

Bayesian Nonparametric Methods for Spatial Stochastic Epidemic Models

Rowland Seymour*, Theodore Kypraios, Philip O'Neill
School of Mathematical Sciences, University of Nottingham, UK

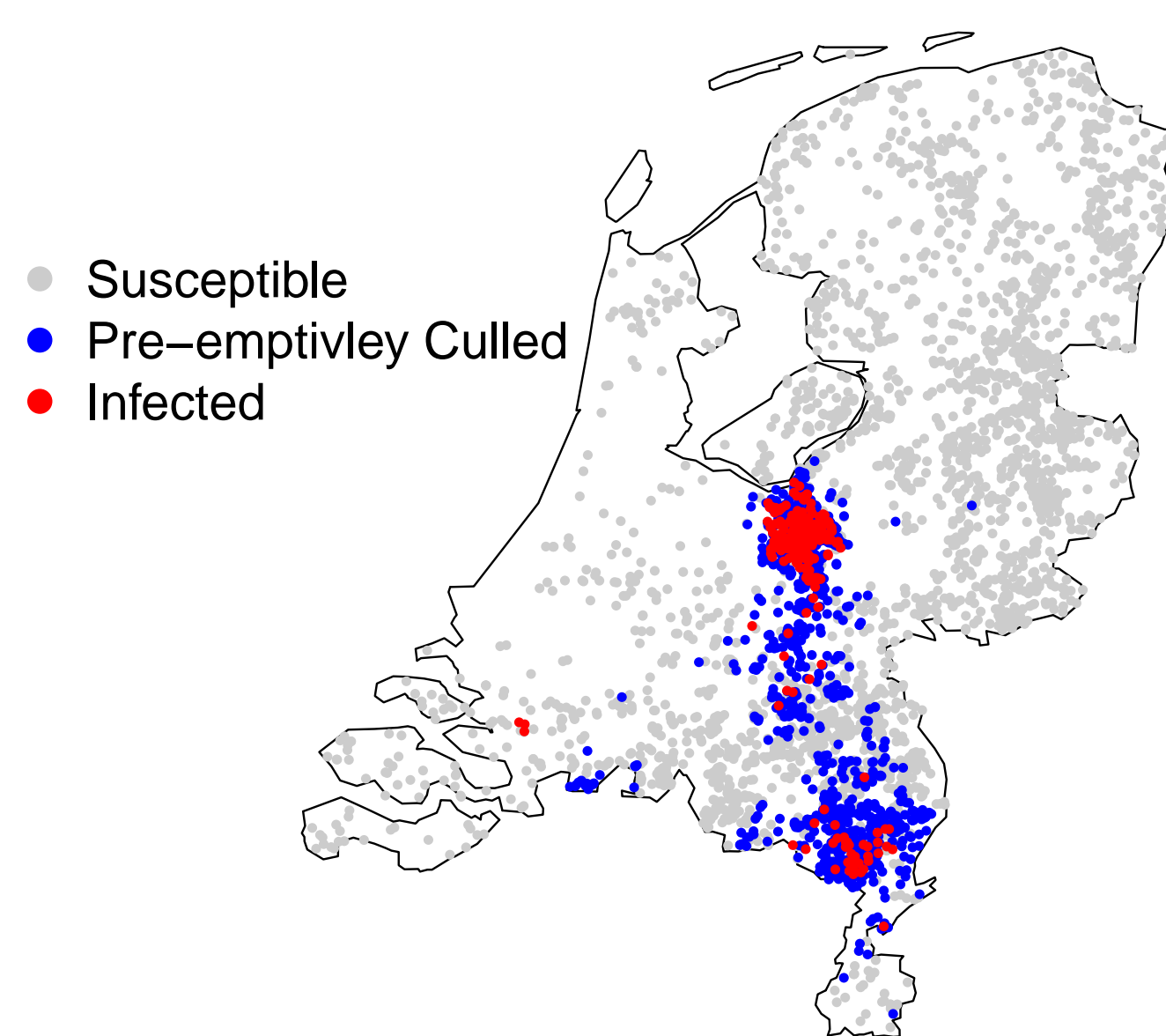
*rowland.seymour@nottingham.ac.uk

Statistical Epidemiology

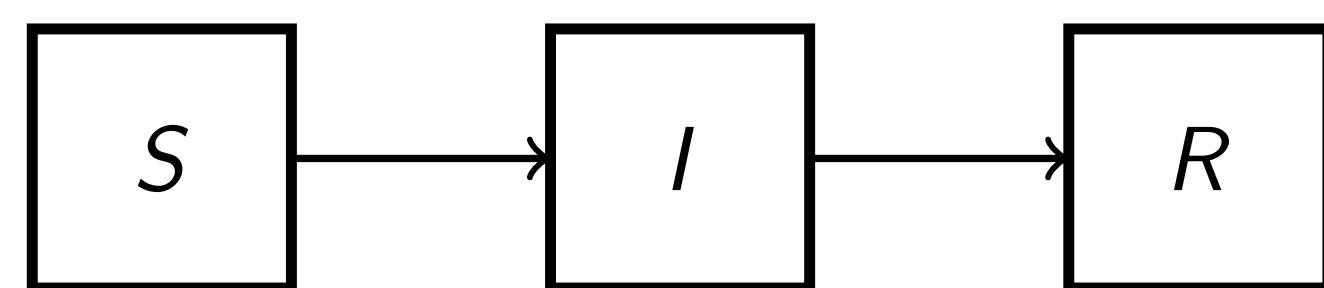
When modelling outbreaks of diseases on farms, we are mainly interested in three questions:

- How did the disease spread between farms,
- Which farms were infected, and
- For long were they infected?

The map below shows an outbreak of Avian Influenza in the Netherlands in 2003, where red farms were culled because they were infected and blue farms were preemptively culled without being tested for the disease.



We model the outbreak by assuming that at any point, each farm is either **Susceptible**, **Infected**, or **Removed**.



The outbreak proceeds as follows:

- An infected farm i infects farm j according to a homogeneous Poisson process with rate $\beta_{i,j}$,
- Once infected, a farm remains so for a time period drawn from a $\Gamma(\lambda, \gamma)$ distribution,
- Once this time has elapsed cull the animals on this farm and all farms within a given radius,
