

Nonparametric Bayesian Methods for Heterogeneously Mixing Epidemic Models

Rowland Seymour, Philip O'Neill and Theodore Kypraios

Nottingham, 14th February 2019



Engineering and Physical Sciences
Research Council



University of
Nottingham
UK | CHINA | MALAYSIA

Table of Contents

1 Introduction

2 Model

3 Parametric Methods

4 Nonparametric Methods

5 Extensions to the Model

6 Avian Influenza

7 Conclusions

Avian Influenza

Avian Influenza Data

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 rd May
4	(4.56, 18.01)	Culled	30 th April
:	:	:	:

Table of Contents

1 Introduction

2 Model

3 Parametric Methods

4 Nonparametric Methods

5 Extensions to the Model

6 Avian Influenza

7 Conclusions

SIR Model

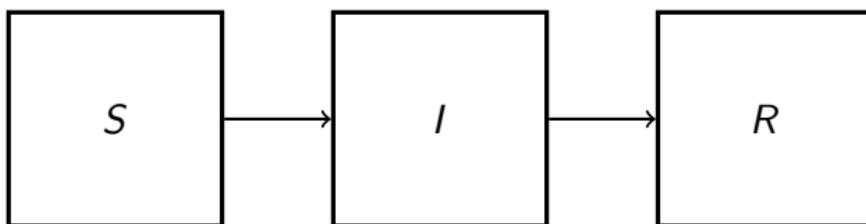


Figure: A compartmental SIR model.

At any time t , farms are either **Susceptible**, **Infected** or **Removed**.

- **Infections:** Infections occur according to a Poisson process with rate $\beta S_t I_t$.
- **Removals:** Farm j remains infectious for a time period drawn from an arbitrary, non-negative distribution \mathcal{D} .

SIR Model

We change the infection rate so the infection rate from farm i to individual j is given by

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

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

- **Infections:** Infections occur according to a Poisson process with a unique rate for each pair of farms, specified by $\beta_{i,j} = f(i,j)$.
- **Removals:** Farm j remains infectious for a time period drawn from an arbitrary, non-negative distribution \mathcal{D} .

Previous Work

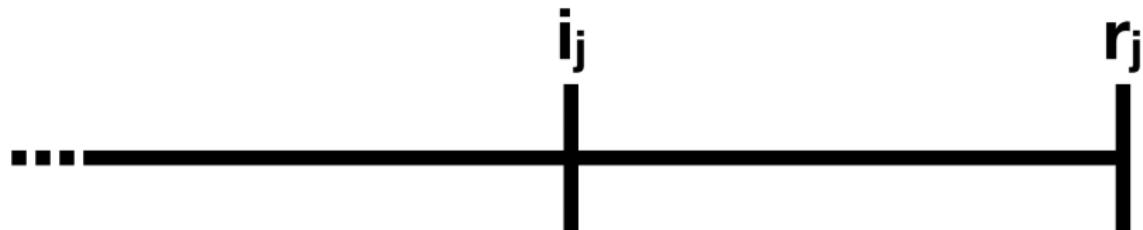
There are currently several methods for modelling heterogeneously mixing outbreaks:

- There are many parametric methods available, for example see Boender *et. al.* (2007), Jewell *et. al.* (2009),
- Becker and Yip (1989) model a time-dependent infection rates nonparametrically using a counting process,
- Lau and Yip (2008) then used kernel estimators for the time dependent infection rate,
- Knock and Kypraios (2014) have used spline based methods, also for modelling these rates, and
- Xu *et. al.* (2016) have also modelled time-dependent infection rates using Gaussian Processes.

There are no nonparametric methods for modelling pair-wise infection rates, so this is where we concentrate our efforts.

The Likelihood Function

To construct the likelihood function, we first consider the contribution of one individual j .



It contributes to the likelihood in several ways:

- By avoiding infection up to time i_j ,
- By becoming infected at time i_j , and
- By being infectious until r_j .

The Likelihood Function

The likelihood function for this model is given by

$$\pi(\mathbf{i}, \mathbf{r} | \boldsymbol{\beta}, \mathcal{D}) \propto \underbrace{\exp \left(- \sum_{j=1}^n \sum_{k=1}^N \beta_{j,k} (\min(r_j, i_k) - \min(i_j, i_k)) \right)}_{\text{Avoiding Infection}} \\ \times \underbrace{\prod_{\substack{j=1 \\ j \neq \kappa}}^n \left(\sum_{k \in \mathcal{Y}_j} \beta_{k,j} \right)}_{\text{Becoming Infected}} \\ \times \underbrace{\prod_{j=1}^n f_{\mathcal{D}}(r_j - i_j)}_{\text{Being infectious}}.$$

Table of Contents

1 Introduction

2 Model

3 Parametric Methods

4 Nonparametric Methods

5 Extensions to the Model

6 Avian Influenza

7 Conclusions

Parametric Methods

Suppose we want to fit a parametric model to the Avian Influenza data. We may assume the infection rate from farm i to farm j depends on the distance between them, $d_{i,j}$.

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

One possible method is to:

- 1 Choose plausible forms for $f(d_{i,j})$,
- 2 Fit the model for each form of $f(d_{i,j})$,
- 3 Use goodness of fit tests to choose the best $f(d_{i,j})$.

Parametric Infection Rates

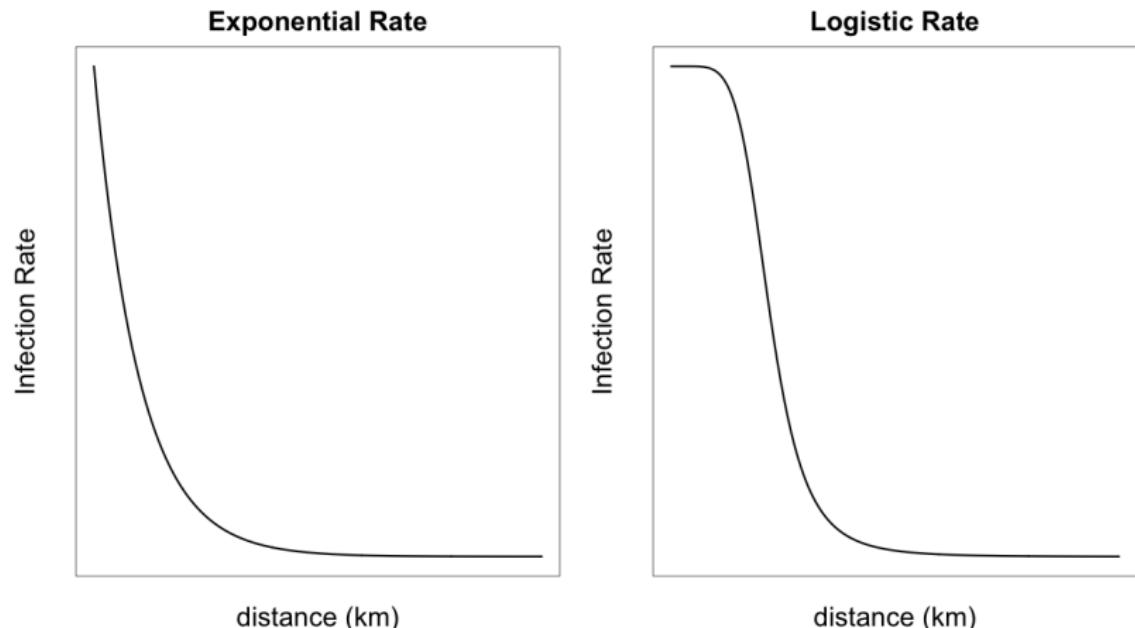


Figure: Two possible functional forms for $f(d_{i,j})$.

Fitting Parametric Models

We follow Boender *et. al.* (2007) and propose the following 5 forms for $f(d_{j,k})$. We compute maximum likelihood estimates for the parameters, and assume farms were infected for 7 days.

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

Model Selection

There are several methods we can use to choose the best model out of our proposed models. For this example, we use Akaike Information Criterion (AIC), which is given by

$$\text{AIC} = 2k - 2 \log \pi(\mathbf{i}, \mathbf{r} | \beta_{j,k}, \lambda, \gamma).$$

Model 3 has the lowest AIC value, closely followed by model 4.

Model	Infection Rate	AIC
1	$\beta_{j,k} = \beta_0$	3977.35
2	$\beta_{j,k} = \frac{\beta_0}{1+d_{j,k}}$	3591.30
3	$\beta_{j,k} = \frac{\beta_0}{1+d_{j,k}^2}$	3117.19
4	$\beta_{j,k} = \frac{\beta_0}{1+d_{j,k}^\alpha}$	3117.97
5	$\beta_{j,k} = \frac{\beta_0}{1+(d_{j,k}/\beta_1)^\alpha}$	3121.17

Parametric Models

For parametric models:

- There are a number of ways to fit the models, and
- A wide variety of methods for model choice and assessment.

