values of maximum and mean R_0 at different levels of mismatch. In general, values for tropical regions were lower than temperate regions at all mismatches. Mismatches of 0.5 led to the lowest mean values

3.5 COVID-19

The COVID-19 model (with mismatch determined by the flu model) suggests one peak of R_0 in the winter of temperate regions although the exact timing does vary (mean is week 9 in northern temperate regions and 29 in southern temperate regions. standard errors are 1.4 and 2.06 weeks respectively). The tropics do not peak at a consistent time (mean is week 29 is and standard error is 2.65 weeks) and often have multiple peaks (Figure?? in Supplementary Information). Shifting the level of mismatch changes the mean and maximum R_0. A shift of 0.5 leads to the lowest R_0 in all but 2 of the 48 temperate regions)



(a) Mean R_0

(b) Maximum R_0

Figure 6: Yearly values of maximum and mean R_0 at different shifted mismatches. A shift of $+0.5$ decreases the mean and maximum R_0 although the maximum R_0 is not very variable. At a shift of 0 (i.e. mismatch as predicted by the influenza model), the tropics are predicted to have less disease than temperate regions.

4 Discussion

4.1 Temperature Findings

This work suggests that in temperate regions, contact rate is highest in cooler temperates (i.e. in Winter). This could be due to overall increased contact rate at this time. This could be related to school terms as winter holidays tend to be shorter than summer holidays and it is clear that schools are a major source of contacts for children. Cauchemez, Valleron, et al. (2008) found that school holidays prevent almost a fifth of influenza cases in France. This study only looked at epidemic periods so was not able to pick up the potential impact of a longer summer holiday on potential for an epidemic. It is worth noting that they did not find a significant difference between Christmas holidays and other holidays which means the effect of a religious festival on contact rates may not be important epidemiologically. It is also possible that the result highlights the role of higher risk or closer contacts which may be more common in winter because more contacts are indoors. For example, Graham and McCurdy (2004) found that more contacts are indoors in cooler and rainy conditions. Of course if this were the case, viral survival at the outdoor temperature may be less relevant. The large disparity between temperate and tropical regions is of interest. In the tropics, there is a strong synchrony between contact rates and virus survival, indicated by a mismatch of 0. In tropical regions, this synchrony is not present in most countries. The temperature range changes with latitude. The tropics have a lower temperature range. In the model, this temperature range effects the difference between that maximum and minimum decay rate. The difference between the maximum and minimum contact rates in the model are assumed to be consistent between countries but in reality, in the tropics there may be less variation in contact rates than in temperate regions due to the decreased temperature range. If this assumption is altered, a more realistic model for contact rate may be obtained. There are many different possibilities for how temperature might effect contact rate both between and within countries. Models exploring these can produce insights but for these models to be accurate, data for how contact rate changes with temperature are needed. The viral survival range does vary with country in this model. As temperature range is lower in the tropics, range in viral survival is also lower in the tropics. This may lead to weaker effects of seasonal forcing in the tropics. Dushoff et al. (2004) predicts that if the intrinsic period of the model is the close to the period of seasonal forcing, resonance between these factors amplifies the effects of seasonal forcing, even if the effect of seasonal forcing is itself small.

In the tropics, there may be differences not included in the model effecting the disparity between countries. Rain was not included in this model but a rainy season is a common feature of tropical countries. It is possible that the timing of the rainy season could be determining the timing of contact

peaks. Bi et al. (2020) find a link between respiratory virus outbreaks and rainy seasons in the tropics, although this is not the case for every country.

At 0 mismatch, both contact rate and and virus survival are highest in low temperatures. This leads to one clear peak in R_0 when temperature is lowest. At 1 mismatch, contact rate peaks in high temperatures while virus survival still peaks at low temperatures. This leads to a peak R_0 in high temperatures. This peak is lower than the peak for 0 mismatch due to the asynchrony. Additionally the R_0 fluctuates over this period. At a mismatch of 0.5 there are two peaks in R_0 and I over the course of a year (in between the amplitude of peaks for 0 and 1 mismatch). These shapes mean that mean R_0 is lowest at an intermediate mismatch and maximum R_0 being lowest at intermediate mismatches. Countries with a mismatch of 0 peak in cool temperatures and may have the most disease. This is consistent with winter peaks in temperate regions, which is very common in flu (J. Tamerius et al. 2011). Additionally, across all levels of mismatch, both maximum and mean R_0 values were lower in the tropics.

