# How well can we predict future epidemics from small scale outbreaks?

Kokouvi Gamado
Glenn Marion Thibaud Porphyre

Wednesday, 13/09/2017


SRUC BioSS The James Hutton Institute Moredun The University of Edinburgh ROSLIN

# Outline

- Real-world motivation
- Spatio-temporal Model
- Bayesian inference
- Model choice
- Conclusions

# Real-world Motivation

- ❑ Emerging and re-emerging pathogens
- ❑ Analysis of historical outbreaks inform control of future outbreaks
- ❑ Data: detection times known
- ❑ Data: At risk premises known
- ❑ Limited/negligible animal movement
- ❑ Infection local: intensity a function of distance
- ❑ Control: depopulation of detected premises

- ❑ No knowledge of times of infection
- ❑ Example of Classical Swine Fever (CSF) in UK

# Homogeneous farms model

- SIR model on closed spatial population, size N
- An individual $i$ makes an infectious contact with a susceptible individual $j$ at rate $\beta_{ij} = \beta_0 h_{ij}$
- $\beta_{ij}$ : the kernel transmission function
- $h_{ij}$ : the distance kernel
- The infectious period follows: $R_i - I_i \approx Ga(\alpha, \gamma)$

- Data augmentation techniques employed for the unknown infection times. Inference more challenging with small scale epidemics coming from emerging and re-emerging pathogens

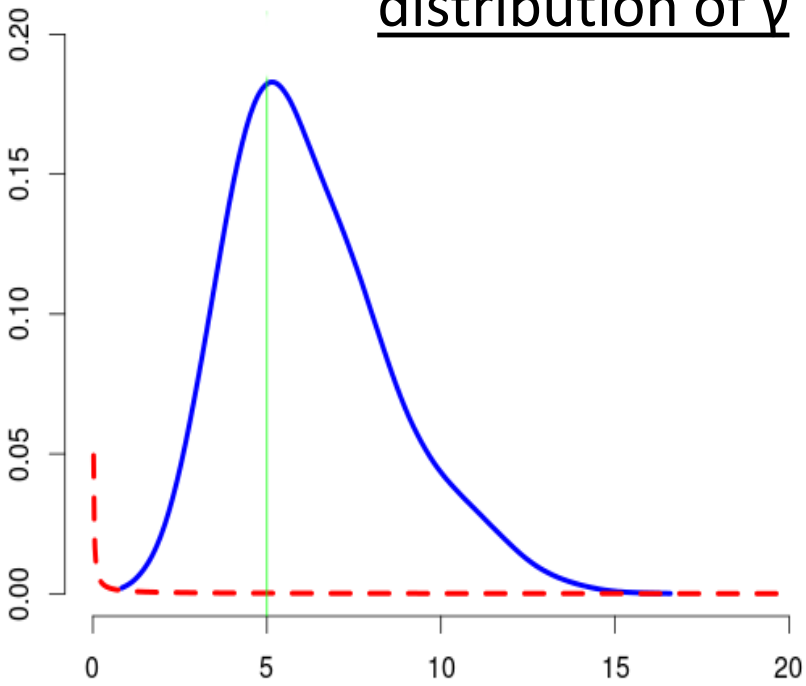
# Bayesian Inference with MCMC framework

- Enables statistical inference of
  - parameter values
  - infection times

- MCMC methods draw successive samples from 'posterior' distribution

- In practice, interests lie in infectious period and kernel transmission functions

