Bayesian Non–Parametric Inference for Stochastic Epidemic Models

Theo Kypraios
University of Nottingham
http://www.maths.nott.ac.uk/∼tk

Mathematical Challenges for Long Epidemic Time Series, University of Warwick

Epidemic Modelling

Mathematical and statistical modelling has become a valuable tool in the analysis of infectious disease dynamics:

- control strategies;
- informing policy–making at the highest levels;
- fundamental role in the fight against disease spread.

Modeling & Estimation

- Enormous attention has been given to the development of:
  - realistic (parametric) model of varying complexity, and
  - methods for efficient parameter estimation (eg infection/removal rates).
- Particular focus has been given to the construction of computationally intensive methods, for example
  - Markov Chain Monte Carlo (MCMC),
  - Sequential Monte Carlo (SMC),
  - Approximate Bayesian Computation (ABC),
  - Plug and play,
  - ...

Research Article

Statistics in Medicine

Introduction and snapshot review: Relating infectious disease transmission models to data

Philip D. O'Neill[*†]

Volume 29, Issue 20, pages 2069–2077, 10 September 2010

Why Non-Parametric?

There has been relatively little activity in the area of non–parametric inference; see, for example, Becker and Yip (1989), Boys and Giles (2007), Kenah (2013).

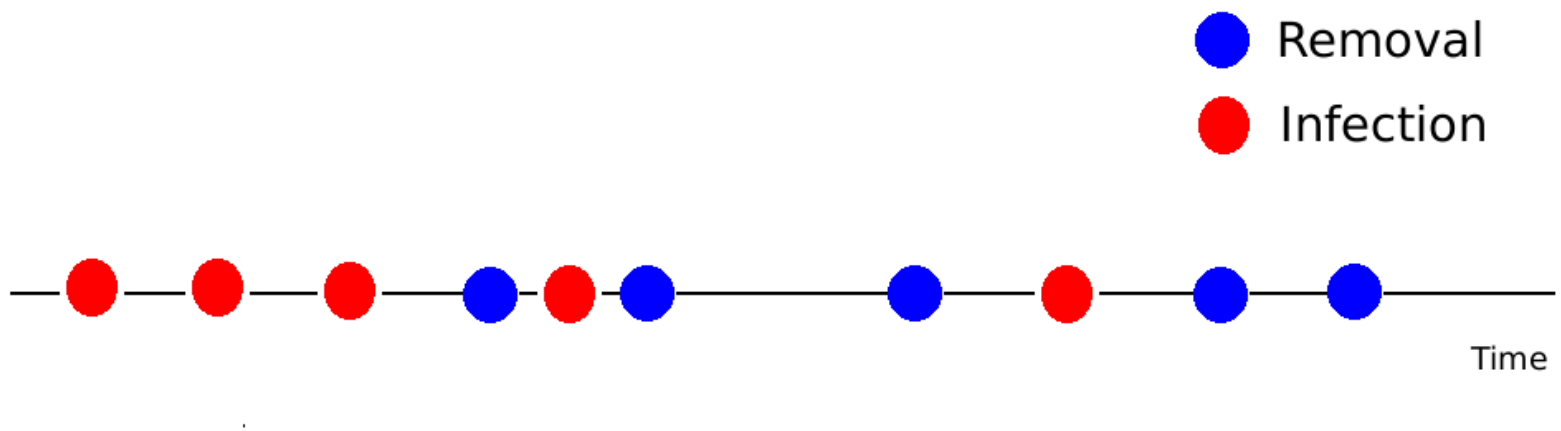
Parametric models are of great value.

However, adopting a non–parametric approach:

- helps to avoid erroneous conclusions . . .
- . . . and biased results arising from the use of parametric models with (perhaps) inappropriate assumptions.
- Offers great modelling flexibility.
- Allows the data to speak for themselves.

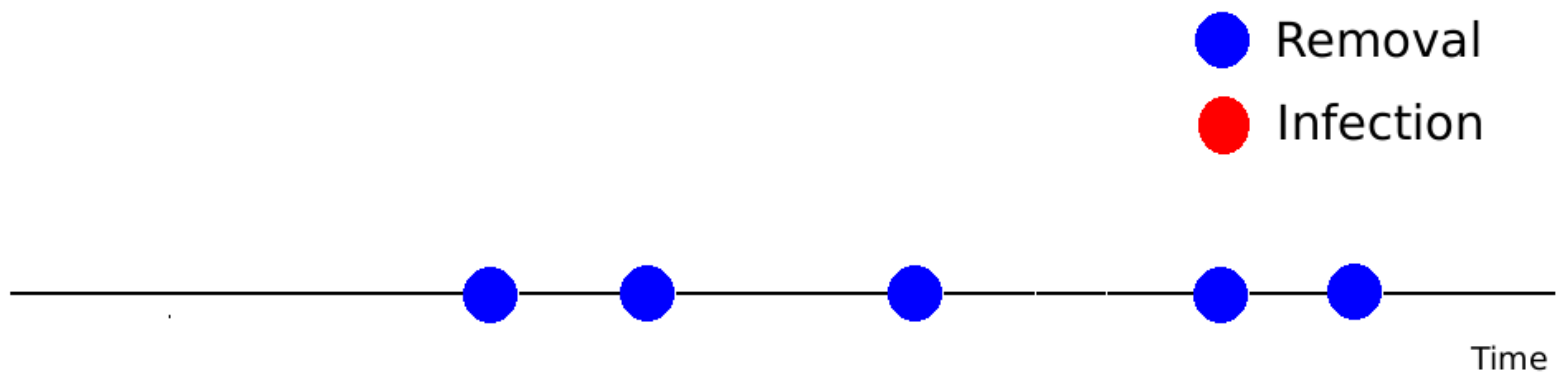
An Ideal Setup

Consider a closed population of size $N = 10$ individuals. Assume that epidemic has ceased and the whole process is observed:



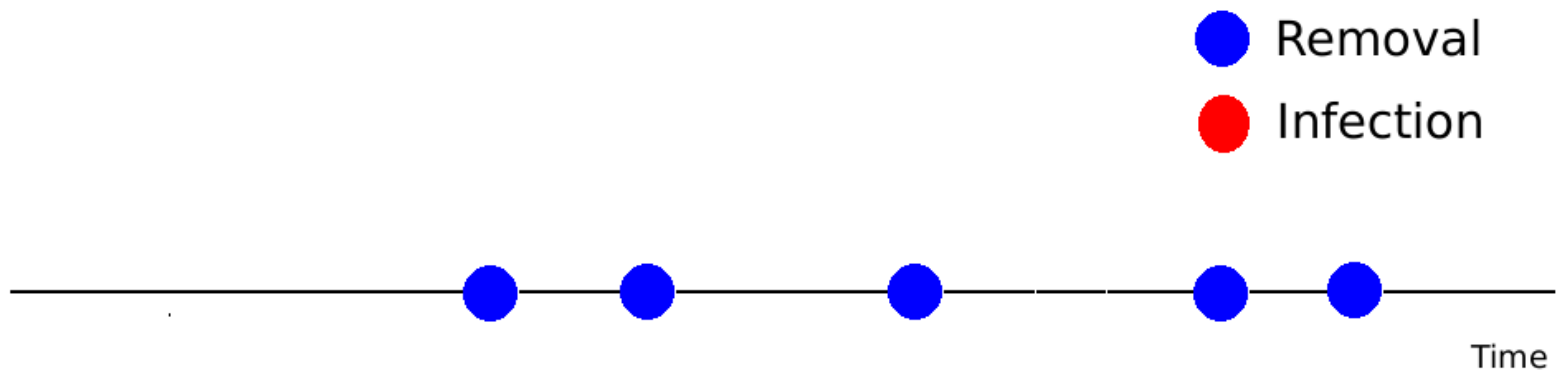
A More Pragmatic Setup

Consider a closed population of size $\mathcal{N}$ individuals.
Assume that that epidemic has ceased and only removal events are observed.



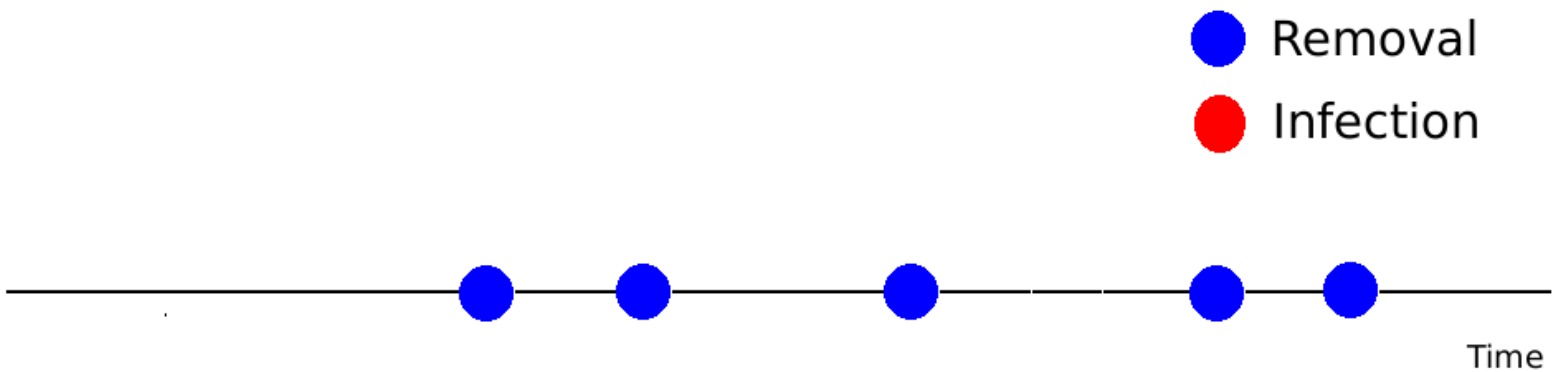
Setup

Consider a closed population of size $\mathcal{N}$ individuals.
Assume that only removal events are observed.
One option: fit a homogeneously mixing Markov S–I–R model.