- The outbreak is over when there are no infected farms remaining.

Inference for this model is challenging as we only observe the dates farms were culled, and not the dates they were infected or whether the preemptively culled farms were infected.

The majority of methods in epidemic modelling are parametric, and our novel BNP framework allows for a more flexible method where we make fewer modelling assumptions.

Acknowledgements

This work is supported by the UK's Engineering and Physical Sciences Research Council. (EP/N50970X/1)

Bayesian Nonparametric Methods

We model the infection rate as a function of the Euclidean distance between farms.

$$\beta_{i,j} = \beta(d_{i,j})$$

We place a Gaussian Process (GP) prior distribution on β such that

$$\beta = \exp(f) \quad f \sim \mathcal{GP}(\mathbf{0}, \Sigma) \quad \Sigma = k(\mathbf{d}, \mathbf{d}; \alpha, l).$$

We think of the GP prior distribution as a prior distribution over the space of all possible functions, and specify this space through the covariance function k .

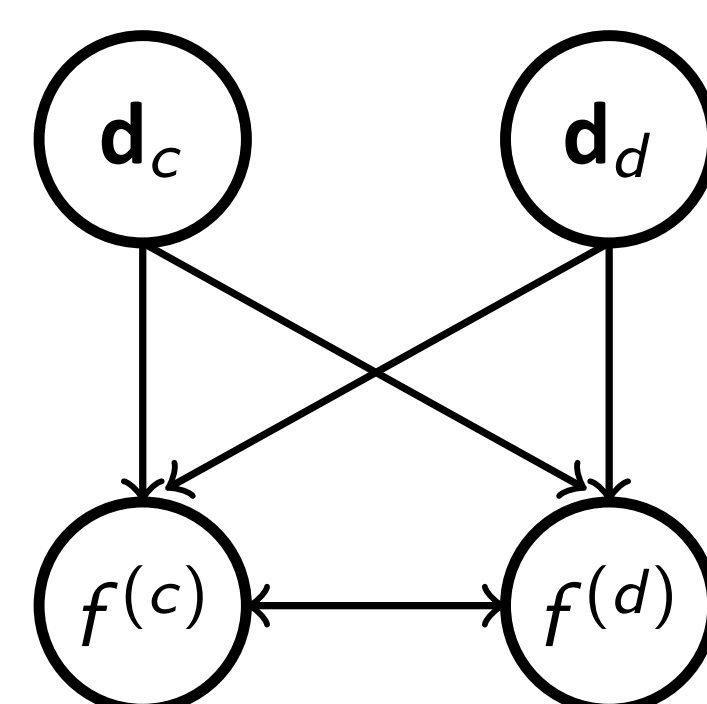
We can extend the model, by allowing the infection rate function to depend the Euclidean distance, as well as the type of susceptible farm, e.g Chicken and Duck.