Inferred (posterior) distribution of γ



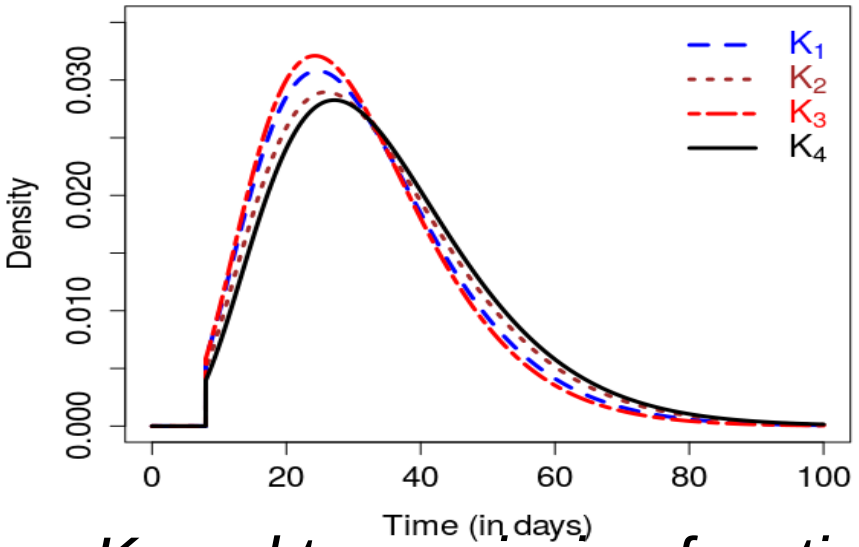

Based on simulated data with true values: $\gamma = 5$, $\alpha = 5$, $\beta_0 = 0.35$, #premises = 201, #infected sites n = 43

# CSF in East Anglia Northfolk

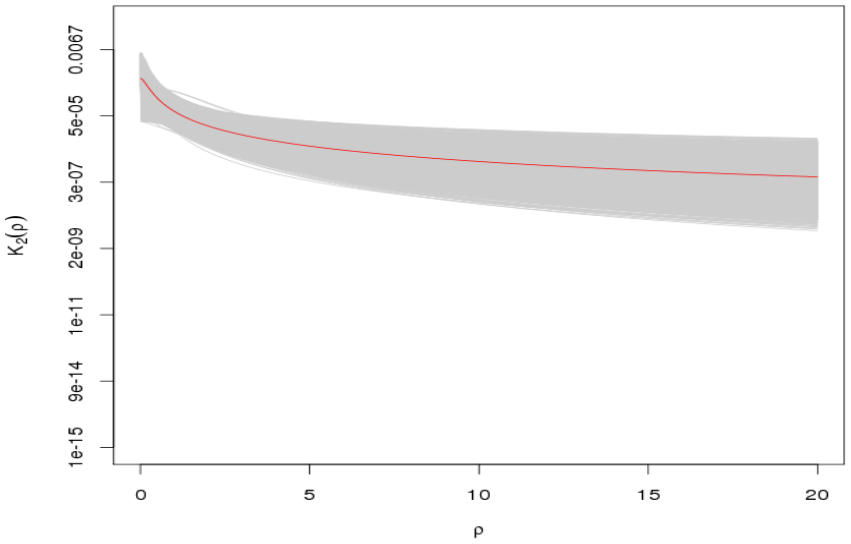
Can apply to real world data even on small outbreaks

- N = 1703 farms with exact location or coordinates
- 16 reported cases
- Can fit multiple models e.g. different kernels
- And infer disease characteristics =>