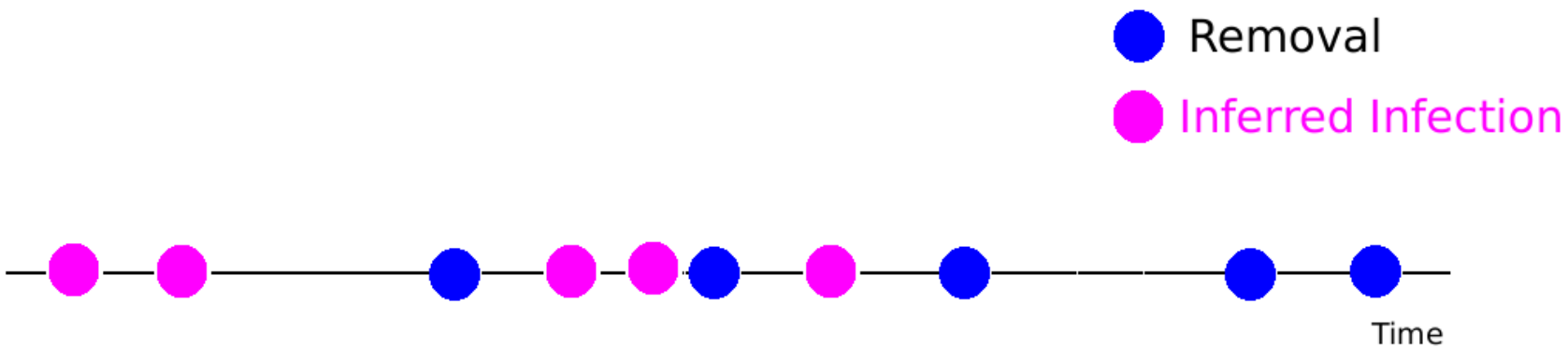
Setup

Consider a closed population of size $\mathcal{N}$ individuals.
Assume that only removal events are observed.
One option: fit a homogeneously mixing Markov S–I–R model.
How? Data augmentation within an MCMC framework.



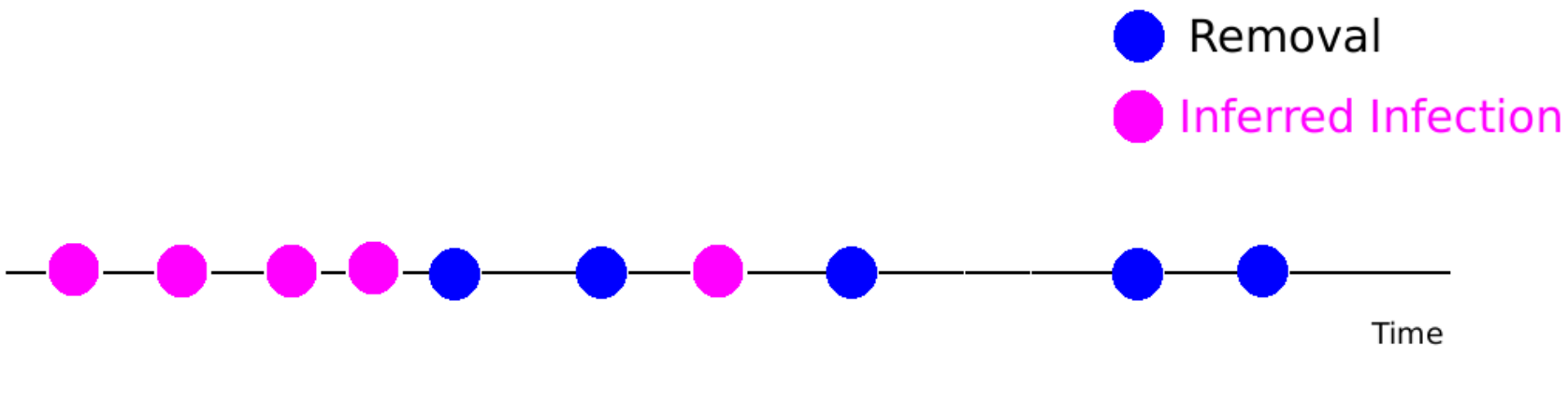
Setup

Consider a closed population of size $\mathcal{N}$ individuals.
Assume that only removal events are observed.
One option: fit a homogeneously mixing Markov S–I–R model.
How? Data augmentation within an MCMC framework.



Setup

Consider a closed population of size $\mathcal{N}$ individuals.
Assume that only removal events are observed.
One option: fit a homogeneously mixing Markov S–I–R model.
How? Data augmentation within an MCMC framework.



Setup (cont.)

- Underlying assumptions:
  - **S** $\rightarrow$ **I**: New infections occur at the points of a time non–homogeneous Poisson process with rate, for example,

$$\beta S_t I_t$$

  - **I** $\rightarrow$ **R**: Infectives become removed after an infectious period which has an Exponential distribution with rate γ

$$R_i - I_i \sim \text{Exp}(\gamma)$$

Setup (cont.)

- Underlying assumptions:
  - **S** $\rightarrow$ **I**: New infections occur at the points of a time non–homogeneous Poisson process with rate, for example,

$$\beta S_t I_t \quad \text{or} \quad \beta S_t I_t^{\delta} \quad \text{or} \quad \beta S_t^{\delta_1} I_t^{\delta_2} \quad \text{or} \quad \ldots$$

  - **I** $\rightarrow$ **R**: Infectives become removed after an infectious period which has an Exponential distribution with rate γ

$$\text{Exp}(\gamma) \quad \text{or} \quad \text{Gamma}(\mu, \nu) \quad \text{or} \quad \text{Weibull}(\mu, \nu) \quad \text{or} \quad \ldots$$

Main Idea

- Underlying Assumptions:
  - **S** $\rightarrow$ **I**: New infections occur at the points of a time non–homogeneous Poisson process with rate

$$h(t) > 0 \quad (t \in \mathbb{R})$$

  - **I** $\rightarrow$ **R**: Infectives become removed after an infectious period which has an arbitrary, but specified distribution, for example:

$$\text{Exp}(\gamma) \quad \text{or} \quad \text{Gamma}(\mu, \nu) \quad \text{or} \quad \text{Weibull}(\mu, \nu) \quad \text{or} \quad \ldots$$

Bayesian Inference

- We wish to infer $h(t)$ from data within a Bayesian framework.
- Surely, there are an uncountably infinite set of possible functions.
- How are we going to compute with this set in finite time?
- How do we place a prior distribution over a function?

Bayesian Inference

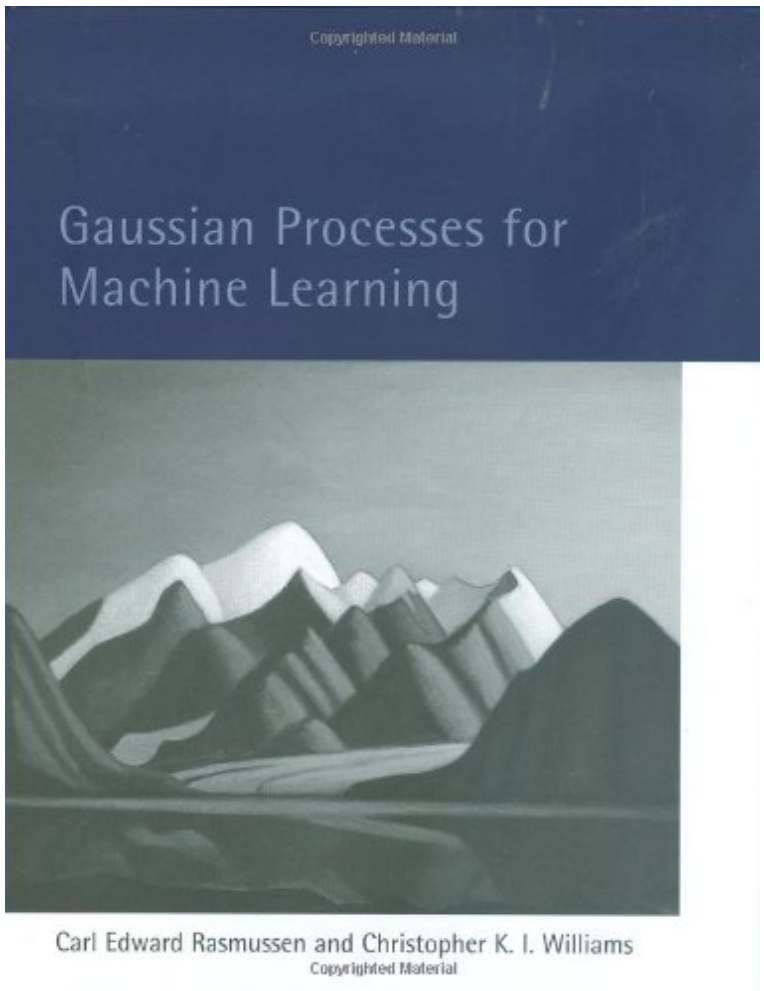
In this talk we discuss three different prior choices /models for the unknown function $h(t)$

- A Gaussian Process (GP)
- A Step Function
- A B-Spline

Gaussian Processes

A Gaussian process (GP) is a generalization of the Gaussian probability distribution.

- The Gaussian distribution is over vectors, whereas the Gaussian process is over functions.
- GPs are used to describe a distribution over functions.



Definition

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

A GP is completely specified by its mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$.

We shall write

$$f \sim \mathcal{GP}\left(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\right)$$

meaning "the function f is distributed as a GP with mean function $\boldsymbol{m}(\cdot)$ and covariance function $\boldsymbol{k}(\cdot, \cdot)$.

Gaussian Processes

Covariance Function

- The covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ is a crucial ingredient in GPs.
- It encodes our assumptions about the function which we wish to learn.
- The notion of similarity between data points is very important;
- under the GP view it is the covariance function that defines nearness or similarity.
- An arbitrary function of input pairs $\boldsymbol{x}$ and $\boldsymbol{x}'$ will not, in general, be a valid covariance function [-positive semidefinite-].

Gaussian Processes

Square Exponential Kernel

One of the most commonly used covariance functions is the *Square Exponential*

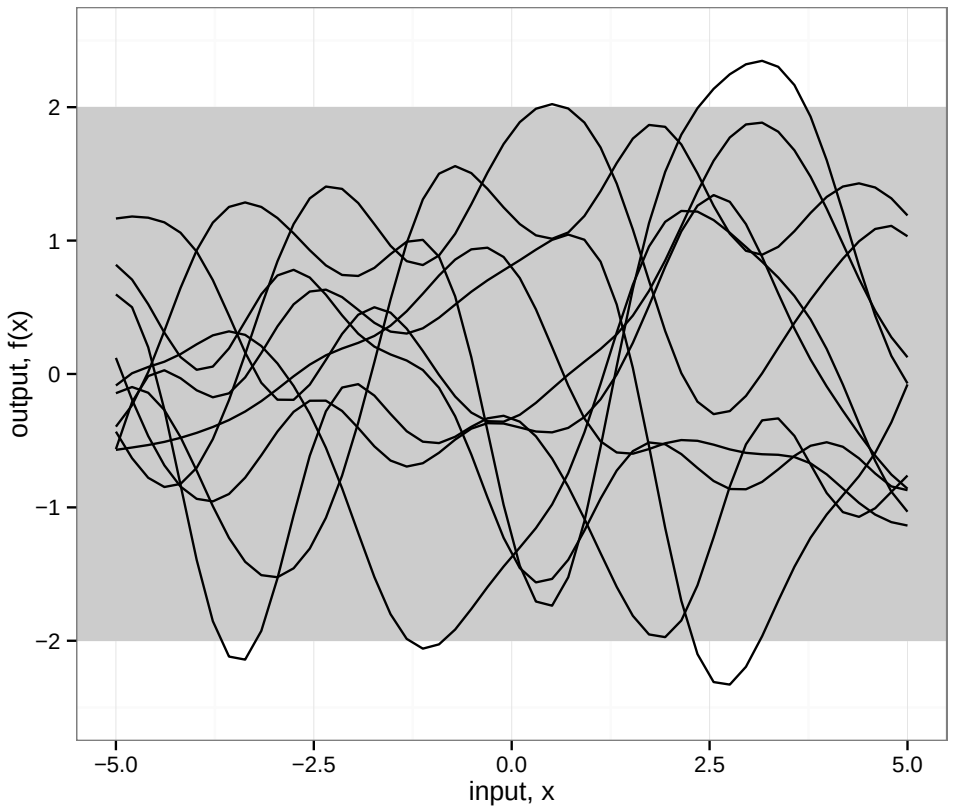
$$\text{cov}(f(\boldsymbol{x}_p), f(\boldsymbol{x}_q)) = k(\boldsymbol{x}_p, \boldsymbol{x}_q) = \exp\left\{-\frac{1}{2}\frac{|x_p - x_q|^2}{l^2}\right\}$$