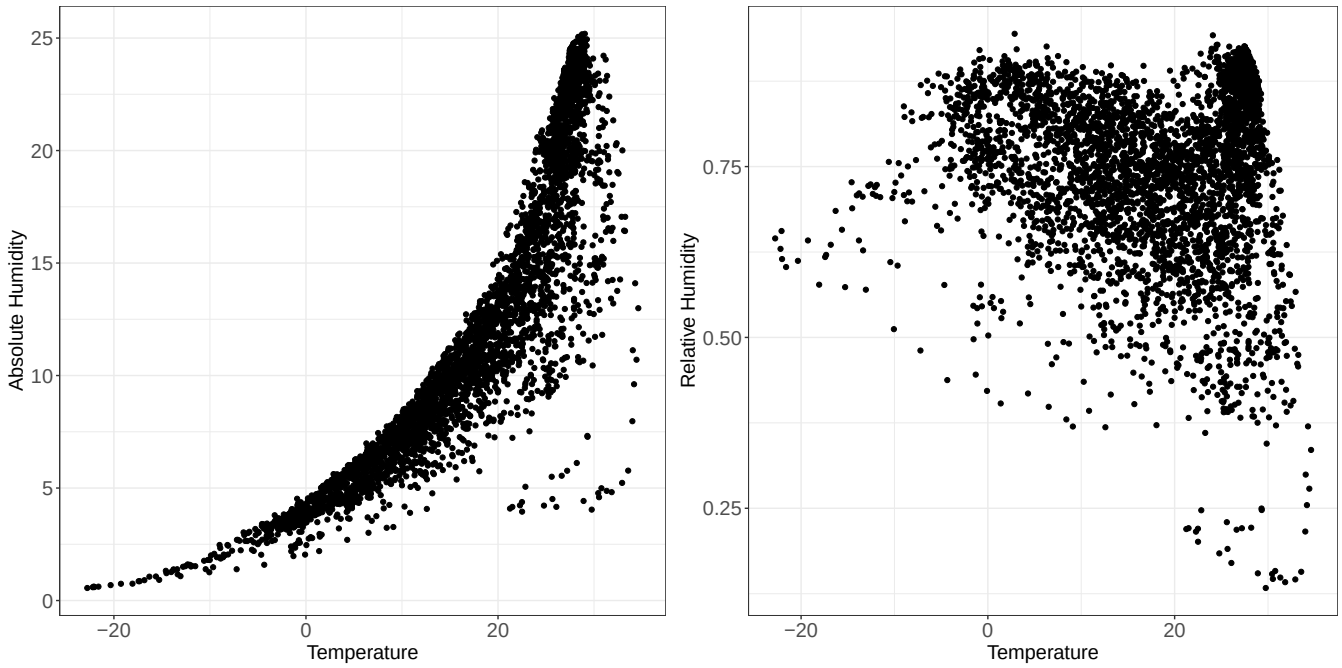
4.2 Different Climate Variables

In this study, the results for absolute humidity were very similar to the results for temperature. Absolute humidity and temperature are linked (Figure 7) and latitude has a strong effect on both temperature and absolute humidity range. The results for relative humidity showed higher levels of mismatch in temperate regions (the opposite relationship to absolute humidity and temperature). The best mismatch in temperate regions was not as consistent. The relationship between temperature and relative humidity may be the cause for this (Figure 7). The results for both relative humidity and absolute humidity were not any more insightful than the results for temperature.

There may be a complex, non-linear relationship between temperature and humidity (Deyle et al. 2016). Combining multiple climatic variables into the virus survival component of this model could be as useful way of understanding this relationship more. Additionally temperature may not be the best climate variable to determine contact rate in the tropics. Rainfall could be incorporated into the contact rate component. Ultimately, different possible predictors and distributions could be incorporated and model selection could be used to find which version of the model is most similar to data.

4.3 COVID-19

As for COVID-19, the model predicts a winter peak in temperate regions. This is at odds with what has been seen in the world, where countries have been affected in all seasons (World Health Organization (WHO) 2020a). This is likely because when there are high numbers of susceptible individuals, the suscep-



(a) Absolute Humidity

(b) Relative Humidity

Figure 7: Mean weekly temperatures and humidities from Deyle et al. (2016) dataset. Absolute humidity seems to grow exponentially with temperature, relative humidity is more mixed.

tible supply exerts a stronger effect than seasonal forcing (Baker et al. 2020). That said, some studies have found links between climate variables and severity of COVID-19 (Wang et al. 2020; Luo et al. 2020; Sajadi et al. 2020; Bannister-Tyrrell et al. 2020; Rahman et al. 2020; Oliveiros et al. 2020; Chen et al. 2020; Ma et al. 2020; Poirier et al. 2020) outbreaks but these are very vulnerable to the effects of confounding factors and poor data. While seasonal forcing does not appear to have had a strong effect in early stages of the outbreak, in the longer term, climate may determine when or where outbreaks occur, so understanding the causes for this is likely to be important.

Figure 6 demonstrates that shifting mismatch may effect the severity of outbreaks. Shifting the timing of contact rates could decrease the impact of COVID-19 R_0 in the future. In temperate regions, this would be by decreasing contact rates during cold temperatures but allowing them to increase during warmer temperatures. This suggests that reducing winter contacts could reduce the impact of COVID-19 in the long term. Lengthening winter holiday and shortening other holidays could have this effect. Additionally large gatherings could be limited in Winter. There are obviously issues with feasibility of these proposals. In the tropics, shift still appears to decrease the contact rate but for this to be helpful a more detailed understanding of how contact rate varies with climate is necessary. The existing timing of peak contact rate would need to be identified for each country prior to interventions.