*Average infectious period*

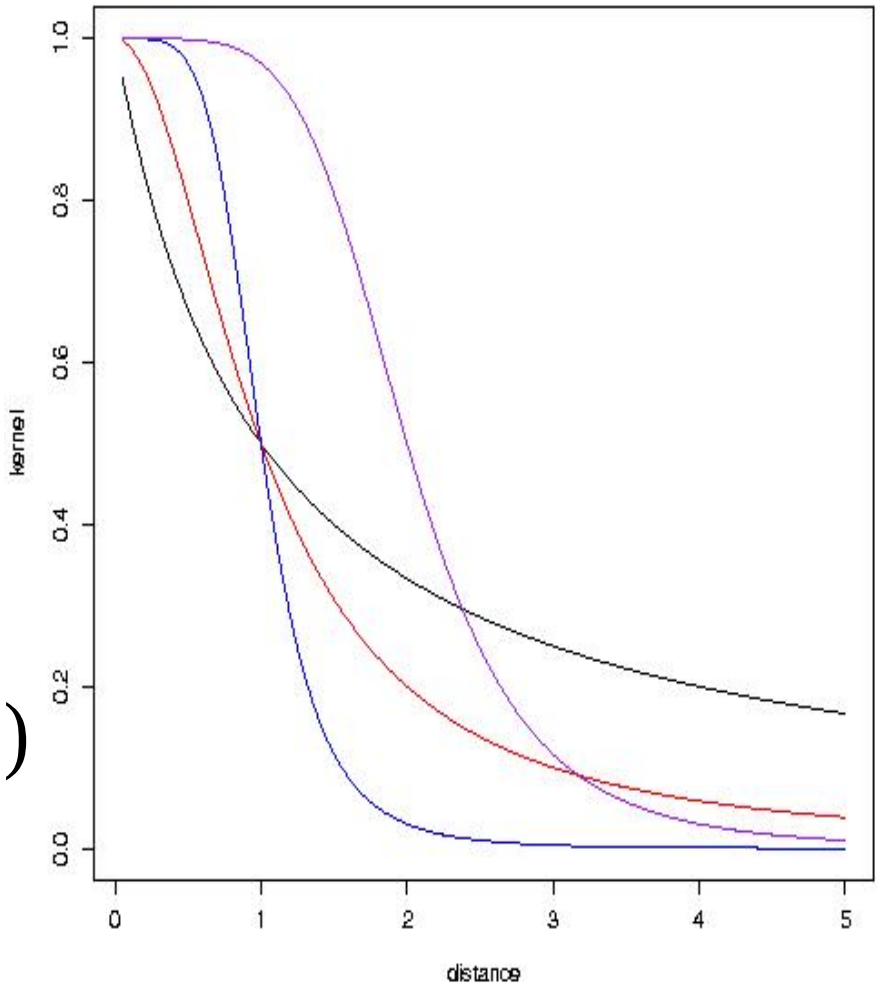



*Kernel transmission function*

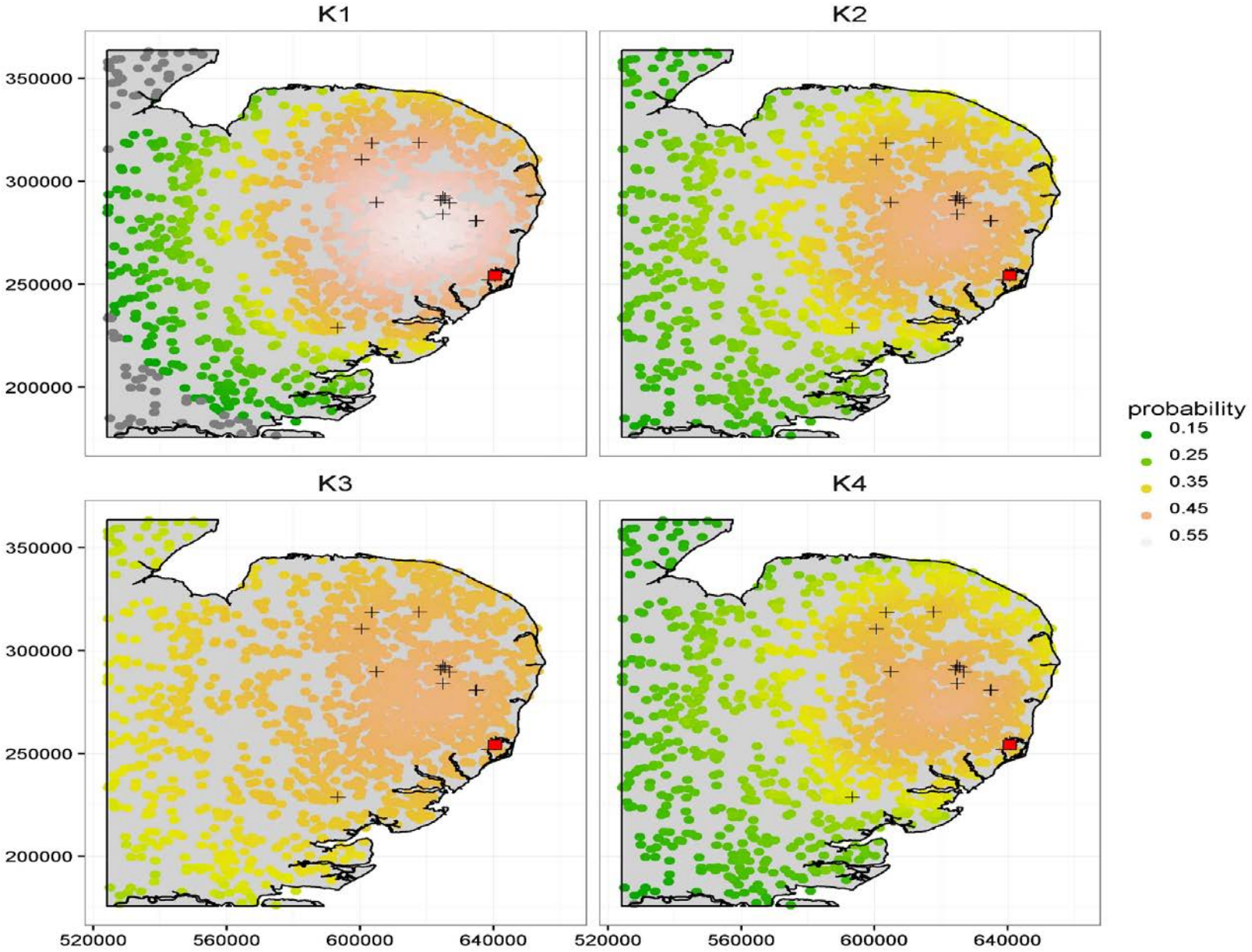


# Fit different spatial kernel transmission functions

- K1: $\exp(-\tau\rho(i,j))$
- K2: $\dfrac{1}{1+(\rho(i,j)/d)^{a}}$
- K3: $\dfrac{1}{1+(\rho(i,j)/d)}$
- K4: $1-\exp(-(\rho(i,j)/d)^{-a})$



# CSF risk assessment: model dependent



# Bayesian model choice tools

- DIC is a model comparison criterion measuring trade-off between the fit of the data and the complexity of the model: DIC = "goodness of fit" + "complexity"

Problems: non invariance to reparameterisation, lack of consistency, weak theoretical justification

- Bayesian Latent residuals: unobserved, iid uniform r.v. that conform with the assumed data generation process

