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

August 2020

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 the Journal of Theoretical Biology

Submitted for the MSc in Computational Methods in Ecology and Evolution

Declaration

Influenza and climate data were from the supplementary information in Deyle et al. 2016. The data was mostly already processed at the point that I accessed it but I did carry out some processing, specifically removing missing data points and reshaping the data for analysis. The mathematical model was developed by me with contributions from Samraat Pawar.

Acknowledgements

I would like to express my thanks to a number of people who helped me during this project:

- My supervisor, Samraat Pawar for being a reliable and thoughtful supervisor. The time you invest in your masters students is really appreciated.

- My second supervisor, Michael Tristem, for suggesting the topic and for valuable virology insight.

- My lab for the opportunity to present and develop this idea and their support and advice.

- My course mates for all of the emotional and practical support. Special thanks go to Henrique Galante, Anne Marie Saunders, Dónal Burns and PokMan Ho. I don't think I could have gotten through this year without your support and friendship.

- My friends for their kindness and (mostly virtual!) company over the last few months. Thankyou for giving me the space to relax and be silly.

- The wonderful people in John Smith for the potlucks, baked goods and kind words.

- My family for living with me while I was doing a masters degree! Thankyou for making me laugh and making my breaks fun.

Abstract

Respiratory viruses can show strong seasonality and are a substantial burden on health services. Understanding their seasonality will help to reduce their impact. Temperature-dependence of viral decay rate and human contact rate may play a role in seasonality but previous work has not included them. I develop a model for transmissibility based on the temperature-dependence of both viral survival and contact rate. This is used in a Susceptible-Exposed-Infectious-Recovered-Susceptible (SEIRS) compartment model. I predict the number of infectious individuals over time, given seasonal fluctuations in temperature. I model the effect of temperature variation mediated by different levels of trait mismatch, the difference between the temperatures where contact rate and virus survival are highest. To assess which level of mismatch is most likely, I find the correlation between average my model and flu data over one year for each mismatch for 77 countries. In temperate regions, the most likely mismatch was consistent with the highest contact rate and virus survival occurring at the same temperature. In the tropics, the most likely mismatch was much more varied and there was high variability between countries. R_0 measures how many people one infected individual would infect in a completely susceptible population. Contact rate peaking at intermediate temperatures (mismatch of 0.5) led to lower mean R_0 values and contact rate peaking at high temperatures (mismatch of 1) led to lower maximum R_0 values. When I used COVID-19 parameters and the mismatch with the best correlation in the influenza model, temperate regions had a consistent winter peak. Positively shifting the mismatch value decreased the mean and maximum R_0. More research into the climate-dependence of contact rate will be essential to fully understand the seasonality of respiratory viruses including COVID-19. This new model has great potential to predict respiratory virus seasonality and could be used to improve current epidemiological interventions. .

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 (WHO 2018b; WHO 2018a). Most recently, COVID-19, a respiratory virus caused by SARS-CoV-2, is having significant impacts across the world. It was first seen in December 2019 and declared a pandemic on the 11th March 2020 (WHO 2020b; WHO 2020a). Previous coronavirus epidemics such as SARS have prompted discussions about the importance of understanding seasonality of 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; Neher et al. 2020). Understanding the dynamics of these diseases helps with planning timing and types of interventions. This can be especially important for seasonal pathogens because periods with fewer cases allow time for planning in advance of the next outbreak.

Temporal and spatial differences in climate have been linked to disease. Seasonal differences are the main source of temporal temperature variation. Influenza is perhaps the most well known seasonal respiratory virus but others include respiratory syncytial virus (Tang et al. 2010), mild betacoronaviruses (Kissler et al. 2020) pneumococcus and rubella (Dowell and Ho 2004). These tend to peak in the winters of temperate regions. Longer-term temporal differences in climate may also affect disease, such as those caused by the El Niño Southen Oscillation (Viboud et al. 2004). In terms of spatial differences, the tropics and temperate regions have different disease patterns. In temperate regions, marked yearly winter peaks are common. In the tropics, disease may peak in the rainy season, have two peaks or remain high all year round (Tamerius et al. 2013).

For the most part, work to understand the impact of climate is statistical work that links climate with disease. While not being mechanistic these can be very valuable in understanding the factors that lead to seasonality and heterogeneity of disease. Increased influenza incidence has been linked to lower temperature and humidity. (Steel, Palese, and Lowen 2011; Barreca and Shimshack 2012). Using convergent cross mapping, a statistical test to find causality, Deyle et al. (2016) find a causal link between absolute humidity and influenza, mediated by temperature. Whether this truly demonstrates causality is questionable because the same approach can be used to demonstrate that influenza causes climatic conditions, which is not the case (Baskerville and Cobey 2017; Sugihara, Deyle, and Ye 2017). As Bjornstad and Viboud (2016) point out, combining such work with mechanistic models will improve disease forecasting.

The causes of seasonality have been discussed elsewhere in extensive detail but can broadly be split

into three potential reasons: human behaviour, human immunity and viral survival (Tamerius et al. 2011). Viral survival during transmission is essential for successful transmission. Numerous studies have explored how climate variables affect the survival of viruses in a lab setting. For lipid enveloped viruses, lower temperatures and lower relative humidities lead to better virus survival (Tang 2009; Tamerius et al. 2011). The effect of temperature on influenza has been reviewed by Irwin et al. (2011) who found that across air, water, faeces and surfaces, temperature is a significant predictor of virus half-life. Additionally, transmission of influenza between guinea pigs in controlled environments is higher in colder, less humid environments (Lowen et al. 2007). Coronaviruses and influenza are both lipid-enveloped viruses so may react similarly in response to climate. Higher temperatures and higher relative humidities have been linked with faster coronavirus decay on surfaces (Biryukov et al. 2020), in aerosols (Doremalen, Bushmaker, and Munster 2013) and in liquids (Chin et al. 2020).

Human behaviour may change with climate due to seasonality in the school year, particularly long summer holidays. Work by Fine and Clarkson (1982) and Finkenstädt and Grenfell (2000) suggests that measles seasonality may be mediated by school terms. Terms are also likely to play a role in influenza (Tamerius et al. 2011; Cauchemez et al. 2008). The weather is also likely to impact human contact rates or type. 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 an attempt to separate the effects of climate on the virus from the effect of climate on human behaviours.

SEIR models are compartmental models with Susceptible, Exposed, Infected and Recovered components. Seasonality has frequently been added to these models by modifying the transmissibility of the virus. Mostly, this is by assuming a sinusoidal shape of transmissibility with time, for example in measles (Bolker and Grenfell 1993) and COVID-19 (Neher et al. 2020). Other work has allowed transmissibility to vary linearly with temperature, humidity or both (Shi et al. 2020; Postnikov 2016). One issue with these models is that parameterisation is necessary for both the intercept and strength of seasonal forcing. These models are helpful when seasonal transmission changes are already known but do not help to explain the underlying causes of seasonality and are less helpful when a disease is novel.

The majority of the potential causes of seasonality are mediated by climate factors. The exact cause of seasonality is not known but a combination of factors may contribute (Lofgren et al. 2007). Dushoff et al. (2004) suggest that only small fluctuations in transmission with climate are required for seasonality of disease based on a simple model where recovered individuals can lose immunity over time. When the period of seasonal changes and the period of the model without seasonal change are similar, they interact, which allows seasonal disease peaks to occur even when the changes in transmission are small. This is

dynamic resonance. Altizer et al. (2006) demonstrate how larger amplitudes of seasonal forcing lead to larger fluctuations in the number of infected individuals and potentially chaotic dynamics. Understanding the interaction of factors is 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 temperature. Understanding the effect of individual components may be useful to understand how interventions can help. In particular, 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 affected by human behaviour and virus properties. Previous work where transmission varies with time tends to assume a sinusoidal shape without including the mechanisms for this. The main aim of this project is to explore the effect of temperature on respiratory viruses when transmissibility is modified by temperature-dependence of viral decay and human contact rate. Research into vector borne-viruses has included temperature-dependence of multiple traits (Mordecai et al. 2013) but this has not yet been done for respiratory viruses. This work will help us to understand the dynamics of COVID-19 if it becomes endemic. 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. Contact rate can encompass different aspects of human behaviour such a social behaviour and school terms. The number of contacts an individual has determines the maximum number of transmission events so it is important to know how human behaviour changes with temperature. Previous work has explored the distribution of contact rates in detail (Mossong et al. 2008; Hoang et al. 2019). However, there is a lack of high quality data exploring temporal or climatic variation in contact rate. 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 growth and human contact rate are the highest (Figure 1). Temperature is focused on as the climate variable of interest here because it has the clearest links with both behaviour and viral survival.

