

# Nonparametric Bayesian Inference for Heterogeneously Mixing Epidemic Models

Rowland Seymour, Prof. Philip O'Neill and Dr Theodore  
Kypraios

University of Nottingham



RSC 2018  
SHEFFIELD



University of  
**Nottingham**  
UK | CHINA | MALAYSIA

# Table of Contents

- 1** Epidemic Modelling
- 2** Homogeneous and Heterogenous Mixing
- 3** Parametric and Non-Parametric Models
- 4** Conclusion

# Avian Influenza

## Approach and Aims

From the outbreak of Avian Influenza, we observe the following data:

ID	Coordinates	Status	Culling Date
1	(5.32, 18.82)	Susceptible	NA
2	(2.90, 15.67)	Susceptible	NA
3	(2.86, 17.99)	Pre-Emptively Culled	3 <sup>rd</sup> May
4	(4.56, 18.01)	Culled	30 <sup>th</sup> April
:	:	:	:

Using this data, we want to work out:

- If there is a spatial element to the spread of the disease?
- If so, how can we quantify this element?

# Table of Contents

- 1** Epidemic Modelling
- 2** Homogeneous and Heterogenous Mixing
- 3** Parametric and Non-Parametric Models
- 4** Conclusion

# Homogeneous Mixing

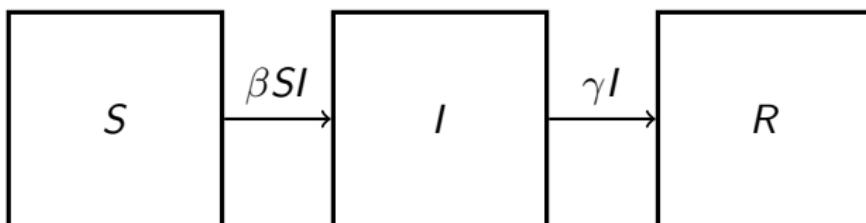


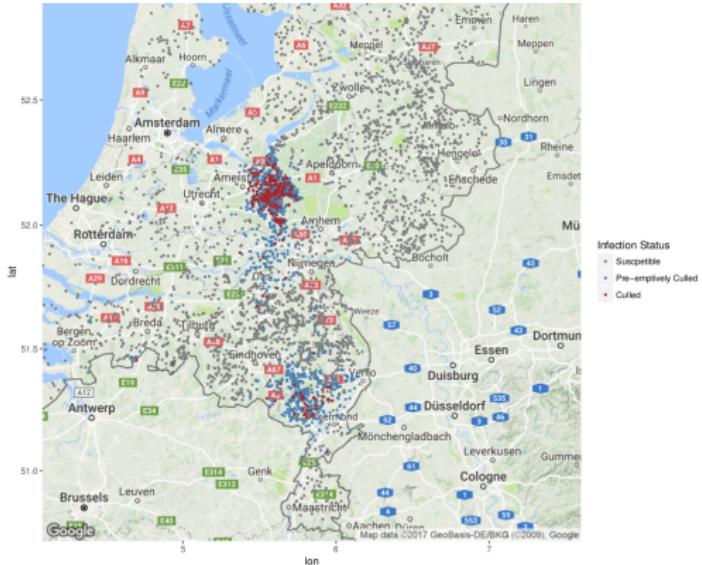
Figure: A homogeneously mixing SIR model.

Farms are either **Susceptible**, **Infected** or **Removed**.

- **Infections:** Infections occur according to a Poisson process with rate  $\beta S_t I_t$ .
- **Removals:** Individual  $j$  remains infectious for a time period drawn from an Exponential distribution, such that  $r_j - i_j \sim \text{Exp}(\gamma)$ .

# Drawbacks of Homogeneous Mixing

- Suppose one farm is infected and  $m$  are susceptible.
- We generate the  $x$ , the number of infections, by sampling from a  $\text{Poisson}(\beta m)$  distribution.
- We then choose  $x$  farms uniformly at random to be infected.



## Heterogeneous Mixing

We need to include the distance between farms in the model. We compute the euclidean distance between each pair of farms farm  $i$  and  $j$ ,  $d_{i,j}$  and we make  $\beta$  a function of the distance.

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

Our heterogeneously mixing model builds on the homogeneously model, as we specify the infection and culling rates as follows:

- **Infections:** Infections occur according to an inhomogeneous Poisson process with a unique rate for each pair of farms, specified by  $\beta_{i,j} = f(d_{i,j})$ .
- **Removals:** Individual  $j$  remains infectious for a time period drawn from an Exponential distribution, such that  $r_j - i_j \sim \text{Exp}(\gamma)$ .