4.4 Limitations and the Future

Due to computing constraints only four different parameters were used as sensitivity analysis when the simulations were carried out. These were the three parameters that had the greatest impact on an analytical sensitivity analysis of R_0 and one parameter that could not be included in R_0. Although the remaining parameters had a very small effect on R_0 during the sensitivity analysis, its possible that interactions between parameters would have been different if they were altered in the SEIR model.

It is important to note that neither the influenza data nor the model were accurate representations of the exact numbers of influenza cases. The models were extremely simple. Adding complexity to the models such as age-dependent contact rates, spatial considerations, variation in mortality and birth rates between countries would improve this.

As J. Tamerius et al. (2011) notes it is imperative to know whether disease transmission is mostly indoors versus outdoors because this is likely to determine the impact of climate. More research is needed into contact patterns with climate and time. This data could be very valuable in mitigating respiratory viruses. If this data could be incorporated into this model, the results would be very valuable. Additionally, with more time, experimental data for the viruses and climatic conditions could have been combined. At present only one study was used for influenza and one for COVID-19. A more robust approach would have been a meta-analysis of all available experimental data. Models are by their nature a simplification of real life. One potential cause for seasonality that was not included was human physiology and immunity. The immune system may vary over they year, potentially mediated by vitamin D or by changes in food availability (J. Tamerius et al. 2011). Additionally, in dryer and potentially cooler environments blood flow and mucociliary clearance in may lower (J. Tamerius et al. 2011) Additionally, temperature may effect the shedding rate (Lowen, Mubareka, Steel, et al. 2007). These factors could not be included because there was not sufficient quantitative information on these effects. More research into these areas is clearly needed.

Understanding these dynamics will particularly useful in predicting the effects of global warming on respiratory viruses. Understanding how contact rate and virus survival interact may help us to determine which areas are more at risk from disease outbreaks and put measures in place to detect and prevent these in the most high risk areas.

This work suggests a new model for transmissability as a function of temperature and highlights the need for more data, particularly the importance of climate-dependent contact data.

5 Conclusion

Previous work acknowledges the importance of seasonality but this is the first work to attempt to untangle the effects of human behaviour and virus physiology. This model provides a good starting point for the importance of contact rate and virus survival. The level of mismatch plays a key role in determining how close the model is to data so this work highlights the need for more research into climate and time-dependent contact rates. This work was able to estimate the temperature at which contact rate peaks for each country by comparing model results to data. The result that in temperate regions contact rate is consistently highest at low temperatures is particularly interesting and could be used to reduce the impact of respiratory viruses. In the tropics, other climatic variables such as precipitation may be more important than temperature in determining contact rates. This work offers plenty of avenues for future work.

References

Altizer, Sonia et al. (2006). "Seasonality and the dynamics of infectious diseases". In: *Ecology Letters* 9.4, pp. 467–484. ISSN: 1461023X. DOI: 10.1111/j.1461-0248.2005.00879.x.

Baker, Rachel E et al. (2020). "Susceptible supply limits the role of climate in the COVID-19 pandemic". In: *Science* 369, pp. 315–319.

Bannister-Tyrrell, Melanie et al. (2020). "Preliminary evidence that higher temperatures are associated with lower incidence of COVID-19, for cases reported globally up to 29th February 2020". In: *medRxiv preprint*.

Baskerville, Edward B. and Sarah Cobey (2017). "Does influenza drive absolute humidity?" In: *Proceedings of the National Academy of Sciences of the United States of America* 114.12, E2270–E2271. ISSN: 10916490. DOI: 10.1073/pnas.1700369114.

Bi, Qifang et al. (2020). "Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study". In: *The Lancet Infectious Diseases* 3099.20, pp. 1–9. ISSN: 14744457. DOI: 10.1016/S1473-3099(20)30287-5.

Biryukov, Jennifer et al. (2020). "Increasing temperature and relative humidity accelerates inactivation of SARS-COV-2 on surfaces". In: *mSphere* 5.4, pp. 1–9. ISSN: 2379-5042. DOI: 10.1128/mSphere.00441-20.

Bjornstad, O.N., Barbel F. Finkenstadt, and Bryan T. Greenfell (2002). "Dynamics of Measles Epidemics : Estimating Scaling of". In: *Ecological Monographs* 72.2, pp. 169–184.

Bjørnstad, Ottar (2018). *Epidemics: Models and data using*, p. 318. ISBN: 9783319974866. DOI: https://doi.org/10.1007/978-3-319-97487-3.

Bjørnstad, Ottar N. and Cecile Viboud (2016). "Timing and periodicity of influenza epidemics". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.46, pp. 12899–12901. ISSN: 10916490. DOI: 10.1073/pnas.1616052113.

Bolker, B. M. and B. T. Grenfell (1993). "Chaos and biological complexity in measles dynamics". In: *Proceedings of the Royal Society B: Biological Sciences* 251.1330, pp. 75–81. ISSN: 14712970. DOI: 10.1098/rspb.1993.0011.