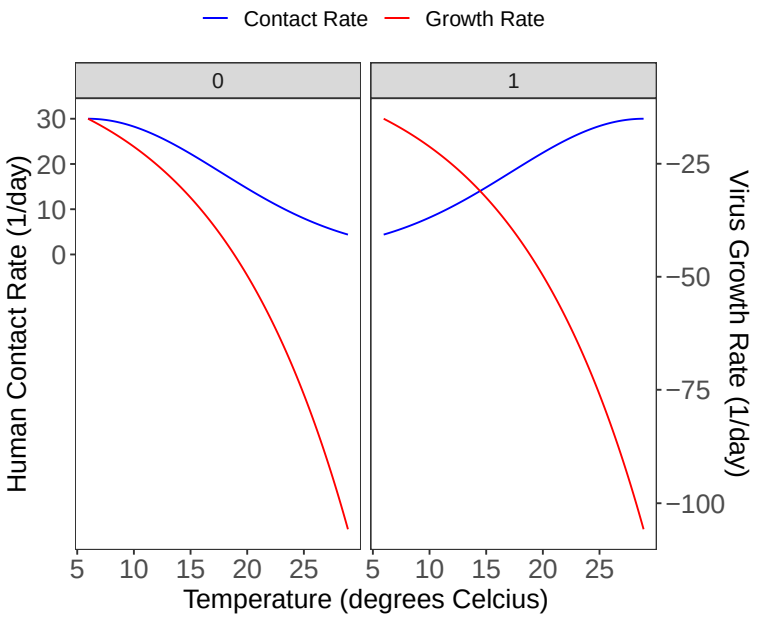


Figure 1: A visualisation of the model contact rates and viral growth rate with temperature for high and low mismatches. Virus growth rate= − Virus decay rate. Increasing value on the y-axis corresponds to better conditions for disease.

1.1 Aims

The main aims of this project are to find out:

i Which level of mismatch is most likely? Does this vary between temperate and tropical regions?

ii How does mismatch affect the severity of disease outbreaks?

iii How might COVID-19 behave as an endemic disease?

I will achieve this by making a basic compartment model with a climate-dependent transmissible term containing both virus survival and contact rates as climate-dependent terms. I will simulate this with influenza parameters at different levels of mismatch (different temperatures where contact rate peaks) and determine what level of mismatch is the most likely for each country by comparing it to long term influenza data. I will explore how mismatch affects influenza R_0. I will investigate the seasonality of COVID-19 and the potential impact of shifting contact rates using the most likely mismatch level for each country.

2 Methods

R 3.3.2 was used for data manipulation and integration. Python 3 was used for sensitivity analysis.

2.1 The Model

The basis of the model is a simple SEIRS compartment model based on Bjornstad (2018) and Keeling and Rohani (2007). SEIR models are a type of compartmental model with 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 system of differential equations, where each differential equation describes the change in the number of individuals in that compartment (Figure 2 and Equation System 1).

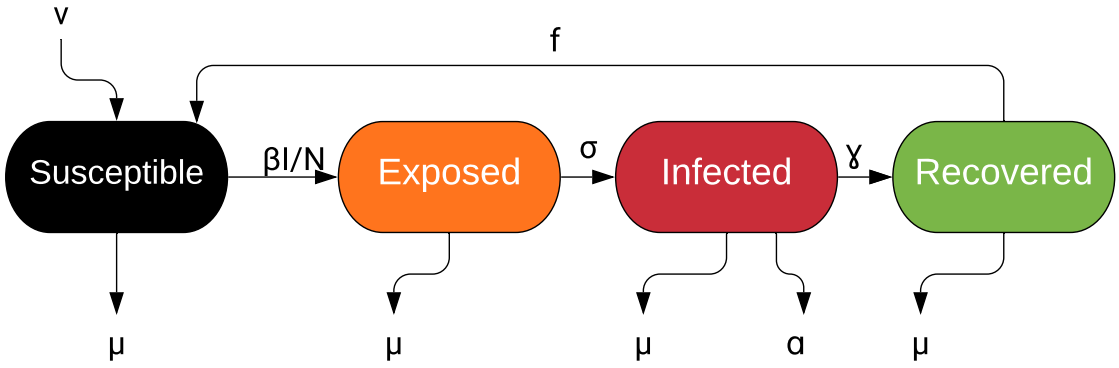


Figure 2: Visual Representation of SEIR model

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

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

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

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

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) (Equation 2a). This modifies Equation 1c to

Equation 2b. It is worth noting that disease-induced mortality was included, unusual for a typical influenza model, due to the relatively high mortality 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.

Table 1: Parameters of SEIRS model. Parameters are described fully (with units and sources) in Table S1

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

The transmissibility (β) used is climate-dependent, obtained by modifying the model in Valle, Hyman, and Chitnis (2013) to add climate-dependence through human contact rate and virus decay rate. Transmissibility is split into the number of contacts each individual has per day (c_r) and a probability of transmission if an infected individual meets a susceptible individual. Within this probability, there is a scaling term (ϵ) and a mechanistic term (Equation 3). The mechanistic term is based on the assumption that at over time, virus in the environment increases by the shedding rate of the infected individual and decreases by the amount of viral decay. This can be integrated to estimate how much virus would be in the environment at a particular time (if one individual is shedding). The amount of virus in the environment at a time equal to the average contact duration can then be compared to the amount of virus that would be shed in the time required for successful infection, which becomes a climate-dependent infection probability, the mechanistic term in Equation 3. The scaling term covers the multitude of other reasons why β is not as large as c_r including the fact that risk from contacts vary depending on reasons other

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

$$\begin{gathered}\beta = \text{contact rate} \times \text{scaling term} \times \text{mechanistic term} \\ \beta(C) = \quad c_r(C) \quad \times \quad \epsilon \quad \times \frac{\frac{1}{b(C)}(1-e^{-b(C)d})}{\frac{1}{b(C)}(1-e^{-b(C)d})+h}\end{gathered} \tag{3}$$

Contact rate as a function of climate was assumed to be based on a normal distribution scaled up so the peak of the distribution is the maximum possible contact rate (c_{ru}) (4). Each country was assumed to have the same maximum contact rate because this work focuses on seasonal differences rather than differences between countries. The value of s, the standard deviation of the contact rate 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 the middle 95% is 1.96, where x is the raw score for that z-score. In this case, x is a climate bound (C_b) which is either the upper (C_u) or lower bound (C_l). Rearranging this gives $s = \frac{x-\mu}{1.96}$, equivalent to Equation 4b. When the climate where the peak occurs (C_m) happens closer to the lower bound of climate, the upper bound (C_u) is used as C_b, otherwise, the lower bound (C_l) is used.

$$c_r(C) = c_{ru} \times e^{\frac{-(C-C_m)^2}{2s^2}} \tag{4a}$$

$$s = \frac{|C_b - C_m|}{1.96} \tag{4b}$$

The climate where the peak occurs (C_m) was not known so the model was run for five different values of C_m, spaced evenly across the climate range. This was converted to the levels of "mismatch" (M), where mismatch is based on the difference between the climate at which contact peaks and the climate at which virus decay is lowest (Equation 5). This is visualised in Figure 3a in Results. The virus decay increases with temperature and humidity so the climate at which virus decay is lowest is C_l.

$$M = \frac{C_m - C_l}{C_u - C_l} \tag{5}$$

A mismatch of 0 means that the contact rate peaks at the lower range of the climate variable, the climate where there is least viral decay. Both of these conditions are "best" for the virus. Therefore the amount of mismatch measures how much difference there is between the best conditions for the virus with respect

to its rate of decay and human contact rates.

Experimental data were obtained for the viability of influenza over time in different temperature (7.5 - 32 degrees Celcius) and relative humidity (20-80%) combinations from Harper (1961). Some pre-processing of the data was carried out. Where the given climate was a range of values (e.g. 21-23 degrees), the mean of this range was used. Additionally, 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)). Then nlsLM in R (which uses the Levenberg-Marquard algorithm to fit non-linear models) was used to fit exponential decay models to the viability over time (Equation 6) to find b, the rate of decay for each environmental condition. b was bounded between 0 and 5000 per day. v_0, the initial viability, was bounded between 0 and 100%. Experimental data for SARS-CoV-2 and climate were also found (Biryukov et al. 2020). This data contained the virus half-life for eight different combinations of temperature and relative humidity conditions. The half-lives were converted into b by rearranging Equation 6 and setting t to the half-life and v to equal $\frac{v_0}{2}$.