# Table of Contents

- 1** Epidemic Modelling
- 2** Homogeneous and Heterogenous Mixing
- 3** Parametric and Non-Parametric Models
- 4** Conclusion

# Parametric Infection Rates

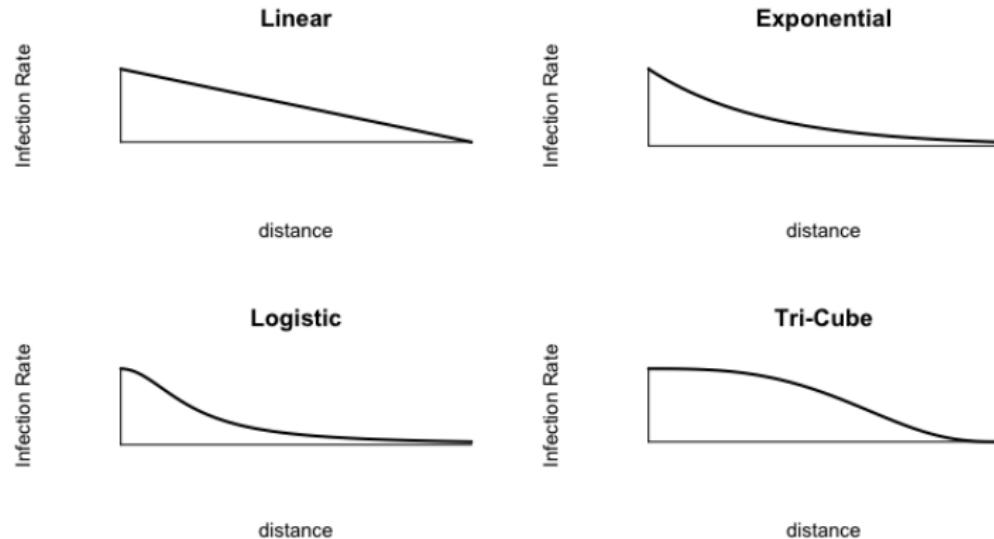


Figure: Four Possible Infection Rates.

# Parametric Models for Avian Influenza

Parametric inference for this data set has already been carried out by Boender *et al.* (2007). They proposed five models and used AIC to choose the best of the proposed models.

Model	Infection Rate
1	$\beta_{i,j} = \beta_0$
2	$\beta_{i,j} = \frac{\beta_0}{1+d_{i,j}}$
3	$\beta_{i,j} = \frac{\beta_0}{1+d_{i,j}^2}$
4	$\beta_{i,j} = \frac{\beta_0}{1+d_{i,j}^\alpha}$
5	$\beta_{i,j} = \frac{\beta_0}{1+(d_{i,j}/\beta_1)^\alpha}$

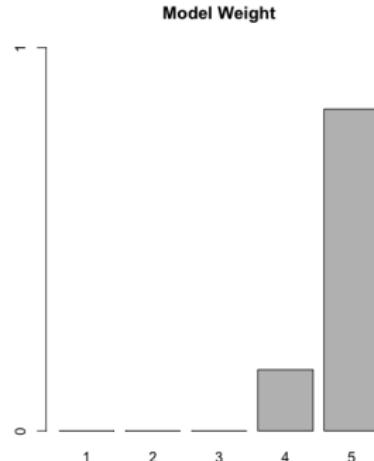


Figure: Model Weights for the different Infection Rates.

