Bayesian nonparametric inference for epidemic models

Rowland Seymour

Lancaster University, 24th March 2022

University of Nottingham
UK | CHINA | MALAYSIA

Engineering and Physical Sciences Research Council

Table of Contents

1 Epidemic modelling

2 Homogeneous and heterogeneous mixing

3 Parametric inference

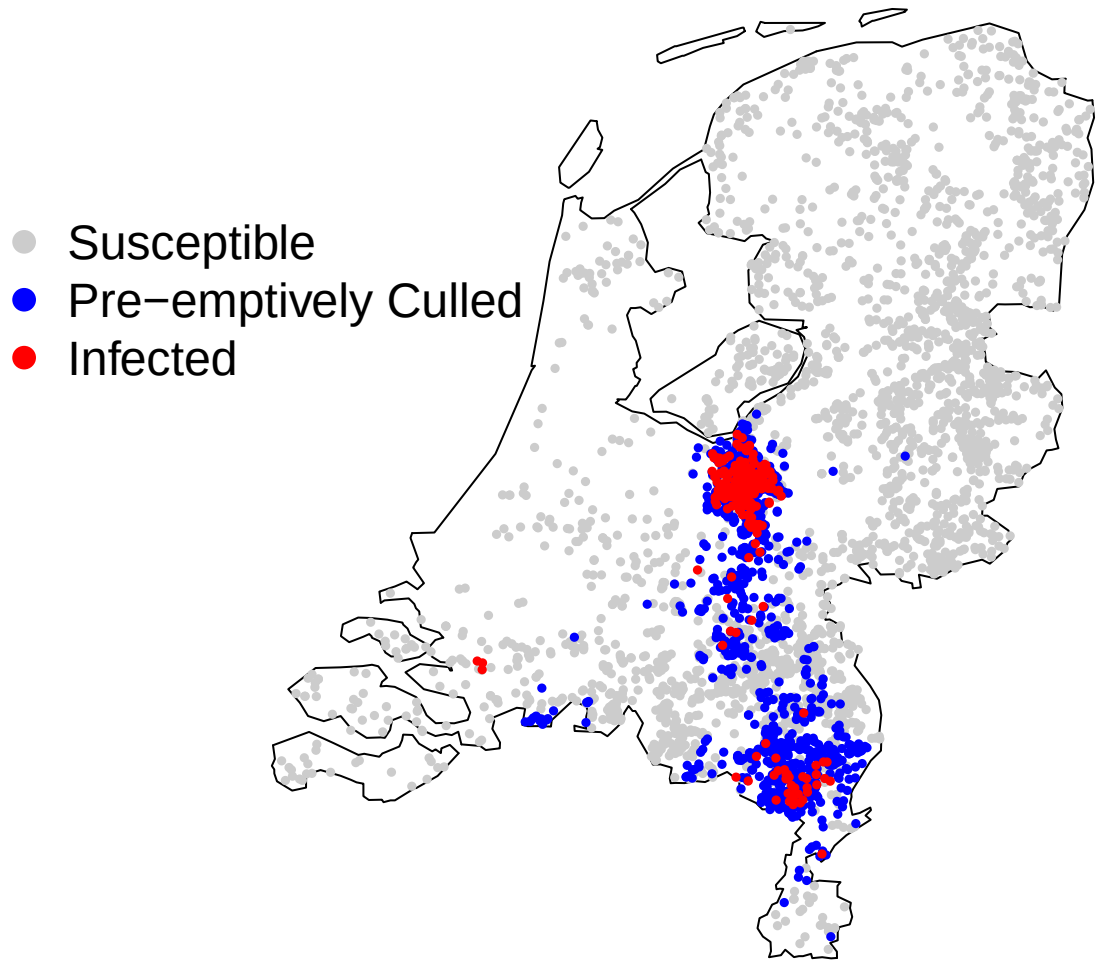
4 Gaussian processes

5 Gaussian processes for epidemic models

6 Foot and mouth disease

7 Avian influenza

Avian Influenza



Data 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^{rd} May
4	(4.56, 18.01)	Culled	30^{th} 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?
- Can we include other information, such as farm type?

Table of Contents

1 Epidemic modelling

2 Homogeneous and heterogeneous mixing

3 Parametric inference

4 Gaussian processes

5 Gaussian processes for epidemic models

6 Foot and mouth disease

7 Avian influenza

Homogeneous mixing

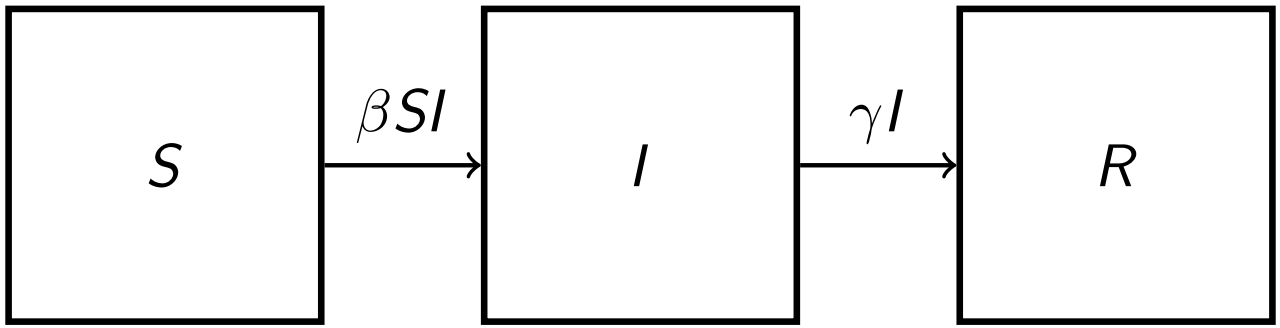


Figure: A homogeneously mixing SIR model.

Farms are either **S**usceptible, **I**nfected or **R**emoved.

- **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 Gamma distribution, such that $r_j - i_j \sim \text{Gamma}(\lambda, \gamma)$.

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 β and 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 $r_j - i_j \sim \text{Gamma}(\lambda, \gamma)$.

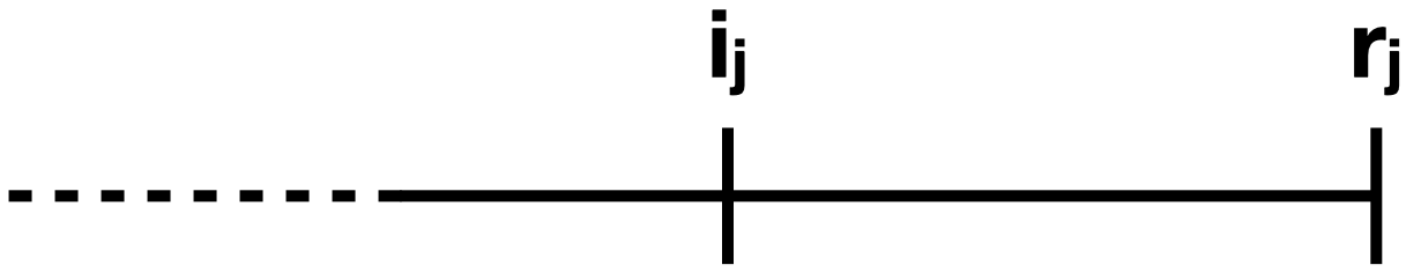
Multitype models

We focus on multitype susceptibility models in which individuals can have varying susceptibility to the disease, but are assumed to be equally infectious if infected. Specifically, if the susceptible farm is type z_j, the infection rate is given by

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

Likelihood function

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



It contributes to the likelihood is several ways:

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

Likelihood function

The augmented likelihood function for this model is given by

$$\pi(\mathbf{i}, \mathbf{r}|\boldsymbol{\beta}, \gamma) \propto \underbrace{\exp\Big(-\sum_{j=1}^{n}\sum_{k=1}^{N} \beta_{j,k}^{(z_k)} \big((r_j \wedge i_k) - (i_j \wedge i_k)\big)\Big)}_{\text{Avoiding infection}}$$

$$\times \underbrace{\prod_{\substack{j=1 \\ j\neq\kappa}}^{n} \Big(\sum_{k\in\mathcal{Y}_j} \beta_{k,j}^{(z_j)} \Big)}_{\text{Becoming infectious}}$$

$$\times \underbrace{\prod_{j=1}^{n} h(r_j - i_j|\gamma)}_{\text{Remaining infected}} .$$

Posterior distribution

Given the removal times, we want to infer:

- the infection rate function,
- the infection times,
- parameters for the infectious period distribution.