For temperature, absolute humidity and relative humidity for influenza and temperature for SARS-CoV-2, the climate conditions and decay rates were then fitted to exponential growth models using nlsLM(Equation 7). This obtained an estimate for the rate of decay when the climate value is 0 (b_0) and the rate of growth (g) of the rate of decay with the climate variable. These values form an equation to find b for each climate variable (C) (Equation 7).

$$v = v_0 e^{-bt} \tag{6}$$

$$b = b_0 e^{gC} \tag{7}$$

2.2 Paramaterisation

A literature search was used to obtain parameter values for COVID-19 and Influenza (Values are available in Table S2).

2.3 Data Sources and Sorting

Climate and influenza data were obtained from Deyle et al. (2016). This contained relative humidity, absolute humidity, temperature and positive flu results per capita for 79 countries. Iraq was removed from analysis because 16 of the relative humidities were much greater than 100%. 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. Years where there was data only for part of the year were removed, leaving only years with 52 or 53 weeks. 11090 rows were removed (about 3% of rows) during this step. This was done because it could not be determined whether missing values were zeros or an absence of data collection. Then the average weekly flu and climate values were found for each country. For each country the minimum (C_l), maximum (C_u) and week of the maximum climate variables (t_p) were found. They were used to create as a sinusoidal function for the climate for each country (Equation 8) which was necessary because the weekly time intervals for climate values were not sufficient for the integration in later steps. Equation 8 was determined by modifying a cos function by setting the period to 365 days, the amplitude to the climate range and shifting the function such that the peak occurs when the climate peaks.

$$C(t) = \frac{C_u - C_l}{2} \times \cos\left(\frac{2\pi}{365} \times (t - t_p)\right) + \frac{C_u + C_l}{2} \tag{8}$$

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 Sensitivity Analysis

The basic reproduction number, R_0, is the average number of individuals an infected individual infects if the population is completely susceptible. The value of R_0 for an SEIR model is shown in Equation 9 (Bjornstad 2018). This was modified by converting from α to p (Equation 2a) and replacing the transmissibility (β) with Equation 3. This could be based on the R_0 for an SEIR model rather than an SEIRS model because R_0 is based on a completely susceptible population so loss of immunity does not affect it.

$$R_0 = \frac{\sigma}{\sigma + \mu} \times \frac{\beta}{\gamma + \mu + \alpha} \tag{9}$$

$$R_0 = \frac{\sigma}{\sigma + \mu} \times \frac{c_r \times \epsilon \times \frac{\frac{1}{b} \times (1 - e^{-bd})}{\frac{1}{b} \times (1 - e^{-bd}) + h}}{(\mu + \gamma)\frac{1}{1-p}} \tag{10}$$

When parameterisation was carried out, there was a lack of strong evidence for the values of parameters. For this reason, a sensitivity analysis was carried out to understand the impact of the parameter values. The most important parameters were then varied in later stages of the analysis. The first step was finding the an equation for partial derivative of R_0 with respect to each parameter using SymPy in Python (Meurer

et al. 2017). Then this equation was divided by R_0 (Equation 10) to find the change in the parameter as a proportion of R_0, giving equation 11. ϕ represents the relative importance of each parameter.

$$\phi = \frac{\left(\frac{\partial R_0}{\partial P}\right)}{R_0} \tag{11}$$

The relative importance (ϕ) was found for each parameter at 900 parameter combinations, by substitution parameter values into equation 11. Specifically, each of the nine parameters in Equation 10 was varied from 50% to 150% of the best estimate of the parameter, at 100 regularly spaced points. q and c_r were assumed to be constant for this sensitivity analysis. They were estimated using temperature as the climate variable in equations 4 and 7. The temperature range was set to the means of the coldest and hottest weeks from all 77 countries. The temperature was set to the midpoint of this. The temperature where contact rate was highest was set to the minimum of the temperature range, so the mismatch was 0.5. As the distribution is symmetrical and the current temperature is in the middle of the range, contact rate is the same whether a minimum or maximum is used.

Over the different parameterisations, h, ϵ and μ were most frequently the most important parameters. For this reason, these were varied during the integration step (described in section 2.5). To further test robustness, the sensitivity analysis was repeated where the temperature was set to the lower, middle and upper bound of the temperature range and c_r and q were otherwise kept constant. The results at these three values were qualitatively very similar to each other and the earlier result.

2.5 Model Simulations

The model was integrated for each country for 5 different levels of mismatch. deSolve in R (with lsoda) Soetaert, Petzoldt, and Setzer (2010) was used for integration. For each country, the climate range was set to the lowest and highest weekly climate values for that country. The climate values where contact peaked (C_m) were set to equally spaced values across the climate range, leading to mismatch levels of 0, 0.25, 0.5, 0.75 and 1.

Due to uncertainty in parameters, this integration was repeated for 100 parameter combinations for each country and mismatch level. The parameters varied were the three most important parameters (h,μ and ϵ) found from the sensitivity analysis and f which could not be included in the sensitivity analysis. Latin hypercube sampling was used to find the 100 parameter combinations. "lhs" in R was used to find a value between 0 and 1 for each parameter. 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. Each SEIRS model was integrated over 11 years at a daily resolution to estimate the number of susceptible, exposed, infected and recovered individuals at each time point. At each time point, the value of the climate variable was found (Equation 8). From this, the contact rate and decay rate were found (Equations 4 and 7). These were used to find the transmissibility (Equation 3). The model started with 0.1% infected and 99.9% susceptible 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 to reduce the impact of the starting time and starting values. R_0 based on β was also found at each time point (Equation 10). This simulation was carried out for temperature, relative humidity and absolute humidity separately.

The minimum and maximum contact rate was not known. For this reason, the integration was repeated with different values for the standard deviation of c_r. s (Equation 4b) was multiplied by 0.75, 1, 1.5, 2, 2.5 and 3. Integration was carried out for these values of s for each country and each mismatch using parameter estimates.

2.6 Analysis of Integration Results

The mean proportion 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 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. The best mismatch for each country was the mismatch where the correlation was the highest. To find the best mismatch over the 100 parameter combinations, the mode was used for each country. Analysis was carried out using temperature as the climate variable. Countries were separated by latitude into temperate and tropics based on whether the absolute value of their latitude was greater than 23.5 degrees.

2.7 COVID-19

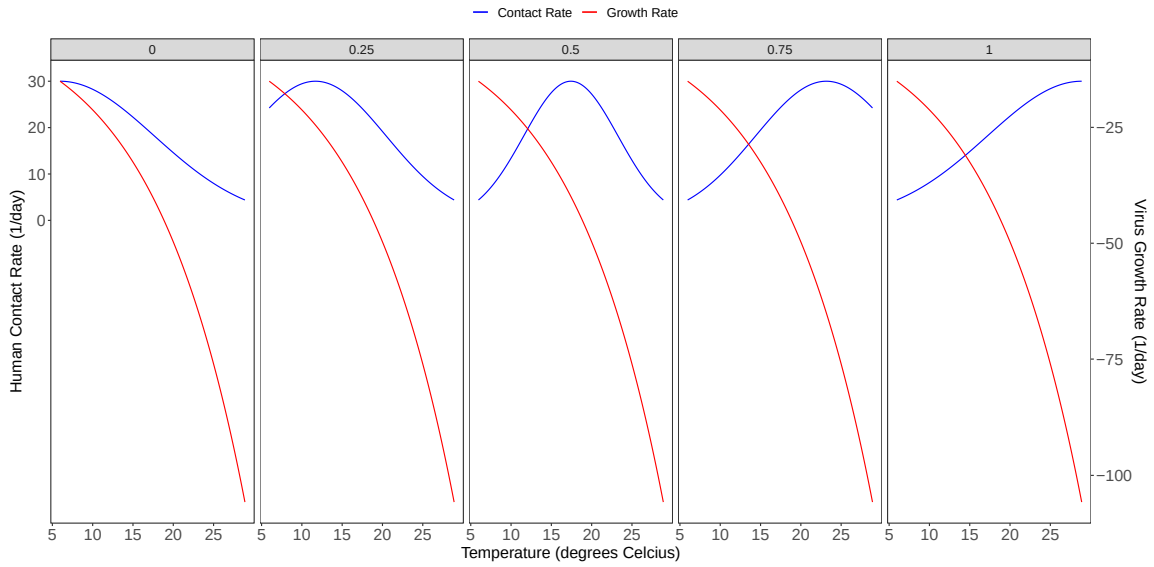
The integration was repeated for COVID-19 with temperature as the climate variable, at 5 different levels of mismatch. The difference between the mismatch and the best mismatch for influenza was found. This mismatch "shift" represents how different each mismatch level was from the most likely estimate.