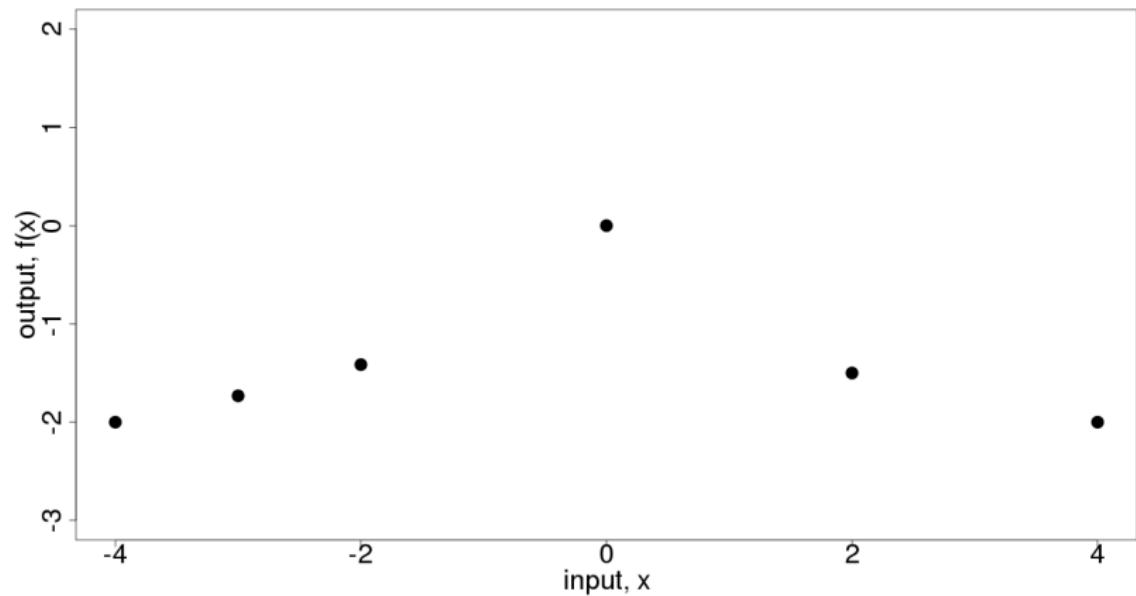
# Inference for Nonparametric Models

Our nonparametric methodology use Gaussian Processes (GPs) to estimate  $\beta_{i,j}$ .

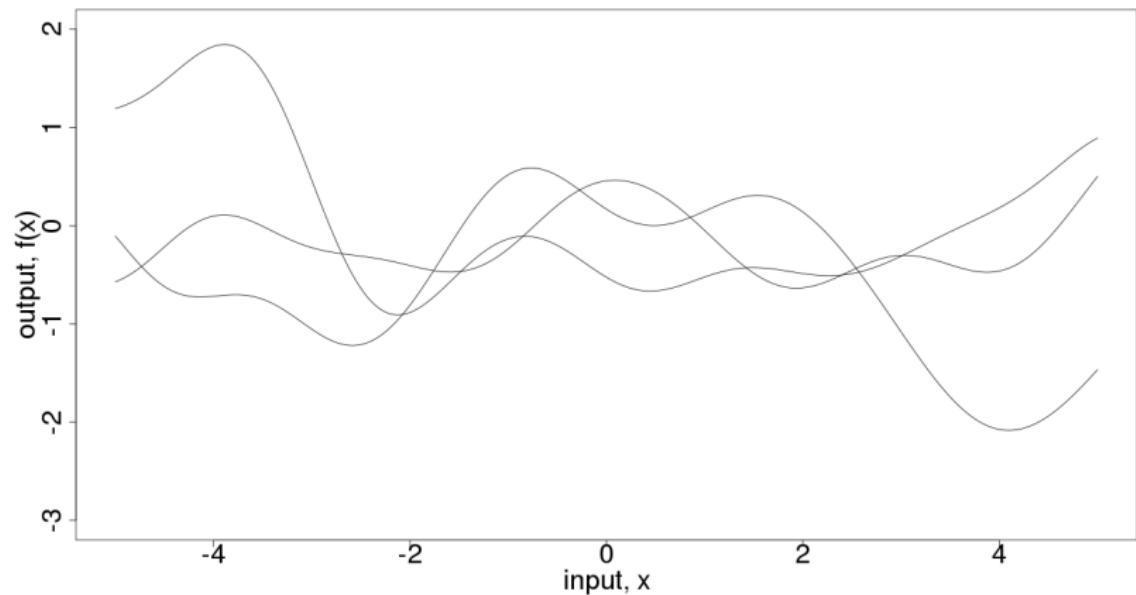
- GPs are a generalisation of the multivariate Gaussian distribution to a function space.
- We use GPs to assign a prior distribution over a function space.
- We need to control the signal variance and volatility.

$$f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, \Sigma)$$

# Data



# GP Prior Distribution



# Posterior Distribution

# Avian Influenza Likelihood

The augmented likelihood function for this model is given by

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

## Avian Influenza Priors

We put a GP prior on the infection rate  $\beta$ , where the covariance matrix is based on the distances between the farms.

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

We put a vague Exponential prior on the infectious period distribution parameter.

$$\gamma \sim \text{Exp}(0.01).$$

We use MCMC to estimate  $\beta$  and  $\gamma$ , as well as estimating the day each farm was infected.