The posterior distribution is given by

$$\pi(\beta, \gamma, \mathbf{i}, \omega, i_\omega | \mathbf{r}, \lambda) \propto \quad \pi(\mathbf{i}, \mathbf{r} | \beta, \lambda, \gamma, \omega, i_\omega) \times \pi(\beta)\pi(\gamma)\pi(i_\omega|\omega)\pi(\omega).$$

Table of Contents

1 Epidemic modelling

2 Homogeneous and heterogeneous mixing

3 Parametric inference

4 Gaussian processes

5 Gaussian processes for epidemic models

6 Foot and mouth disease

7 Avian influenza

Parametric infection rates

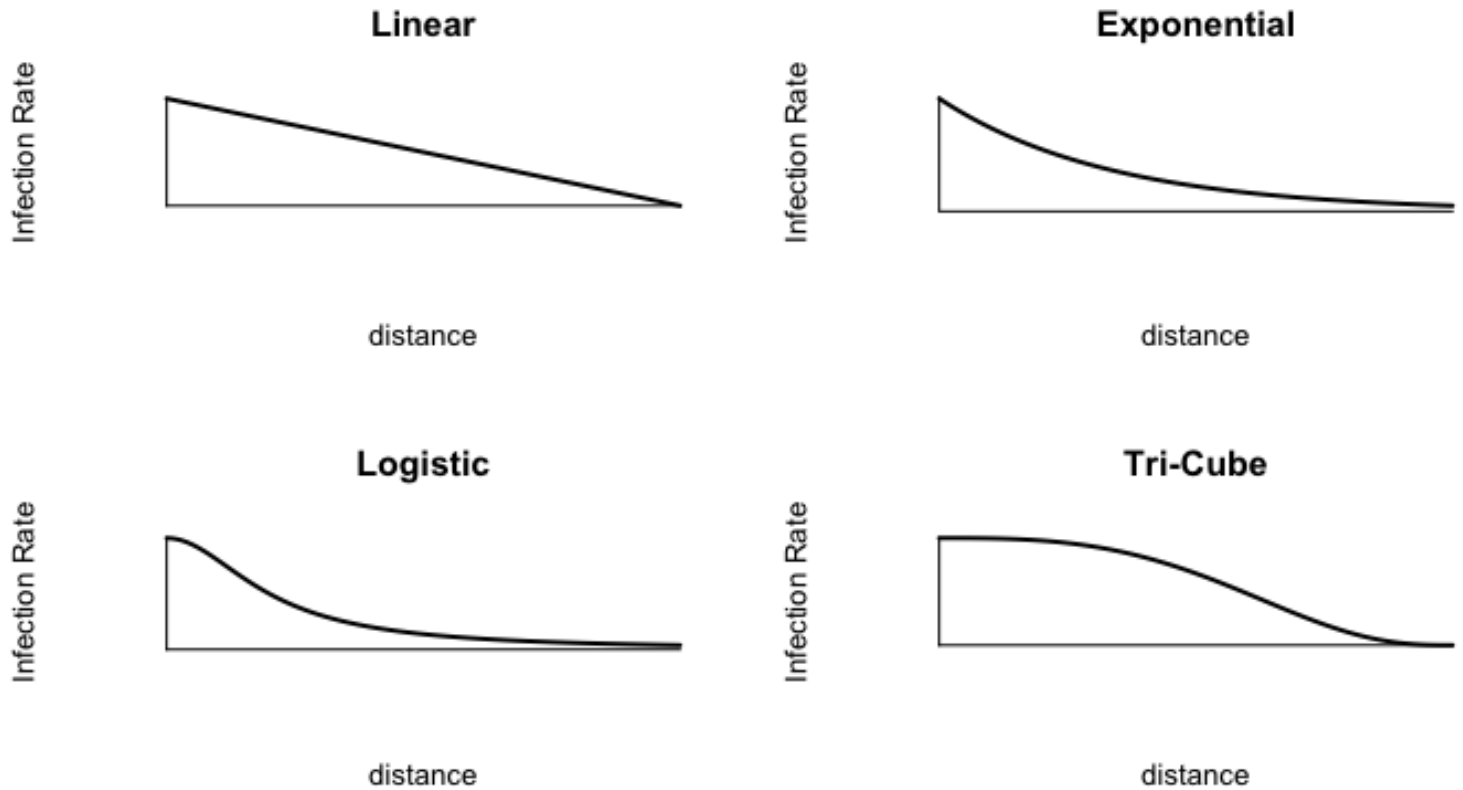


Figure: Four Possible Infection Rates.

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

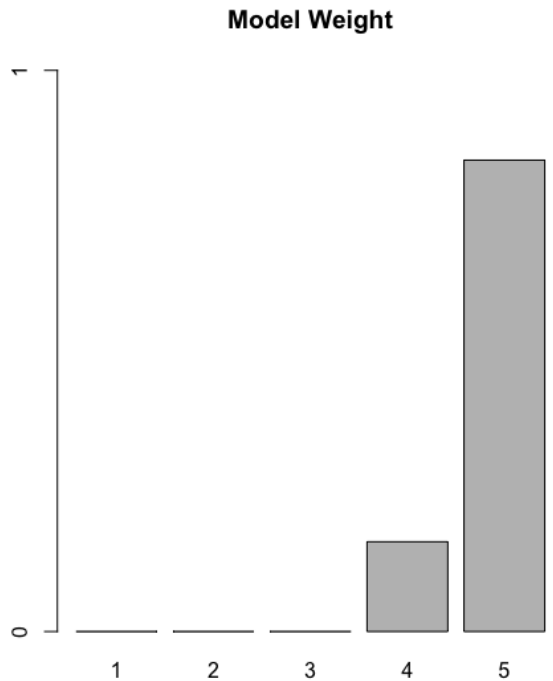


Figure: Model Weights for the different Infection Rates.

Parametric methods

If you want to do parametric inference, there are three main steps:

1. write down a suite of parametric functions,
2. fit the models,
3. assess the models and choose the most suitable one.

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 Epidemic modelling

2 Homogeneous and heterogeneous mixing

3 Parametric inference

4 Gaussian processes

5 Gaussian processes for epidemic models

6 Foot and mouth disease

7 Avian influenza

Introduction to Gaussian processes

Our Bayesian nonparametric methodology use Gaussian Processes (GPs) to estimate f.