However:

- It can be difficult to propose parametric functions given the observed data, and
- These forms are often based on strict assumptions about the infection rate.

Table of Contents

1 Introduction

2 Model

3 Parametric Methods

4 Nonparametric Methods

5 Extensions to the Model

6 Avian Influenza

7 Conclusions

Bayesian Nonparametric Methods

We wish to determine the form of $f(d_{i,j})$ without making assumptions about its parametric form.

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

We need to place a prior distribution over the space of functions for $f(d_{i,j})$, which makes few assumptions about the parametric form.

We could use:

- Piecewise constant functions,
- Splines, or
- Gaussian Processes.

Gaussian Processes

Definition: Gaussian Processes. A Gaussian Process (GP) is a collection of random variables such that any finite number subset of which has a multivariate normal distribution. We write this as

$$\mathbf{f} \sim \mathcal{GP}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where

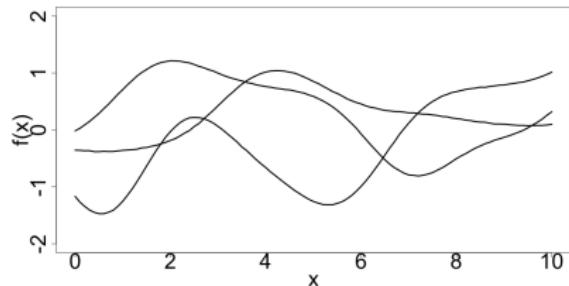
- $\mathbf{f} = \{f(d_1), \dots, f(d_{N(N-1)/2})\}$,
- $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_n\}$ is the mean vector, and
- $\boldsymbol{\Sigma}_{i,j} = \text{Cov}(f_i, f_j)$.

Covariance Functions

We can put assumptions about the function into the model through the covariance function.

The Squared Exponential Covariance Function:

$$k(x, x') = \alpha^2 \exp \left\{ -\frac{(x - x')^2}{l^2} \right\}$$

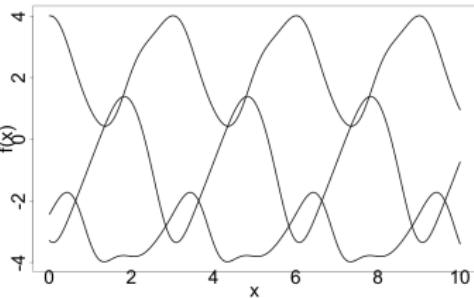


Covariance Functions

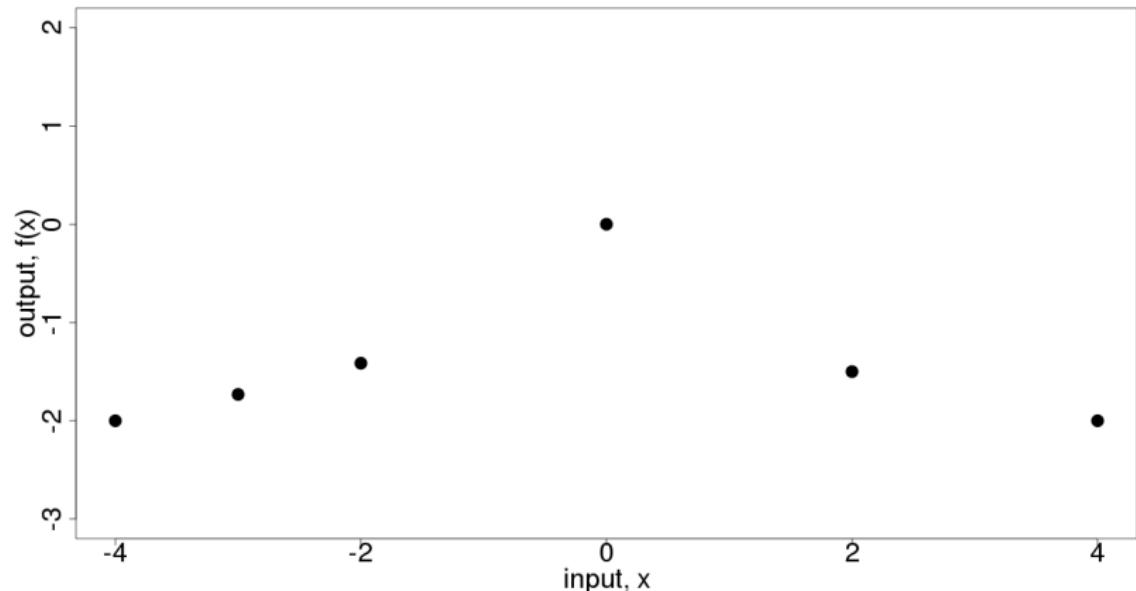
We can put assumptions about the function into the model through the covariance function.

The Periodic Covariance Function:

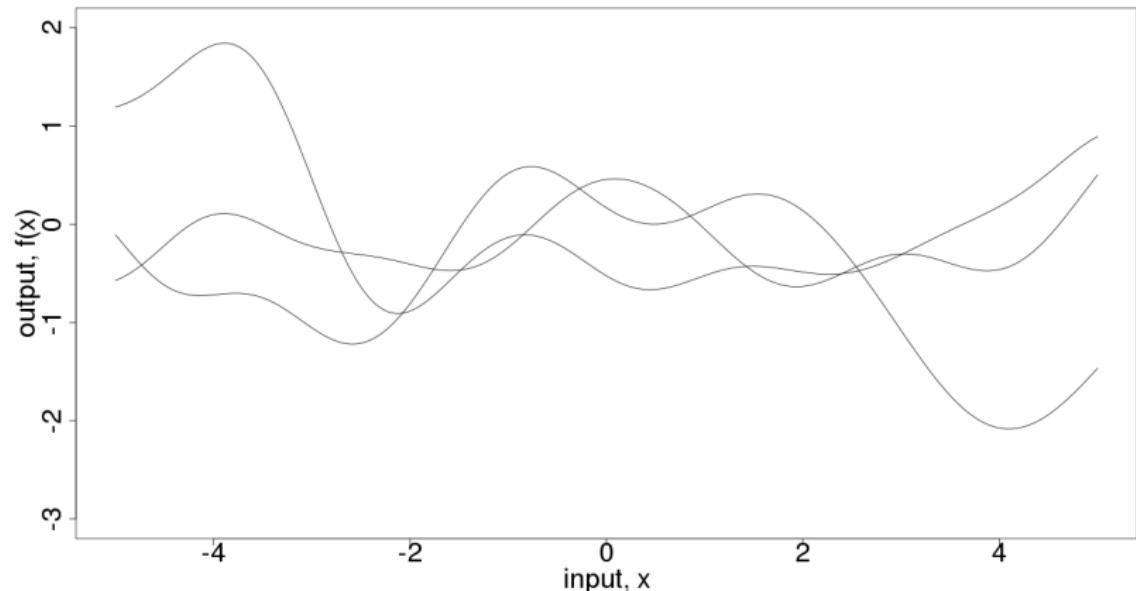
$$k(x, x') = \alpha^2 \exp \left\{ \frac{-\sin^2(\nu\pi|x - x'|)}{\ell^2} \right\}$$



GP Regression



GP Regression



GP Regression

Epidemic Model Prior Distributions

We put a GP prior on the infection rate β , where is constructed out of the pair-wise distances.

$$\log \beta_{j,k} = f(d_{j,k}) \quad \mathbf{f} \sim \mathcal{GP}(\mathbf{0}, \Sigma), \quad \Sigma = k(\mathbf{d}, \mathbf{d}; \alpha, l)$$

We put a vague Exponential prior distribution on the GP length scale.

$$k(x, x') = \alpha^2 \exp\left\{-\frac{(x - x')^2}{l^2}\right\}, \quad l \sim \text{Exp}(0.01).$$

Although we can put a prior distribution on α , we fix this to a large enough value so that the proposed functions cover a large range.

Epidemic Model Prior Distributions

We assume the infectious period is a $\Gamma(\lambda, \gamma)$ distribution, and we put a vague Exponential prior distribution on the rate parameter.

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

We assume λ is fixed to some value.

We place a uniform prior on the label of the first infected individual, and place an exponential prior distribution on their infection time such that

$$\kappa \sim U[1, \dots, n], \quad -i_\kappa \sim \text{Exp}(0.01).$$

Data Augmentation

To evaluate the likelihood function, we require both the removal times, r , and the infection times, i . Instead, we observe:

- The removal times, r ,
- The dates farmers notified the authorities their farm might be infected, and
- The dates laboratory results confirmed a farm was infected.