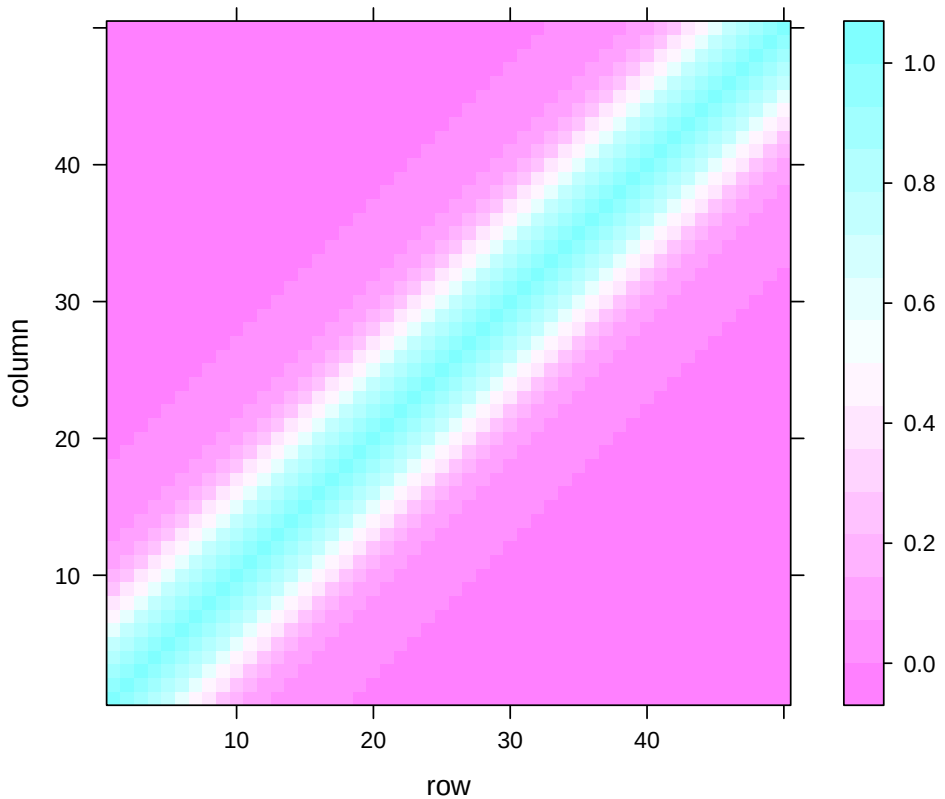
We see that

- the covariance is almost unity between variables whose corresponding inputs are very close;
- decreases as their distance in the input space increases;
- l is a characteristic length–scale parameter which informally can be thought of as roughly the distance you have to move in input space before the function value can change significantly.

Gaussian Processes

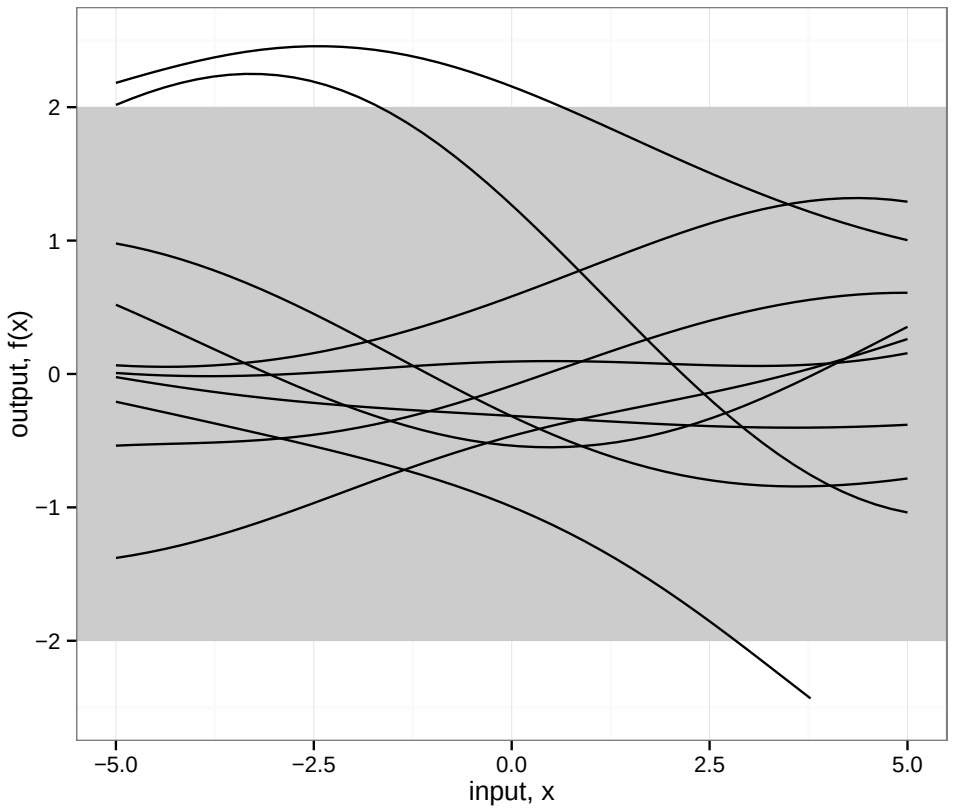
Square Exponential ($l = 1$)

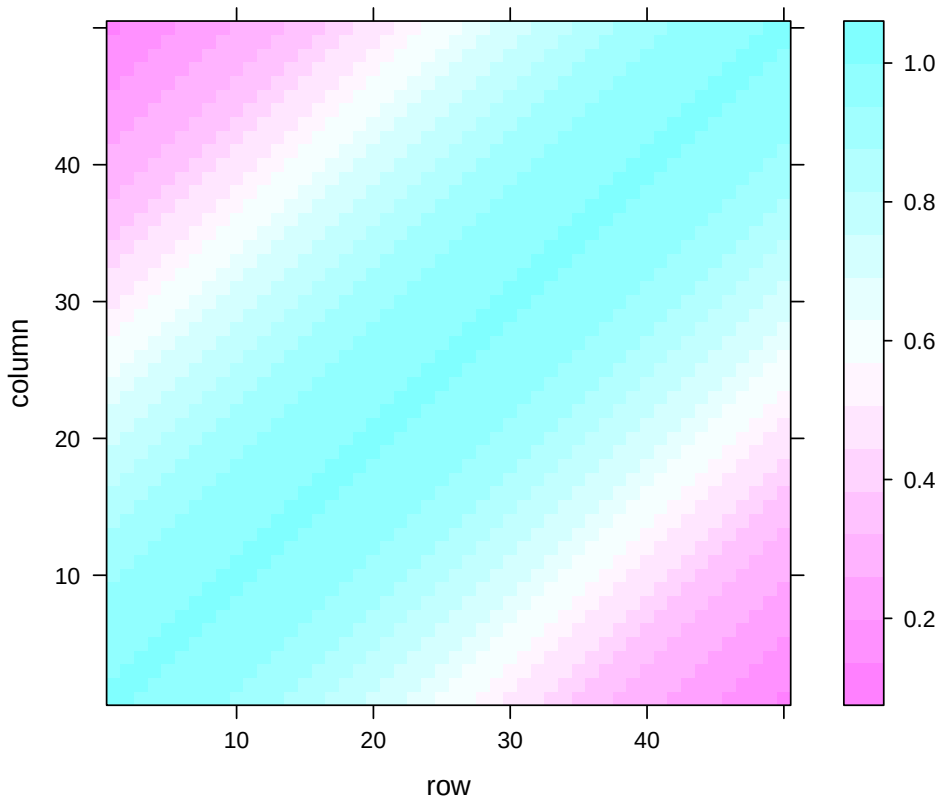




Gaussian Processes

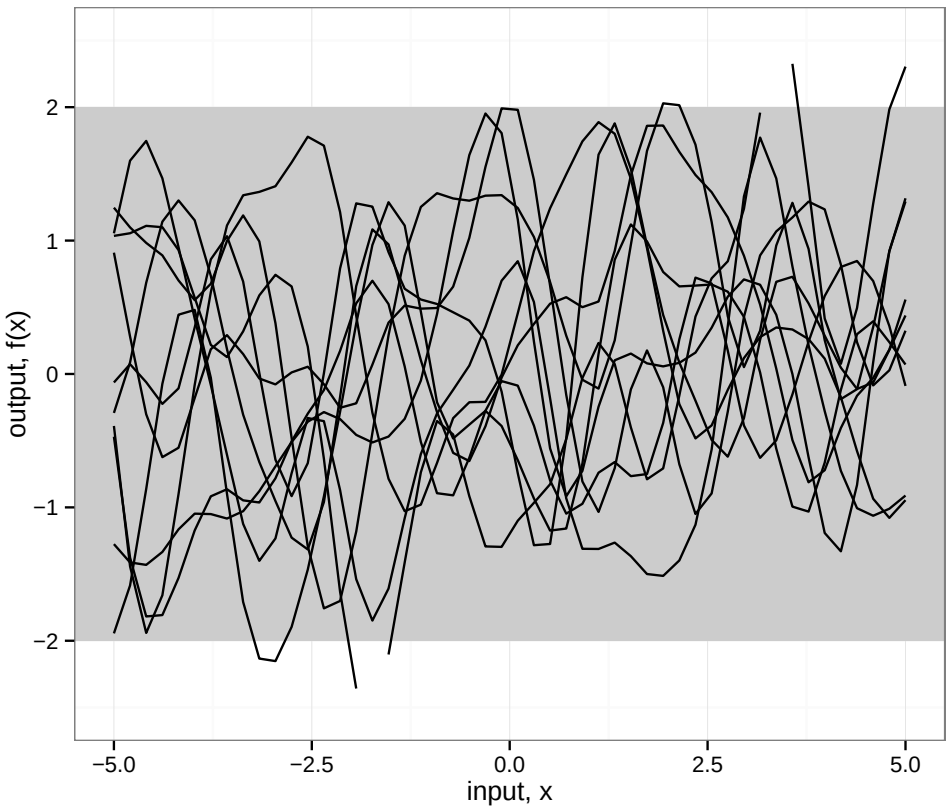
Square Exponential ($l = 5$)