# MCMC Algorithm

---

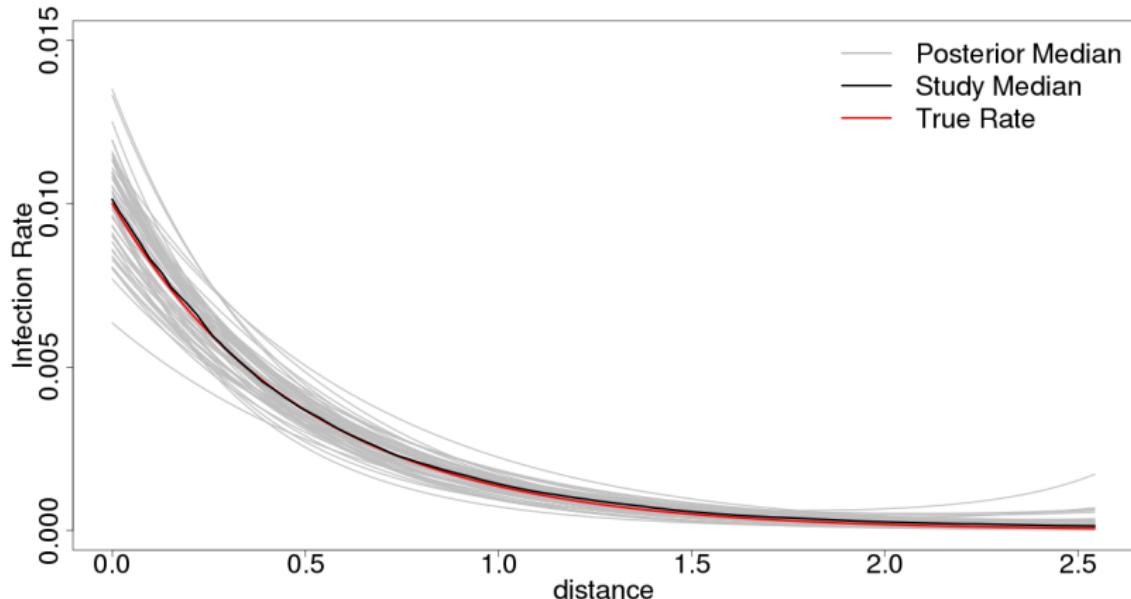
**Algorithm 1** Structure of the MCMC algorithm

---

- 1: Initialise the chain with estimates  $\gamma^{(0)}$ ,  $\mathbf{f}^{(0)}$ ,  $I^{(0)}$ , and  $\mathbf{i}^{(0)}$   
*Repeat the following steps*
  - 2: Sample  $\gamma$  from the conditional distribution  $\pi(\gamma|\lambda, \mathbf{i}, \mathbf{r}, \chi_\gamma)$  using a Gibbs sampler
  - 3: Sample  $\mathbf{f}$  using an underrelaxed proposal method for a Metropolis Hastings algorithm
  - 4: Sample  $I$  using a Metropolis Hastings algorithm
  - 5: Update  $\mathbf{i}$ , the vector of infection times
-