3 Results

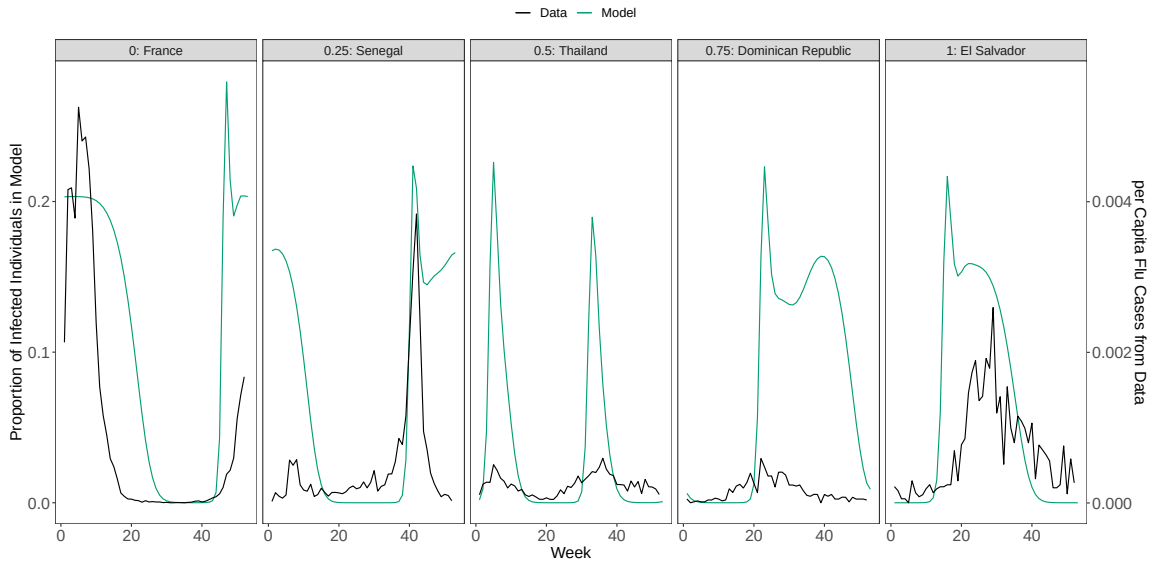
3.1 Models and Data

The temperature range is lower in tropical regions than in temperate regions. The virus decay 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 in 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 3a). Virus decay rate is always lowest when the climate variable is lowest. A mismatch of 0 means that the decay rate is lowest when the contact rate is highest, so the contact rate peaks at low climate values. A mismatch of 1 means that the decay rate is highest when the contact rate is highest, so the contact rate peaks at high climate values. Mismatch measures the difference between ideal conditions for the virus with respect to contact rate and viral survival. Mismatch affected how correlated each model was to the data so the climate at which contact rate peaked played a key role. The mismatch level with the highest correlation varied between countries implying that in different countries, contact rate has a different relationship with climate (Figure 3).

The correlation for the best mismatch ("best correlation") for each set of country and parameter combinations was higher for temperature (mean: 0.536, standard deviation: 0.206, n: 7700) than absolute (mean: 0.516, standard deviation: 0.239, n: 7700) or relative humidities (mean: 0.477, standard deviation: 0.158, n: 7700). For temperature, the parameter combinations resulted in more variation in best correlation than the country. The standard deviation when averaged across combinations (0.196, n: 100) was greater than the standard error when averaged across countries (0.130, n: 77).



(a) Mismatch Visualisation



(b) Data and best model for selected countries

Figure 3: Visualisation for different levels of mismatch.
(a) is a visualisation of the model contact rates and viral growth rate with temperature for different levels of mismatch. Contact range and virus growth rate across the temperature range for Italy is shown here, but the effect of mismatch is the same shape for other climate variables. Virus growth rate $= -$ Virus decay rate. Increasing value on the y-axis corresponds to better conditions for disease. Mismatch values are along the top panel.
(b) is the number of infectious individuals according to the best mismatch model for that country(green) and the mean number of positive flu tests per capita (black). The mismatch and country are along the top panel. Note: the model and results have different axes because the models and data show very different results, due to both the simplicity of the model and flu data collection method (lab results).

3.2 Mismatch and Region

In temperate regions, models were most similar to data when the contact rate peaked at low temperatures. 46 countries had a mismatch of 0, 2 countries had a mismatch of 0.25 (Figure 4). In tropical regions, the temperature at which contact rate was most likely to peak varied between countries (Figure 4). 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). Further, these results were robust to the variation in parameter values. Figure 4 shows the mean and error bars for 100 combinations of parameters. This shows the same effect as when just the parameter estimates are used. Using absolute humidity as the climate variable had the same effect as temperature (Figure S2). Using relative humidity as the climate variable led to a weaker effect of mismatch and in temperate regions, models with mismatches of 1 were most similar to the data (Figure S2). Increasing the standard deviation for the contact range did not change the overall relationship between region and mismatch but the effect of mismatch became less strong (Figure S3).

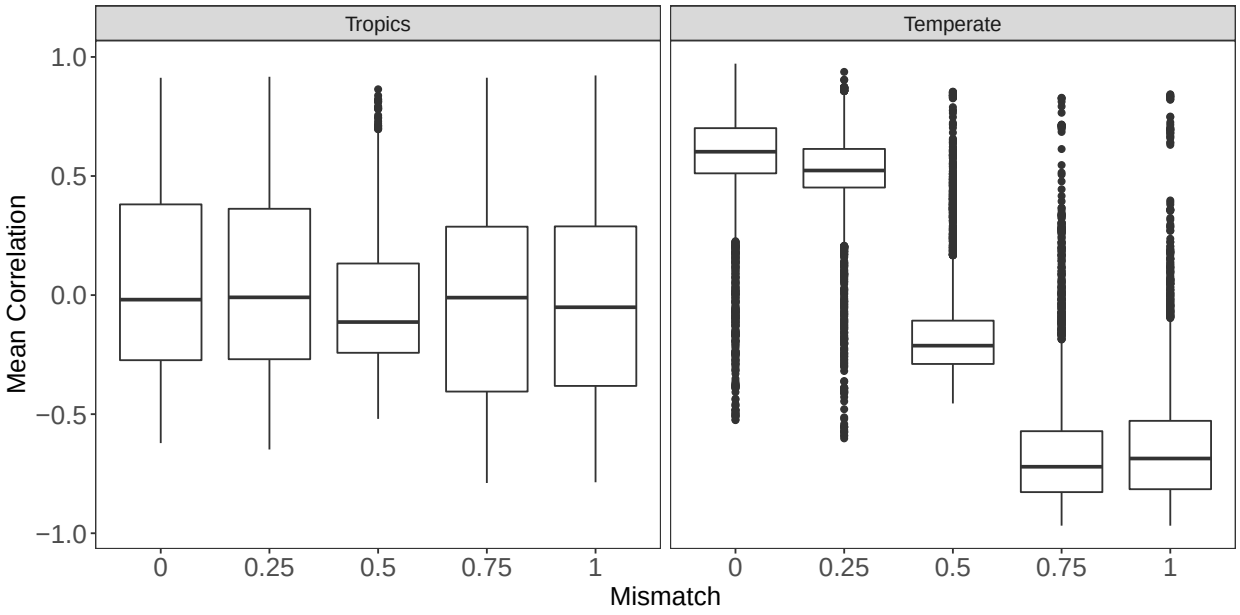


Figure 4: Box and whisker plot of the correlation between data and the model for 5 different mismatches split by whether the country is tropical or temperate. Higher correlations imply that the model is more similar to the data. The data is the weekly mean per capita positive influenza tests and the model is the weekly mean proportion of infected individuals. A mismatch of 0 has a higher correlation in temperate regions. In temperate regions, there is not a strong effect of mismatch. n=29 in tropics, 48 in temperate regions.

3.3 Sensitivity Analysis

Sensitivity analysis found that the most important parameters for R_0 were h, μ and ϵ (Figure 5).