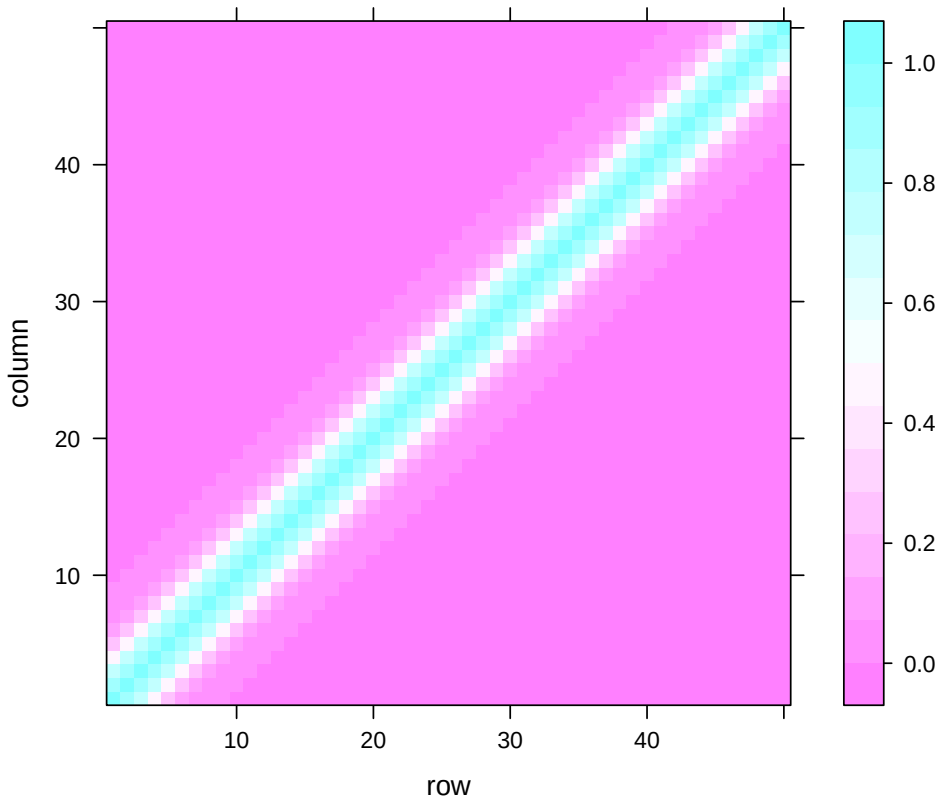




Gaussian Processes

Square Exponential ($l = 0.5$)





Non Parametric Estimation of a Poisson's Process Intensity

- Suppose (for now) that we have a fully observed epidemic (i.e. both infection and removal times).
- We wish to infer the rate at which infections occur and hence estimate $h(t)$ without assuming a parametric form for it.
- This is equivalent to estimating the intensity function of a time−inhomogeneous Poisson process.
- Likelihood−based inference is generally intractable due to the need to integrate an infinite-dimensional random function.
- Various approximations have been introduced to deal with this intractability (e.g. Diggle, 1985).

GPs on Stochastic Epidemic Models

Our approach is to

1. adopt a Bayesian framework;
2. assign a GP prior on $h(t)$;
3. overcome the intractability by incorporating a data−augmentation framework ...
4. ... and develop efficient Markov Chain Monte Carlo algorithms to explore the posterior distribution of interest.

.

GPs on Stochastic Epidemic Models

Our approach is to

1. adopt a Bayesian framework;
2. assign a GP prior on $h(t)$;
3. overcome the intractability by incorporating a data−augmentation framework ...
4. ... and develop efficient Markov Chain Monte Carlo algorithms to explore the posterior distribution of interest.

Such an approach avoids any approximations.

Log Gaussian Cox Process

GP ingredients

- Input space $\mathcal{X}$ (e.g. $\mathbb{R}^D$)
- Output space $\mathcal{Y} = \mathbb{R}$
- Positive-definite covariance function $k(\boldsymbol{x}, \boldsymbol{x}'; \theta) \to \mathbb{R}$
- mean function $m(\boldsymbol{x}; \theta) : \mathcal{X} \to \mathcal{Y}$

The idea is to use a GP as a prior for the Poisson process intensity $h(t) \quad (t > 0)$. But the output space is $\mathbb{R}$ ...

Do the natural thing, i.e. exponentiate $f(\cdot)$:

$$\begin{aligned} f(t) &\sim \mathcal{GP}(t, \theta) \\ g(t) &= \exp(f(t)) \end{aligned}$$

Doubly Intractable Inference

The likelihood of events $\{t_p\}_{p=1}^P$ between 0 and T

$$P\left(\{t_p\}_{p=1}^P \,|\, g(t) = \boldsymbol{g}\right) = \exp\left\{-\int_0^T \exp\left(g(t)\right)\,\mathrm{d}t + \sum_{p=1}^P g(t_p)\right\}$$

Problems

- $g(t)$ is infinite dimensional;
- the posterior distribution is doubly intractable in the sense that the likelihood is only known up to a constant which depends on the parameters of interest.
- Inference is hard and routine MCMC algorithms do not work.
- Recent advances in MCMC enable inference for such problems [Möller et al., 2004, Murray et al., 2006] given that you are able to generate *exact realisations* from the model.

Exact Simulation a Time−Inhomgeneous Poisson Process

Aim: Simulate points from a PP with intensity $\lambda\phi(t)$

Assume intensity λ on region $\mathcal{V}$

1. Find the measure of $\mathcal{V}$, i.e. $\mu(\mathcal{V})$
2. Sample the number of events $N(\mathcal{V}) \sim \text{Poisson}(\lambda\mu(\mathcal{V}))$
3. Distribute the $N(\mathcal{V})$ points, say $\{t_p\}_{p=1}^{P}$, independently and uniformly on $\mathcal{V}$
4. Remove t_p with probability $1 - \phi(t_p)$

The remaining events are points from the desired Poisson process.

Remarks

- The data are exactly drawn from a Poisson process with the desired intensity.
- We did not have to discover the function at more than a finite number of locations.
- We did not have to integrate the function.

Inference via Latent History

Adams, Murray and MacKay, 2009

Given the P events on $\mathcal{V}$ and the GP prior, the posterior is still intractable.

However, if we augment the state with the "latent history" of the generative procedure ... and assume there were K thinned events, $\{s_k\}_{k=1}^K$ we can write down the full joint distribution:

$$\pi\left(\{t_p\}_{p=1}^P, \{s_k\}_{k=1}^K, \boldsymbol{g}|\lambda, \theta\right) =$$

$$\lambda^{P+K} \exp\{-\lambda\mu(\mathcal{V})\} \quad \text{[homogeneous Poisson process]}$$

$$\times \prod_{p=1}^{P} g(t_p) \times \prod_{k=1}^{K}(1-g(t_p)) \quad \text{[probability of unthinned/thinned events]}$$

$$\times \mathcal{GP}\{g(t_p)_{p=1}^P\}, \{g(s_k)_{k=1}^K\}|\theta \quad \text{[GP prior]}$$

Exploring the Posterior Distribution

$\pi\left(\{t_p\}_{p=1}^{P}, \{s_k\}_{k=1}^{K}, \boldsymbol{g}|\lambda, \theta\right)$ is not pleasant but tractable and can sample from it using MCMC:

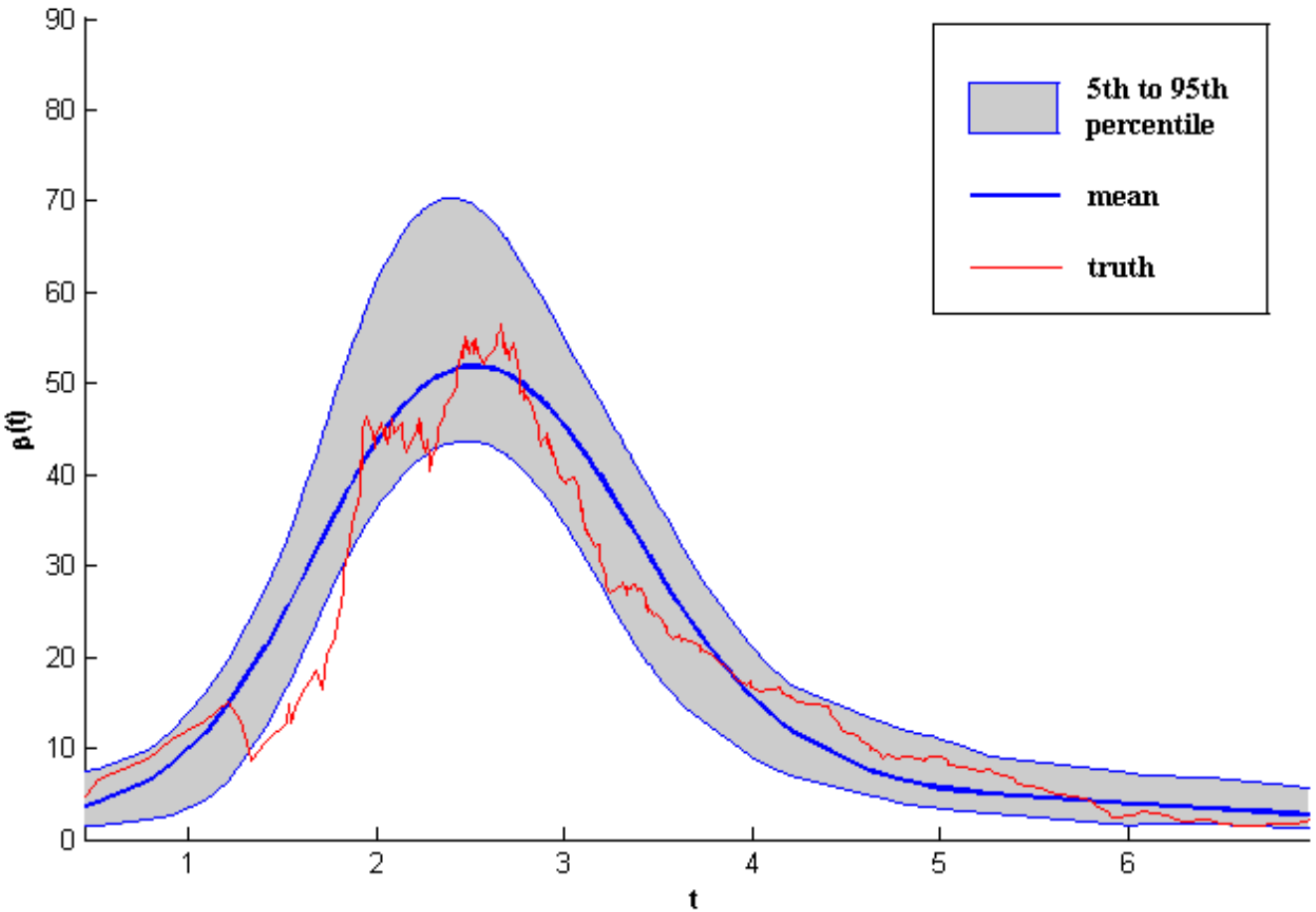
- We update each part of the latent state separately, conditioned on the others using a Gibbs−like sampler.
- Insert and remove latent thinned events via Metropolis−Hastings.
- Move latent thinned events around via Metropolis−Hastings.
- Sample the latent function Metropolis−Hastings (or Hamiltonian Monte Carlo for more efficiency)
- The hyperparameters of the GP can also be updated.