We use a data augmentation technique, treating the infection times as augmented data. We then estimates the infection times alongside the other model parameters.

Posterior Distribution

The posterior distribution is given by:

$$\begin{aligned}\pi(\exp(\mathbf{f}), \gamma, \mathbf{i}, i_\kappa, \kappa | \mathbf{r}, \mathbf{d}, \lambda, \alpha) &\propto \pi(\mathbf{i}, \mathbf{r} | \exp(\mathbf{f}), \lambda, \gamma, \kappa, i_\kappa, \chi_\gamma) \pi(\mathbf{f}|I) \pi(I) \\ &\times \pi(\gamma) \pi(\kappa) \pi(i_\kappa | \kappa).\end{aligned}$$

$$\begin{aligned}&\propto \exp \left\{ - \sum_{j=1}^n \sum_{k=1}^N \exp(f(d_{j,k})) (\min(r_j, i_k) - \min(i_j, i_k)) \right\} \\ &\times \prod_{\substack{j=1 \\ j \neq \kappa}}^n \left(\sum_{k \in \mathcal{Y}_j} \exp(f(d_{k,j})) \right) \prod_{j=1}^n f_{\mathcal{D}}(r_j - i_j | \lambda, \gamma) \\ &\times N(\mathbf{f}; \mathbf{0}, \Sigma) \exp \{-0.01I\} \exp \{-0.01\gamma\} \exp \{0.01i_\kappa\}\end{aligned}$$

Sampling from the Posterior Distribution

The posterior distribution is difficult to work with analytically, so we generate samples from it using Monte Carlo Markov Chain (MCMC) methods.

Algorithm 1 Structure of the MCMC algorithm

- 1: Initialise the chain with values $\gamma^{(0)}$, $\mathbf{f}^{(0)}$, $\mathbf{l}^{(0)}$, and $\mathbf{i}^{(0)}$
Repeat the following steps
 - 2: Sample γ 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 \mathbf{l} using a Metropolis Hastings random walk algorithm
 - 5: Update \mathbf{i} , the vector of infection times
-

Sampling the Posterior Distribution

This posterior is difficult to work with, especially as the infection times are unobserved. The conditional distribution for \mathbf{f} is:

$$\begin{aligned}\pi(\mathbf{f}|\mathbf{i}, \mathbf{r}, I, \kappa, i_\kappa) &\propto \exp \left\{ - \sum_{j=1}^n \sum_{k=1}^N \exp\{f(d_{j,k})\} (\min(r_j, i_k) - \min(i_j, i_k)) \right\} \\ &\times \prod_{\substack{j=1 \\ j \neq \kappa}} \left(\sum_{k \in \mathcal{Y}_j} \exp\{f(d_{k,j})\} \right) N(\mathbf{f}; \mathbf{0}, \Sigma).\end{aligned}$$

To update \mathbf{f} , we could:

- Update each value of \mathbf{f} individually using a MH Random Walk,
- Use a block update random walk for the entire GP, or
- Split the GP into sections and sample each section conditionally on the remaining sections.

Sampling from the Posterior Distribution

We follow Neal (1995) and to update the function \mathbf{f} , we propose a new function by

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

The advantage to this is that the acceptance probability is given by

$$\begin{aligned} p_{acc} &= \min \left\{ \frac{\pi(\mathbf{i}, \mathbf{r} | \exp(\mathbf{f}'), \lambda, \gamma)}{\pi(\mathbf{i}, \mathbf{r} | \exp(\mathbf{f}), \lambda, \gamma)} \frac{N(\mathbf{f}'; \mathbf{0}, \Sigma)}{N(\mathbf{f}; \mathbf{0}, \Sigma)} \frac{q(\mathbf{f} | \mathbf{f}')}{q(\mathbf{f}' | \mathbf{f})}, 1 \right\} \\ &= \min \left\{ \frac{\pi(\mathbf{i}, \mathbf{r} | \exp(\mathbf{f}'), \lambda, \gamma)}{\pi(\mathbf{i}, \mathbf{r} | \exp(\mathbf{f}), \lambda, \gamma)}, 1 \right\}. \end{aligned}$$

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.

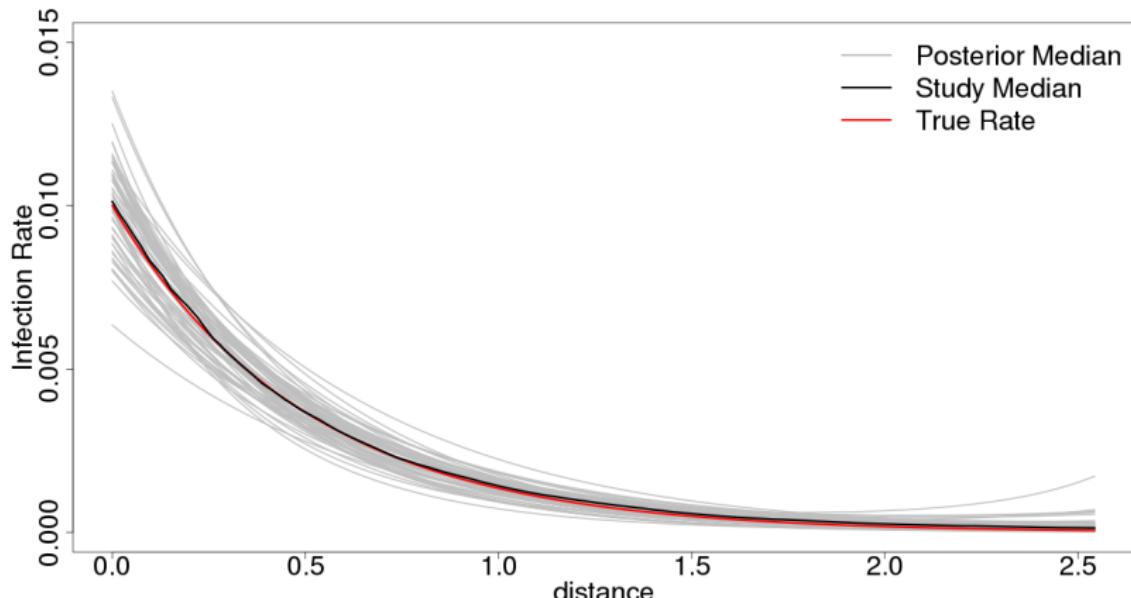


Table of Contents

1 Introduction

2 Model

3 Parametric Methods

4 Nonparametric Methods

5 Extensions to the Model

6 Avian Influenza

7 Conclusions

Extensions to the Model

We have now derived the posterior distribution and developed an MCMC framework. We can extend this model further to:

- 1 allow for modelling outbreaks in large populations,
- 2 include monotonicity assumptions in the function,
- 3 include information about the type of farm.

Mean Projection Approximation

One difficulty with GPs is the size of the covariance matrix. To reduce the dimension of the covariance matrix we use the mean projection approximation (Rasmussen and Williams, 2006).

Definition: The Conditional Normal Distribution. Suppose the vectors \mathbf{x} and \mathbf{y} follow a multivariate normal distribution

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim N \left(\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \Sigma_{x,x} & \Sigma_{x,y} \\ \Sigma_{y,x} & \Sigma_{y,y} \end{pmatrix} \right).$$

The distribution of \mathbf{x} given \mathbf{y} is

$$\mathbf{x}|\mathbf{y} \sim N \left(\boldsymbol{\mu}_x + \Sigma_{x,y} \Sigma_{y,y}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y), \Sigma_{x,x} - \Sigma_{x,y} \Sigma_{y,y}^{-1} \Sigma_{y,x} \right).$$

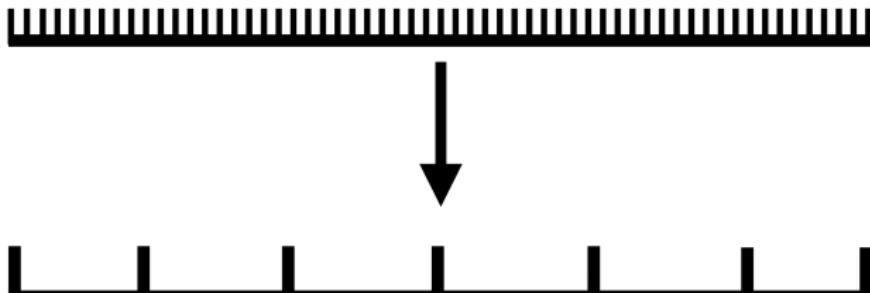
Mean Projection Approximation

We can use this distribution to reduce the size of the covariance matrix.