Brenner, Frank, Norbert Marwan, and Peter Hoffmann (2017). "Climate impact on spreading of airborne infectious diseases: Complex network based modeling of climate influences on influenza like illnesses". In: *European Physical Journal: Special Topics* 226.9, pp. 1845–1856. ISSN: 19516401. DOI: 10.1140/epjst/e2017-70028-2.

Cauchemez, Simon, F. Carrat, et al. (2004). "A Bayesian MCMC approach to study transmission of influenza: Application to household longitudinal data". In: *Statistics in Medicine* 23.22, pp. 3469–3487. ISSN: 02776715. DOI: 10.1002/sim.1912.

Cauchemez, Simon, Alain Jacques Valleron, et al. (2008). "Estimating the impact of school closure on influenza transmission from Sentinel data". In: *Nature* 452.7188, pp. 750–754. ISSN: 14764687. DOI: 10.1038/nature06732.

CDC (2020a). *Past Seasons Estimated Influenza Disease Burden*. URL: https://www.cdc.gov/flu/about/burden/past-seasons.html.

— (2020b). *Public Health Guidance for Community-Related Exposure*. URL: https://www.cdc.gov/coronavirus/2019-ncov/php/public-health-recommendations.html.

Chen, Biqing et al. (2020). "Roles of meteorological conditions in COVID-19 transmission on a worldwide scale". In: *medRxiv preprint*.

Chin, Alex W.H. et al. (2020). "Stability of SARS-CoV-2 in different environmental conditions". In: *medRxiv preprint*.

Deyle, Ethan R. et al. (2016). "Global environmental drivers of influenza". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.46, pp. 13081–13086. ISSN: 10916490. DOI: 10.1073/pnas.1607747113.

Doremalen, N. van, T. Bushmaker, and V. J. Munster (2013). "Stability of middle east respiratory syndrome coronavirus (MERS-CoV) under different environmental conditions". In: *Eurosurveillance* 18.38, pp. 1–4. ISSN: 15607917. DOI: 10.2807/1560-7917.ES2013.18.38.20590.

Dowell, Scott F. and Mei Shang Ho (2004). "Seasonality of infectious diseases and severe acute respiratory syndrome - What we don't know can hurt us". In: *Lancet Infectious Diseases* 4.11, pp. 704–708. ISSN: 14733099. DOI: 10.1016/S1473-3099(04)01177-6.

Dushoff, Jonathan et al. (2004). "Dynamical resonance can account for seasonality of influenza epidemics". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.48, pp. 16915–16916. ISSN: 00278424. DOI: 10.1073/pnas.0407293101.

Ferguson, Neil M. et al. (2005). "Strategies for containing an emerging influenza pandemic in Southeast Asia". In: *Nature* 437.7056, pp. 209–214. ISSN: 00280836. DOI: 10.1038/nature04017.

Fine, Paul E.M. and Jacqueline A. Clarkson (1982). "Measles in England and Wales - I: An analysis of factors underlying seasonal patterns". In: *International Journal of Epidemiology* 11.1, pp. 5–14. ISSN: 03005771. DOI: 10.1093/ije/11.1.5.

Finkenstädt, Bärbel F. and Bryan T. Grenfell (2000). "Time series modelling of childhood diseases: A dynamical systems approach". In: *Journal of the Royal Statistical Society. Series C: Applied Statistics* 49.2, pp. 187–205. ISSN: 00359254. DOI: 10.1111/1467-9876.00187.

Google (2019). *countries*. URL: https://developers.google.com/public-data/docs/canonical/countries_csv.

Graham, Stephen E. and Thomas McCurdy (2004). "Developing meaningful cohorts for human exposure models". In: *Journal of Exposure Analysis and Environmental Epidemiology* 14.1, pp. 23–43. ISSN: 10534245. DOI: 10.1038/sj.jea.7500293.

Harper, G. J. (1961). "Airborne micro-organisms: Survival tests with four viruses". In: *Journal of Hygiene* 59.4, pp. 479–486. ISSN: 00221724. DOI: 10.1017/S0022172400039176.

Hoang, Thang et al. (2019). "A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections". In: *Epidemiology* 30.5, pp. 723–736. ISSN: 15315487. DOI: 10.1097/EDE.0000000000001047.

Irwin, C. K. et al. (2011). "Using the systematic review methodology to evaluate factors that influence the persistence of influenza virus in environmental matrices". In: *Applied and Environmental Microbiology* 77.3, pp. 1049–1060. ISSN: 00992240. DOI: 10.1128/AEM.02733-09.

Keeling, Matt J. and Pejman Rohani (2007). *Modeling Infectious Diseases in Humans and Animals*. Princeton, New Jersey: Princeton University Press.

Kissler, Stephen et al. (2020). "Projecting the transmission dynamics of SARS-C0V-2 through the post-pandemic period". In: *Science*, pp. 1–13.

Lauer, Stephen A. et al. (2020). "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application". In: *Annals of Internal Medicine* 2019. ISSN: 0003-4819. DOI: 10.7326/m20-0504.

Lofgren, E. et al. (2007). "Influenza Seasonality: Underlying Causes and Modeling Theories". In: *Journal of Virology* 81.11, pp. 5429–5436. ISSN: 0022-538X. DOI: 10.1128/jvi.01680-06.