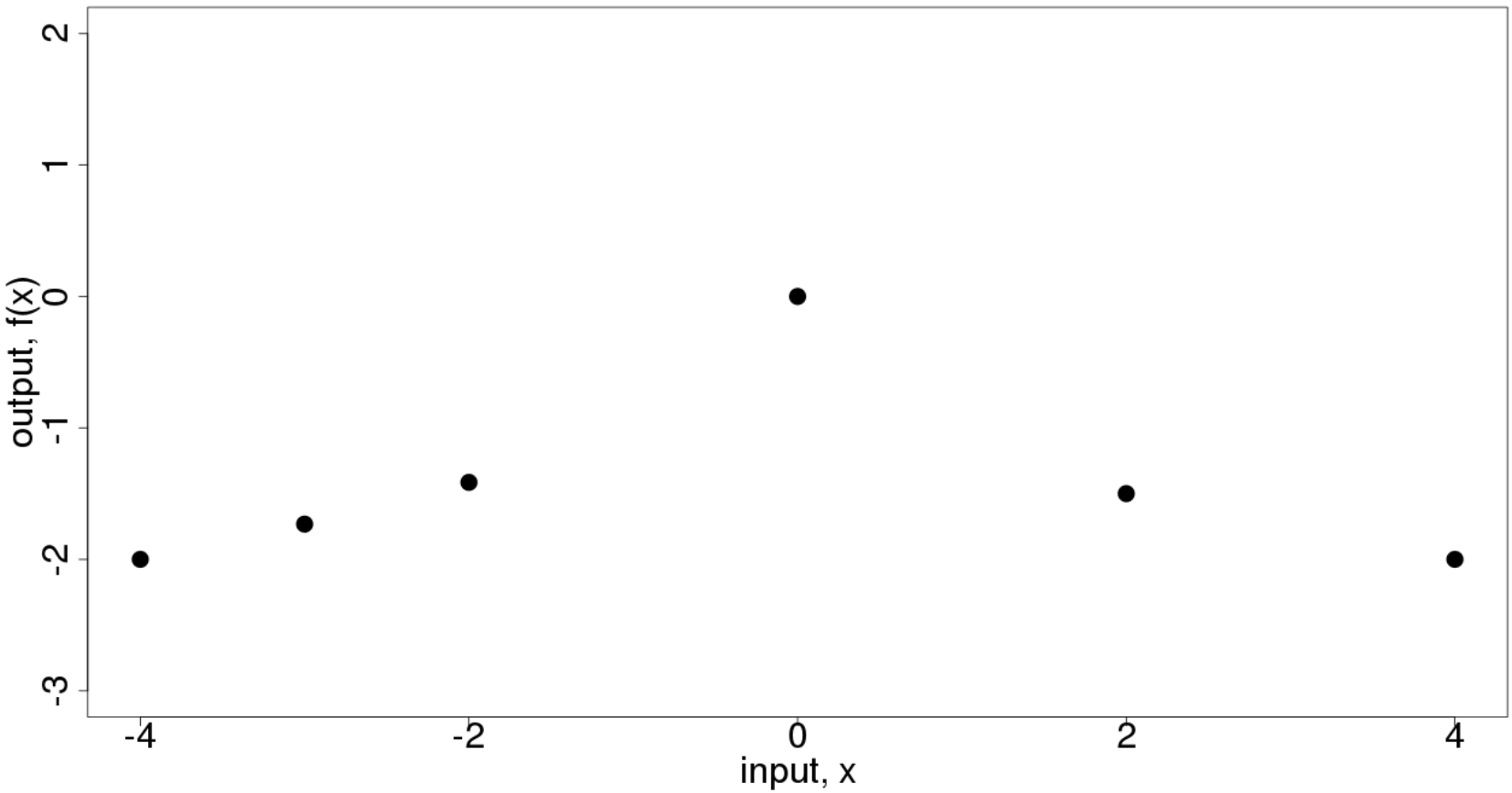
- GPs are a generalisation of the multivariate Gaussian distribution to a function space.
- We use GPs to assign a prior distribution over a function space.
- We need to control the signal variance and volatility.

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

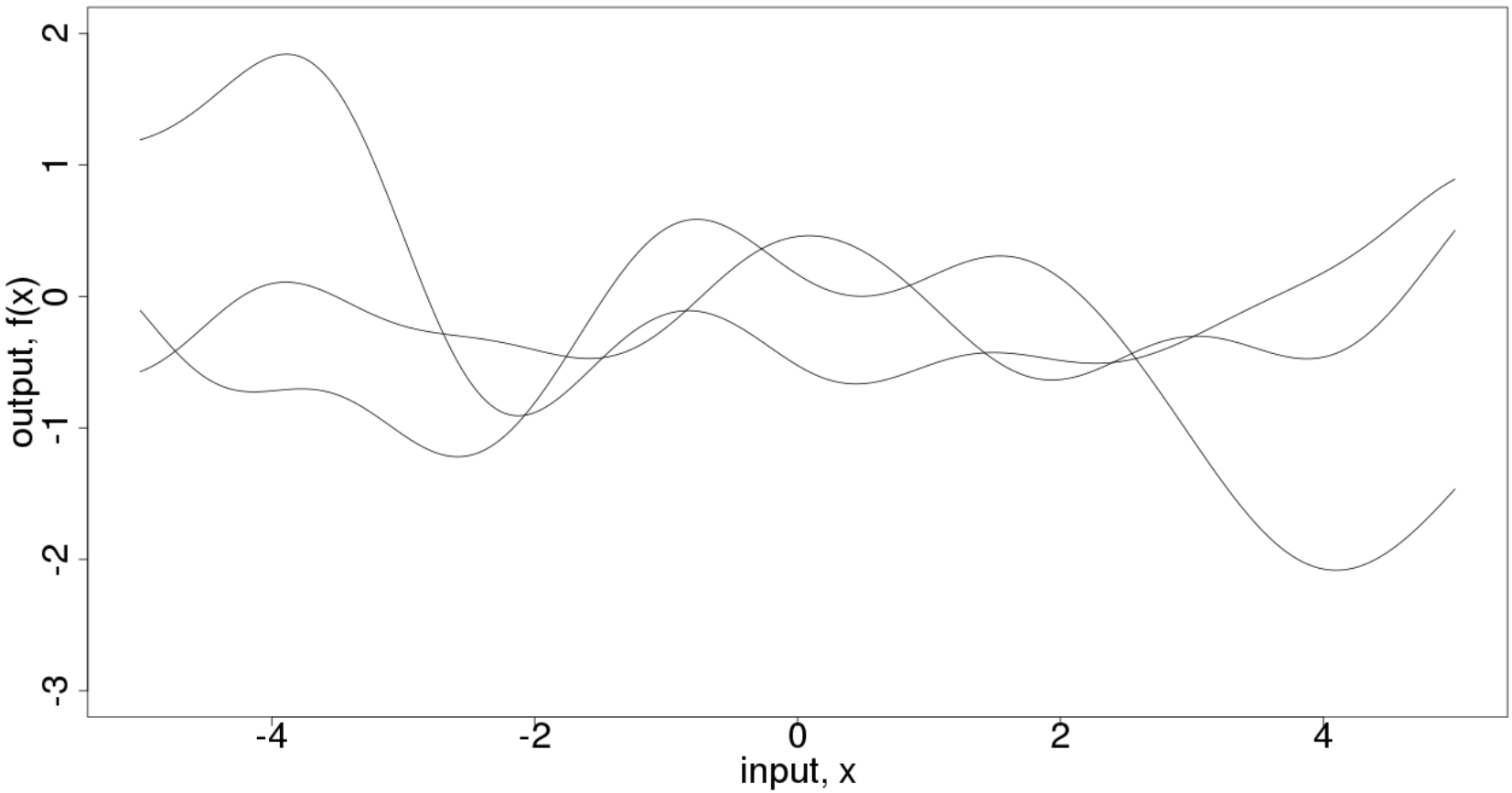
We use the squared exponential covariance function