Suppose we have a large input space \mathbf{d} , we create a pseudo subset $\bar{\mathbf{d}}$, and a pseudo function \bar{f} .

$$\mathbf{d} = \{d_1, \dots, d_{N(N-1)/2}\}, \quad \mathbf{f} = \{f(d_1), \dots, f(d_{N(N-1)/2})\},$$
$$\bar{\mathbf{d}} = \{d_1^*, \dots, d_m^*\}, \quad \bar{\mathbf{f}} = \{f(d_1^*), \dots, f(d_m^*)\}.$$

We can choose $\bar{\mathbf{d}}$ a subset of \mathbf{d} , or we can construct a grid over \mathbf{d} .



Mean Projection Approximation

We place a prior distribution on $\bar{\mathbf{f}}$

$$\bar{\mathbf{f}} \sim \mathcal{GP}(\mathbf{0}, \Sigma_{\bar{d}, \bar{d}}).$$

The joint distribution of $\bar{\mathbf{f}}$ and \mathbf{f} is

$$\begin{pmatrix} \mathbf{f} \\ \bar{\mathbf{f}} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma_{d,d} & \Sigma_{d,\bar{d}} \\ \Sigma_{\bar{d},d} & \Sigma_{\bar{d},\bar{d}} \end{pmatrix} \right).$$

The expected value of \mathbf{f} given $\bar{\mathbf{f}}$ is

$$\mathbb{E}[\mathbf{f}|\bar{\mathbf{f}}] = \Sigma_{d,\bar{d}} \Sigma_{\bar{d},\bar{d}}^{-1} \bar{\mathbf{f}}.$$

We use this as our estimate for \mathbf{f} .

Mean Projection Approximation

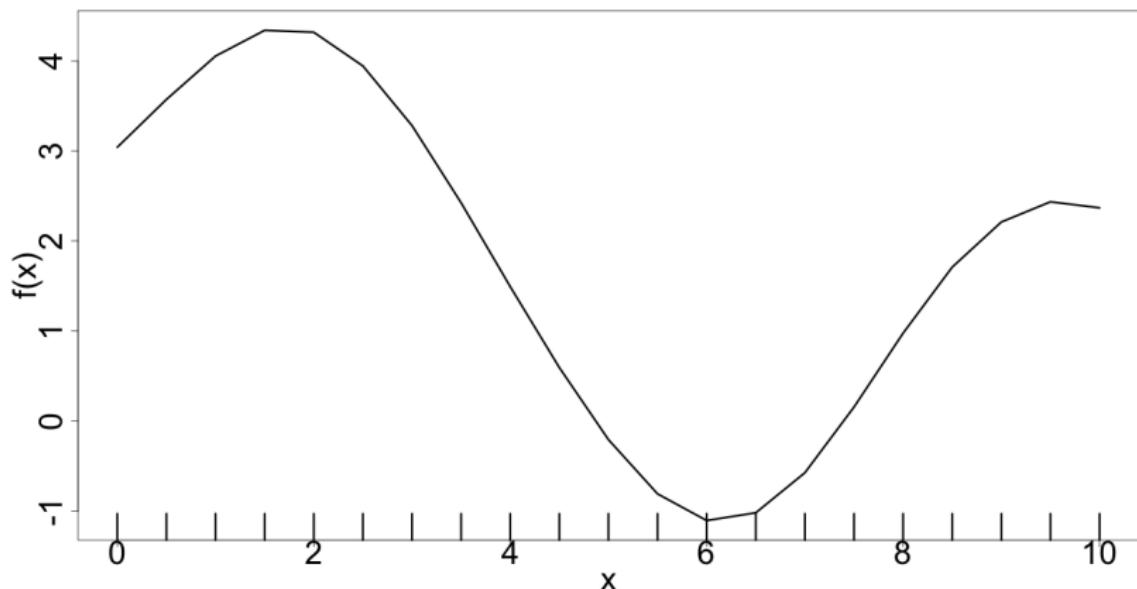


Figure: A sample of a GP, \bar{f} , drawn on 26 points.

Mean Projection Approximation

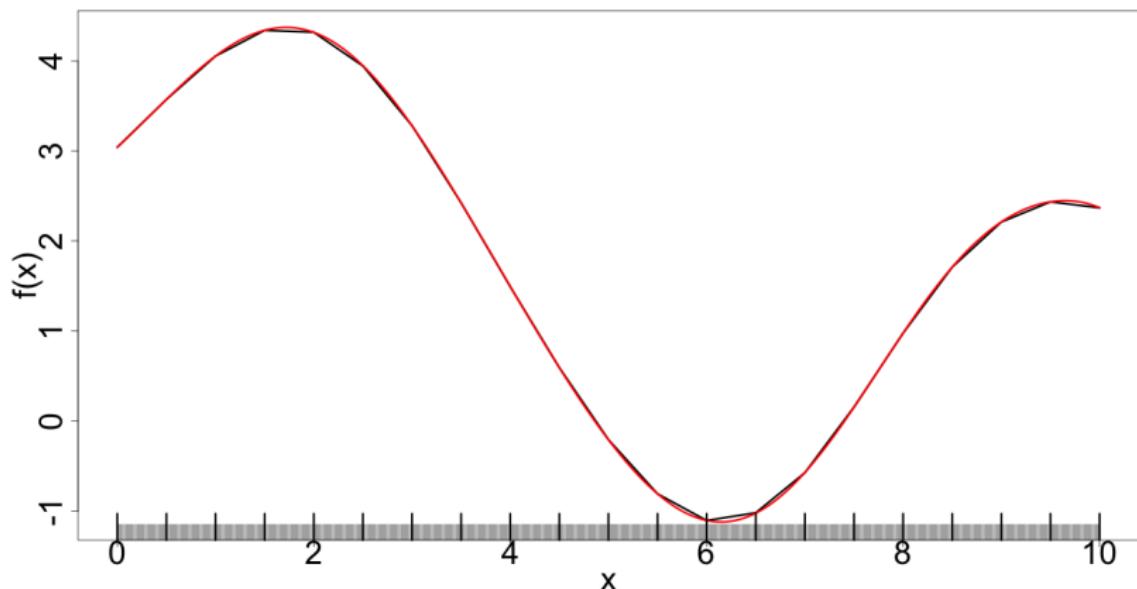


Figure: The sample \bar{f} projected onto the whole dataset, which contains over 10,000 points.

Inducing Monotonicity

We may want to include assumptions in our model, for example monotonicity of $f(\cdot)$. In our Avian Flu model, we may want to assume the infection rate is a decreasing function of the distance.

$$\begin{aligned}\text{cov}(f_i, f_j) &= k(x_i, x_j) \\ &= \alpha^2 \exp\left\{-\frac{(x_i - x_j)^2}{l^2}\right\}.\end{aligned}$$
$$\begin{aligned}\text{cov}(f_i, f'_j) &= \frac{\partial k(x_i, x_j)}{\partial x_j} \\ &= -\frac{2}{l^2}(x_i - x_j)k(x_i, x_j).\end{aligned}$$
$$\begin{aligned}\text{cov}(f'_i, f'_j) &= \frac{\partial^2 k(x_i, x_j)}{\partial x_i \partial x_j} \\ &= \frac{2}{l^2} \left[1 - \frac{2}{l^2}(x_i - x_j)^2\right] k(x_i, x_j).\end{aligned}$$

Inducing Monotonicity

We evaluate the gradient at a small number of points,
 $\mathbf{f}' = \{f'(x_1), \dots, f'(x_p)\}$. The joint prior distribution is

$$\begin{pmatrix} \mathbf{f} \\ \mathbf{f}' \end{pmatrix} \sim \mathcal{GP} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} k(\mathbf{x}, \mathbf{x}') & \frac{\partial k(\mathbf{x}, \mathbf{x}')}{\partial \mathbf{x}} \\ \frac{\partial k(\mathbf{x}, \mathbf{x}')}{\partial \mathbf{x}'} & \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')}{\partial \mathbf{x} \partial \mathbf{x}'} \end{pmatrix} \right).$$