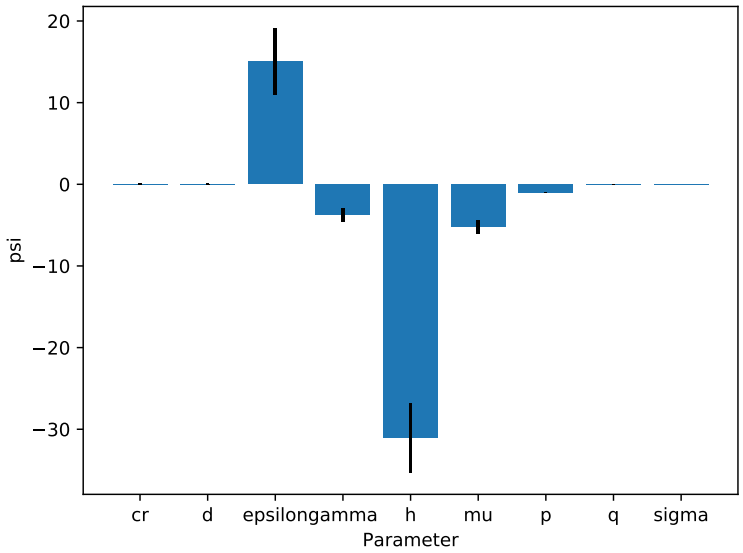


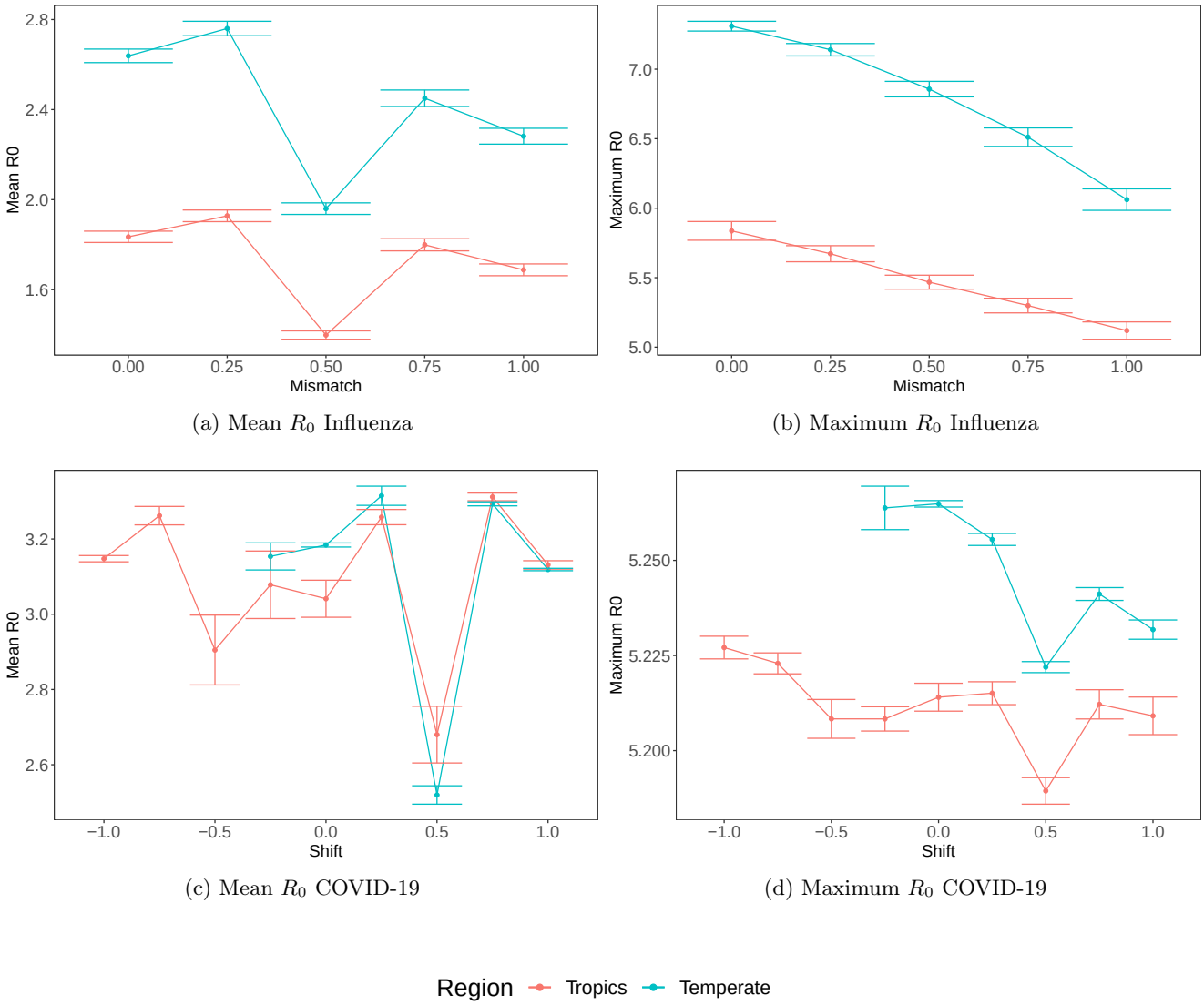
Figure 5: Results of Sensitivity analysis for each parameter in equation 10. ψ (Equation 11) was calculated for each parameter for 900 parameter combinations. The value of ψ shows the relative importance of each parameter. This plot shows the mean and standard deviation of ψ for each parameter over the 900 parameter combinations. μ, h and ϵ are the most important parameters.

3.4 Mismatch and Disease Prevalence

In general, the tropics had lower R_0 values than in temperate regions for all levels of mismatch (Figure 6). Models where the contact rate was highest at intermediate temperatures ($M = 0.5$) had the lowest mean R_0 and models where the contact rate was highest at high temperatures ($M = 1$) had the lowest maximum R_0 (Figure 6a and 6b).

3.5 COVID-19

The COVID-19 model suggests one R_0 peak in the winter of temperate regions although the exact timing does vary. The 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. The mean is week 29 and standard error is 2.65 weeks (Figure S4a). There are often multiple peaks per year in the tropics (Figure S4b). 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 (Figure 6c and 6d).



(a) Mean R_0 Influenza

(b) Maximum R_0 Influenza

(c) Mean R_0 COVID-19

(d) Maximum R_0 COVID-19

Figure 6: Mean and Maximum R_0 at different mismatches for influenza and shifts for COVID-19. In general, values for tropical regions were lower than in temperate regions. For influenza, mismatches of 0.5 led to the lowest mean values and mismatches of 1 led to the lowest maximum values. For COVID-19, a shift of $+0.5$ decreases the mean and maximum R_0 the most, although differences are small for maximum R_0.

4 Discussion

In temperate regions, there is strong synchrony between contact rates and virus survival, indicated by a mismatch of 0. This is consistent with the highest contact rate occurring in cooler temperatures (i.e. in winter). This could be due to more contacts occurring in cold temperatures. 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 occur indoors in cooler and rainy conditions. Alternatively, the mismatch could be related to differences in time in school over the year. Winter holidays tend to be shorter than summer holidays and it is clear that schools are a major source of contacts for children. If the length of the holiday modifies the transmission rate, shorter winter holidays could help to explain seasonality. Cauchemez et al. (2008) found that school holidays prevent almost a fifth of influenza cases in France , however, they only looked at epidemic periods so were not able to pick up the potential impact of a long summer holiday. In countries with a mismatch of 0, contact peaks in cool temperatures and the maximum R_0 is highest (Figure 6a and 6b). This is consistent with winter peaks in temperate regions, which is very common in flu (Tamerius et al. 2011). However, research into these seasonal changes has mostly focused on influenza. COVID-19 has a higher R_0 than influenza and the types of contact important for influenza and age groups with high transmission may be very different. For example, Davies et al. (2020) find children are less susceptible to COVID-19 and interventions aimed at children may only have a small impact on transmission.

The large disparity between temperate and tropical regions is of interest. In tropical regions, contact rate and viral survival are less synchronous. This suggests a systematic difference in temperature-dependence of human contact behaviour between the tropics and temperate regions. The tropics have a lower temperature range. This affects the difference between the maximum and minimum virus decay rate. However, the difference between the maximum and minimum contact rates in the model are assumed to be consistent between countries; in reality, the range of contact rates may depend on the temperature range. The difference between tropics and temperate regions could be to do with differing contact rate ranges between these regions. There are many different possibilities for how temperature might affect contact rate both between and within countries. Until accurate temperature-dependent contact rate data are obtained, models based on contact rate will be limited by accuracy.

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 (Marr et al. 2019). The models were better (highest best correlation) when temperature was the climate variable. In this study, the results for absolute humidity were very similar to the results for temperature. Absolute humidity grows exponentially with temperature (Figure S1) and absolute humidity range increase with latitude. Unlike temperature and absolute humidity, in relative humidity models, high mismatches were more similar to data. The results for absolute humidity and relative humidity may be caused by their relationship with temperature (Figure S1). There may be a complex, non-linear relationship between the effects of temperature and humidity on disease (Deyle et al. 2016). Combining multiple climatic variables into the virus decay component of this model could make it more accurate. Additionally, in the tropics, rainfall may have more of an effect on contact than temperature.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. Rainfall could be incorporated into the contact rate component. This may help us to understand the mechanisms underlying the differences between tropical and temperature countries.