$$\Sigma_{ij} = \alpha^2 \exp\left\{ -\frac{(x_i - x_j)^2}{l^2} \right\}.$$

Regression example: data



Regression example: GP prior distribution



Regression example posterior distribution

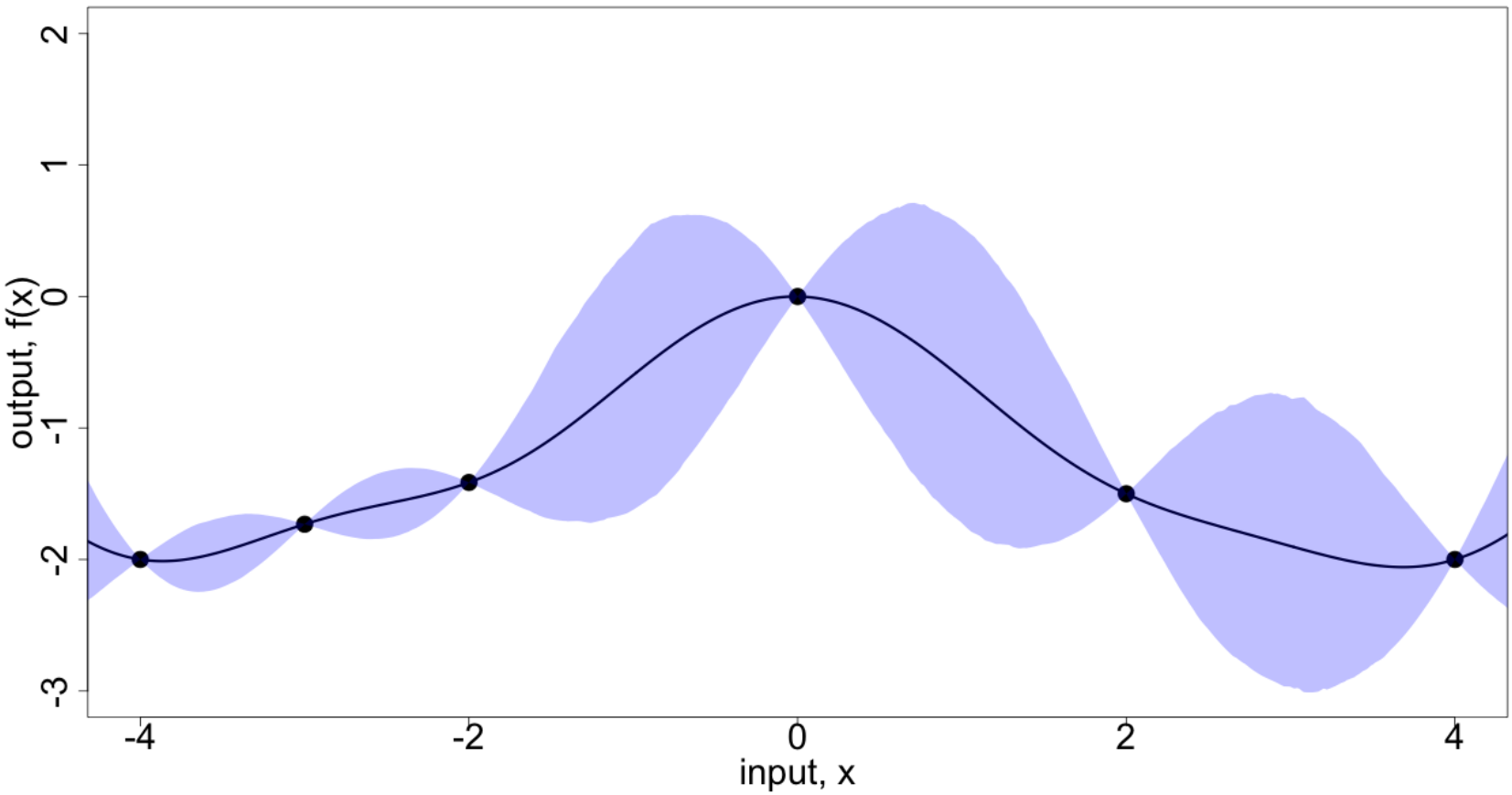


Table of Contents

1 Epidemic modelling

2 Homogeneous and heterogeneous mixing

3 Parametric inference

4 Gaussian processes

5 Gaussian processes for epidemic models

6 Foot and mouth disease

7 Avian influenza

Prior distributions

The posterior distribution is given by

$$\pi(\beta, \gamma, \mathbf{i}, \omega, i_\omega, l|\mathbf{r}, \lambda) \propto \quad \pi(\mathbf{i}, \mathbf{r}|\beta, \lambda, \gamma, \omega, i_\omega) \times \pi(\beta \mid l)\pi(l)\pi(\gamma)\pi(\omega)\pi(i_\omega|\omega).$$

We assume *a priori* that $l \sim \text{Exp}(\chi_l)$, $\gamma \sim \text{Exp}(\chi_\gamma)$, that ω is uniformly distributed on $\{1, \ldots, n\}$ and that $-i_\omega \sim \text{Exp}(\chi_\omega)$.

Independent model

The independent model is the simplest model:

$$\beta^{(j)} = \exp(f^{(j)}), \qquad f^{(j)} \sim \mathcal{GP}\left(0, \Sigma^{(j)}\right).$$

Multi-output covariance model

We can allow for correlation between the infection rate functions

$$\begin{pmatrix} f^{(1)} \\ f^{(2)} \\ \vdots \\ f^{(p)} \end{pmatrix} \sim \mathcal{GP}\left(0, \begin{pmatrix} \Sigma^{(1,1)} & \dots & \rho_{1,p}\Sigma^{(1,p)} \\ \rho_{2,1}\Sigma^{(2,1)} & \dots & \rho_{2,p}\Sigma^{(2,p)} \\ \vdots & & \vdots \\ \rho_{p,1}\Sigma^{(p,1)} & \dots & \Sigma^{(p,p)} \end{pmatrix}\right),$$

$$\beta^{(j)} = \exp(f^{(j)})$$

Discrepancy based model

We can set $f^{(1)}$ as a baseline, to which we assign a GP prior with mean zero and covariance matrix $\Sigma^{(1)}$. For $j = 2, \ldots p$ we then assume that

$$f^{(j)} = f^{(1)} + u^{(j)}, \quad u^{(j)} \sim \mathcal{GP}\left(0, \Sigma^{(j)}\right),$$

where $u^{(j)}$ represents the discrepancy between $f^{(j)}$ and $f^{(1)}$, with $f^{(1)}, u^{(2)}, \ldots, u^{(p)}$ assumed to be mutually independent.

MCMC algorithm

Algorithm 1 Basic Structure of the MCMC Algorithms

1: Initialize the chain with values $\gamma^{(0)}$, $\beta^{(0)}$, $I_1^{(0)}, \ldots, I_m^{(0)}$, $\mathbf{i}^{(0)}$, $\omega^{(0)}$ and $i_\omega^{(0)}$

Repeat the following steps

2: Update β using a Metropolis-Hastings step;
3: Update GP hyperparameters using a Metropolis-Hasting step;
4: Update γ using a Gibbs step;
5: Update ω and an infection time $i_\omega|\omega$ using a Metropolis-Hastings step;
6: Choose an infection time at random and update it using a Metropolis-Hastings step.