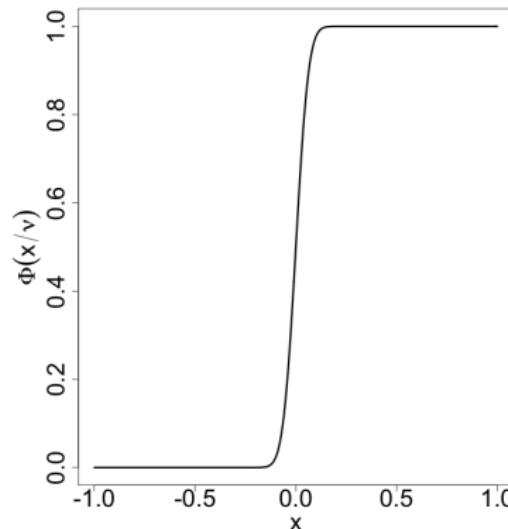
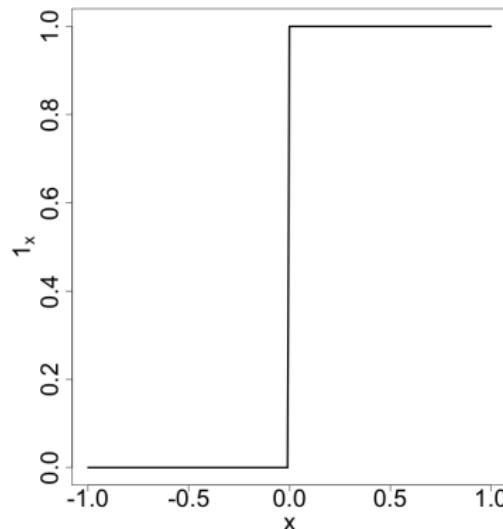
Given observations \mathbf{y} , Riihimaki and Vehtari (2010) write the posterior as

$$\pi(\mathbf{f}, \mathbf{f}' | \mathbf{x}, \mathbf{y}) \propto \pi(\mathbf{f}, \mathbf{f}') \pi(\mathbf{y} | \mathbf{x}, \mathbf{f}) \pi(\pm \mathbf{1} | \mathbf{f}'),$$

Inducing Monotonicity

We can write the contributions from the derive observations by

$$\pi(\mathbf{1}|\mathbf{f}') = \prod_{j=1}^p 1_{f'_j} \quad \text{or} \quad \pi(\mathbf{1}|\mathbf{f}') = \prod_{j=1}^p \Phi\left(f'_j \frac{1}{\nu}\right)$$



Inducing Monotonicity

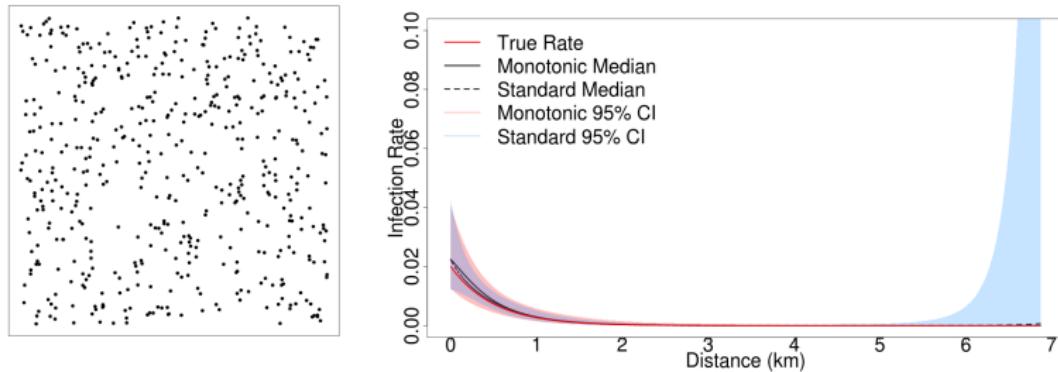


Figure: The positions of the farms, and a standard GP compared to the Monotonic GP for this data set.

Multi-Type Epidemics

We now extend our model to allow for individuals to be of different types. This might be

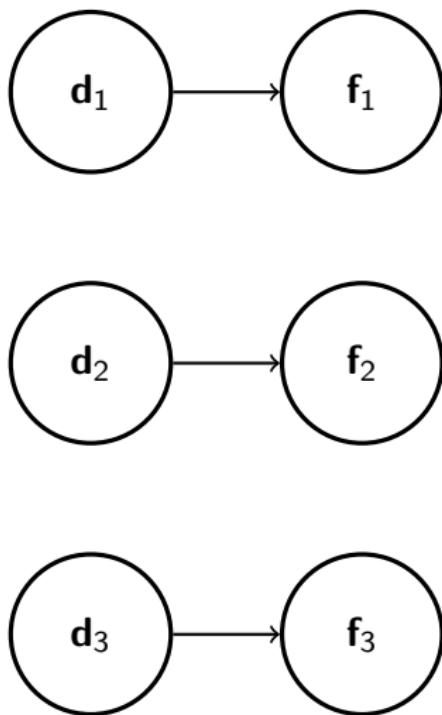
- Chicken Farms,
- Duck Farms, and
- Turkey Farms.

We model the infection rate by

$$\log \beta_{j,k} = \begin{cases} f_1(d_{j,k}) & \text{if } k \text{ type 1,} \\ f_2(d_{j,k}) & \text{if } k \text{ type 2,} \\ f_3(d_{j,k}) & \text{if } k \text{ type 3.} \end{cases}$$

We model this using Multi-Output Gaussian Processes, a way of modelling several functions at once, allowing for correlation between the functions.

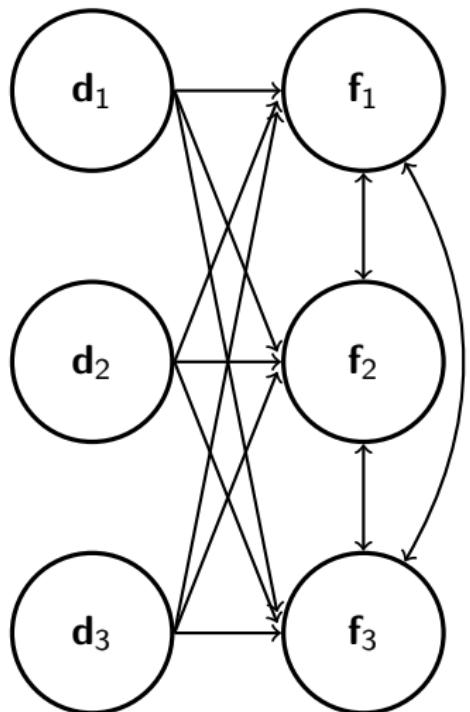
Independent GP Model



$$\begin{pmatrix} \log \beta_1 \\ \log \beta_2 \\ \log \beta_3 \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{pmatrix} \sim \mathcal{GP}(\mathbf{0}, \Sigma).$$

$$\Sigma = \begin{pmatrix} \Sigma_{1,1} & 0 & 0 \\ 0 & \Sigma_{2,2} & 0 \\ 0 & 0 & \Sigma_{3,3} \end{pmatrix}.$$

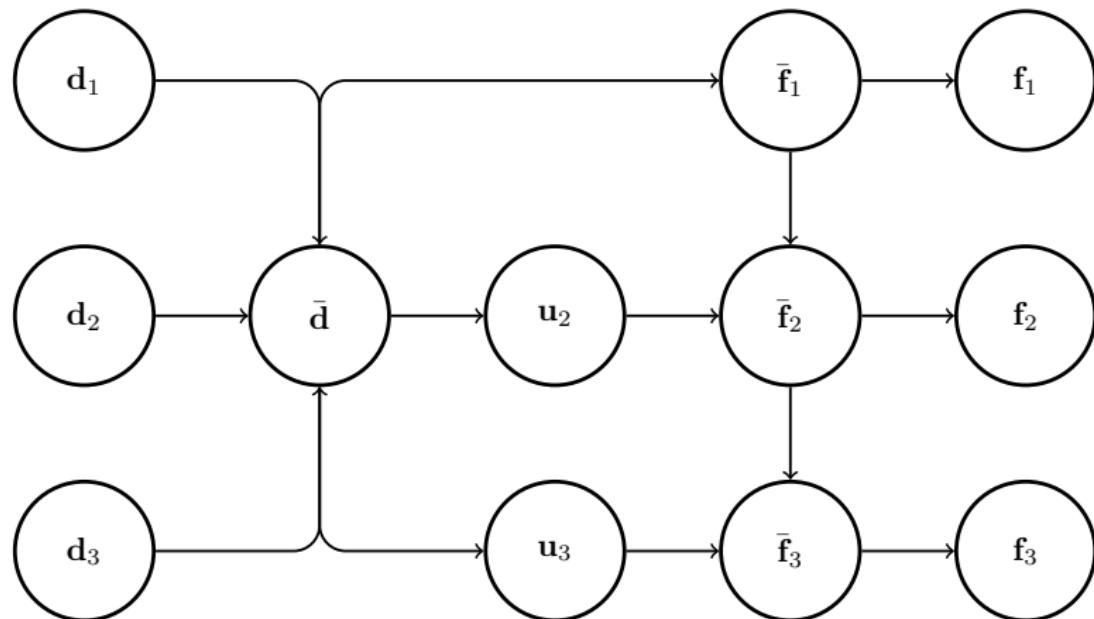
Multi-Output Covariance Model



$$\begin{pmatrix} \log \beta_1 \\ \log \beta_2 \\ \log \beta_3 \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{pmatrix} \sim \mathcal{GP}(\mathbf{0}, \Sigma).$$

$$\Sigma = \begin{pmatrix} \Sigma_{1,1} & \rho_{1,2}\Sigma_{1,2} & \rho_{1,3}\Sigma_{1,3} \\ \rho_{1,2}\Sigma_{2,1} & \Sigma_{2,2} & \rho_{2,3}\Sigma_{2,3} \\ \rho_{1,3}\Sigma_{3,1} & \rho_{2,3}\Sigma_{2,3} & \Sigma_{3,3} \end{pmatrix}$$