As for COVID-19, the model predicts a winter peak in temperate regions. This is at odds with what has been seen in reality where countries have been affected in all seasons (WHO 2020a). When there are high numbers of susceptible individuals, the susceptible supply exerts a stronger effect than seasonal forcing (Baker et al. 2020). Implementing social distancing and restriction of mass gatherings were associated with lower epidemic growth in March but the effects of climate were weak (Jüni et al. 2020). In the longer term, once there are fewer susceptible individuals, climate may determine when or where outbreaks occur as is the case for other respiratory viruses (Deyle et al. 2016; Kissler et al. 2020). This means that understanding disease dynamics will be important for informing public health policy, for example by increasing testing at certain times or places. Kissler et al. (2020) and Neher et al. (2020) looked into how COVID-19 may be affected by seasonal forcing based on seasonal variation in the transmission of seasonal coronaviruses. They found that if immunity is permanent, in the short term, seasonal changes in transmission are expected to reduce the disease incidence but not eliminate the disease. If immunity is not permanent, recurrent winter outbreaks are possible, shifting from pandemic to seasonal epidemics as was seen in H1N1. This is in agreement with the model in this work.

Shifting mismatch may affect the severity of outbreaks (Figure 6c), so modifying timing of contact rates could decrease the impact of COVID-19 R_0. 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. This could be by lengthening winter holidays and shortening other holidays or limiting large gatherings in winter. These proposals would require substantial effort by governments and individuals but could reduce the burden of disease. In the tropics, shifting mismatch also appears to decrease the contact rate but for this to be helpful, the best mismatch would need to be identified for each country first.

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 model can easily be modified to other diseases as long as experimental data is available for virus decay with temperature. It would be particularly useful for endemic diseases, where susceptible supply is no longer having a large effect.

4.1 Limitations

Parameter values make more of a difference to the accuracy of the model than the country does. This suggests that the sampling around parameter values is likely to be important in similar models, particularly as many parameters have to be estimated.

As 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 how climate affects contact rate. Specifically, diary-based social contact studies over long periods could be used. Current work has high temporal resolution (Mossong et al. 2008) but the period of each study is short.

With with more time, data from multiple experiments for viral decay could be combined. At present only one study was used for influenza (Biryukov et al. 2020) and one for COVID-19 (Harper 1961). 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. Differences in human physiology over the year may affect disease transmission. The immune system may vary over the year, potentially mediated by vitamin D or by changes in food availability (Tamerius et al. 2011). Additionally, in dryer and potentially cooler environments blood flow and mucociliary clearance may be lower (Tamerius et al. 2011). This would reduce the white blood cell activity in these areas and slow the removal of pathogens from airways. Additionally, lower temperatures may increase the shedding rate in guinea pigs (Lowen et al. 2007). These factors could not be included because there was not sufficient quantitative information on these effects but if data become available in the future, physiological differences could be included.

5 Conclusion

Previous work acknowledges the importance of seasonality but this is the first attempt to untangle the effects of human behaviour and virus physiology. This model provides a good starting point for exploring the importance of contact rate and virus survival. The level of mismatch effects how close the model is to data so this work highlights the need for more observational studies of human contact rate in different climates. I estimate the temperature at which contact rate peaks for each country by comparing model results to data. In temperate regions, contact rate is consistently highest at low temperatures. This information 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. I present a new model for transmissibility and suggest a difference in synchrony between viral growth rate and contact rate between tropics and temperate regions. This will be useful in improved disease modelling and public health policy, particularly during the current COVID-19 pandemic.

Data and Code Availability

Data and Code are available at GitHub : https://github.com/rbak54/Climate_Virus

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.

Barreca, Alan I. and Jay P. Shimshack (2012). "Absolute humidity, temperature, and influenza mortality: 30 years of county-level evidence from the united states". In: *American Journal of Epidemiology* 176.SUPPL. 7, pp. 114–122. ISSN: 00029262. DOI: 10.1093/aje/kws259.

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. ISSN: 2379-5042. DOI: 10.1128/mSphere.00441-20.

Bjornstad, O.N. (2018). *Epidemics: Models and Data using R*. Switzerland: Springer Nature.

Bjornstad, 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, pp. 75–81. ISSN: 14712970. DOI: 10.1098/rspb.1993.0011.

Cauchemez, Simon 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 et al. (2008). "Estimating the impact of school closure on influenza transmission from Sentinel data". In: *Nature* 452, 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.

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

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

Davies, Nicholas G. et al. (2020). "Age-dependent effects in the transmission and control of COVID-19 epidemics". In: *Nature Medicine* 26.August, pp. 1205–1211. DOI: 10.1038/s41591-020-0962-9.

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, 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, 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.

Jüni, Peter et al. (2020). "Impact of climate and public health interventions on the COVID-19 pandemic: A prospective cohort study". In: *CMAJ* 192.21, E566–E573. ISSN: 14882329. DOI: 10.1503/cmaj.200920.

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

Kissler, Stephen M et al. (2020). "Projecting the transmission dynamics of SARS-CoV-2 through the post-pandemic period". In: *medRxiv*, p. 2020.03.04.20031112. DOI: 10.1101/2020.03.04.20031112.

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, pp. 1083–1087. ISSN: 00368075. DOI: 10.1126/science.1115717.

Lowen, Anice C. 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.

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.

Mordecai, Erin A. et al. (2013). "Optimal temperature for malaria transmission is dramatically lower than previously predicted". In: *Ecology Letters* 16.1, pp. 22–30. ISSN: 1461023X. DOI: 10.1111/ele.12015.

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.

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

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.

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.

Steel, J., P. Palese, and A. C. Lowen (2011). "Transmission of a 2009 Pandemic Influenza Virus Shows a Sensitivity to Temperature and Humidity Similar to That of an H3N2 Seasonal Strain". In: *Journal of Virology* 85.3, pp. 1400–1402. ISSN: 0022-538X. DOI: 10.1128/jvi.02186-10.

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, S737–S746. 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.

Viboud, Cécile et al. (2004). "Association of influenza epidemics with global climate variability". In: *European Journal of Epidemiology* 19, pp. 1055–1059. ISSN: 03932990. DOI: 10.1007/s10654-004-2450-9.

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. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

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

Imperial College London

Imperial College London

Life Sciences

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

Supplementary Information

Author:

Ruth Keane

August 2020

Parameter Meaning Table

Table S1: Table of model parameters

Parameter	Meaning	Unit	Origin
S	Number of Susceptible Individuals	number of individuals(n)	Bjornstad (2018)
E	Number of Exposed Individuals	number of individuals(n)	Bjornstad (2018)
I	Number of Infected Individuals	number of individuals(n)	Bjornstad (2018)
R	Number of Recovered Individuals	number of individuals(n)	Bjornstad (2018)
N	Total Number of Individuals	number of individuals(n)	Bjornstad (2018)
β	transmissibility (number of infected individuals per infected individual per day)	day^{-1}	Bjornstad (2018)
f	rate of loss of immunity	day^{-1}	
ν	natural per capita birth rate	day^{-1}	Keeling and Rohani (2007)
σ	rate of movement from E to I (reciprocal of latent period)	day^{-1}	Bjornstad (2018)
μ	natural per capita death rate	day^{-1}	Bjornstad (2018)
γ	recovery rater	day^{-1}	Bjornstad (2018)
α	disease induced mortality rate	day^{-1}	Bjornstad (2018)
p	case fatality rate	%	Keeling and Rohani (2007)
cr	contact rate per day	day^{-1}	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	day^{-1}	
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	gm^{-3}	
RH	Relative humidity	%	
s	standard deviation of contact rate	day^{-1}	
v	Viability of virus or measure of concentration of virus	%	Harper (1961)
t	time	days	
g	growth constant of decay rate with temperature	T^{-1}	

Parameter Values Table

Table S2: 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).