Table of Contents

1 Epidemic modelling

2 Homogeneous and heterogeneous mixing

3 Parametric inference

4 Gaussian processes

5 Gaussian processes for epidemic models

6 Foot and mouth disease

7 Avian influenza

FMD in the UK

In 2001, there was a large outbreak of Foot and Mouth Disease (FMD) in the UK.
In Cumbria, which was the most affected area, there were 5,436 farms consisting of

- 1,061 sheep farms,
- 1,064 cattle farms, and
- 3,253 farms with both sheep and cattle.

Of these farms, n= 1,021 were infected including 8% of sheep farms, 13% of cattle farms, and 24% of farms where both sheep and cattle were present.

Multioutput for FMD

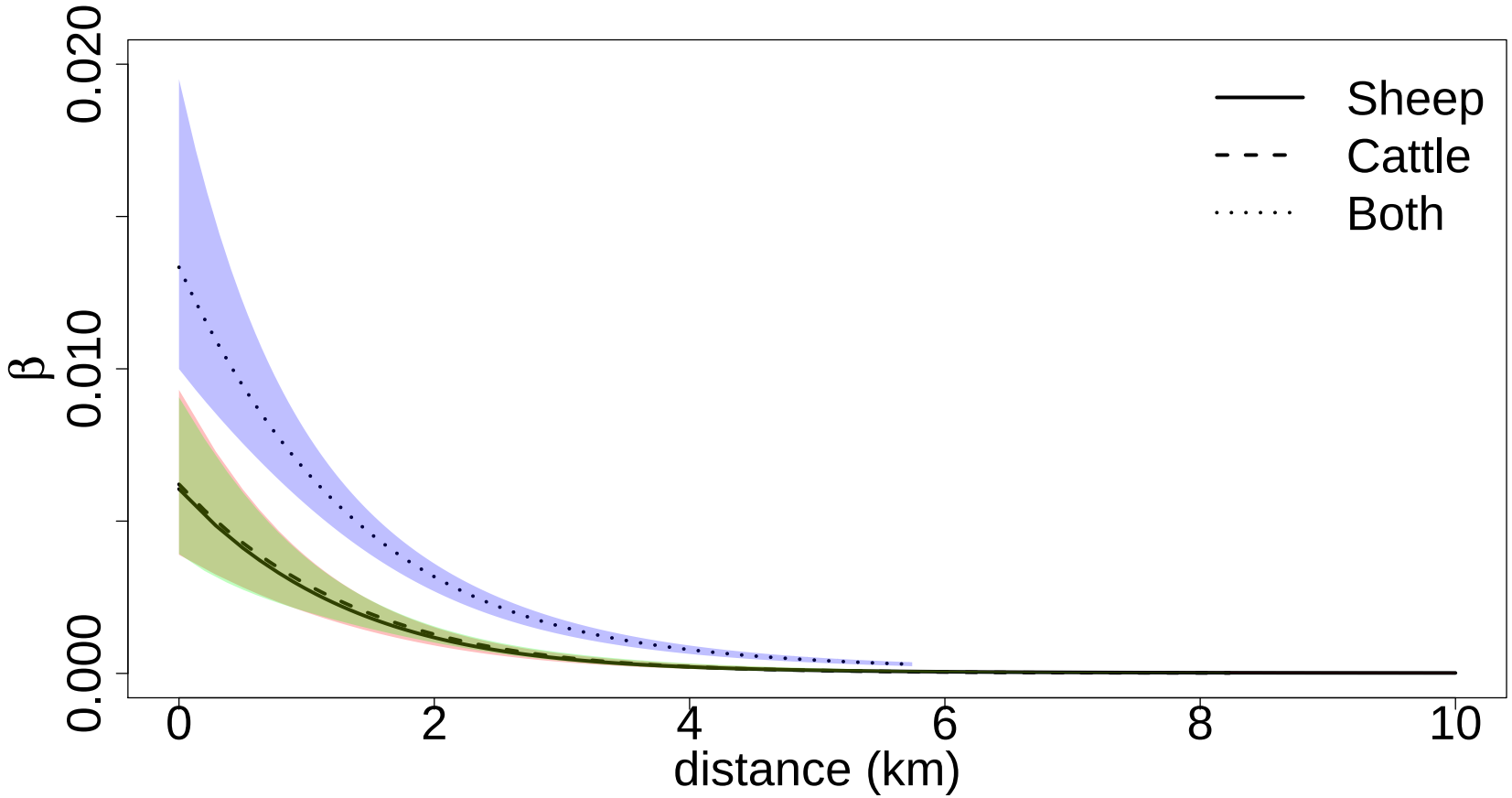


Figure: Results of the MOC model applied to the FMD data set.

Multioutput for FMD

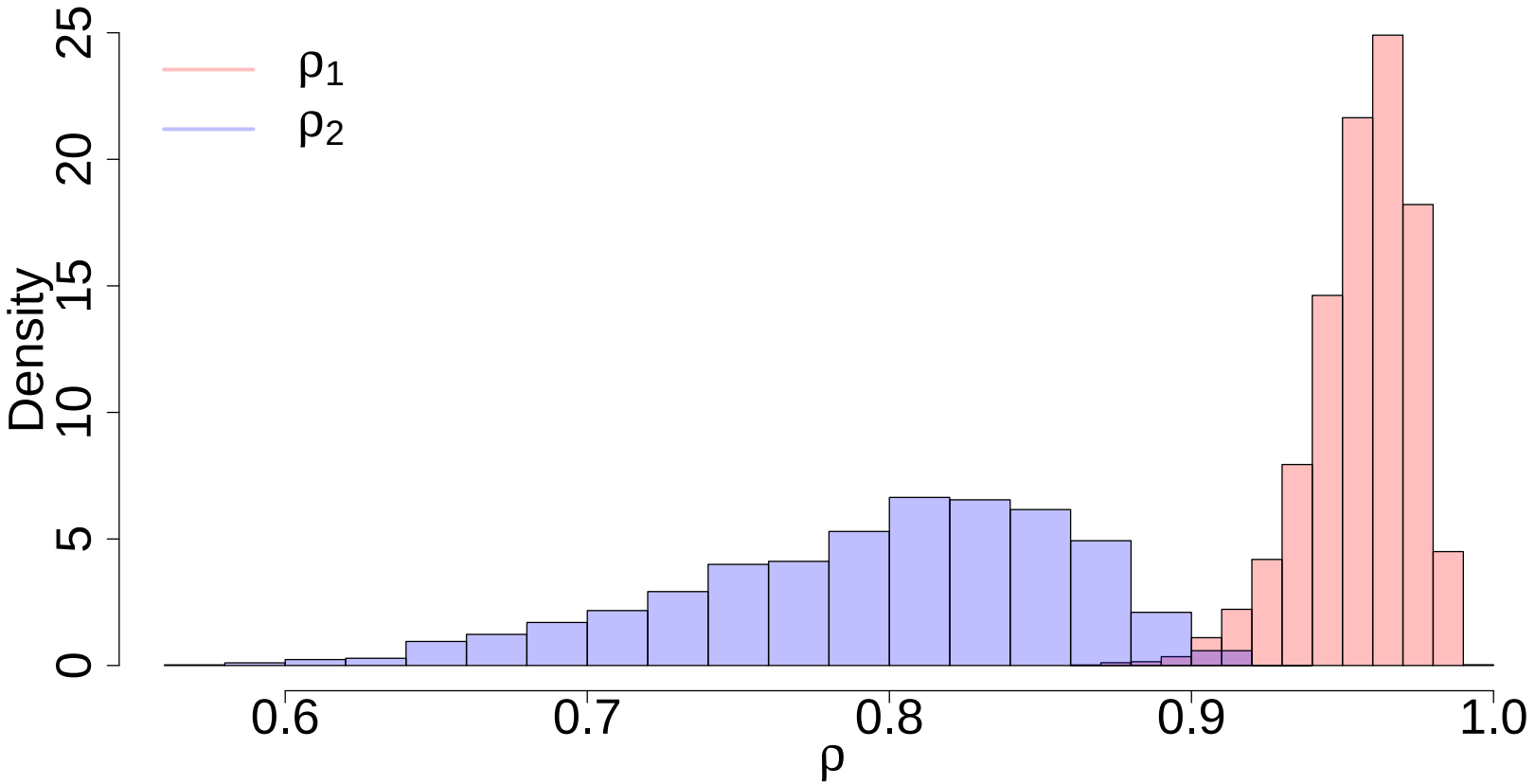


Figure: Results of the MOC model applied to the FMD data set.

