Understanding the 2003 Outbreak of Avian Influenza in the Netherlands

Rowland Seymour, Dr Theodore Kypraios, Prof Philip O'Neill
School of Mathematical Sciences, University of Nottingham

University of Nottingham
UK | CHINA | MALAYSIA

Abstract

In 2003, an outbreak of the H7N7 Avian Influenza virus occurred in the Netherlands, and as a result 30 million birds were culled, 89 people were infected and one person died. We show how we can estimate how the disease spreads between bird farms, without having to specify the exact form of the infection rate. Our nonparametric method requires the use of very large and dense matrices, which need to be repeatedly updated and inverted, techniques that are well suited to High Performance Computing (HPC). As we only observe the culling dates of the farm, we use data augmentation to infer the infection dates of each farm, and the infection status of farms that were not tested for the virus. This method is well suited to parallel computation using the HPC service. We also show how we can use the HPC system to compute the infectious force each infected farm exerts on the susceptible farms. Finally, will explain how the HPC service can be used to simulate hundreds of outbreaks of Avian Influenza in order to test our method for estimating the infection rate.

The Outbreak

Highly Pathogenic Avian Influenza is a virus which poses a large threat to animals on poultry farms. In 2003, over 1200 of the 5000 farms in the Netherlands were affected by the virus, leading to large scale culling, the introduction of biocontainment measures, and the death of a veterinarian.

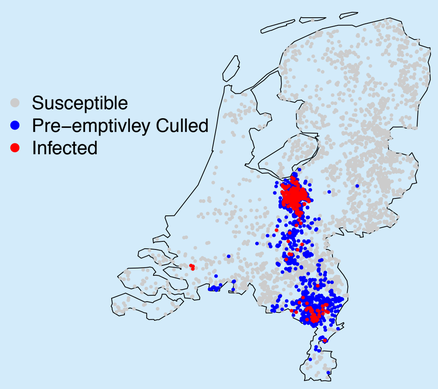

Figure 2. The extent of the Avian Influenza outbreak. Red farms were confirmed to be infected and culled, blue farms were culled without being tested, and grey farms were not infected.

To model the outbreak we build a compartmental model, where farms are either **S**usceptible, **I**nfected or **R**emoved.

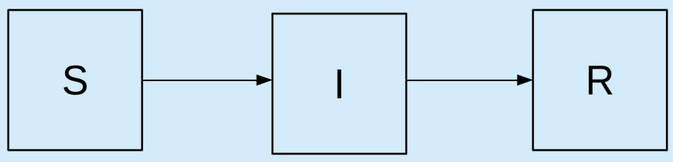

Figure 3. A typical SIR model.

We are interested in the rate at which farms move between classes. As the infected farms are clustered together, we assume the infection rate between any two farms depends on the distance between them. Standard modelling methods involve assuming a given form of the infection rate and inferring the model parameters. An example of this is given below.

$$\beta_{i,j} = \beta_0 \exp\{-\beta_1 d(i,j)\}.$$

This can be restrictive as we need to write down the exact form of the function, or repeat this with several functions and use model selection methods to choose the best one.

We have developed a new method to do this nonparametrically. This method however involves large dense matrices and uses Markov Chain Monte Carlo (MCMC) methods, where we need to repeatedly evaluate the likelihood function. As we do not observe the dates on which the farms are infected, we also need to use data augmentation to infer these dates.

Acknowledgements

This work is supported by EPSRC (EP/N50970X/1).

Bayesian Nonparametric Methods

To construct the likelihood function, we use the infectious pressure paradigm. Each infected farm exerts infectious pressure onto susceptible farms for the length of their infection. We write the likelihood function as

$$\pi(\mathbf{i},\mathbf{r}|\boldsymbol{\beta},\gamma) \propto \underbrace{\exp\Big(-\sum_{j=1}^{n}\sum_{k=1}^{N}\beta_{j,k}\big((r_j \wedge i_k)-(i_j \wedge i_k)\big)\Big)}_{\text{Total Infectious Pressure}} \times \underbrace{\prod_{j=1}^{n}\Big(\sum_{k\in\mathcal{Y}_j}\beta_{k,j}\Big)}_{\text{Pressure infectives put on each susceptible}} \times \underbrace{\prod_{j=1}^{n} f(r_j - i_j|\gamma)}_{\text{Infectious period distribution}}.$$