Test various components of models and "true" assumptions present uniformly distributed residuals

# Discussion

- Inferred kernel transmission functions for CSF
- Applied novel model selection tools to assess and select kernels

- Assessing the amount of data needed for correct kernel identification in real-time epidemic outbreaks
- Account for socio-economic effects in modelling epidemics
- Make the models more dynamic by accounting for control strategies at different points in time

## Reference

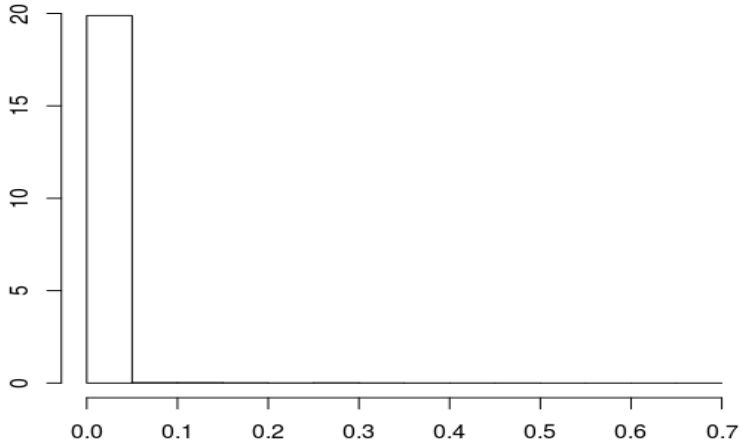
K Gamado, G Marion & T Porphyre. Data-Driven Risk Assessment from Small Scale Epidemics: Estimation and Model Choice for Spatio-Temporal Data with Application to a Classical Swine Fever Outbreak. Frontiers in Veterinary Science (2017) 4:16

Thanks!