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 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 (2020b)	0.25/24	Definition of close contact by 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)		

Transmission Model Details

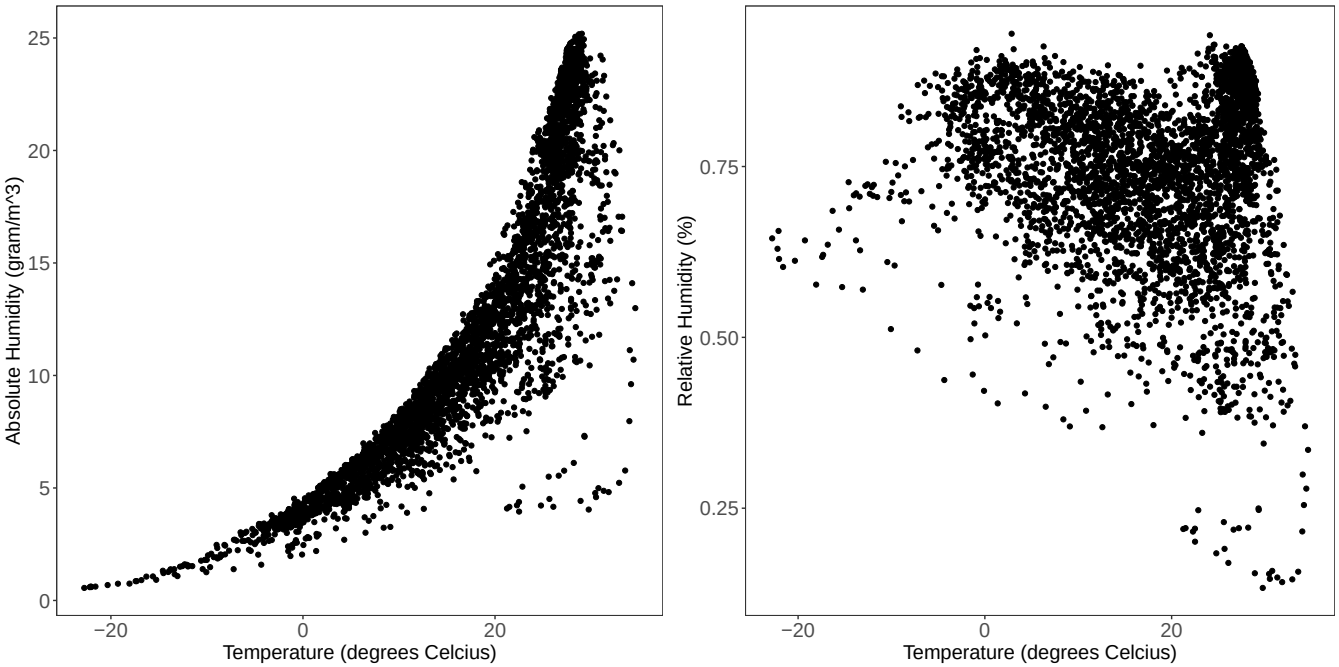
$$\frac{dV}{dt} = a - bV$$

where V is the virus in the environment, a is the shedding rate, b is the rate of decay.

$$V = \frac{a}{b}(1 - e^{-bt})$$

t becomes the average duration of contact d. Then V is divided by itself plus ah where h is the time needed for infection to occur. This cancels and becomes Equation 3

Temperature and Humidity

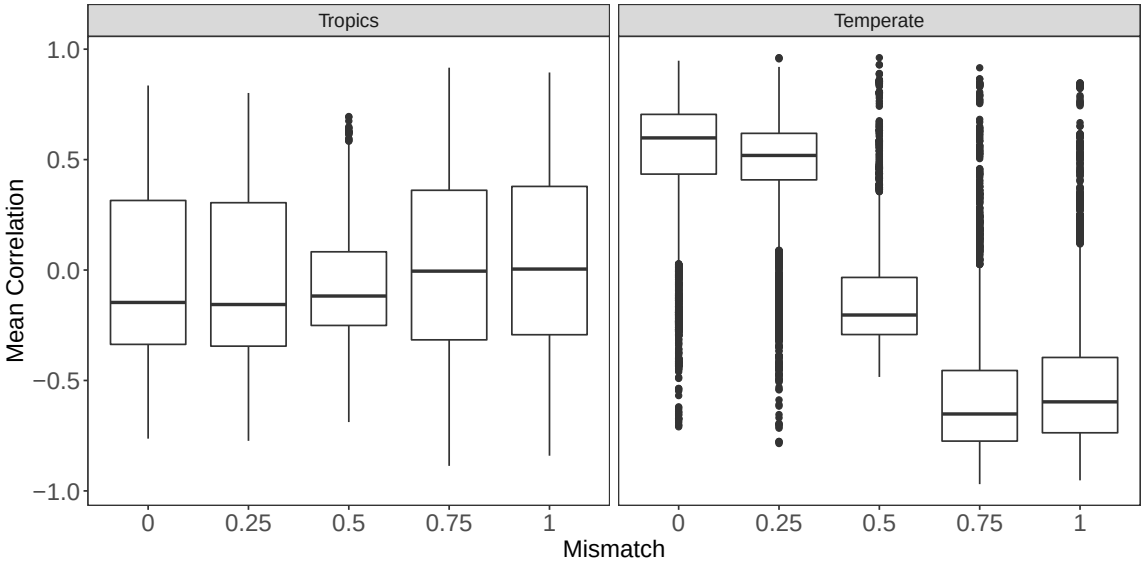


(a) Absolute Humidity

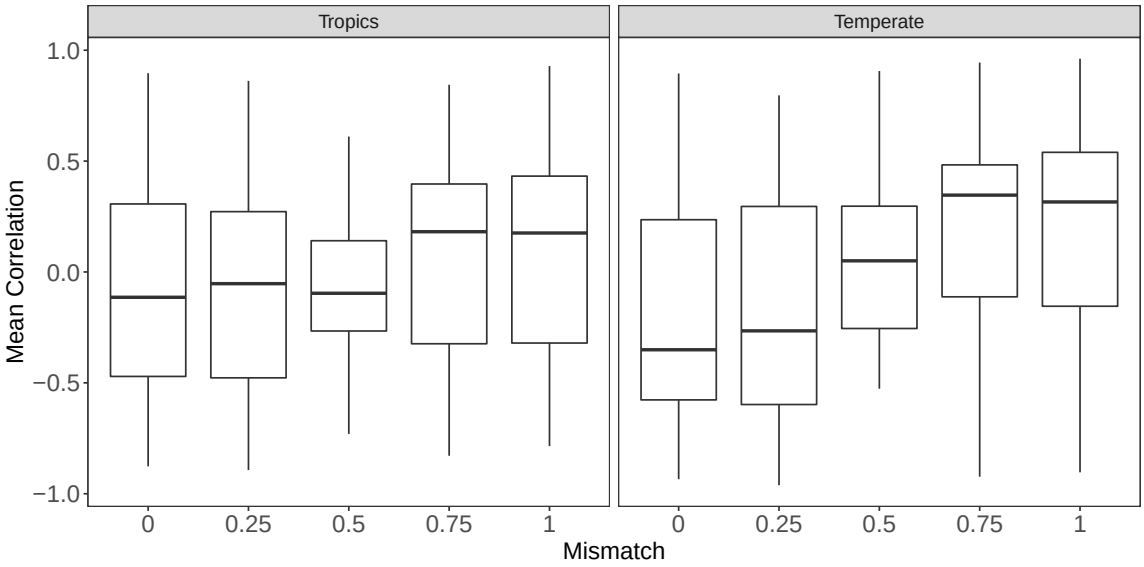
(b) Relative Humidity

Figure S1: Mean weekly temperatures and humidities from Deyle et al. (2016) dataset. Absolute humidity grows exponentially with temperature, relative humidity is more mixed.

