Contagion Reproduction

Joel Winterton

October 2024

Code and sources for data: https://github.com/joel-winterton/Contagion/

1 Introduction

This documents and reproduces the model given in the Contagion! paper [2]. The aim of this is to simulate a realistic example disease spread through the UK, using collected mobility data.

2 Data

Mobility data for this model were obtained from the BBC Pandemic dataset. Details about how this dataset was curated are found in [1], but briefly the data is collated from event logs obtained using an app, which log contact events along with the age of the contact and contactee. For this project, the raw data are used, instead of mobility data imputed from models and data.

The UK is divided into local authority districts, and then the number of people whose "home" is in patch j and who make contact with someone in patch i is given in a matrix B_{ij}.

Age data is also considered in this model, and since I could not find the BBC Pandemic age contact data, age mixing data from the POLYMOD study [3] was used instead.

3 Model

The model is split into two different models, within-patch and between-patch. This relies on the assumption that once the chain of infection has been established in a patch, reintroductions from other patches will not drastically affect the within-patch behaviour. This might not be true if the disease leaves a sufficient number of susceptibles in a patch after dying out.

Care is taken to normalise things in such a way that the R_0 of the disease can be explicitly set.

3.1 Within-patch model

Fix patch i, with total population N_i. Divide the population into j age sub-populations, and let n_{ij} be the number of people in sub-population j. Let G_{jk} be the mean number of contacts an individual in the age group k makes with individuals in the age group j. We introduce an infectious profile $\beta(\tau)$ that describes the transmissive behaviour of the disease at time τ. This profile is obtained from an estimated profile for influenza:

τ	1	2	3	4	5
$\beta(\tau)$	0	1.6	0.8	0.2	0.2

An SEIR-like discrete-time model is then created. Let $S_{ij}(t)$ be the proportion of susceptible individuals from patch i and age group j, also denote the proportion of new cases at time t by $I_{ij}(t)$. Then this age group j experiences a force of infection:

$$\Lambda_{ij}(t) = \sum_{\tau=1}^{5} R_0 \beta(\tau) \sum_k A_{jk} I_{ik}(t-\tau)$$

The age mixing matrix A_{jk} is obtained from the mean mixing matrix and then rescaled to have a dominant eigenvalue of 1, so that (using the Next-Generation method) the threshold condition inside the patch is normalised to 1 so

$$A_{jk} = \frac{1}{\rho(\hat{G})} \hat{G}$$

where

$$\hat{G}_{jk} = G_{jk} \frac{n_k}{n_j}$$

and $\rho(G)$ is the dominant eigenvalue of G. The initial conditions are then that each age group has 1% infected in the 5 days leading up to the infection, and that these infections are spread evenly throughout these 5 days.

A quantity that we need is some force of infection that an infected patch exerts:

$$\phi_i(t) = \sum_{\tau=1}^{5} \beta(\tau) \sum_k I_{ik}(t-\tau) n_k$$

The time unit here is local (so t is the time since the start of the infection $t = 0$).

3.2 Mobility components

3.2.1 Spatial kernel

The base mobility kernel was chosen to discriminate between rural and urban movement, which (as the data shows) differs largely. As such let $ru(i)$ be an indicator that gives r if patch i is rural and u if it is urban. The pairwise distances between region centroids was calculated, and the number of visits from home to another location was counted at each distance (rounded up to the

nearest kilometre), with the counts being separated by if the home patch was rural or urban. Each of these lists was then normalised to give a proportion $F_{ru(i)}(d)$, so if $ru(i) = r$ then $F_{ru(i)}(d)$ gives the proportion of all visits from rural homes that go d km.

3.2.2 Long range jumps

Events are only included in the data if centroids were within 100 km of each other, so to account for longer distance jumps, all patches with more than $10,000$ people were selected, and then all pairs i, j of these patches that were over 100km of each other. To represent this an indicator matrix is create J_{ij} that is 1 if these two conditions are met and is 0 otherwise. A small long range jump term can then be added to the force of infection using this indicator matrix.