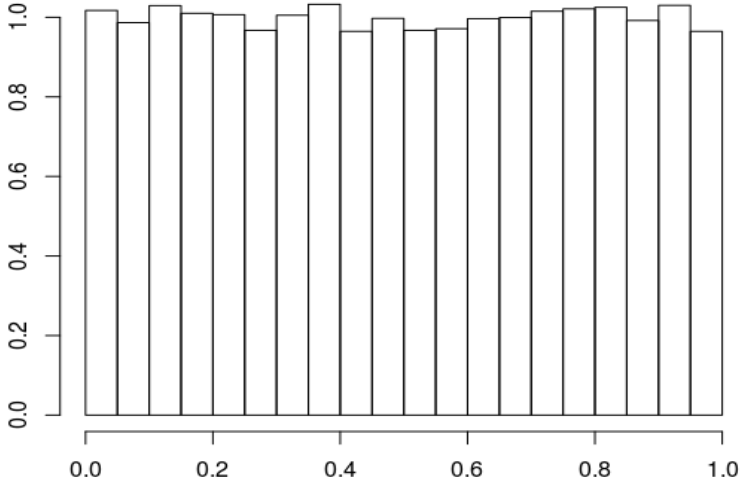
# Summary of work on heterogeneity

- Groups of detection
- Hierarchy or random effects on removal rates
- Groups of susceptibility
- Hierarchy on susceptibility