The Discrepancy Based Model



Input Data

Pseudo Data
Set

GP Prior
Distribu-
tions

Discrepancy
Calculations

Projection
onto Input
Data Set

The Discrepancy Based Model

The prior distribution for the Discrepancy Based Model is given by:

$$\bar{\mathbf{f}}_1 \sim \mathcal{GP}(\mathbf{0}, \Sigma_1), \quad \Sigma_1 = k(\bar{\mathbf{x}}, \bar{\mathbf{x}}; \alpha^2, l_1^2)$$

$$\mathbf{u}_2 \sim \mathcal{GP}(\mathbf{0}, \Sigma_2), \quad \Sigma_2 = k(\bar{\mathbf{x}}, \bar{\mathbf{x}}; \alpha^2, l_2^2), \quad \bar{\mathbf{f}}_2 = \bar{\mathbf{f}}_1 + \mathbf{u}_2$$

$$\mathbf{u}_3 \sim \mathcal{GP}(\mathbf{0}, \Sigma_3), \quad \Sigma_3 = k(\bar{\mathbf{x}}, \bar{\mathbf{x}}; \alpha^2, l_3^2), \quad \bar{\mathbf{f}}_3 = \bar{\mathbf{f}}_2 + \mathbf{u}_3$$

This model is asymmetric as we have to draw samples from the prior sequentially.

Prior Distributions for Remaining Parameters

The Multi Output Covariance Model: We put a vague Exponential prior distributions on the GP length scale.

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

We put a uniform prior distribution on the covariance parameter.

$$\rho \sim U[-1, 1].$$

Independent GPs and Discrepancy Based Models: We place independent exponential priors on the length scale parameters.

$$l_1, l_2, l_3 \sim \text{Exp}(0.01).$$

Simulation Studies

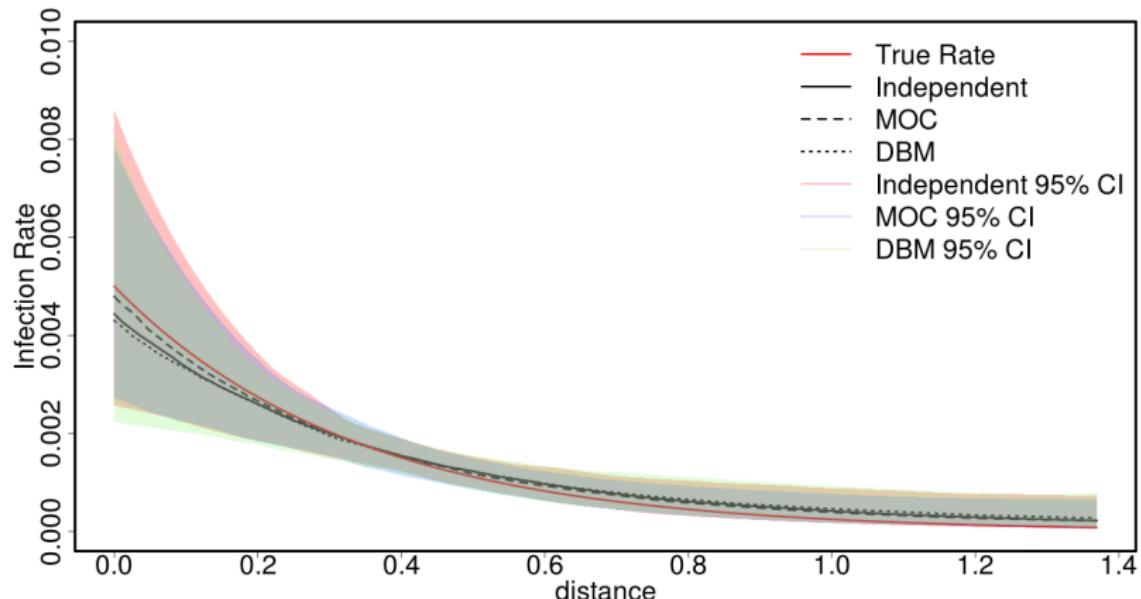


Figure: Estimates for the infection rate for type one individuals

Simulation Studies

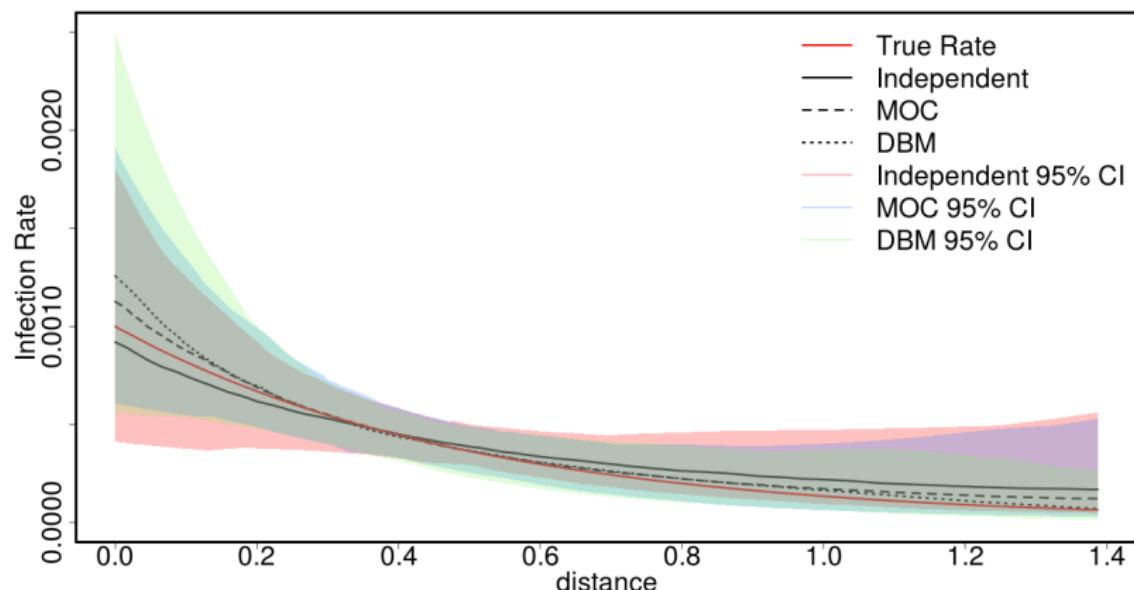
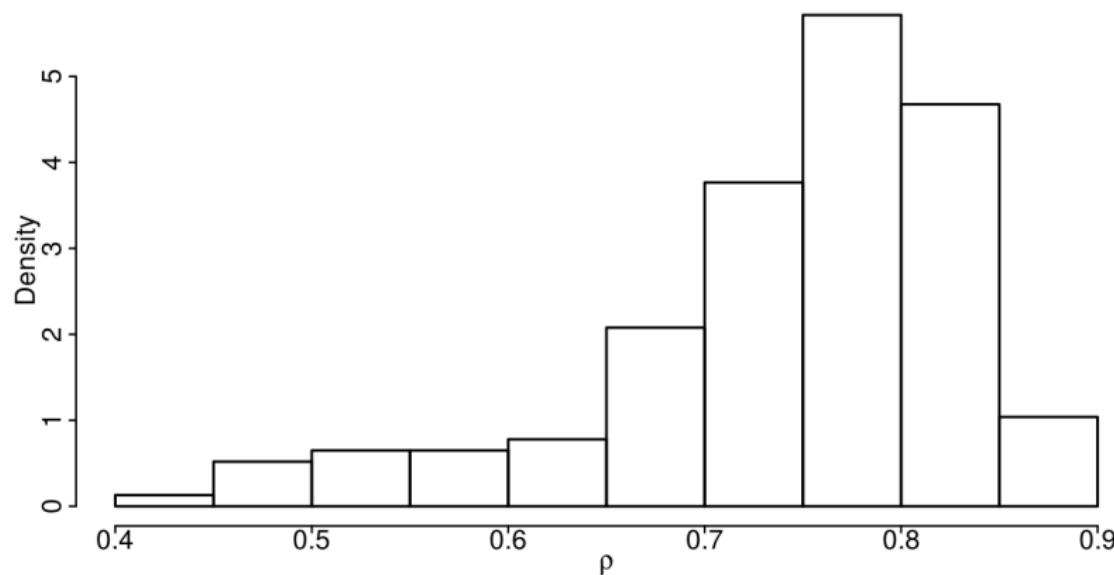


Figure: Estimates for the infection rate for type two individuals

Model Comparisons

The multi-output covariance model allows us to infer the correlation between the GPs.