Accounting for Unobserved Infection Times

- So far we have assumed that we observe both the infection and removal times.
- Infection times are rarely observed.
- Augment the space (even further) with them and update them as well via an MCMC scheme.

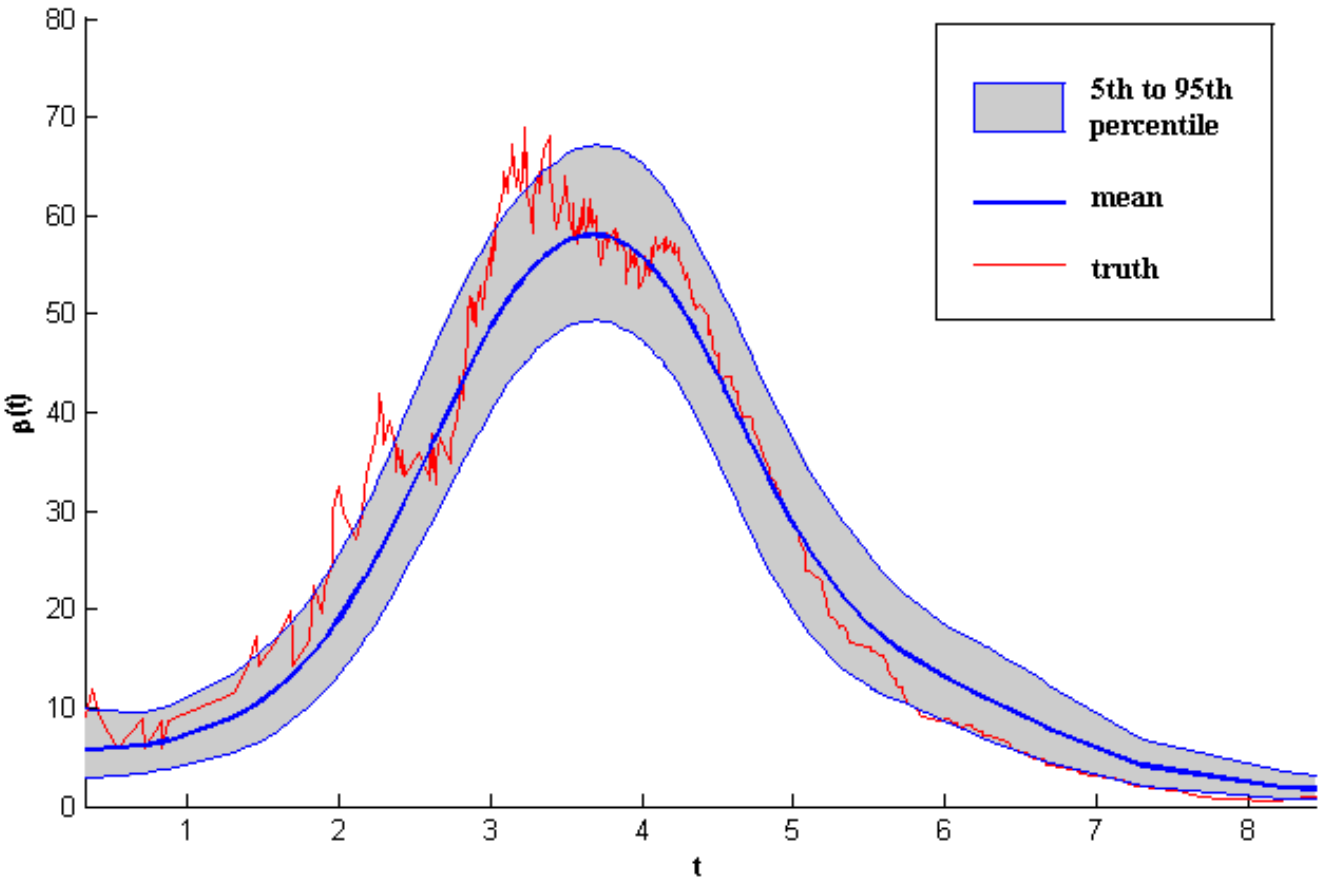
Does this Really Work?

SIR model, $N = 150$ true intensity: $\beta S_t I_t$, synthetic data



Does this Really Work?

SIR model, $N = 200$ true intensity: $\beta S_t I_t$, synthetic data



GPs for Epidemic Models

Reflections

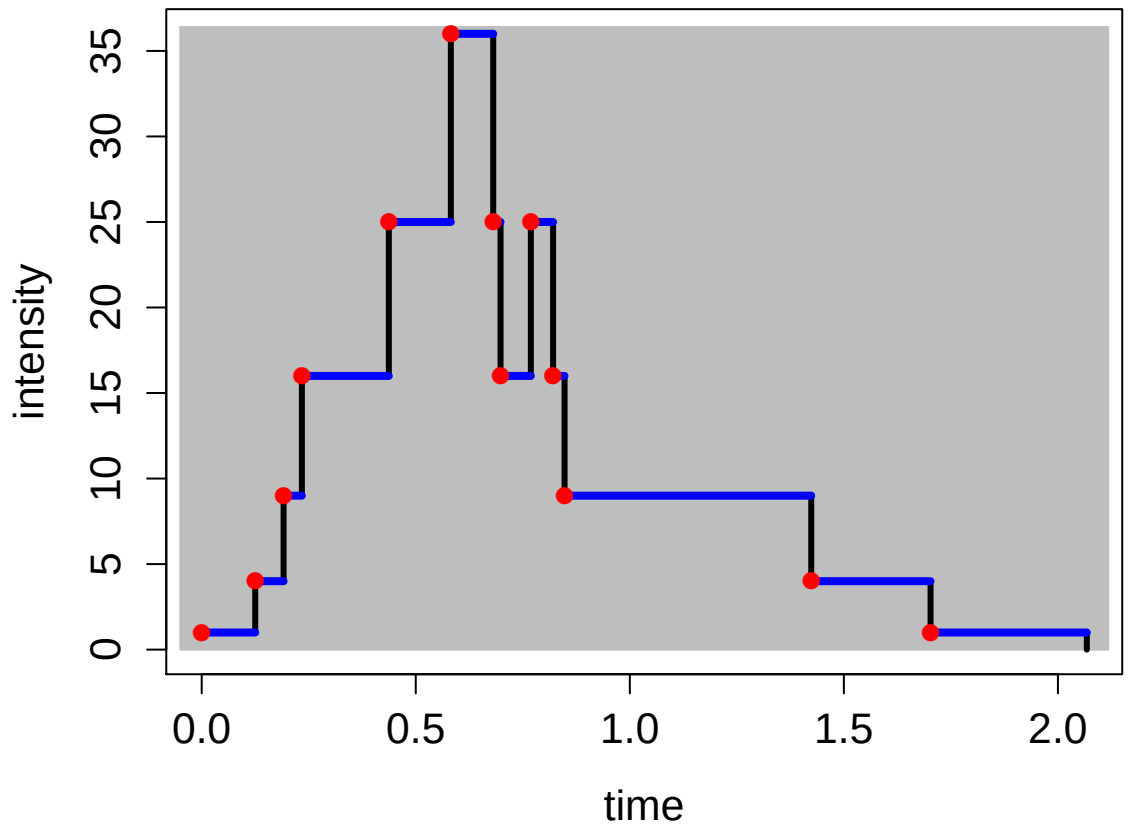
- Flexible models
- Covariance functions.
- Need efficient MCMC samplers.
- Computational cost can be quite high.

Model the infection rate $h(t)$ via Step−Functions

An alternative approach (model) for $h(t)$ is to treat it as a *step function*.

Ingredients:

- changepoints (number + locations)
- the heights.



Model the infection rate $h(t)$ via Step−Functions

- Modelling $h(t)$ enables straightforward calculation of $\int_0^T h(t)\,\mathrm{d}t$ required for the likelihood function.
- However:
  - the number of change points, say k,
  - their locations, $s_1, \ldots, s_k$
  - and the height of the function at these points $(h_0, h_1, \ldots, h_k)$

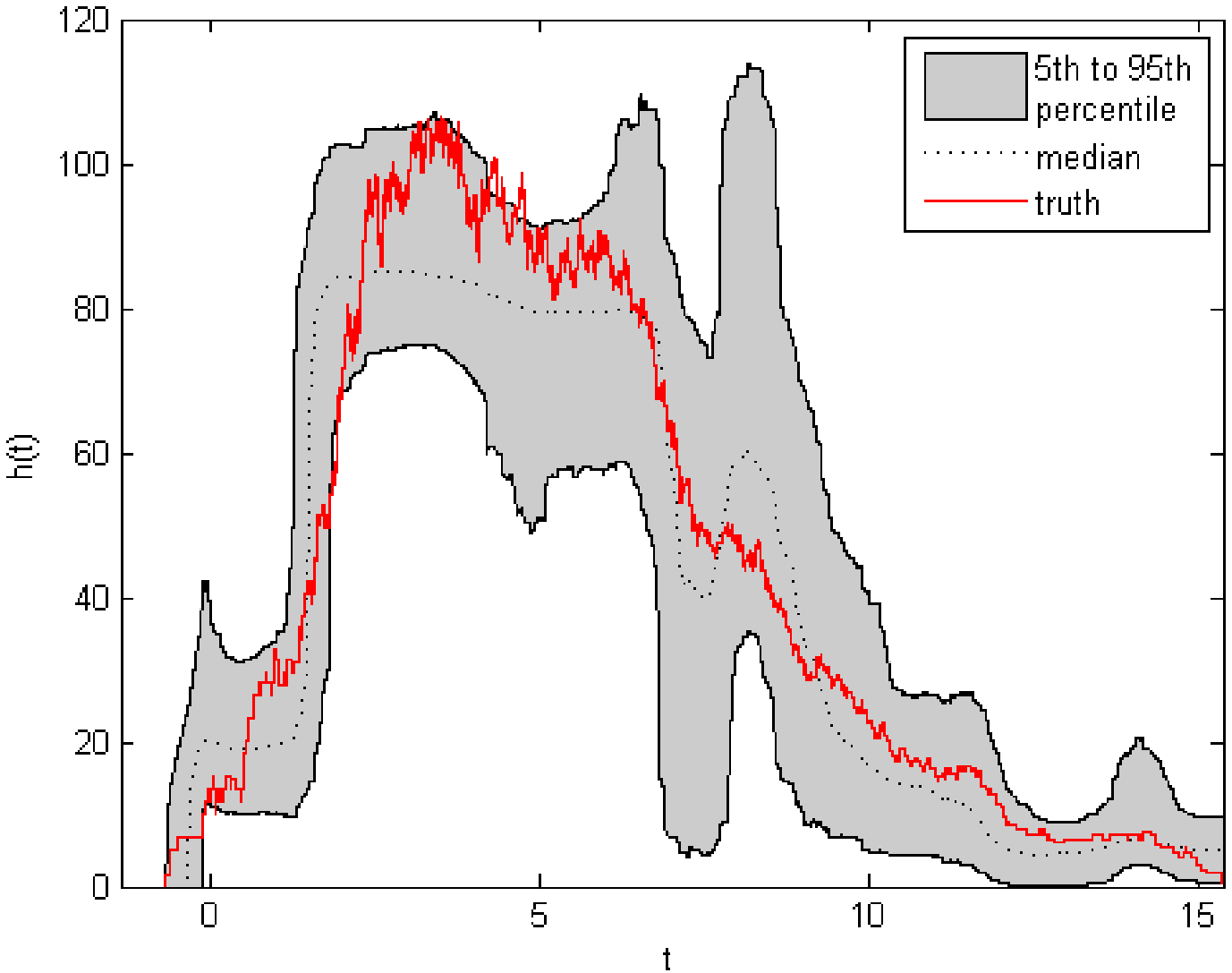
  are unknown and needed to be estimate from the observed data (i.e. removal times).
- Prior distributions:
  - $k \sim \text{Poisson}(\lambda)$
  - $s_1, \ldots, s_k$ are distributed as the even-numbered order statistics of $2k+1$ points uniformly i.i.d. on T.
  - $h_0, h_1, \ldots, h_k$ have independent $\Gamma(\kappa, \mu)$ and $\mu \sim \Gamma(a, b)$.