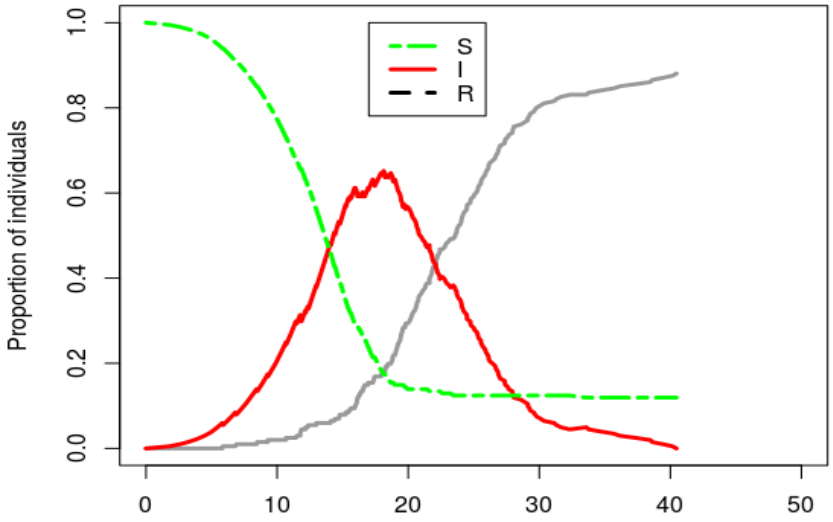
*Homogeneous assumption*



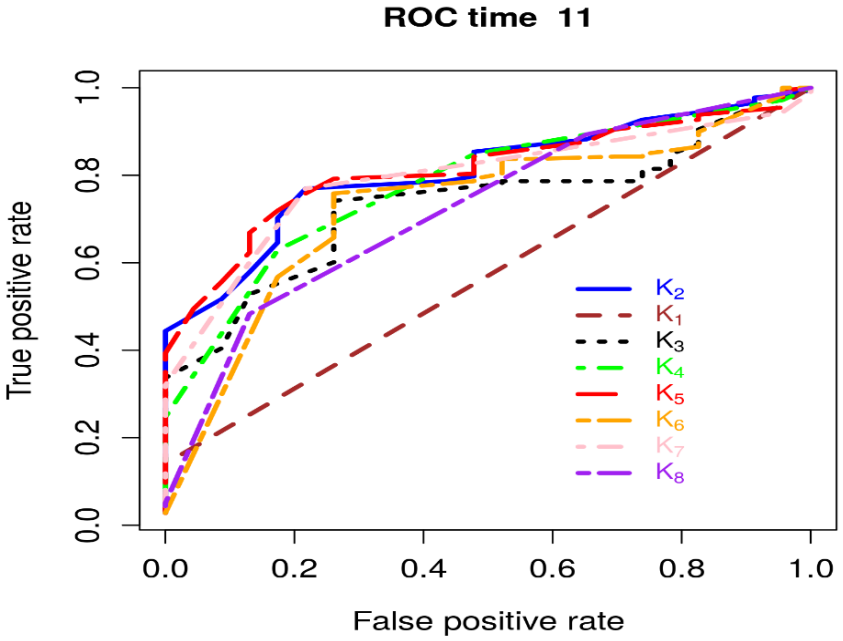

*Heterogeneous assumption*

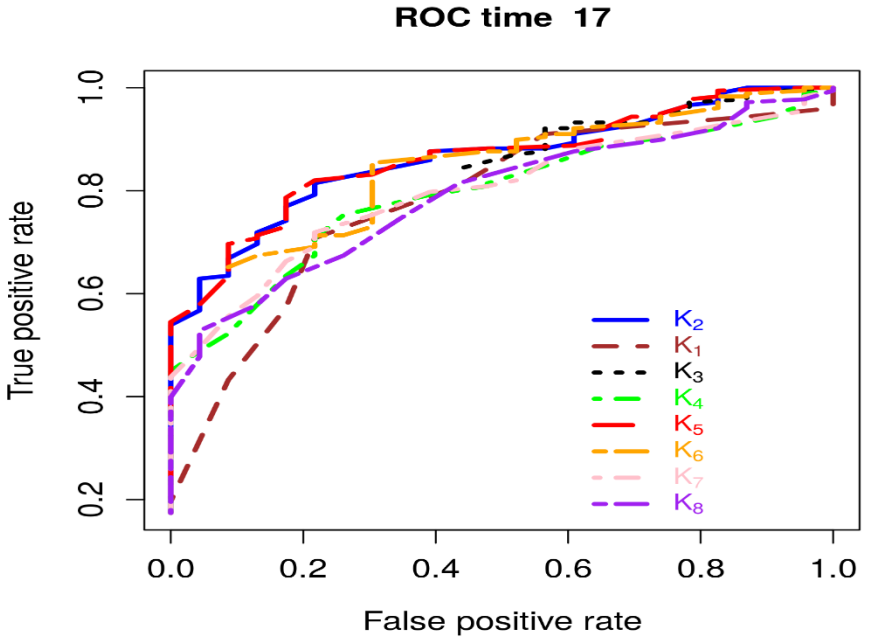


# Ongoing epidemics example




Observe the epidemic daily
Estimate parameters & LV
prediction of the future
ROC curve

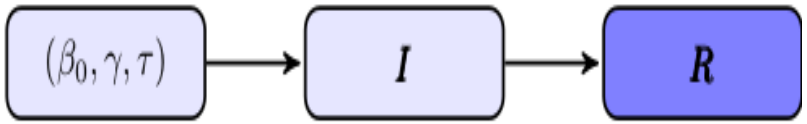


# Computationally efficient inference

## Centered and non-centered schemes

- Standard algorithms adopt a ‘centred’ approach.

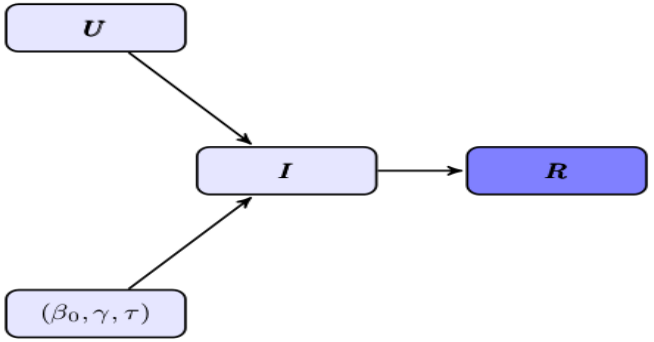



- updating the model parameters
- then updating the latent or unobserved variables