3.3 Between patch model

When a patch i is infected, it's day of infection τ_i is recorded. If d_{ij} is the distance between region centroids (rounded up to the nearest km), then the force of infection on each patch is modelled as:

$$\lambda_i(t) = \sum_{j \in \mathcal{I}(t)} \varepsilon \phi_j(t - \tau_j) N_i^{\mu} \Big[F_{ru(i)}(d_{ij}) + \epsilon J_{ij} \Big]$$

Then at each time-step we can simulate outbreaks, since the probability of an outbreak is then $P(\tau_i = t) = 1 - \exp(-\lambda_i(t))$.

4 Simulating

The simulation was implemented in Python using the parameters given below:

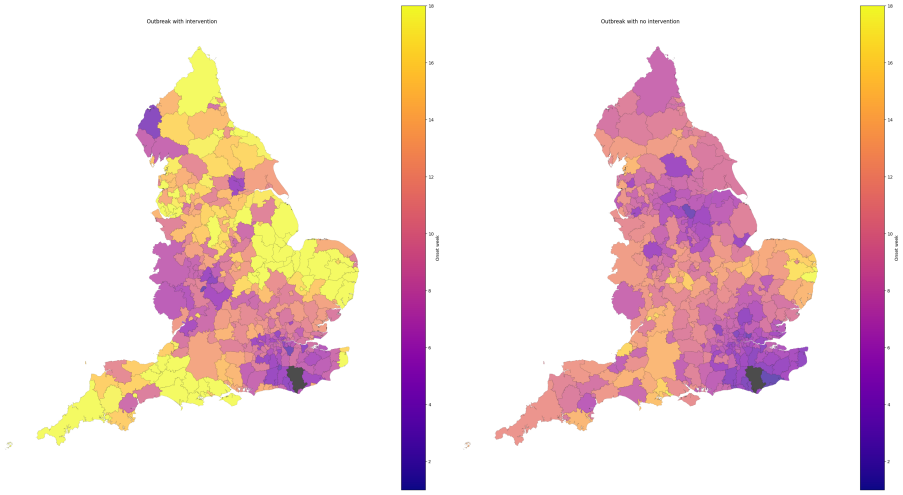
R_0	1.8
$\beta(\tau)$	As in previous table
ε	10
μ	0.32

5 Results

References

[1] Andrew Jk Conlan, Petra Klepac, Adam J Kucharski, Stephen Kissler, Maria L Tang, Hannah Fry, and Julia R Gog. Human mobility data from the BBC Pandemic project, February 2021.

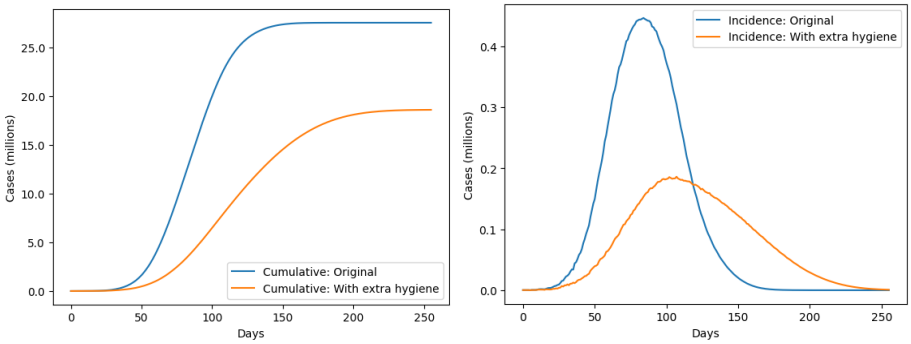
[2] Petra Klepac, Stephen Kissler, and Julia Gog. Contagion! The BBC Four Pandemic – The model behind the documentary. *Epidemics*, 24:49–59, September 2018.



(a) $R_0 = 1.41$ (b) $R_0 = 1.8$

Figure 1: Spatial visualisations of onset with and without interventions (lower R_0 is intervention). Outbreak location in black in lower right corner.

[3] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine*, 5(3):e74, March 2008.



(a) Cumulative infections over time.

(b) New cases over time.

Figure 2: Outbreak in numbers, showing the estimated difference in basic hygiene measures. Control measures prevent about 10 million cases, and the peak in infections is later and much smaller in basic hygiene scenario, but seems to last slightly longer.