Sampling from $\pi(h(t)|\text{data})$

- Employ a transdimensional Markov Chain Monte Carlo (RJMCMC) (Green, 1995)
- At each iteration we make one of three types of updates:
  - `birth of a changepoint;`
  - `death of a changepoint;`
  - `within-model updates, i.e. move existing changepoints and (propose to) change heights.`
- As before, infection times are also updated within the above MCMC scheme.

Illustration

True model: mass action, $N = 1000$, 657 infectives, $R_i - I_i \sim Exp(\gamma)$



Introducing Smoothness

Alternatively to assuming that the heights are a priori independent we can assume that they a priori follow a *martingale structure*.

We assume that $h_0 \sim \text{Gamma}(\alpha_0, \beta_0)$ and that, given $h_0, \ldots, h_{i-1}$, $\lambda_i \sim \text{Gamma}(\alpha_i, \beta_i)$ where

$$\alpha_i = \alpha \quad \text{and} \quad \beta_i = \alpha / h_{i-1}$$

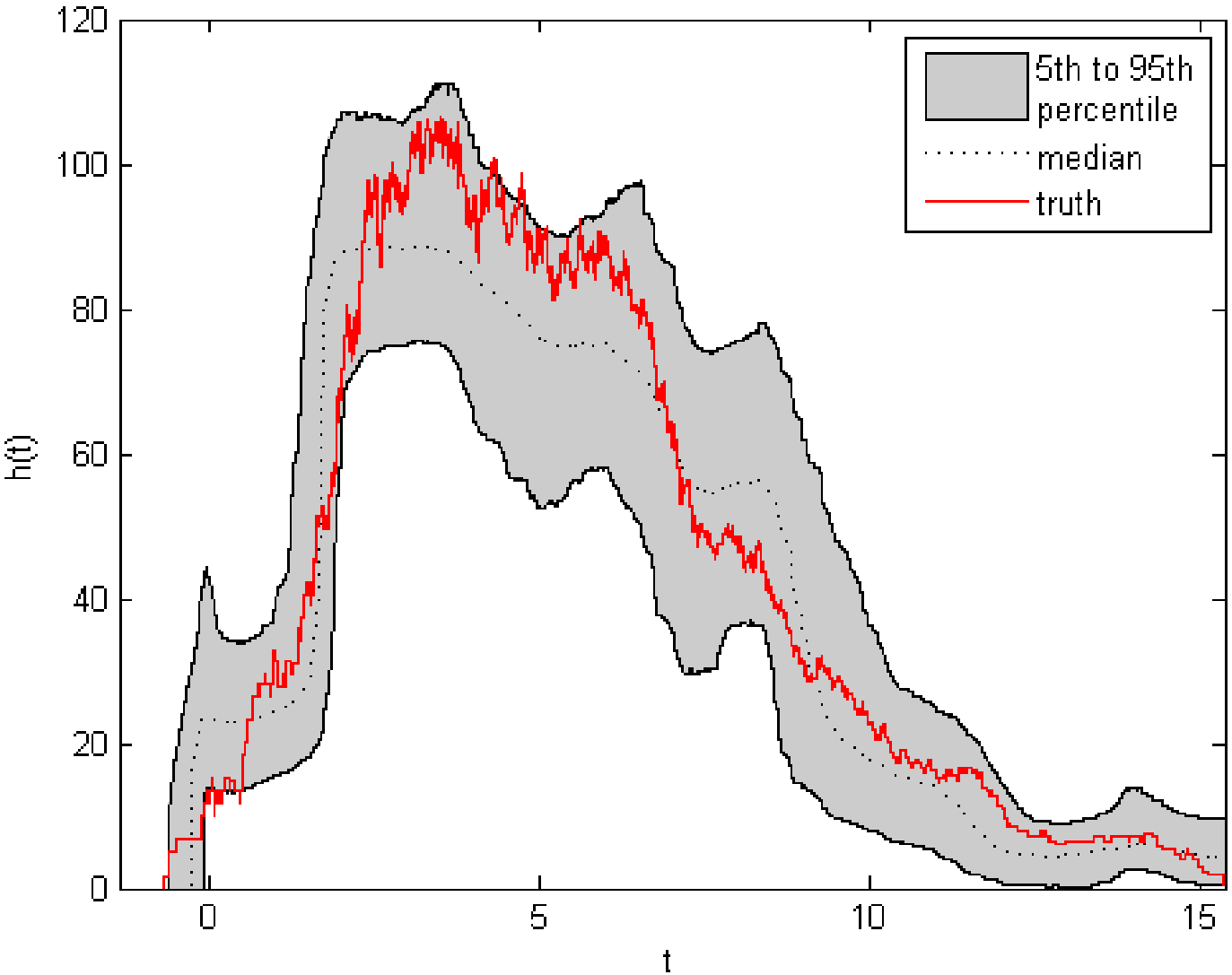
so that

$$\mathsf{E}[h_i \mid h_{i-1}] = h_{i-1}.$$

[Similar to Arjas and Gasbarra, 1994]

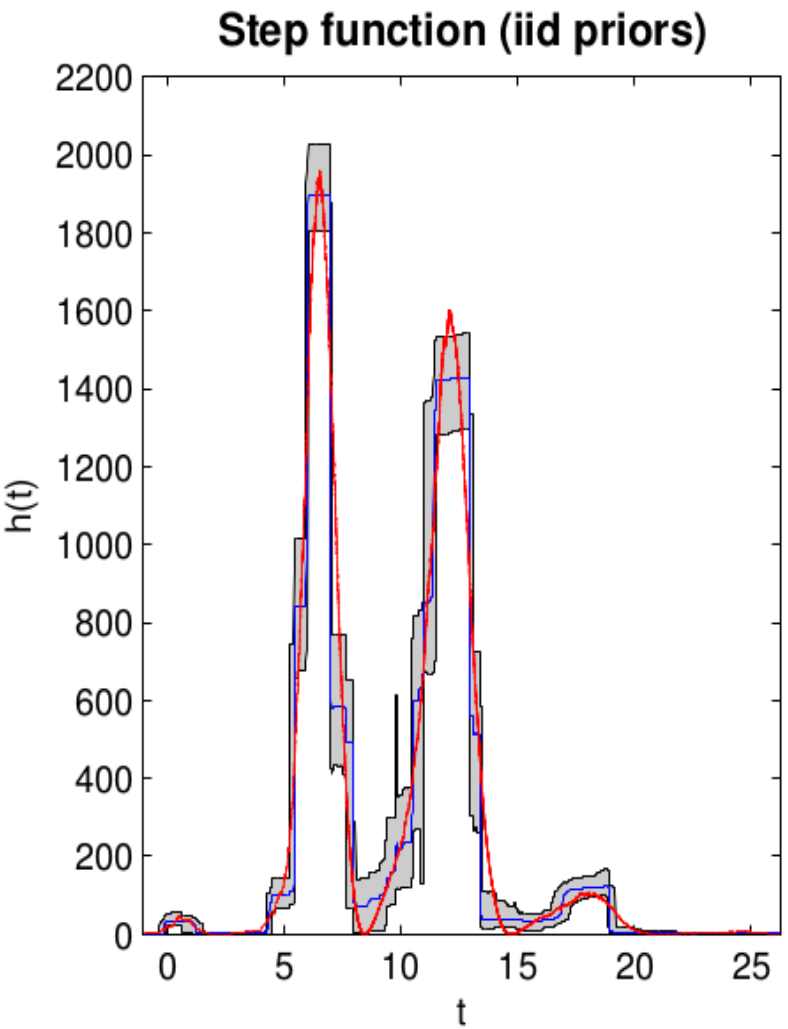
Illustration

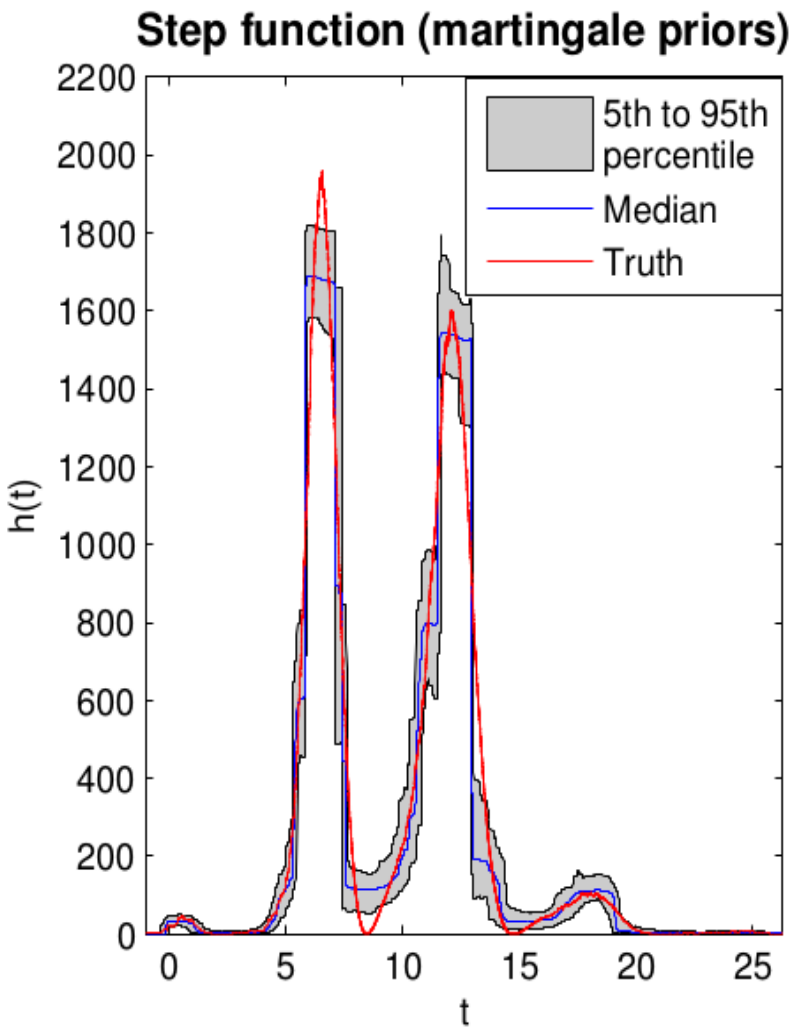
True model: mass action, $N = 1000$, 657 infectives, $R_i - I_i \sim Exp(\gamma)$



Illustration

True model: $h(t) = \beta(1 + \cos(t - I_{(1)}))S_t I_t$, $N = 10,000$, 6971 infectives





Introducing (Even) More Smoothness

An alternative to step functions we consider is a 2nd-order B-spline, which is a continuous, piecewise quadratic function.