Our first method is to allow for functions to be correlated in the joint prior distribution:

$$\beta^{(c)} = \exp(f^{(c)}), \quad \beta^{(d)} = \exp(f^{(d)})$$

$$\begin{pmatrix} f^{(c)} \\ f^{(d)} \end{pmatrix} \sim \mathcal{GP} \left(\mathbf{0}, \begin{pmatrix} k(\mathbf{d}_c, \mathbf{d}_c) & \rho k(\mathbf{d}_c, \mathbf{d}_d) \\ \rho k(\mathbf{d}_d, \mathbf{d}_c) & k(\mathbf{d}_d, \mathbf{d}_d) \end{pmatrix} \right)$$

We can represent this as:

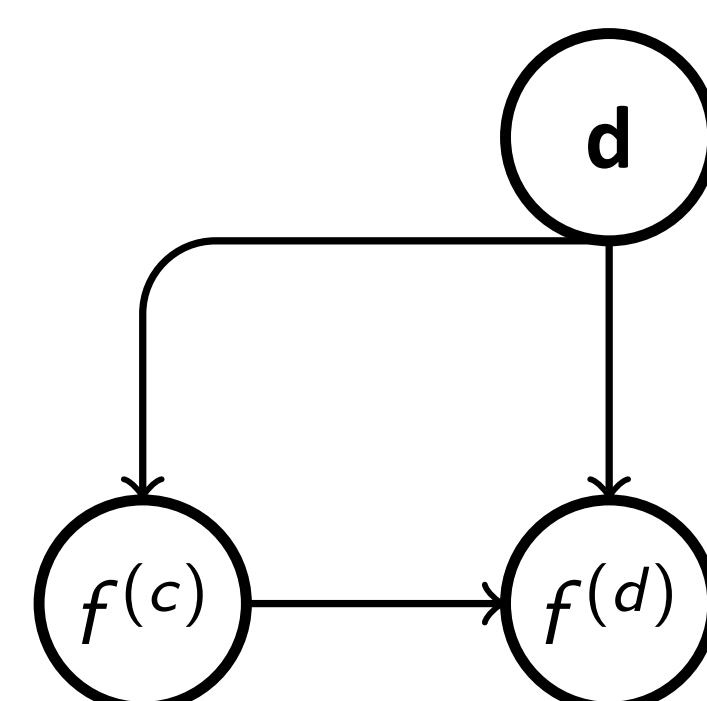


Our second method is to allow the infection rate function for ducks to depend on that of chickens:

$$\beta^{(c)} = \exp(f^{(c)}), \quad f^{(c)} \sim \mathcal{GP}(\mathbf{0}, \Sigma_c)$$

$$\beta^{(d)} = \exp(f^{(d)}), \quad f^{(d)} \sim \mathcal{GP}(f^{(c)}, \Sigma_d)$$

This requires an approximation method to compute the discrepancy between two functions on heterotropic data. The graphical version of the model is:



For large or heterotropic data sets, we use a projection approximation. We fit a GP prior distribution to a smaller pseudo-data set:

$$f^* \sim \mathcal{GP}(\mathbf{0}, \Sigma^*) \quad \Sigma^* = k(\mathbf{d}^*, \mathbf{d}^*).$$

We then compute the expected function given the pseudo function:

$$f = k(\mathbf{d}, \mathbf{d}^*) \Sigma^{*-1} f^*.$$

Markov Chain Monte Carlo

The likelihood function is given by

$$\pi(\mathbf{i}, \mathbf{r}, B, C, D | \beta, \lambda, \gamma, \phi, i_\phi) =$$

$$\exp \{ -\Psi \} \prod_{\substack{j=1 \\ j \neq \phi}}^n \left(\sum_{k \in \mathcal{Y}_j} \beta(d_{k,j}) \right)$$

$$\times \prod_{j \in B} f_D(r_j - i_j | \lambda, \gamma) \prod_{j \in C} S_D(r_j - i_j | \lambda, \gamma),$$

$$\Psi = \sum_{j=1}^n \left[\sum_{k \in A \cup B \cup C} \beta(d_{j,k}) ((r_j \wedge i_k) - (i_j \wedge i_k)) \right.$$

$$\left. + \sum_{k \in D} \beta(d_{j,k}) ((r_j \wedge r_k) - (i_j \wedge r_k)) \right],$$