Longini, Ira M. et al. (2005). "Containing pandemic influenza at the source". In: *Science* 309.5737, pp. 1083–1087. ISSN: 00368075. DOI: 10.1126/science.1115717.

Lowen, Anice C., Samira Mubareka, John Steel, et al. (2007). "Influenza virus transmission is dependent on relative humidity and temperature". In: *PLoS Pathogens* 3.10, pp. 1470–1476. ISSN: 15537366. DOI: 10.1371/journal.ppat.0030151.

Lowen, Anice C., Samira Mubareka, Terrence M. Tumpey, et al. (2006). "The guinea pig as a transmission model for human influenza viruses". In: *Proceedings of the National Academy of Sciences of the United States of America* 103.26, pp. 9988–9992. ISSN: 00278424. DOI: 10.1073/pnas.0604157103.

Luo, Wei et al. (2020). "The role of absolute humidity on transmission rates of the COVID-19 outbreak". In: *medRxiv*, p. 2020.02.12.20022467. DOI: 10.1101/2020.02.12.20022467. URL: http://medrxiv.org/content/early/2020/02/17/2020.02.12.20022467.abstract%0Ahttps://www.medrxiv.org/content/10.1101/2020.02.12.20022467v1.

Ma, Yueling et al. (2020). "Effects of temperature variation and humidity on the mortality of COVID-19 in Wuhan". In: *medRxiv preprint*. DOI: 10.1101/2020.03.15.20036426.

Marr, Linsey C. et al. (2019). "Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence". In: *Journal of the Royal Society Interface* 16.150. ISSN: 17425662. DOI: 10.1098/rsif.2018.0298.

Meurer, Aaron et al. (Jan. 2017). "SymPy: symbolic computing in Python". In: *PeerJ Computer Science* 3, e103. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.103. URL: https://doi.org/10.7717/peerj-cs.103.

Mossong, Joël et al. (2008). "Social contacts and mixing patterns relevant to the spread of infectious diseases". In: *PLoS Medicine* 5.3, pp. 0381–0391. ISSN: 15491277. DOI: 10.1371/journal.pmed.0050074.

Neher, Richard A. et al. (2020). "Potential impact of seasonal forcing on a SARS-CoV-2 pandemic". In: *Swiss medical weekly* 150.3, w20224. ISSN: 14243997. DOI: 10.4414/smw.2020.20224.

Oliveiros, B et al. (2020). "Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases". In: *medRxiv preprint*. ISSN: 0217751X. DOI: 10.1142/S0217751X20500220.

Onder, Graziano, Giovanni Rezza, and Silvio Brusaferro (2020). "Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy". In: *JAMA - Journal of the American Medical Association* 2019, pp. 2019–2020. ISSN: 15383598. DOI: 10.1001/jama.2020.4683.

Poirier, Canelle et al. (2020). "The Role of Environmental Factors on Transmission Rates of the COVID-19 Outbreak: An Initial Assessment in Two Spatial Scales." In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3552677.

Postnikov, Eugene B. (2016). "Dynamical prediction of flu seasonality driven by ambient temperature: Influenza vs. common cold". In: *European Physical Journal B* 89.1. ISSN: 14346036. DOI: 10.1140/epjb/e2015-50845-7.

Rahman, Arifur et al. (2020). "A Retrospective Analysis of Influence of Environmental / Air Temperature and Relative Humidity on SARS-CoV-2 Outbreak". In: *preprints* March. DOI: 10.20944/preprints202003.0325.v1.

Sajadi, Mohammad M et al. (2020). "Temperature, humidity, and latitude analysis to predict potential spread and seasonality for COVID-19". In: *SSRN Electronic Journal* 410, pp. 6–7.

Shaman, Jeffrey and Melvin Kohn (2009). "Absolute humidity modulates influenza survival, transmission, and seasonality". In: *Proceedings of the National Academy of Sciences of the United States of America* 106.9, pp. 3243–3248. ISSN: 00278424. DOI: 10.1073/pnas.0806852106.

Shi, Peng et al. (2020). "Impact of temperature on the dynamics of the COVID-19 outbreak in China". In: *Science of the Total Environment* January.

Soetaert, Karline, Thomas Petzoldt, and R Woodrow Setzer (2010). "Solving Differential Equations in {R}: Package de{S}olve". In: *Journal of Statistical Software* 33.9, pp. 1–25. ISSN: 1548-7660. DOI: 10.18637/jss.v033.i09. URL: http://www.jstatsoft.org/v33/i09.

Sugihara, George, Ethan R. Deyle, and Hao Ye (2017). "Misconceptions about causation with synchrony and seasonal drivers". In: *Proceedings of the National Academy of Sciences of the United States of America* 114.12, E2272–E2274. ISSN: 10916490. DOI: 10.1073/pnas.1700998114.

Tamerius, James D. et al. (2013). "Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates". In: *PLoS Pathogens* 9.3. ISSN: 15537366. DOI: 10.1371/journal.ppat.1003194.