We assume given $k + 6$ knots that $h(t)$ is a linear combination of B-spline basis functions:

$$h(t) = \sum_{i=0}^{k+2} P_{i+1} b_{i,2}(t),$$

where $b_{i,j}(t)$ is the ith B-spline basis function of order j.

These basis functions can be defined recursively by

$$b_{i,0}(t) = 1_{[t_i, t_{i+1})}(t),$$

and

$$b_{i,j}(t) = \frac{t - t_i}{t_{i+j} - t_i} b_{i,j-1}(t) + \frac{t_{i+j+1} - t}{t_{i+j+1} - t_{i+1}} b_{i+1,j-1}(t).$$

Modelling $h(t)$ via 2nd Order B-Splines

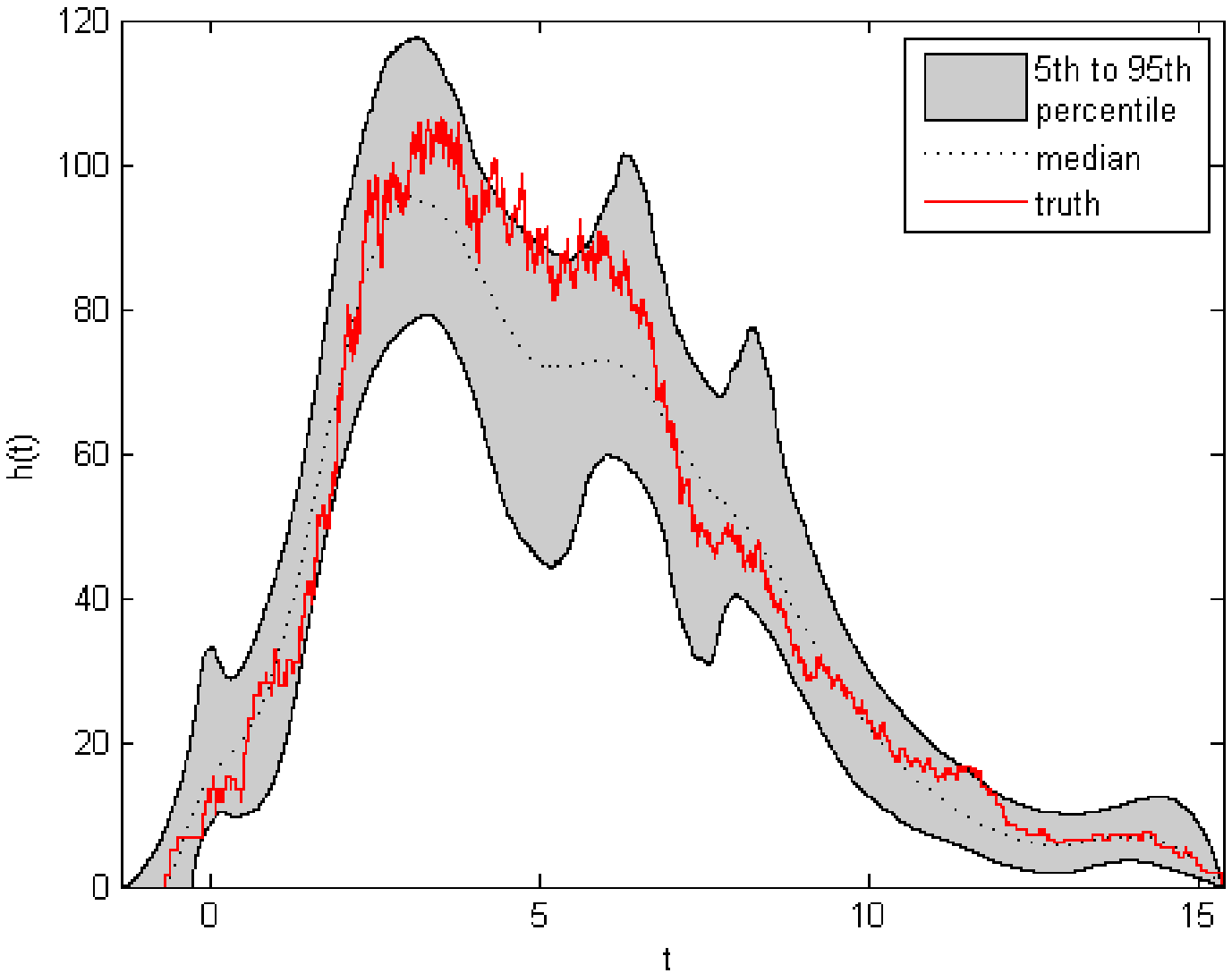
By modelling $h(t)$ via 2nd order B-splines we can then evaluate the desired integral

$$\int_0^T h(t)dt = \frac{1}{3}\sum_{j=1}^{k+3} P_j(t_{j+2} - t_{j-1}).$$

- P_j are coefficients that will be estimated;
- We assume that k has an a priori Poisson distribution with rate λ;
- the k interior knots are distributed as the even-numbered order statistics of $2k + 1$ points uniformly and independently distributed on $[0, T]$.
- Ensure that the B-spline is non-negative; assuming that the coefficients P_j are all positive makes life a little bit easier.

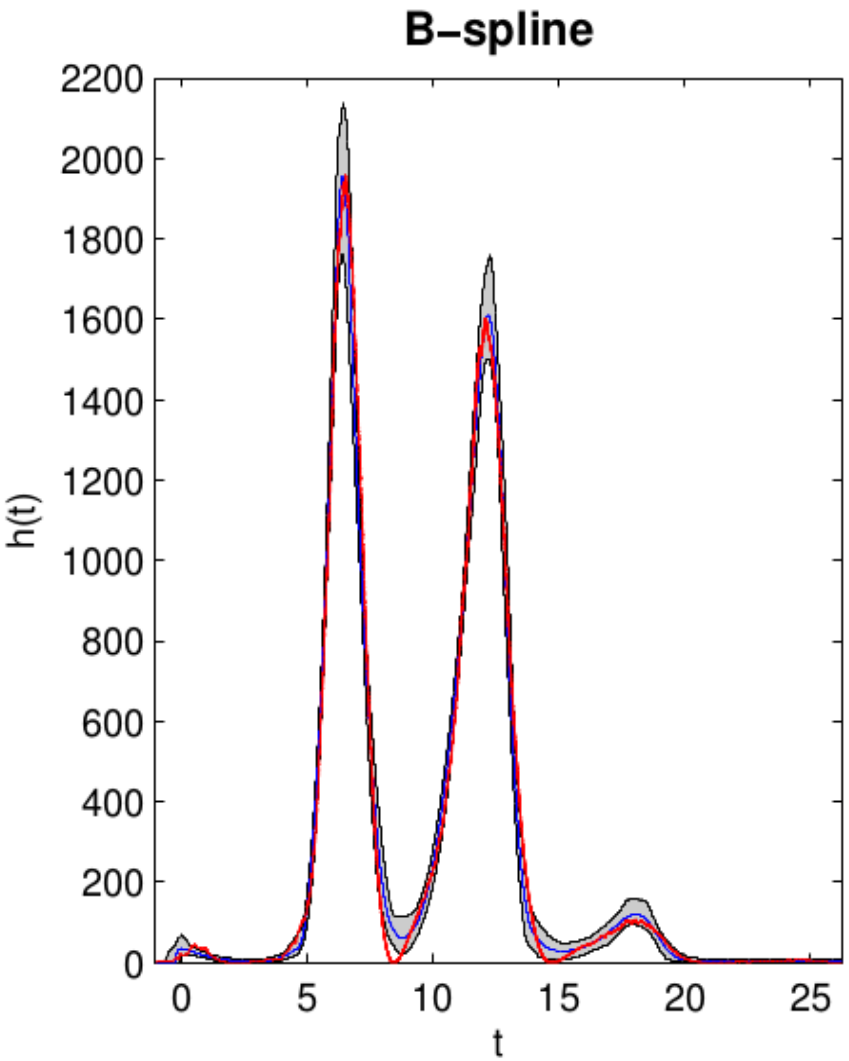
Illustration

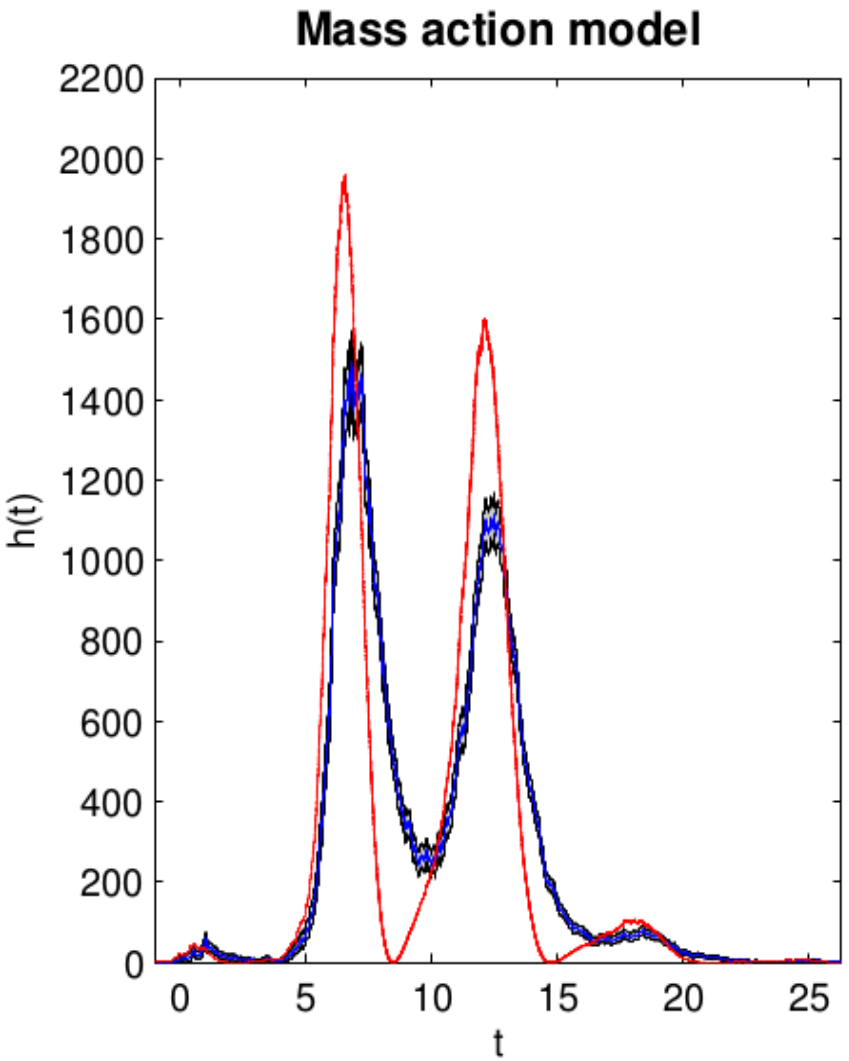
True model: mass action, $N = 1000$, 657 infectives, $R_i - I_i \sim Exp(\gamma)$