Humidity



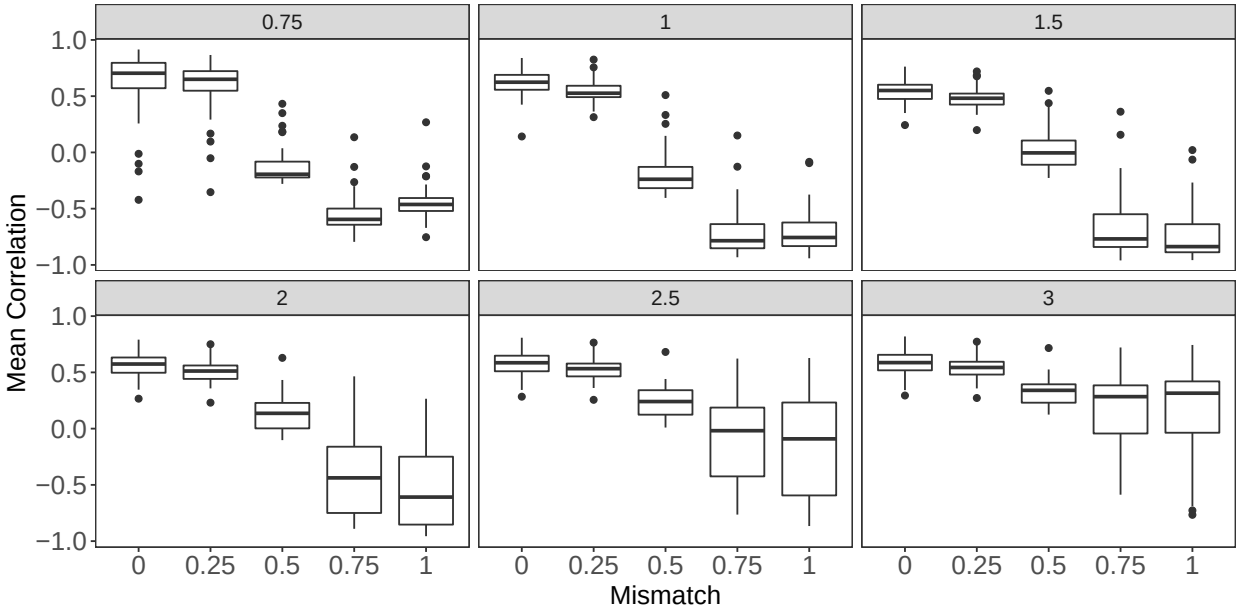
(a) Absolute Humidity: 1 sample



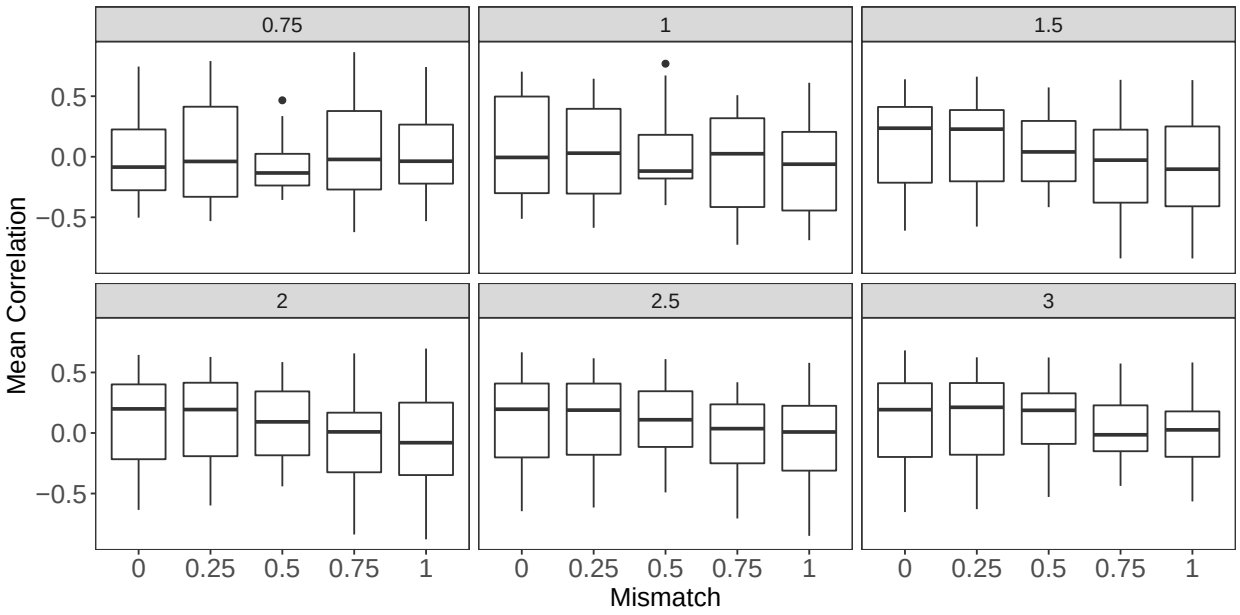
(b) Relative Humidity: 1 sample

Figure S2: 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 is best for absolute humidity in temperate regions. A mismatch of 1 is best for relative humidity in temperate regions. n=29 in tropics, 48 in temperate regions.

Variance



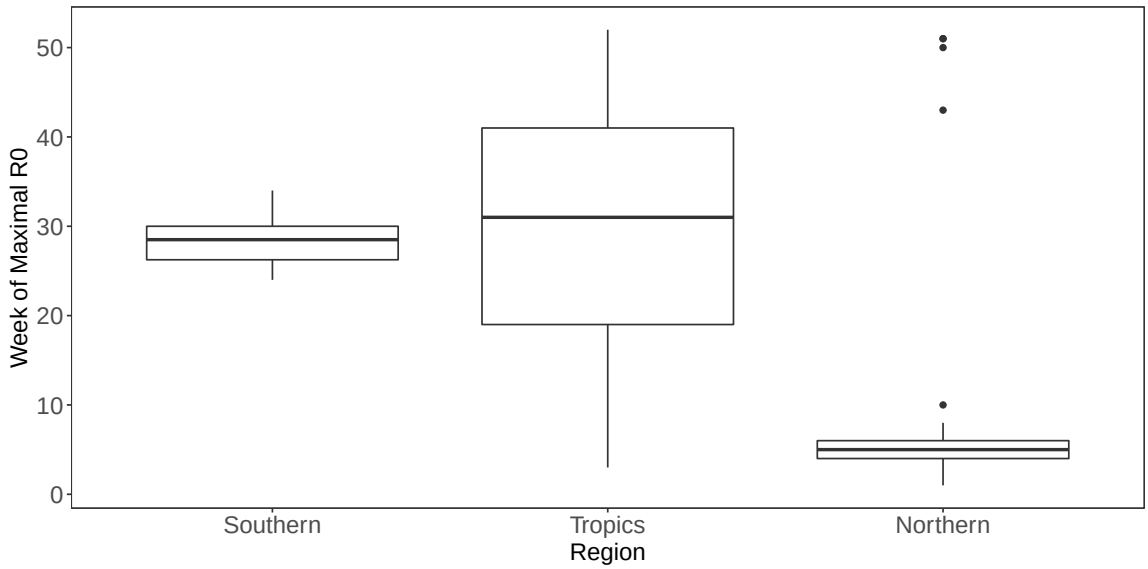
(a) Temperate



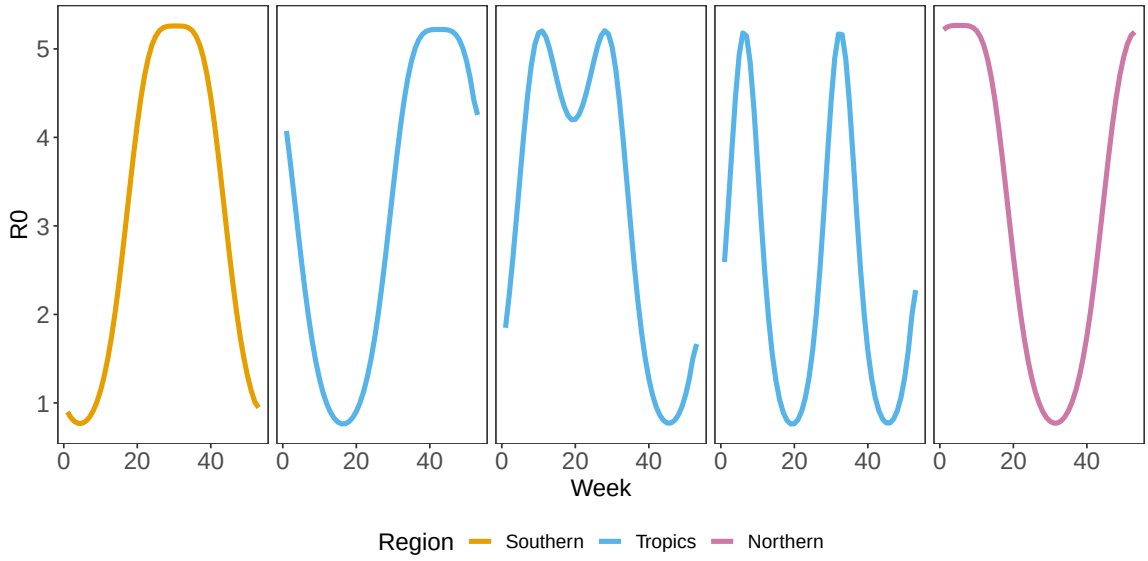
(b) Tropics

Figure S3: Box and whisker plots for correlation and mismatch in temperate and tropical regions where the spread of contact rates has been multiplied by different constants (0.75 to 3). In tropical regions, increasing the constant reduces the difference between mismatches but the overall pattern remains the same. In tropical regions there is no clear pattern.

COVID-19 over the year



(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 each week for five countries. United Kingdom in the Northern Hemisphere, Australia in the Southern Hemisphere, and Cambodia, India and Niger in the Tropics. Northern and Southern Hemisphere have the same basic shape but in the tropics, there are three different shapes of R_0 over time.

Figure S4: COVID-19 Seasonality