Discrepancy for FMD

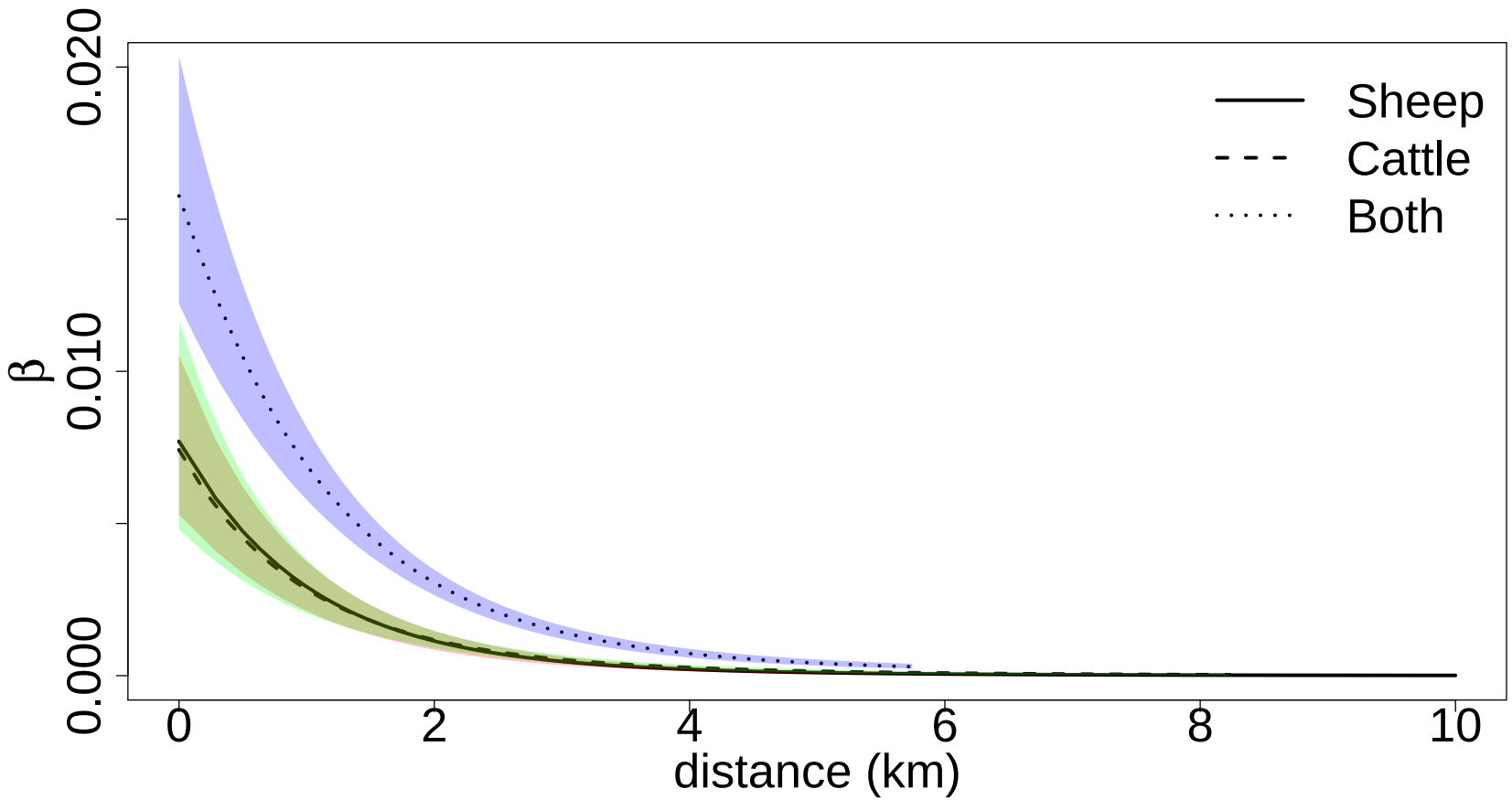


Figure: Results of the DB model applied to the FMD data set: posterior medians and 95% credible intervals for the infection rate functions.