We put a Gaussian Process prior on the infection rate β, this allows us to non parametrically estimate this parameter. To do this, we construct a covariance matrix using the distances between every pair of farms in the Netherlands, giving us a dense matrix with over 14 million elements.

$$\beta \sim \mathcal{GP}(\mathbf{0},\, \boldsymbol{\Sigma}_d).$$

We use MCMC to estimate the parameters in the likelihood function alongside the infection times. For each outbreak, we want to infer the infection rate, infection times, and length of infectious period. For each MCMC iteration, we

- suggest changing 200 infection times,
- decompose the covariance matrix and suggest new parameters to control the covariance,
- and propose a change to the infection rate.

At each of these steps we recompute the likelihood function.

Using HPC and Results

We first run a simulation study to asses our new method. We use the HPC to simulate 250 outbreaks of a disease with a known infection rate across 1,000 farms. We then run our algorithm on each outbreak to infer the model parameters and infection rate. Using one node this takes roughly four days, however using OpenMP we can reduce this to by about 35%. The results for this study are shown in figure 4.

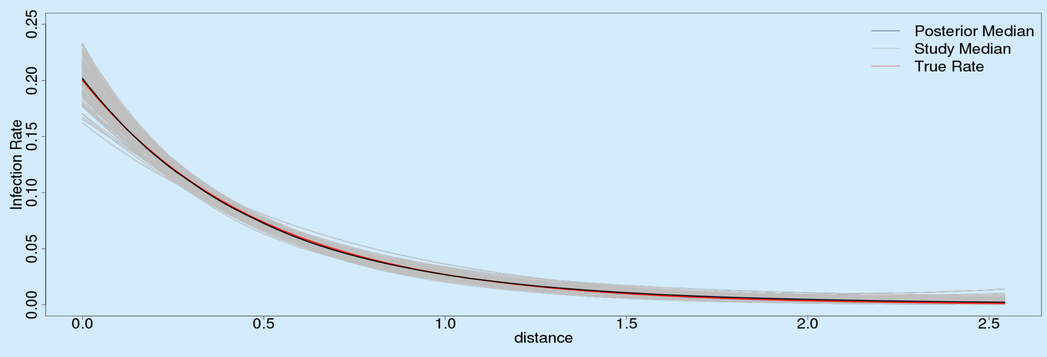

Figure 4. The results of the simulation study compared against the true infection rate

We now run our algorithm on the Avian Influenza data set. These are preliminary results assuming a fixed infectious period. We can compare this to our best parametric model. The results of the parametric and nonparametric results are similar but differ slightly suggesting that our new nonparametric method allows greater flexibility.

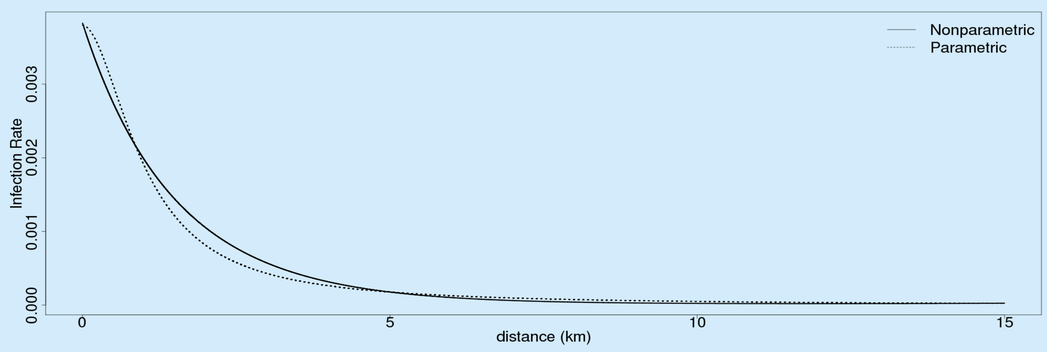

Figure 5. Nonparametric and parametric estimates for the Avian Influenza Infection rate.