Tamerius, James et al. (2011). "Global influenza seasonality: Reconciling patterns across temperate and tropical regions". In: *Environmental Health Perspectives* 119.4, pp. 439–445. ISSN: 00916765. DOI: 10.1289/ehp.1002383.

Tang, J. W. et al. (2010). "Incidence of common respiratory viral infections related to climate factors in hospitalized children in Hong Kong". In: *Epidemiology and Infection* 138.2, pp. 226–235. ISSN: 09502688. DOI: 10.1017/S0950268809990410.

Tang, Julian W. (2009). "The effect of environmental parameters on the survival of airborne infectious agents". In: *Journal of the Royal Society Interface* 6.SUPPL. 6. ISSN: 17425662. DOI: 10.1098/rsif.2009.0227.focus.

The World Bank (2019). *Population*. URL: https://data.worldbank.org/indicator/SP.POP.TOTL.

United Nations (2019). *World Population Prospects 2019*. URL: ttps://population.un.org/wpp/.

Valle, Sara Y.Del, James M. Hyman, and Nakul Chitnis (2013). "Mathematical models of contact patterns between age groups for predicting the spread of infectious diseases". In: *Mathematical Biosciences and Engineering* 10.5-6, pp. 1475–1497. ISSN: 15471063. DOI: 10.3934/mbe.2013.10.1475.

Wang, Jingyuan et al. (2020). "High Temperature and High Humidity Reduce the Transmission of COVID-19". In: *arXiv Populations and Evolution*. DOI: arXiv:2003.05003v1. URL: http://arxiv.org/abs/2003.05003?utm_source=researcher_app&utm_medium=referral&utm_campaign=RESR_MRKT_Researcher_inbound.

Wickham, H (2016). *ggplot2: Elegant graphics for data analysis*. New York.

World Health Organization (WHO) (2018a). *Influenza (avian and other zoonotic)*. URL: https://www.who.int/news-room/fact-sheets/detail/influenza-(avian-and-other-zoonotic).

— (2018b). *Influenza (seasonal)*. URL: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal).

— (2020a). *Coronavirus disease 2019 Situation Report -51*. Tech. rep. March. DOI: 10.1001/jama.2020.2633. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

— (2020b). *Novel Coronavirus (2019-nCoV) Situation Report - 1*. Tech. rep.

Supplementary Information

Parameter Meaning Table

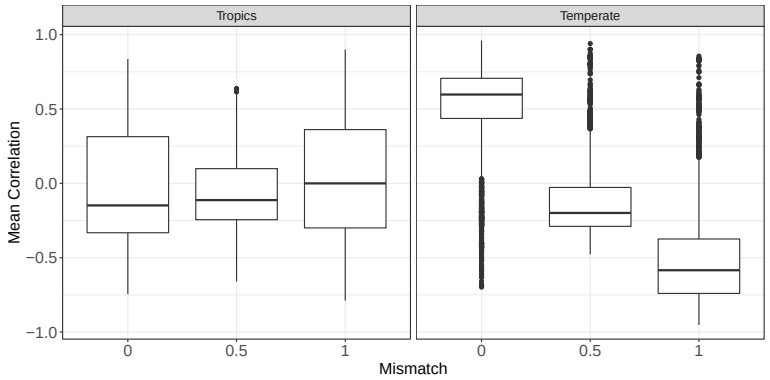
Parameter	Meaning	Unit	Origin
S	Number of Susceptible Individuals	number of individuals(n)	O. Bjørnstad (2018)
E	Number of Exposed Individuals	number of individuals(n)	O. Bjørnstad (2018)
I	Number of Infected Individuals	number of individuals(n)	O. Bjørnstad (2018)
R	Number of Recovered Individuals	number of individuals(n)	O. Bjørnstad (2018)
N	Total Number of Individuals	number of individuals(n)	O. Bjørnstad (2018)
β	transmissibility (number of infected individuals per infected individual per day)	n/n/day	O. Bjørnstad (2018)
f	rate of loss of immunity	1/day	
ν	natural per capita birth rate	1/day	Keeling and Rohani (2007)
σ	rate of movement from E to I (reciprocal of latent period)	1/day	O. Bjørnstad (2018)
μ	natural per capita death rate	1/day	O. Bjørnstad (2018)
γ	recovery rater	1/day	O. Bjørnstad (2018)
α	disease induced mortality rate	1/day	O. Bjørnstad (2018)
p	case fatality rate	-	Keeling and Rohani (2007)
cr	contact rate per day	n/n/day	Valle, Hyman, and Chitnis (2013)
ϵ	scaling factor as not all contacts lead to successful transmission	-	Valle, Hyman, and Chitnis (2013)
b	rate of decay of virus	1/day	
d	average contact duration	days	Valle, Hyman, and Chitnis (2013)
h	contact duration needed for transmission	days	reciprocal of ζ Valle, Hyman, and Chitnis (2013)
T	Temperature	degrees Celcius	
AH	Absolute humidity	+++	
RH	Relative humidity	-	
s	standard deviation of contact rate	n/n/day	
v	Viability of virus or measure of concentration of virus	%	Harper (1961)
t	time	day	
g	growth constant of decay rate with temperature	1/degrees C	