Model Comparisons

Whereas the discrepancy based model allows us to infer the discrepancy between the functions.

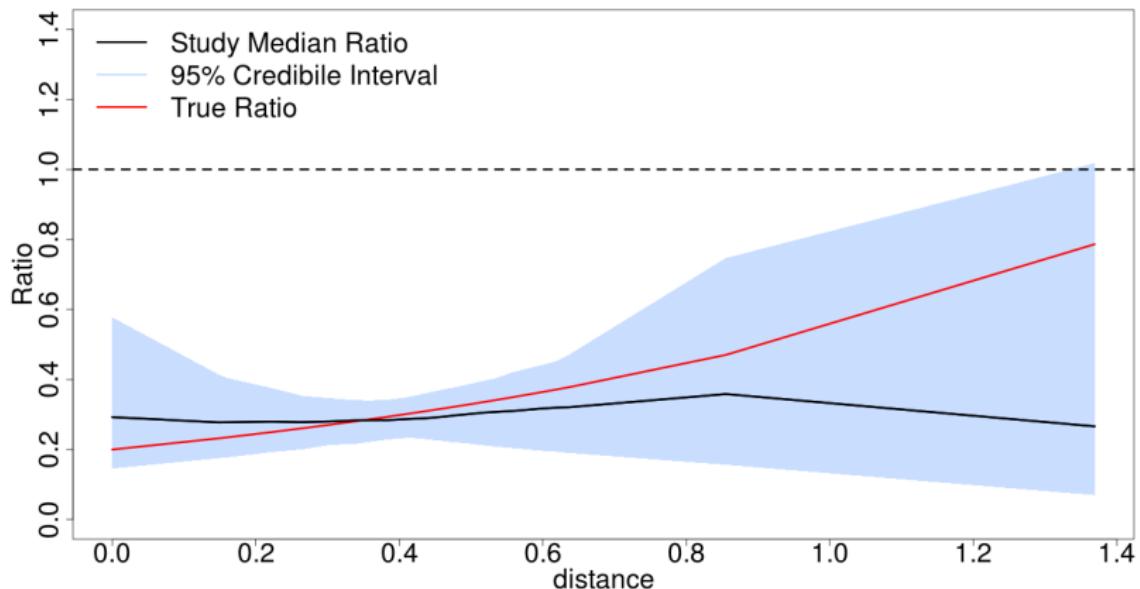


Table of Contents

1 Introduction

2 Model

3 Parametric Methods

4 Nonparametric Methods

5 Extensions to the Model

6 Avian Influenza

7 Conclusions

Avian Influenza Data

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 rd May
4	(4.56, 18.01)	Culled	30 th April
:	:	:	:

Adjustments for Culling

During the outbreak, the authorities were trying to control the spread of the disease by culling infected farms preemptively.

Set	Infected	Culled	Preemptively Culled
A	✗	✗	✗
B	✓	✓	✗
C	✓	✓	✓
D	✗	✓	✓

We use these labels to rewrite the double sum in our likelihood function as:

$$\begin{aligned}\Psi = \sum_{j=1}^n & \left[\sum_{k \in A \cup B \cup C} \beta_{j,k} ((r_j \wedge i_k) - (i_j \wedge i_k)) \right. \\ & \left. + \sum_{k \in D} \beta_{j,k} ((r_j \wedge r_k) - (i_j \wedge r_k)) \right].\end{aligned}$$

Adjustments for Culling

If a farm is infected but culled preemptively, we do not observe its full infectious period. Its contribution to the likelihood function is instead given by the survivor function.

$$S_{\mathcal{D}}(r_j - i_j; \alpha, \lambda) = \int_{r_j - i_j}^{\infty} f_{\mathcal{D}}(r_j - i_j; \alpha, \lambda).$$

The full likelihood function is therefore:

$$\begin{aligned} \pi(\mathbf{i}, \mathbf{r} | \boldsymbol{\beta}, \lambda, \gamma, \kappa, i_\kappa, \mathbf{s}_A, \mathbf{s}_B, \mathbf{s}_C, \mathbf{s}_D) &= \exp \{-\Psi\} \prod_{\substack{j=1 \\ j \neq \kappa}}^n \left(\sum_{k \in \mathcal{Y}_j} \beta_{k,j} \right) \\ &\times \prod_{j \in B} f_{\mathcal{D}}(r_j - i_j | \lambda, \gamma) \prod_{j \in C} S_{\mathcal{D}}(r_j - i_j | \lambda, \gamma). \end{aligned}$$

Results

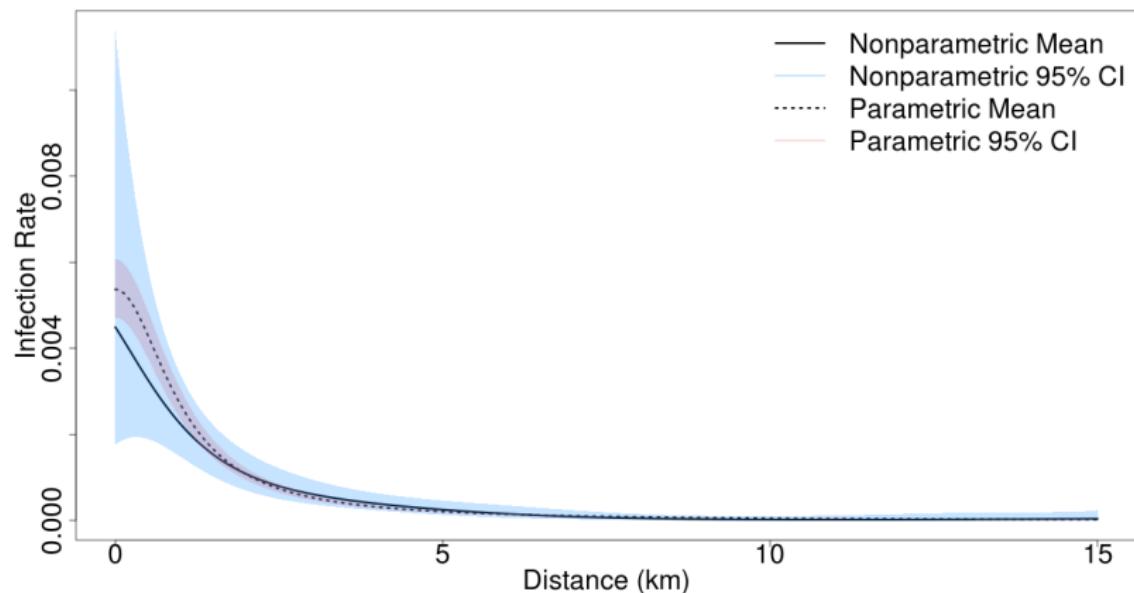


Figure: Comparison of the parametric and parametric models.