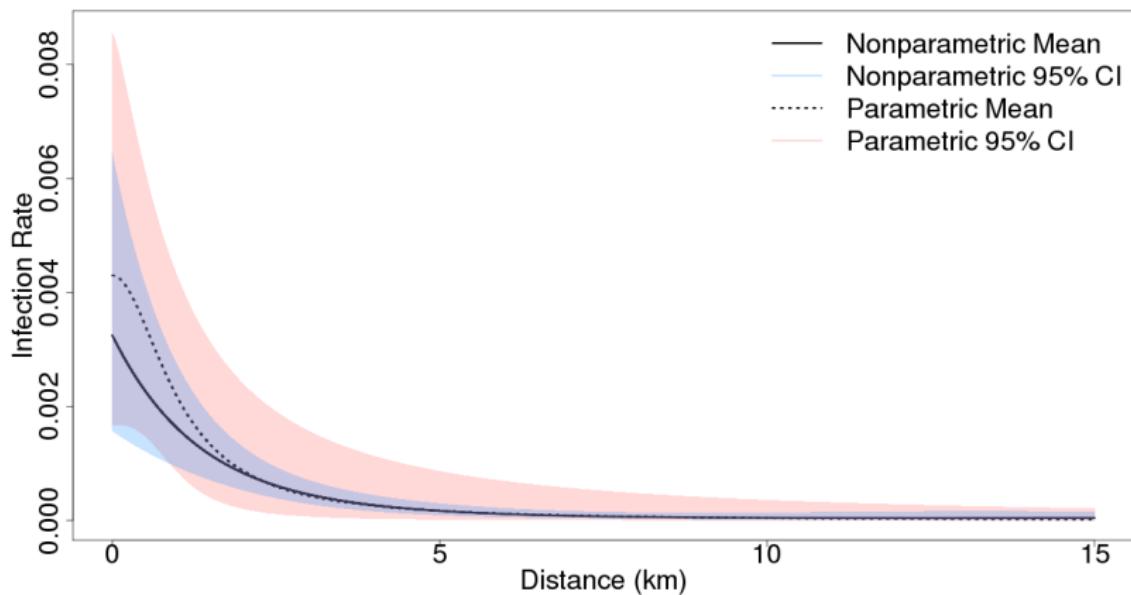
# Simulation Study

We generated the positions of 1000 farms, and simulated 250 outbreaks with the same infection rate. Only using the positions and removal times, we estimated the infection rate.



# Avian Influenza Results

We then applied the GP inference method to the Avian Influenza data set and compared our model to the parametric model.



# Table of Contents

- 1** Epidemic Modelling
- 2** Homogeneous and Heterogenous Mixing
- 3** Parametric and Non-Parametric Models
- 4** Conclusion

# Conclusion

Throughout this talk, we have shown that:

- We can model infection rates non-parametrically with GPs,
- We do not need to make restrictive assumptions about the parametric form of the infection rate,
- We can do this for large data sets without having to deal with large covariance matrices.

We can improve and build on this method by:

- Including more covariates, such as the number of animals on each farm,
- Extend this to include information about the type of animals on each farm.

## References

- Boender, G. J., Hagenaars, T. J., Bouma, A., Nodelijk, G., Elbers, A. R. W., de Jong, M. C. C. and van Boven, M. (2007). Risk Maps for the Spread of Highly Pathogenic Avian Influenza in Poultry. *PLoS Computational Biology*, 3(4), pp. 704-712.
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. USA: MIT Press.
- Xu, X., Kypraios, T. and O'Neill, P. D. (2016). Bayesian non-parametric inference for stochastic epidemic models using Gaussian Processes. *Biostatistics*, 17(4), pp. 619-633.
- Jewell, C. P., Kypraios, T., Christley, R. M. and Roberts, G. O. (2009). A novel approach to real-time risk prediction for emerging infectious diseases: A case study in Avian Influenza H5N1. *Preventive Veterinary Medicine*, 91(1), pp. 19-28.

# Gaussian Processes Overview

- A Gaussian Process is a collection of random variables, any finite number of which have a joint Gaussian Distribution. (Rasmussen and Williams, 2007).
- We describe the log of the infection rate fully through the GP, this ensure the infection rate is always positive.

$$\log \beta_{i,j} \sim \mathcal{GP}(\mathbf{0}, \Sigma).$$

- We use the squared exponential covariance function to define the covariance matrix.

$$\Sigma = k(\mathbf{d}, \mathbf{d}'; \alpha, l) = \alpha^2 \exp \left( -\frac{(\mathbf{d} - \mathbf{d}')^2}{l^2} \right).$$

# GPs and Epidemic Models

We put a GP prior on  $\beta_{i,j}$  and use under-relaxed MCMC to give estimate this rate.

- 1 Propose new infection rate based on current one.

$$\log \beta'_{i,j} = \sqrt{1 + \delta^2} \log \beta_{i,j} + \delta \nu,$$

$$\nu \sim \mathcal{GP}(\mathbf{0}, \Sigma).$$

- 2 Evaluate likelihood of proposed infection rate.
- 3 Accept new rate based on probability of proposed against current rate being the true rate.

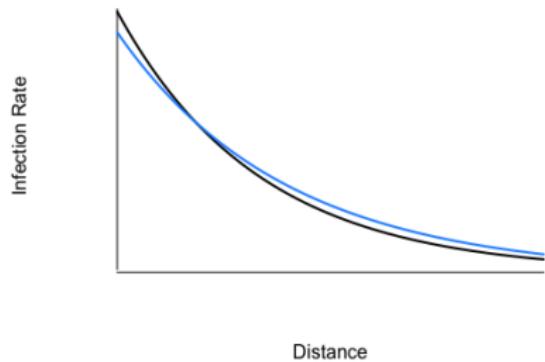


Figure: Current infection rate (black) and proposed rate (blue).