Table 2: Table of model parameters

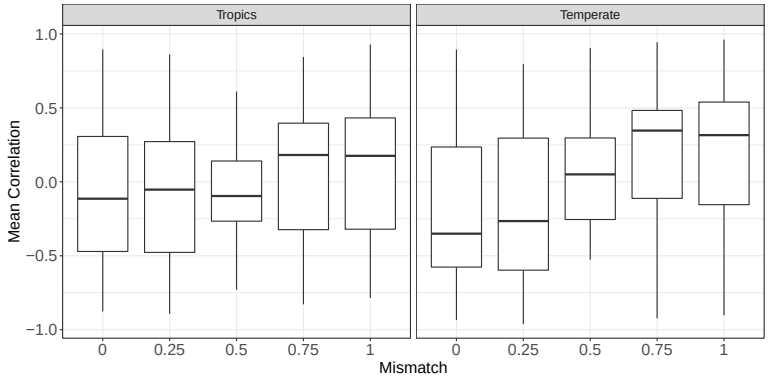
Parameter Values Table

Parameter	Influenza	Source	COVID-19	Source
f	0.1	Estimated by trial and error	0.05	Estimated by trial and error
ν	5.07e-5	Combined mean of countries United Nations (2019)	5.07e-5	combined mean of countries United Nations (2019)
σ	0.68	Ferguson et al. (2005)	0.2	Lauer et al. (2020)
μ	2.05e-5	Mean of countriesUnited Nations (2019)	2.05e-5	United Nations (2019)
γ	0.25	Median of Cauchemez, Carrat, et al. (2004) and Longini et al. (2005)	0.048	Bi et al. (2020)
p	0.001	Estimated from CDC 2018-2019 data CDC (2020a)	0.01	Onder, Rezza, and Brusaferro (2020)- South Korea
ϵ	0.05	Estimated by trial and error	0.05	Estimated by trial and error
d	4/24	Most frequent category of contact duration in polymod study Mossong et al. (2008)	4/24	Most frequent category of contact duration in polymod study Mossong et al. (2008)
h	0.25/24	Definition of close contact by CDC CDC (2020b)	0.25/24	Definition of close contact by CDC CDC (2020b)
max $\_cr$	26.97	Hoang et al. (2019)	26.97	Hoang et al. (2019)
g for T	0.095	From analysis of Harper (1961)	0.078	From analysis of Biryukov et al. (2020)
b_0 for T	9.079	From analysis of Harper (1961)	0.256	From analysis of Biryukov et al. (2020)
g for RH	0.0209	From analysis of Harper (1961)		
b_0 for RH	21.98	From analysis of Harper (1961)		
g for AH	30.16	From analysis of Harper (1961)		
b_0 for AH	0.062	From analysis of Harper (1961)		

Table 3: Table of parameter values. From literature searching and analysis of data in literature. b_0 and g depend on climate variable so only calculated for influenza (as COVID-19 analysis focuses on temperature).



(a) Absolute Humidity: 1 sample



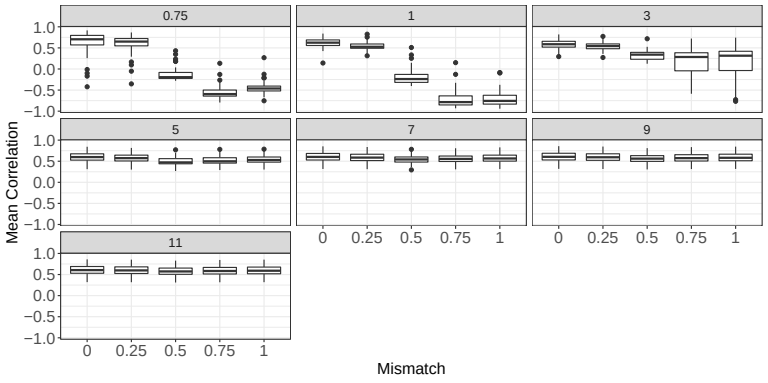
(b) Relative Humidity: 1 sample

Figure 8: Box and whisker plots of the correlation between data and the model for 5 different mismatches split by whether the country is tropic or temperate. The data is the weekly mean per capita positive influenza tests and the model is the weekly mean proportion of infected individuals. A mismatch of 0 has a higher correlation for absolute humidity. n=29 in tropics, 48 in temperate region