- Non-centering scheme

$$U_i = \gamma(R_i - I_i)$$

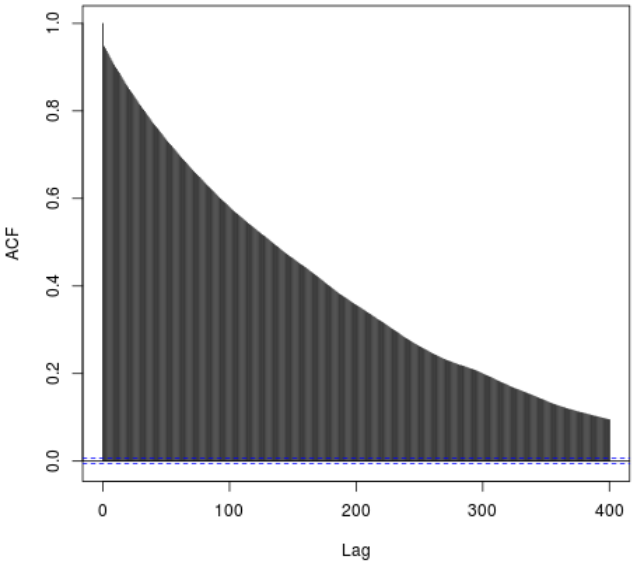
$$(I, \beta_0, \gamma, \alpha, R) \rightarrow (U, \beta_0, \gamma, \alpha, R)$$



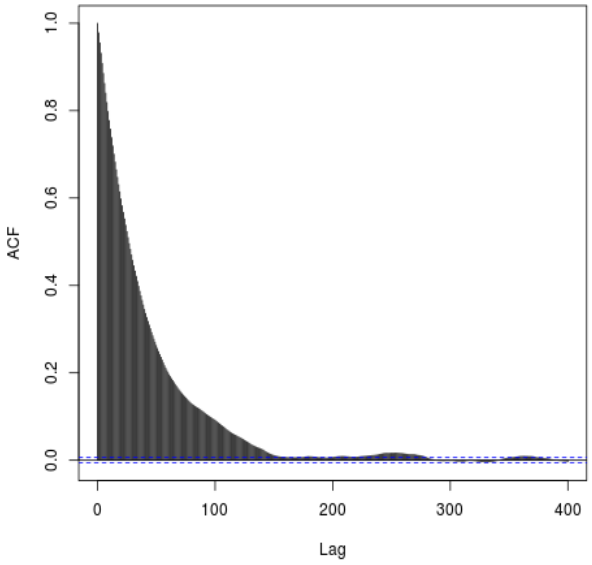

- Update simultaneously parameters and latent variables

# Non-centred and hybrid schemes are valuable improvement over standard scheme

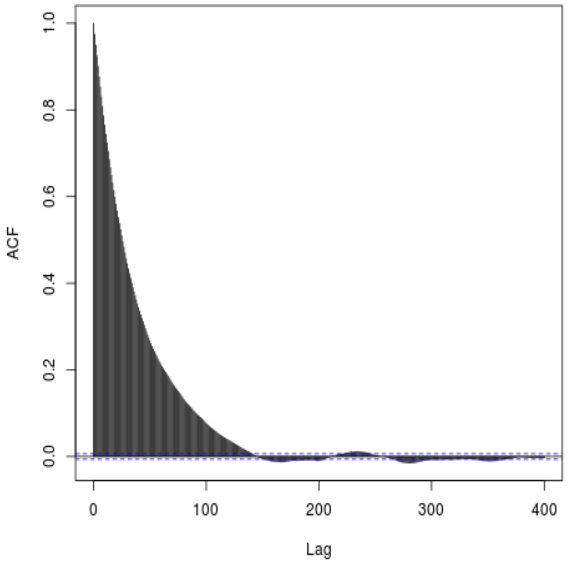
Standard scheme




Non-centred scheme




Hybrid scheme




**Non centred parameterisation enables**

- correlated moves in Markov chain
- more efficient exploration of posterior
- faster MCMC for complex models
- Non-centered & hybrid schemes reduce autocorrelation

# Variability between farms

- Most models assume homogeneity of farms
- In reality, each farm is likely to behave differently depending on farm type
- Bio-security practices, areas of location, reaction to epidemic...
- Generally one aim of our research is to move away from the assumptions of constant susceptibility, infectivity and detection, all to be functions of farms' characteristics