Discrepancy for FMD

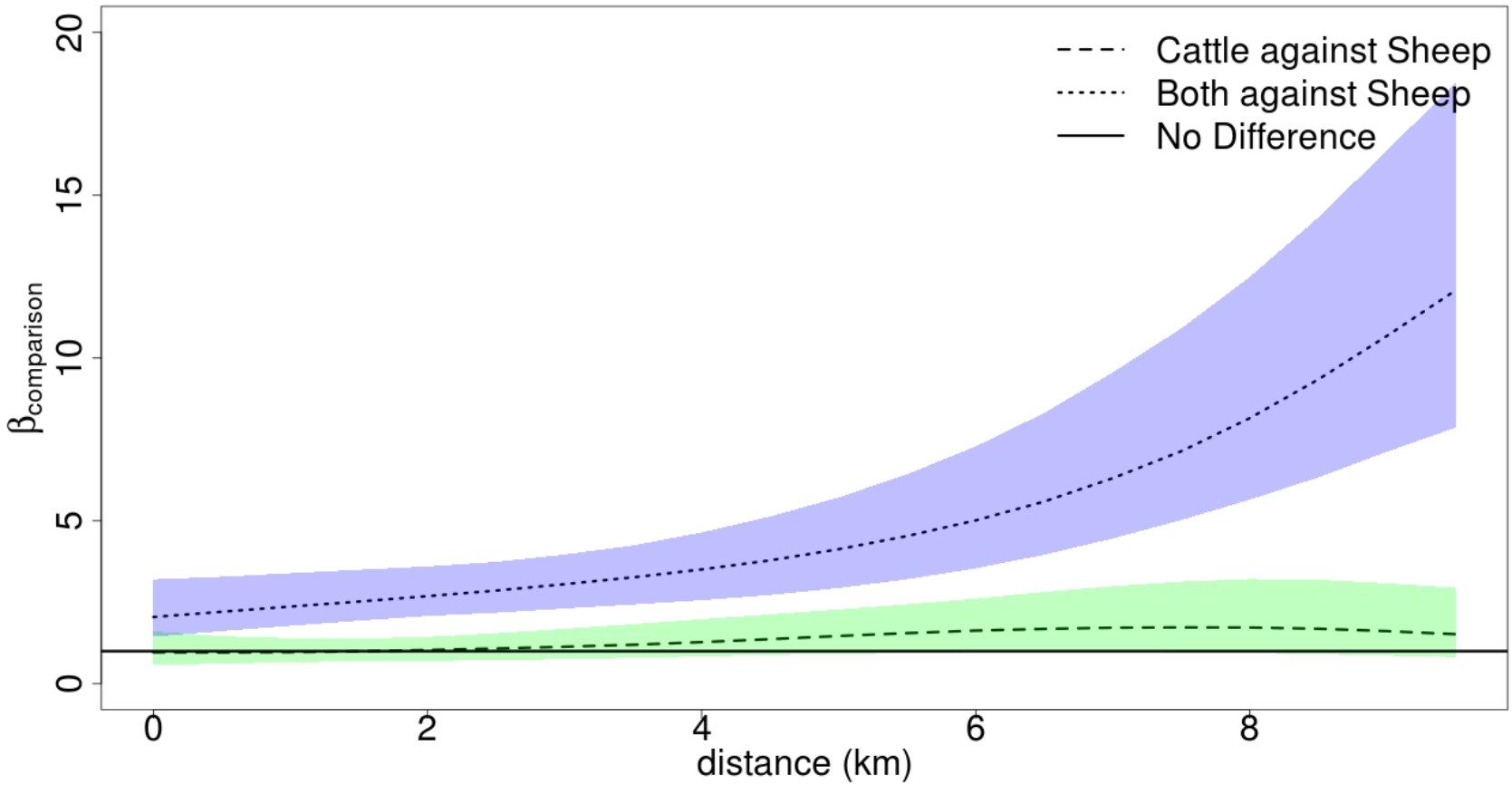


Figure: Results of the DB model applied to the FMD data set: posterior median and 95% credible intervals for the discrepancies.

Table of Contents

1 Epidemic modelling

2 Homogeneous and heterogeneous mixing

3 Parametric inference

4 Gaussian processes

5 Gaussian processes for epidemic models

6 Foot and mouth disease

7 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
⋮	⋮	⋮	⋮

Including culling

We define the following sets of farms based on their state at the end of the epidemic.

Set	Infected	Culled	Pre-emptively Culled
$\mathcal{A}$	$\times$	$\times$	$\times$
$\mathcal{B}$	✓	✓	$\times$
$\mathcal{C}$	✓	✓	✓
$\mathcal{D}$	$\times$	✓	✓

Note that if a farm has been pre-emptively culled, we are unable to distinguish whether it belongs to set $\mathcal{C}$ or $\mathcal{D}$ unless its infection status is known.

Including culling

The likelihood function including culling is given by

$$\pi(\mathbf{i}, \mathbf{r}^c, \mathcal{B}, \mathcal{C}, \mathcal{D} \mid \beta, \lambda, \gamma, \omega, i_\omega, \mathbf{r}^p)$$

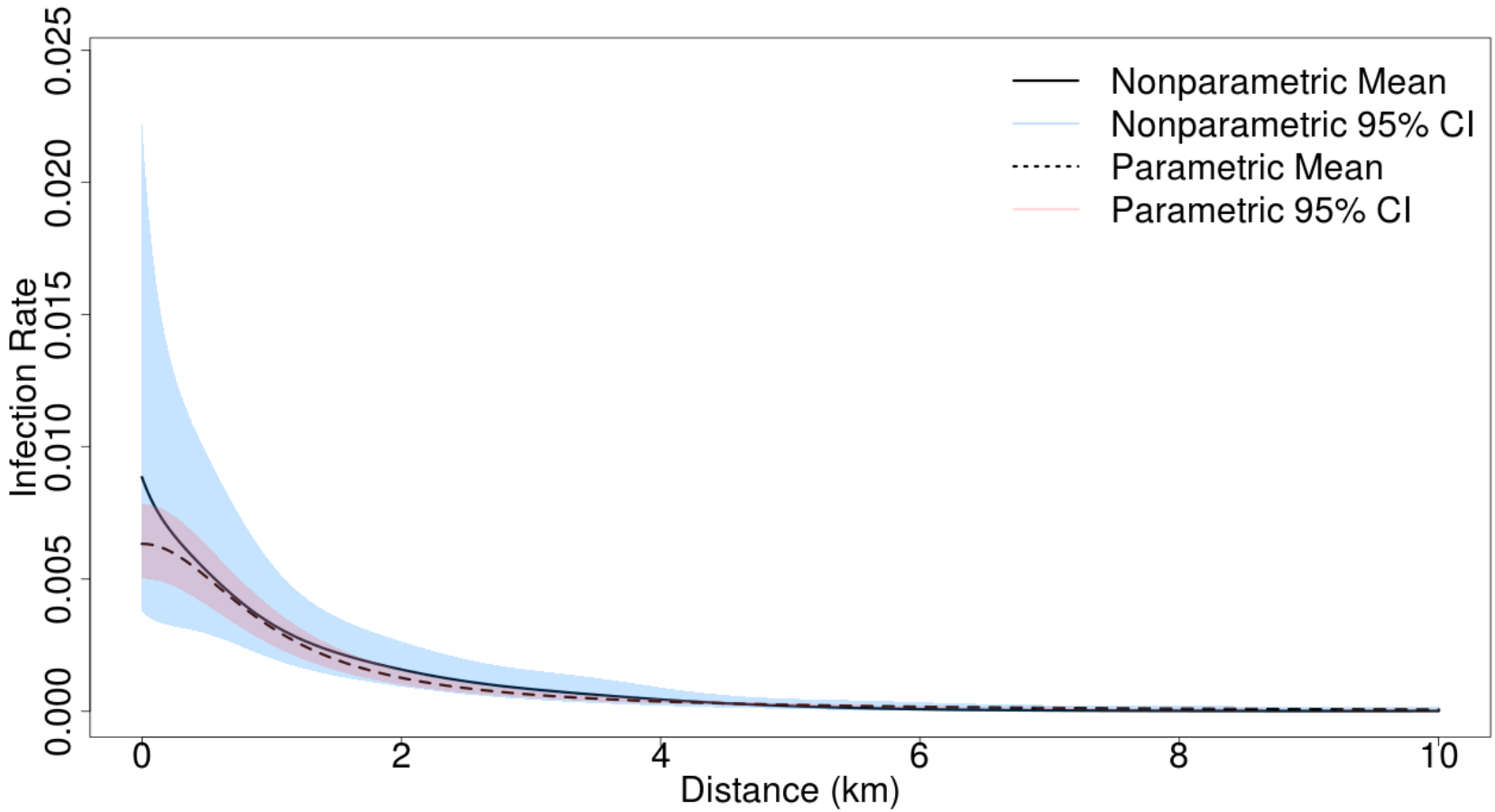
$$= \exp\{-\Psi\} \prod_{\substack{j \in \mathcal{B} \cup \mathcal{C} \\ j \neq \omega}} \left(\sum_{k \in \mathcal{Y}_j} \beta(d_{k,j}) \right) \prod_{j \in \mathcal{B}} h(r_j - i_j \mid \lambda, \gamma) \prod_{j \in \mathcal{C}} S(r_j - i_j \mid \lambda, \gamma).$$