Table of Contents

1 Introduction

2 Model

3 Parametric Methods

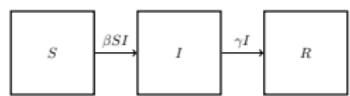
4 Nonparametric Methods

5 Extensions to the Model

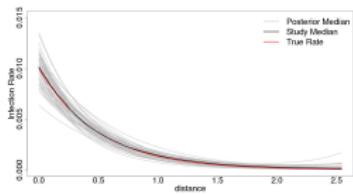
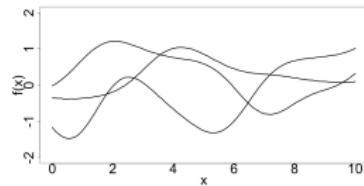
6 Avian Influenza

7 Conclusions

Conclusions



$$\beta_{i,j} = \frac{\beta_0}{1 + d_{i,j}^{\beta_1}}$$



$$\pi(\mathbf{1}|\mathbf{f}') = \prod_{j=1}^p \Phi\left(f_j' \frac{1}{\nu}\right)$$



Further Work

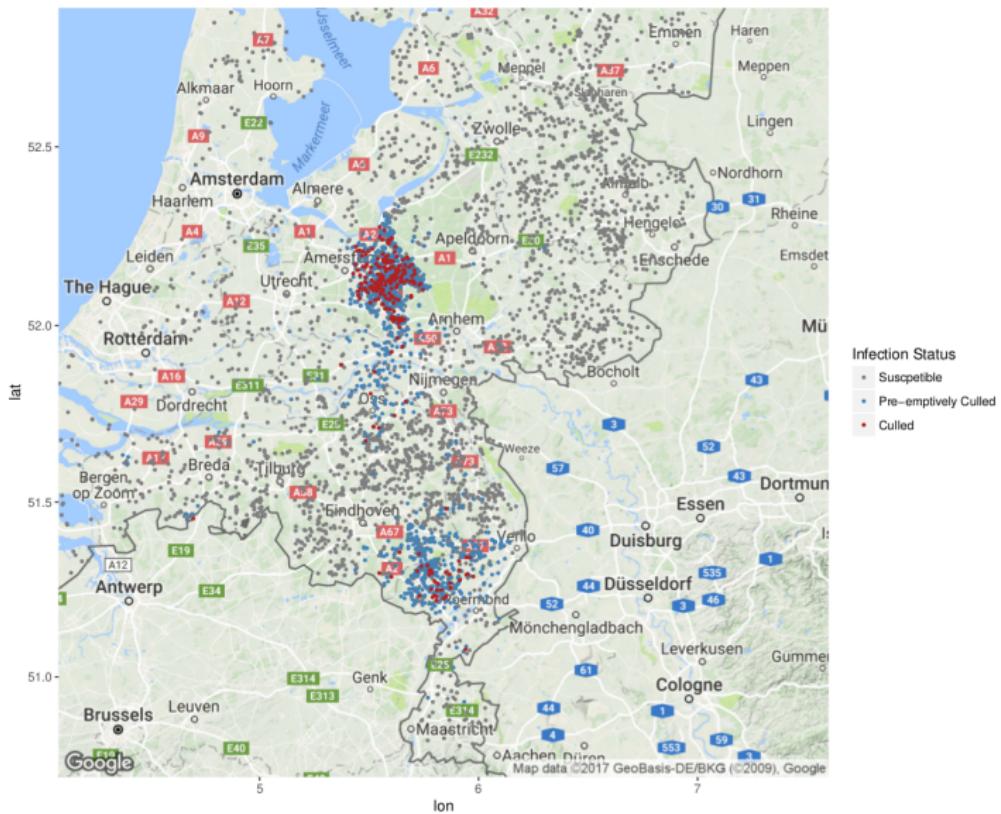
We are still investigating this method, and looking to

- Extend the model to include the information about a second continuous covariate,
- Use the model to investigate disease control strategies, and
- Investigate the Avian Influenza data set to look at how farm size and type affected the outbreak.

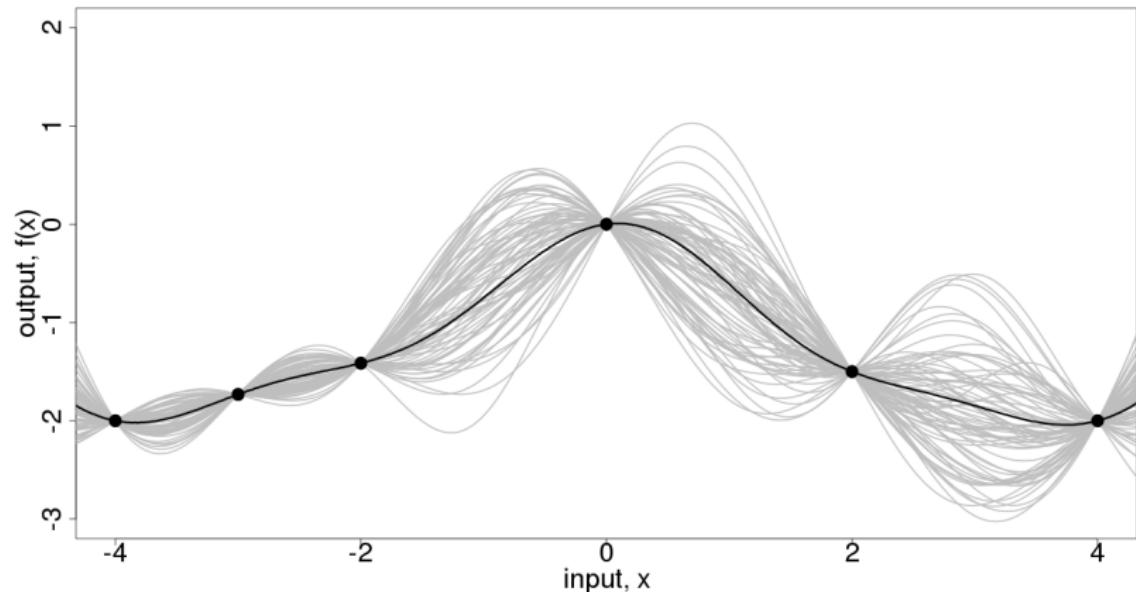
Selected References

- Becker, N. G. and Yip, P. (1989). Analysis of variation in an infection rate. *Australian and New Zealand Journal of Statistics*, 31:42–52.
- Neal, R. M. (1995). Suppressing random walks in Markov Chain Monte Carlo using ordered overrelaxation. Technical report, Department of Statistics, University of Toronto.
- Rasmussen, C. E. and Williams, C. K. I. (2006) *Gaussian Processes for Machine Learning*. MIT Press.
- Riihimaki, J. and Vehtari, A. (2010). Gaussian Processes with Monotonicity Information. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, volume 9 of PMLR, pages 645-652.
- Boender, G. J., Hagenaars, T. J., Boumma , A., Nodelijk , G., Elbers, A. R. W., de Jong, M. C. M., and van Boven, M. (2007). Risk maps for the spread of highly pathogenic avian influenza in poultry. *PLOS Computational Biology*, 3(4):704-712.
- Nguyen, T. V. and Bonilla , E. V. (2014). Collaborative Multi-output Gaussian Processes. In Zhang, N. and Tian, J., editors, *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pages 643-652

Avian Influenza



GP Regression



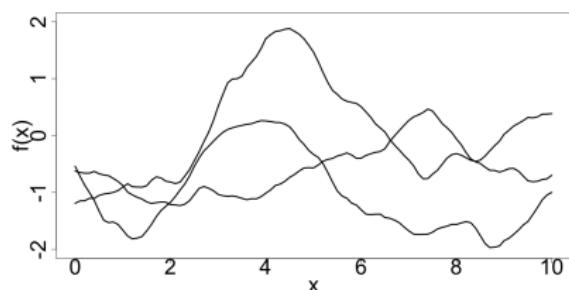
Covariance Functions

We can put assumptions about the function into the model through the covariance function.

The Matérn Covariance Function:

$$k(x, x') = \frac{2^\nu}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}(x - x')}{l^2} \right)^\nu \times K_\nu \left(\frac{\sqrt{2\nu}(x - x')}{l^2} \right),$$

where K_ν is the modified Bessel Function of the second kind



Sampling from the Posterior Distribution

Sampling γ : The conditional distribution for γ is

$$\begin{aligned}\pi(\gamma|\lambda, \mathbf{i}, \mathbf{r}) &\propto \prod_{j=1}^n f_{\mathcal{D}}(r_j - i_j | \lambda, \gamma) \exp \{-0.01\gamma\} \\ &\propto \gamma^{(n\lambda+1)-1} \exp \left\{ -\gamma \left(0.01 + \sum_{i=1}^n (r_i - i_j) \right) \right\}.\end{aligned}$$

Updating infection times: For each infection time i_j , we propose a new value by $i'_j = r_j - t$, where $t \sim \Gamma(\lambda, \gamma)$. We accept this with probability:

$$p_{acc} = \min \left\{ \frac{f_{\mathcal{D}}(r_j - i'_j | \lambda, \gamma)}{f_{\mathcal{D}}(r_j - i_j | \lambda, \gamma)} \frac{\pi(\mathbf{i} - i_j + i'_j, \mathbf{r} | \exp(\mathbf{f}), \lambda, \gamma)}{\pi(\mathbf{i}, \mathbf{r} | \exp(\mathbf{f}), \lambda, \gamma)}, 1 \right\}.$$