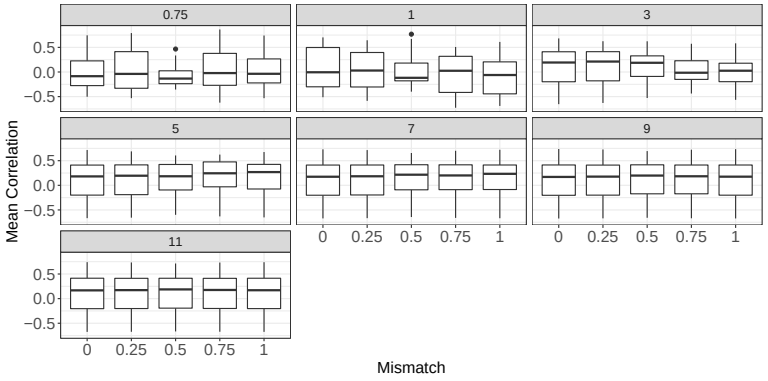
Humidity

Variance

5.1 COVID-19 over the year

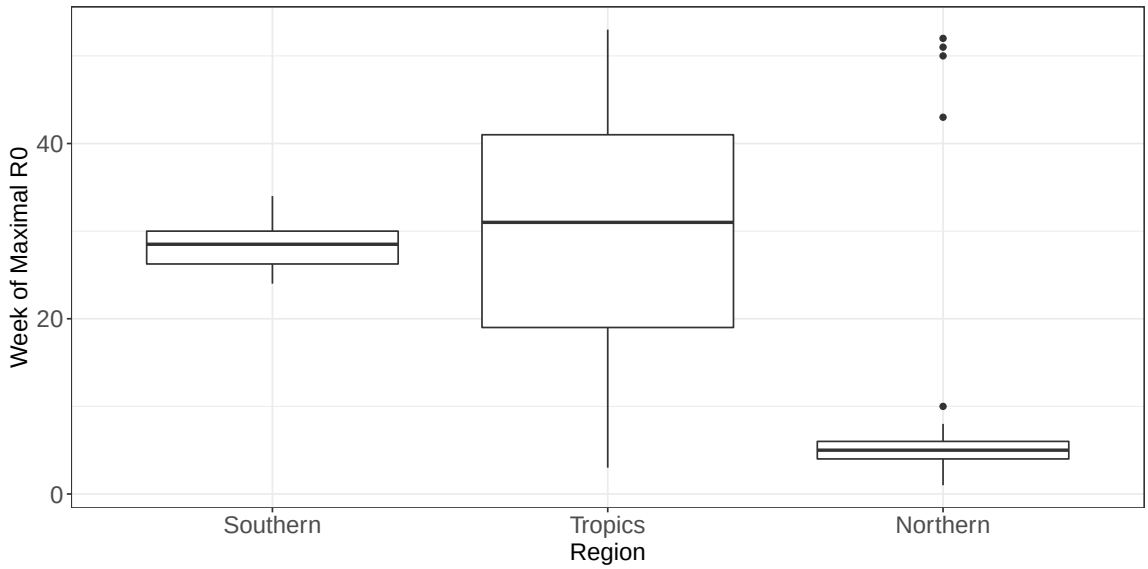


(a) Temperate

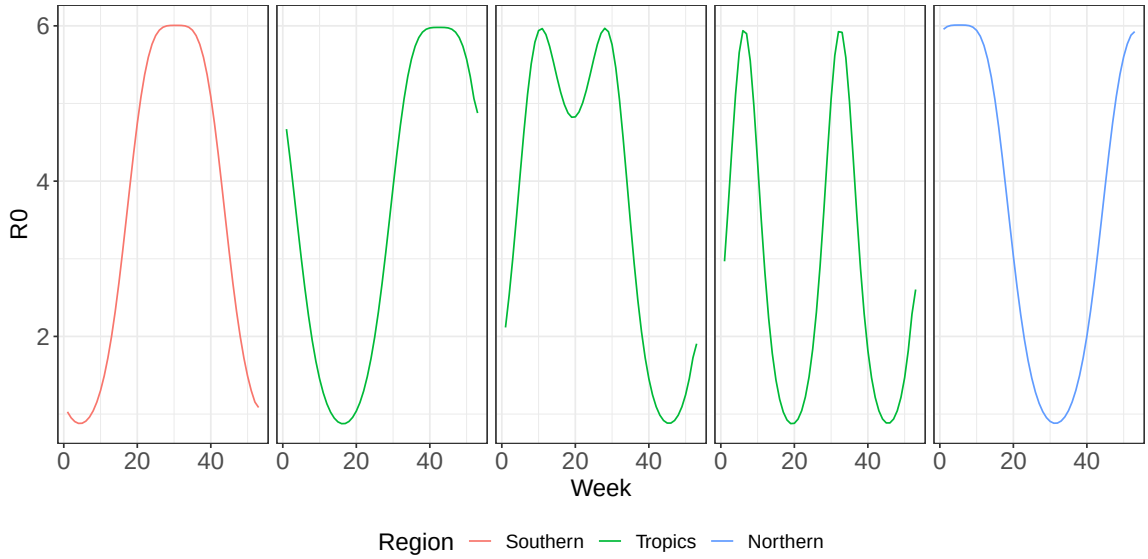


(b) Tropics

Figure 9:



(a) Mean of the week where R_0 is highest for each country for the tropics, northern temperate and southern temperate region. n=29 in tropics, 42 in northern temperate region and 6 in southern temperate region.



(b) Mean R_0 for each week for 5 different countries, representative of each region. Northern and Southern Hemisphere have the same basic shape but in the tropics, there are three different shapes of R_0 over time.

Figure 10: R_0 over time in different regions