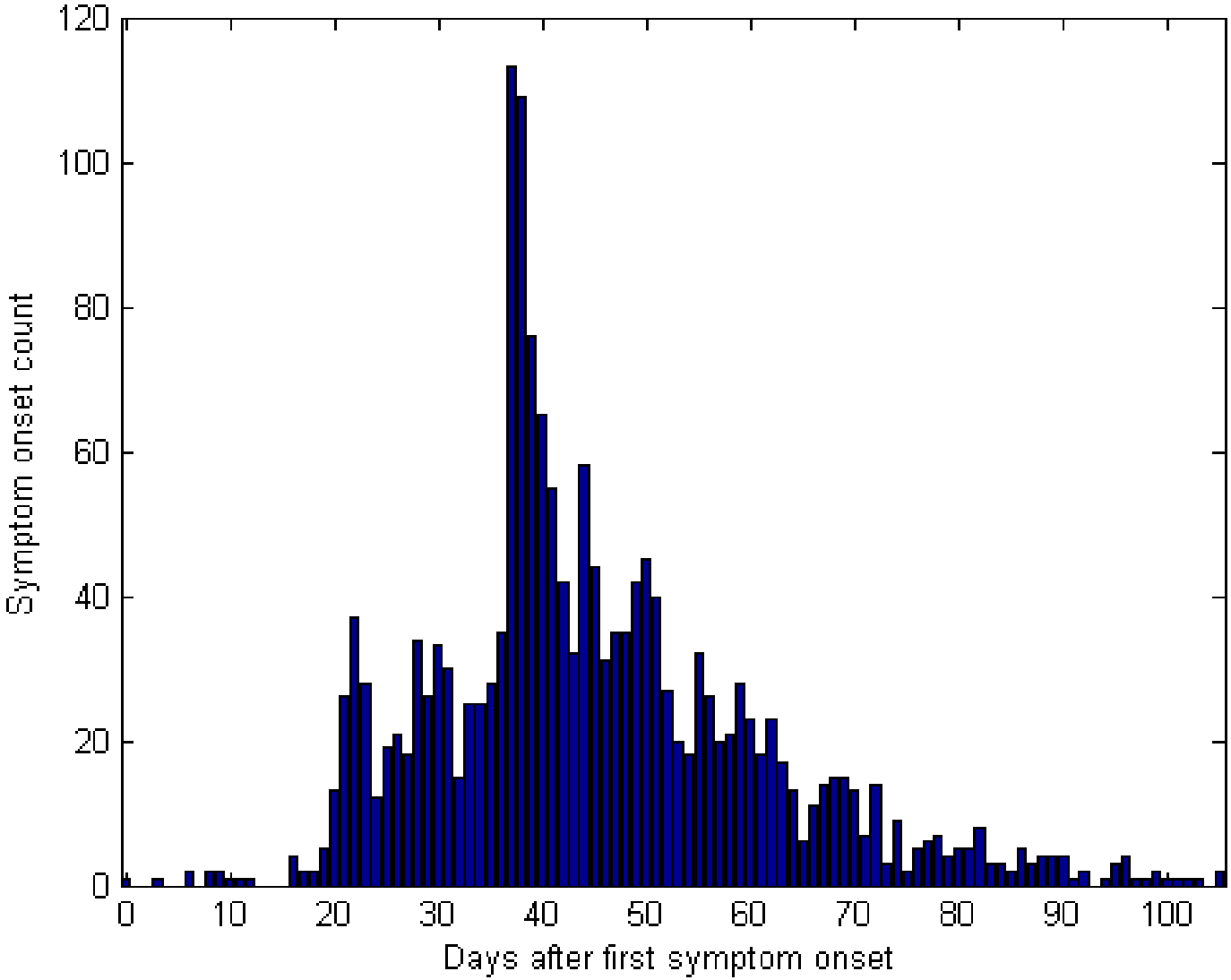
Illustration

True model: $h(t) = \beta(1 + \cos(t - I_{(1)}))S_t I_t$, $N = 10,000$, 6971 infectives



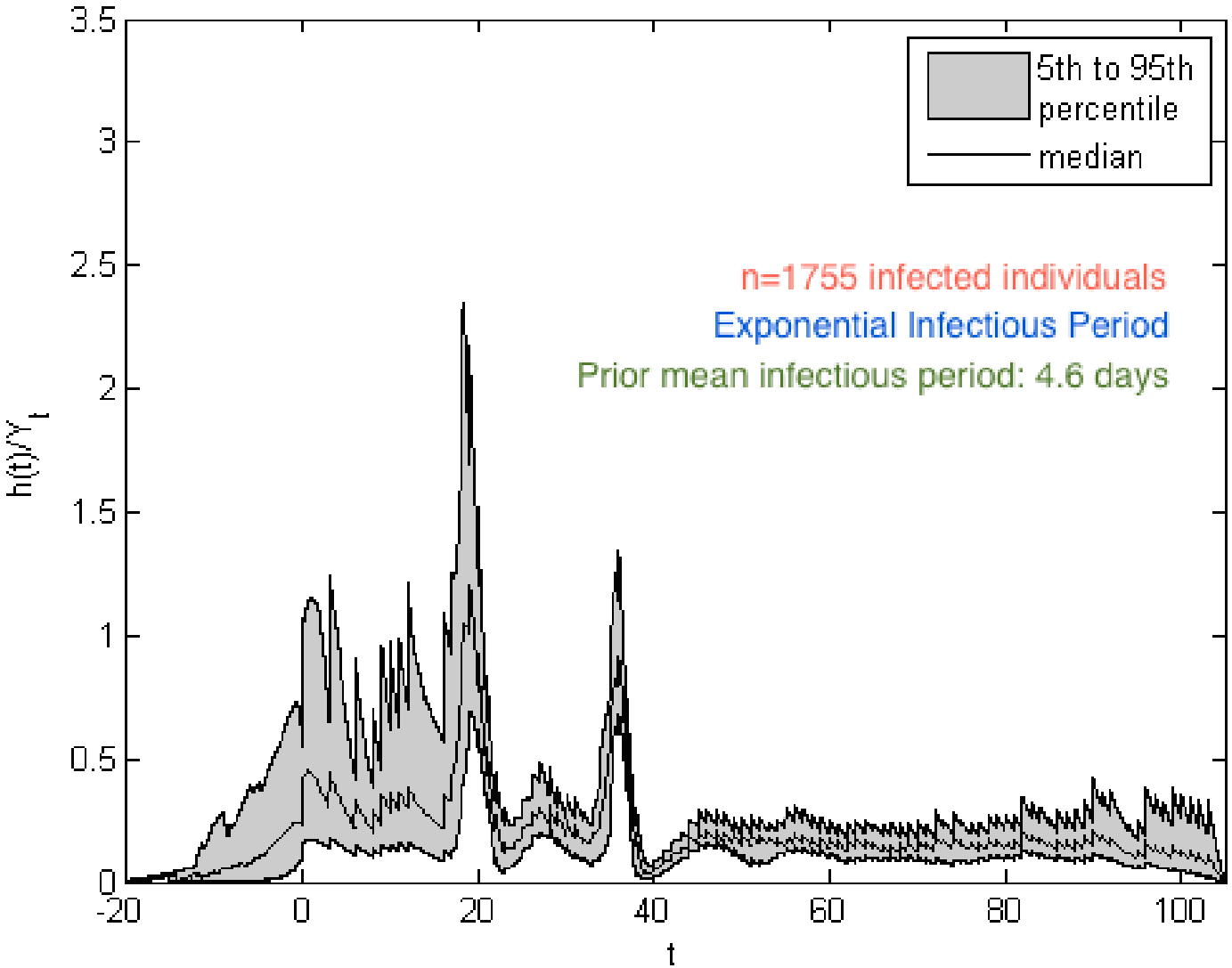


Application to 2003 Hong Kong SARS symptom onset data



Application to 2003 Hong Kong SARS symptom onset data

Bayesian non−parametric estimation of the rate per infective $(h(t)/I_t)$



Conclusions

- Non−parametric estimation offers more flexibility and we avoid making unnecessary/unrealistic assumptions.
- Assume that the infection rate $h(t)$ only depends on t and draw inference within a Bayesian framework:
  - Step function;
  - B−splines.
  - Gaussian process;
- Data−augmentation using efficient Markov Chain Monte Carlo.
- Assessing the goodness of fit as a by product? Is there really any infectious disease outbreak there?

Discussion

- Are we completely ignoring key quantities that have been used in epidemic modelling? ($\beta S_t I_t \to h(t)?$)
  - $\beta S_t I_t \to \beta\, h_1(S_t) \cdot h_2(I_t)$
  - $\beta S_t I_t \to \beta\, h(S_t I_t)$
  - $\beta S_t I_t \to \beta(t) S_t I_t$; see Xu's thesis (2014).
  - ...
- Discrete time models (e.g. Reed−Frost type of models, infectiousness); some promising results already.
- Model the population structure non−parametrically (e.g. a prior distribution over networks); see Ashley's thesis.
- Need to be careful with the choice of priors!
- Links to back-projection/calculation methodology.

Acknowledgements

- **The team**
  - GPs: Xiaoguang (Allen) Xu @ University of Nottingham
  - Splines & Step Functions: Edward Knock @ University of Nottingham
  - Phil O'Neill @ University of Nottingham
- **Funding**

The University of Nottingham

THE ROYAL SOCIETY

EPSRC
Engineering and Physical Sciences Research Council