# Examples of heterogeneity

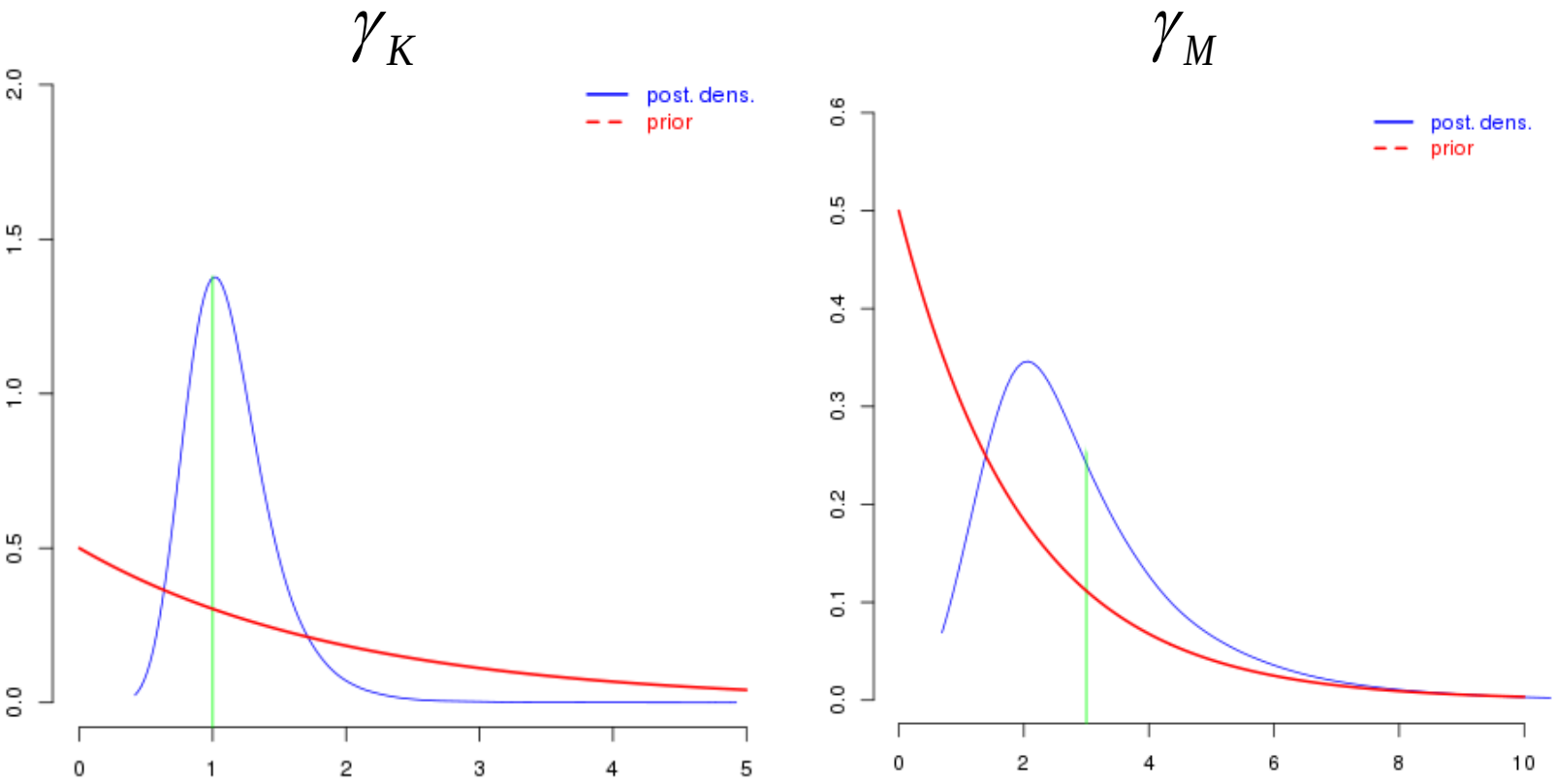
- Different detection rates per group

Assume there are n groups of farms: G1 to Gn
The infectious period follows: $R_i - I_i \approx Ga(\alpha, \gamma_s)$

$s \in$ Gs

- Different groups of susceptibility
  Kernel transmission function: $\beta_0^s h_{ij}$

# Inferring the rates per group




- The data consist of the detection times of the infected premises and the groups of each farm
- LR helps to select models with heterogeneity when it is the "truth"