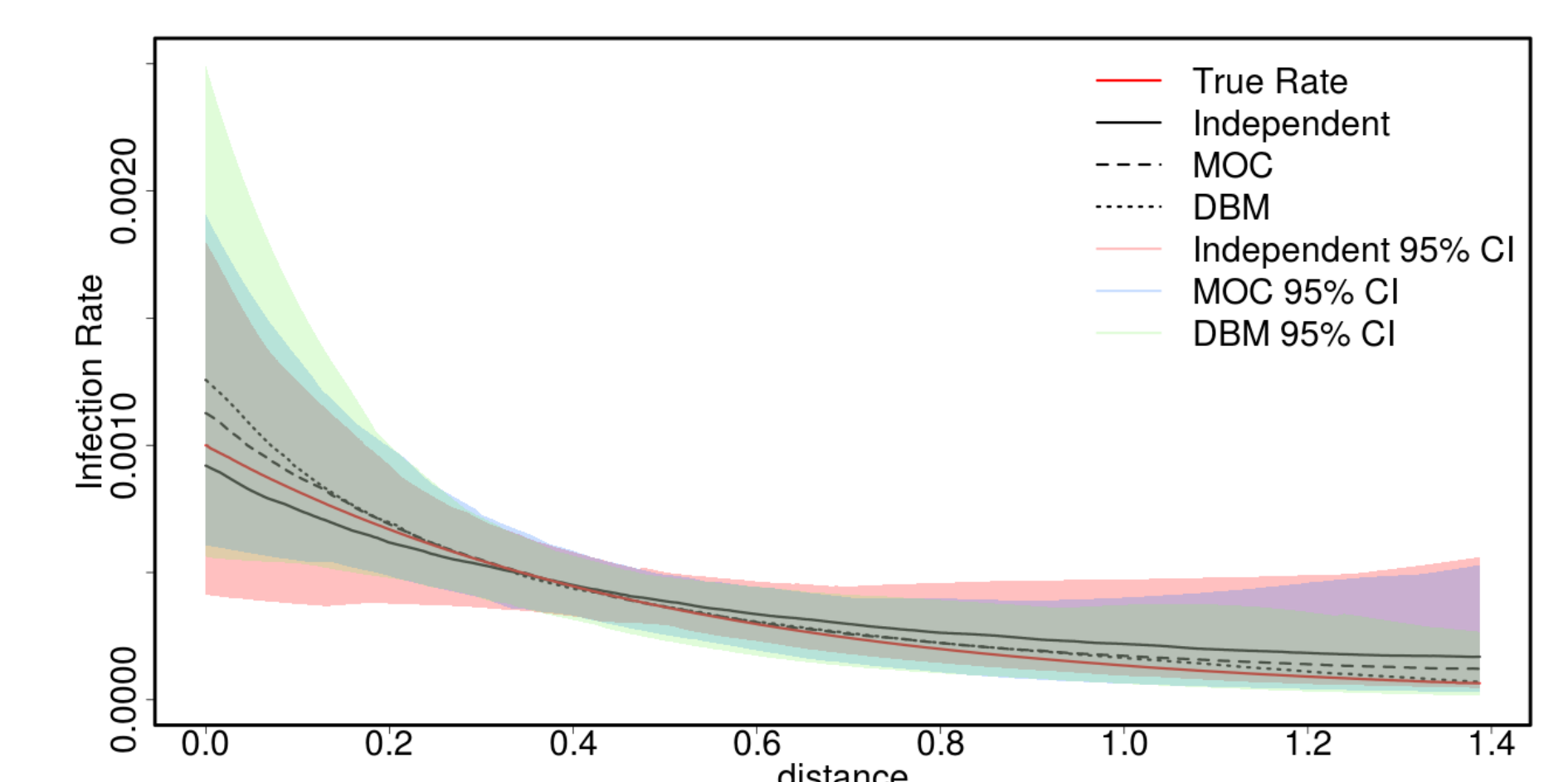
where B is the set of farms which were confirmed to be infected, C is the set of infected farms which were preemptively culled, D is the set of farms which were not infected but preemptively culled, and f_d and S_D are the probability density survival functions of the Gamma distribution.