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

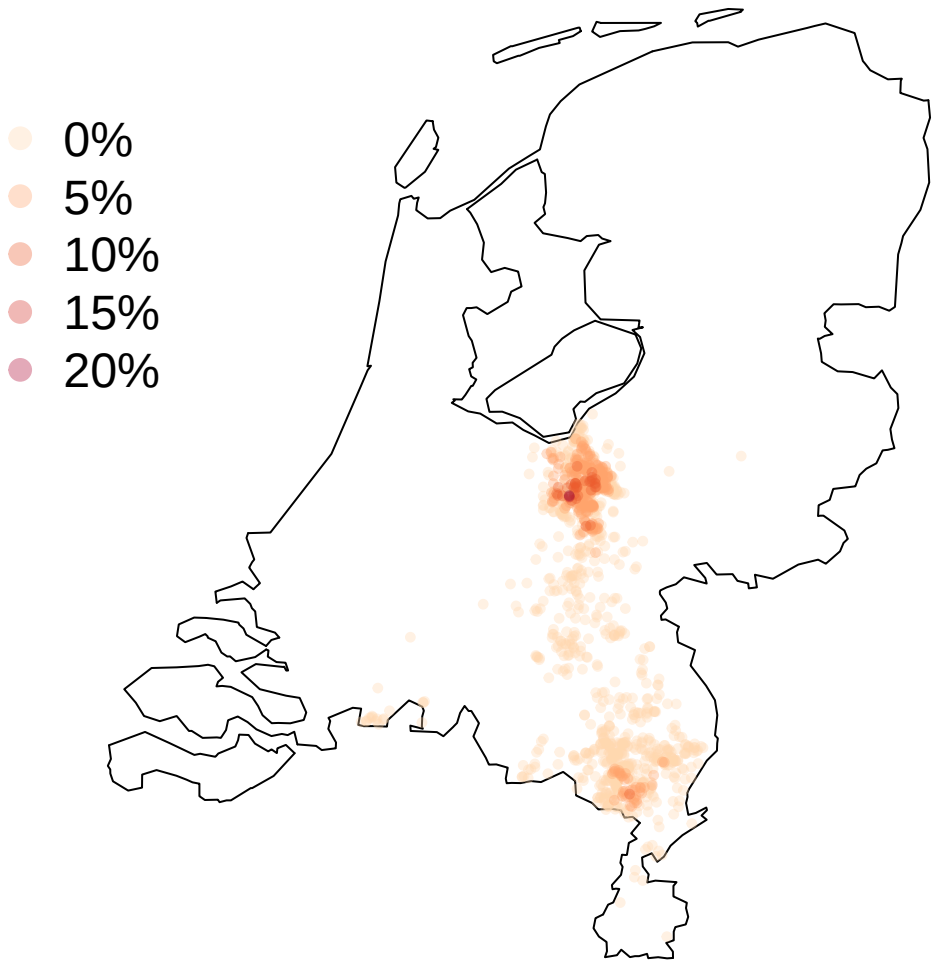
$$\left. + \sum_{k \in \mathcal{D}} \beta(d_{j,k}) \left((r_j \wedge r_k) - (i_j \wedge r_k) \right) \right].$$

Avian Influenza Results

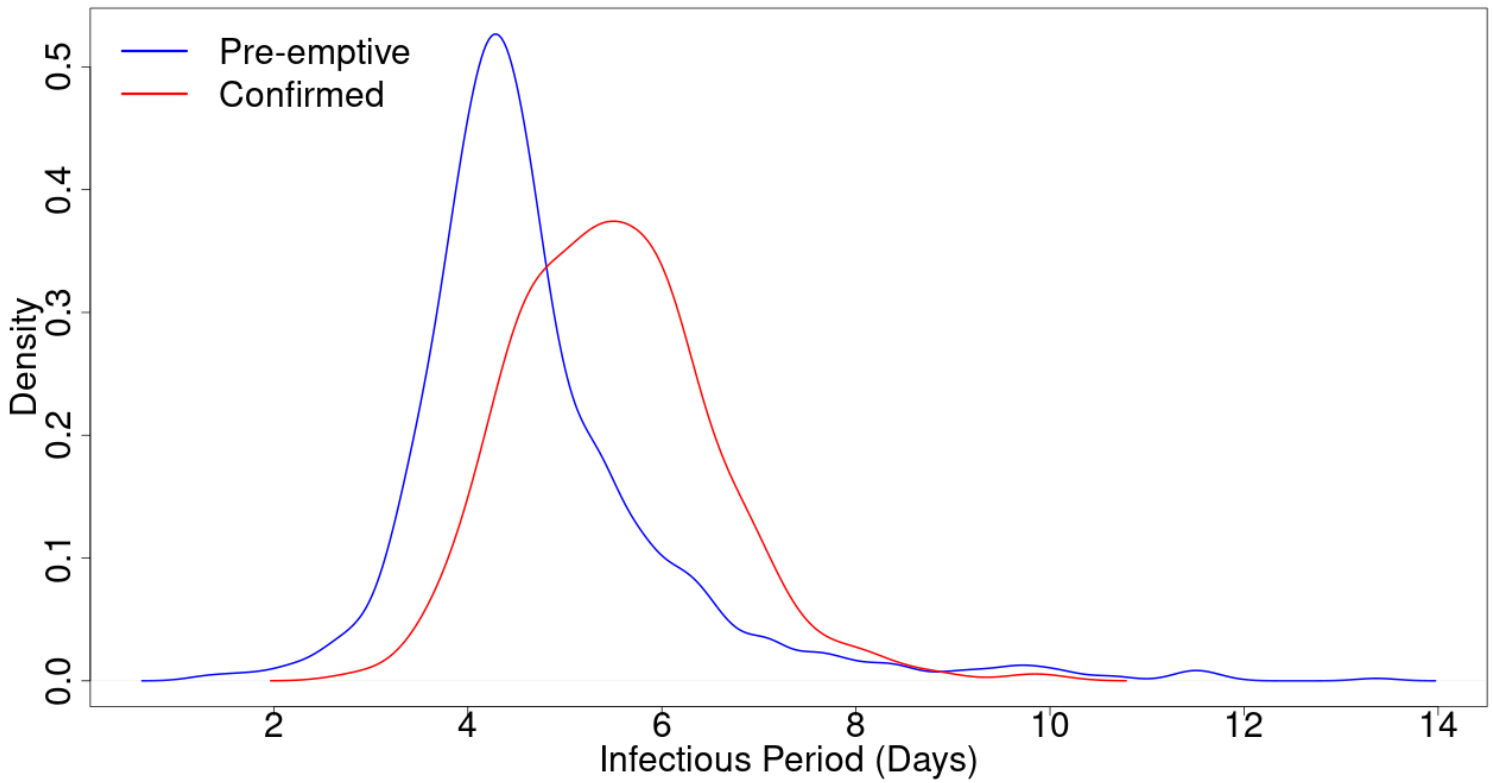
We then applied the GP inference method to the Avian Influenza data set and compared our model to the parametric model.



Avian Influenza Results



Avian Influenza Results



Culling Analysis

Table: Posterior predictive medians (95% probability intervals) for the number of infected and culled farms and the amount of compensation paid.

Radius (km)	Infected Farms	Culled Farms	Compensation (mil)
0	443 (151, 644)	443 (151, 644)	24.8 (8.62, 35.9)
1	297 (110, 535)	489 (215, 709)	27.2 (12.2, 38.9)
2	283 (108, 608)	488 (217, 740)	27.5 (12.2, 41.7)
3	283 (112, 582)	517 (242, 775)	29.0 (13.2, 43.1)
4	274 (105, 564)	512 (228, 793)	28.5 (12.3, 43.9)
5	280 (109, 549)	527 (226, 797)	39.2 (12.4, 41.9)

Table of Contents

1 Epidemic modelling

2 Homogeneous and heterogeneous mixing

3 Parametric inference

4 Gaussian processes

5 Gaussian processes for epidemic models

6 Foot and mouth disease

7 Avian influenza

Conclusion

- 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 use this to analyse control strategies.

For more information:

- Seymour, R.G., Kypraios, T., O'Neill, P.D. and Hagenaars, T.J. (2021). A Bayesian nonparametric analysis of the 2003 outbreak of highly pathogenic avian influenza in the Netherlands. J R Stat Soc Ser C 70 (5).
- Seymour, R.G., Kypraios, T. and O'Neill, P.D. (2022). Bayesian nonparametric inference for heterogeneously mixing infectious disease models. Proc Natl Acad Sci. 119 (10).