To sample from the posterior distribution, we use an MCMC algorithm. To update the infection rate function, we use an underrelaxed proposal mechanism, where

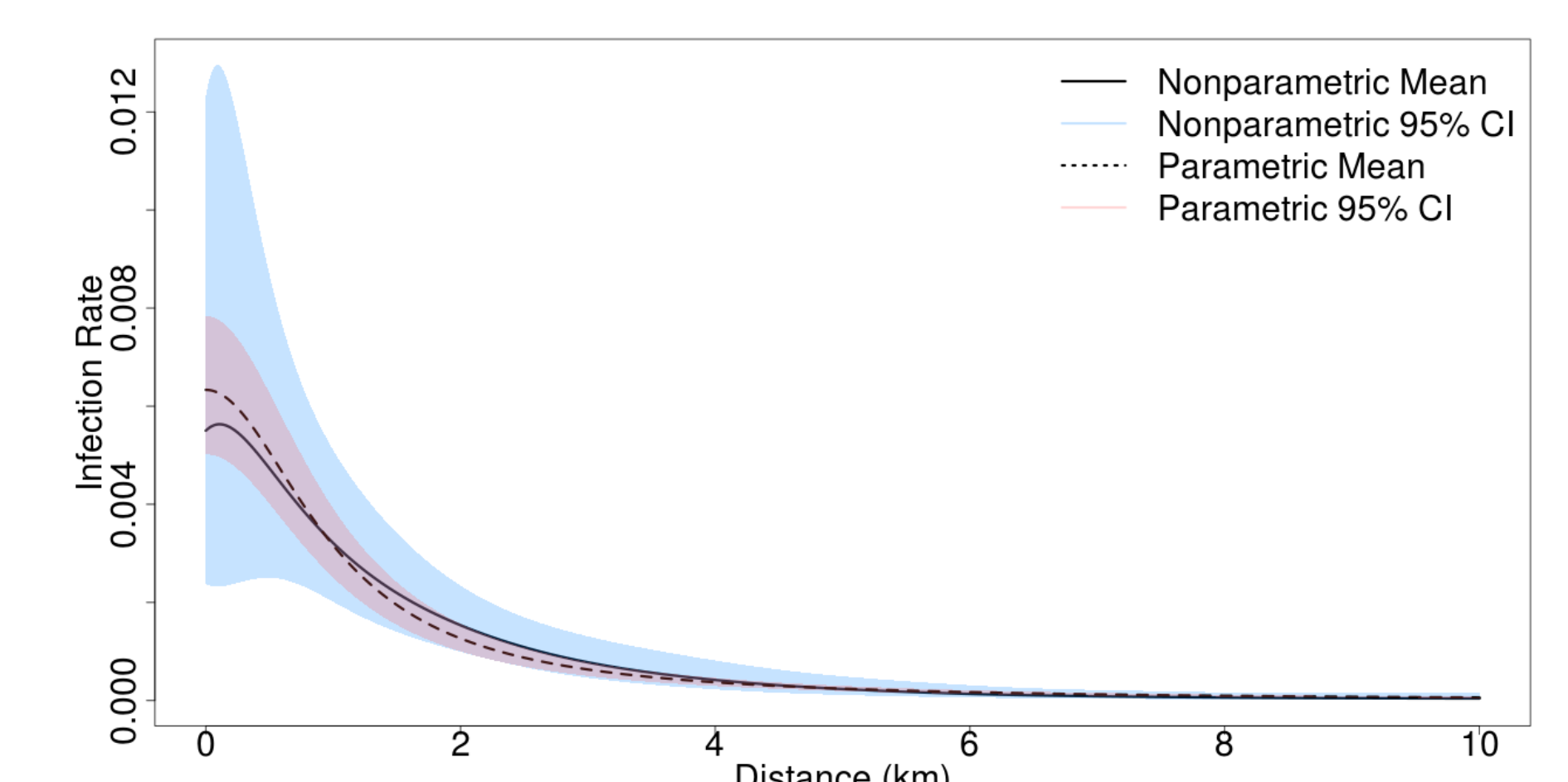
$$f' = \sqrt{1 - \delta^2} f + \delta \nu, \quad \nu \sim \mathcal{GP}(\mathbf{0}, \Sigma), \quad \delta \in (0, 1]$$

Results

We compare the models using a simulated data set of 1,000 individuals. The figure below shows the average infection rate for ducks for 250 outbreaks.



We now infer the infection rate for the Avian Influenza outbreak, assuming there was no difference between the types of farms.



We also estimate the mean infectious period to be 6.4 days, and around 20 farms to have been infected before the authorities declared an outbreak. Compared to the parametric method, the uncertainty for the nonparametric method is much larger.