InFER: Inference For Epidemic related Risk

Chris Jewell

Dept of Statistics, University of Warwick
chris.jewell@warwick.ac.uk

InFER2011

28th March 2011

What is InFER?

- £600k grant from BBSRC

- Application-driven development of statistical methodology for epidemics
  - Computationally intensive Bayesian methods (eg MCMC, SMC)
  - Models for datasets with varying degrees of missingness and resolution
  - Development of developer- and user-level software

- Collaboration between Statistics and Life Sciences @ Warwick: Gareth Roberts, Laura Green, Matt Keeling, Chris Jewell, Judith Brown

Overview

Outline

Motivation

Welfare and Economics

- Foot and Mouth Disease
  - 2001: £8 billion, 6.5 million slaughtered
  - 2007: £100 million, 2610 slaughtered
- Avian Influenza
  - worth >3.5 billion
  - 40% UK primary meat market (2004)
- bTB?
- Endemic diseases?
- Human diseases - Influenza H1N1?

Epidemic Control

Framework Response Plan for Exotic Animal Diseases:

1. Minimise the number of animals which need to be culled either to control the disease or on welfare grounds, and which keep animal welfare problems to a minimum.
2. Protect public health.
3. Cause the least possible disruption to the food, farming and tourism industries, to visitors to the countryside, and to rural communities in the wider economy.
4. Minimise damage to the environment.
5. Minimise the burden on taxpayers and the public.

Defra 2007

Aims

- Model-based analysis aims to provide a prediction of the RISK posed by an epidemic in real-time
  - Who is likely to be infected next?
  - Who presents the greatest risk to the population if they get infected?
  - How many occult (undetected) infections are there?
  - Incorporate parameter uncertainty into any predictions
- What is the relative importance of each parameter in propagating the epidemic?
- Incorporate PROBABILITY to better inform control policies

A job for the statistician...?

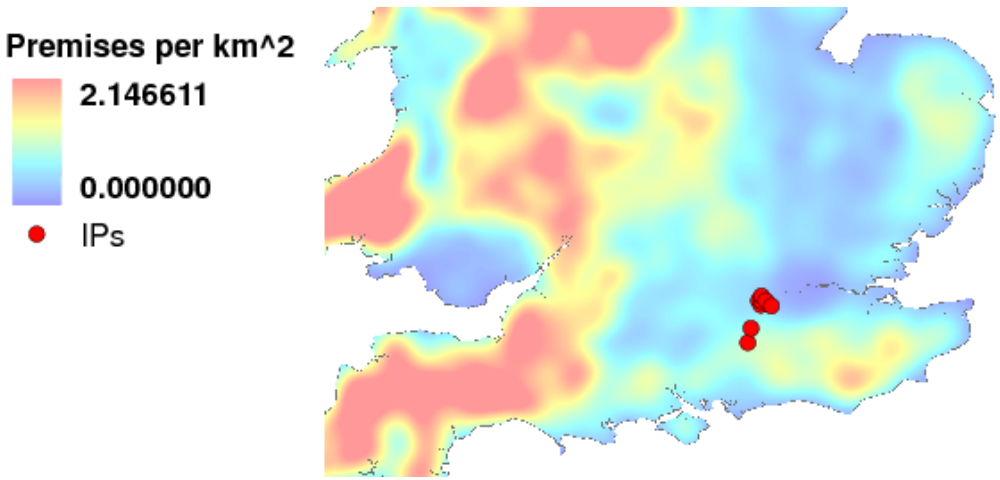
- Results of forward simulation depend on model parameters
- A statistical approach provides formal estimates for parameters given the model
- Historically:
  - Estimation of R_0 is relatively easy
    - Relevance to heterogeneous populations?
  - Difficult to estimate infection and removal rates together
  - Due to missing data

References: Bailey (1975); Becker (1989)

Outline

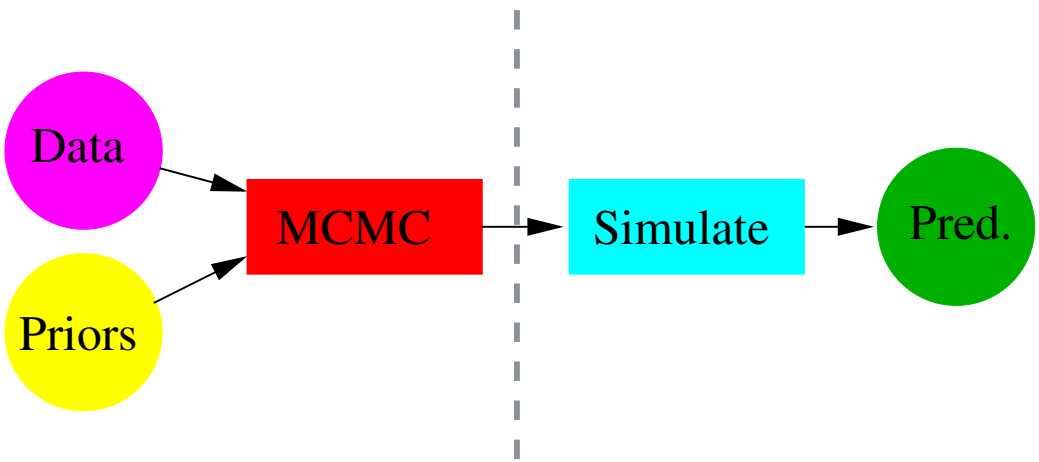
Available data

- Covariate data
  - Location, number and type of animals, contact networks
- Epidemiological data
  - Detection times
  - Cull times



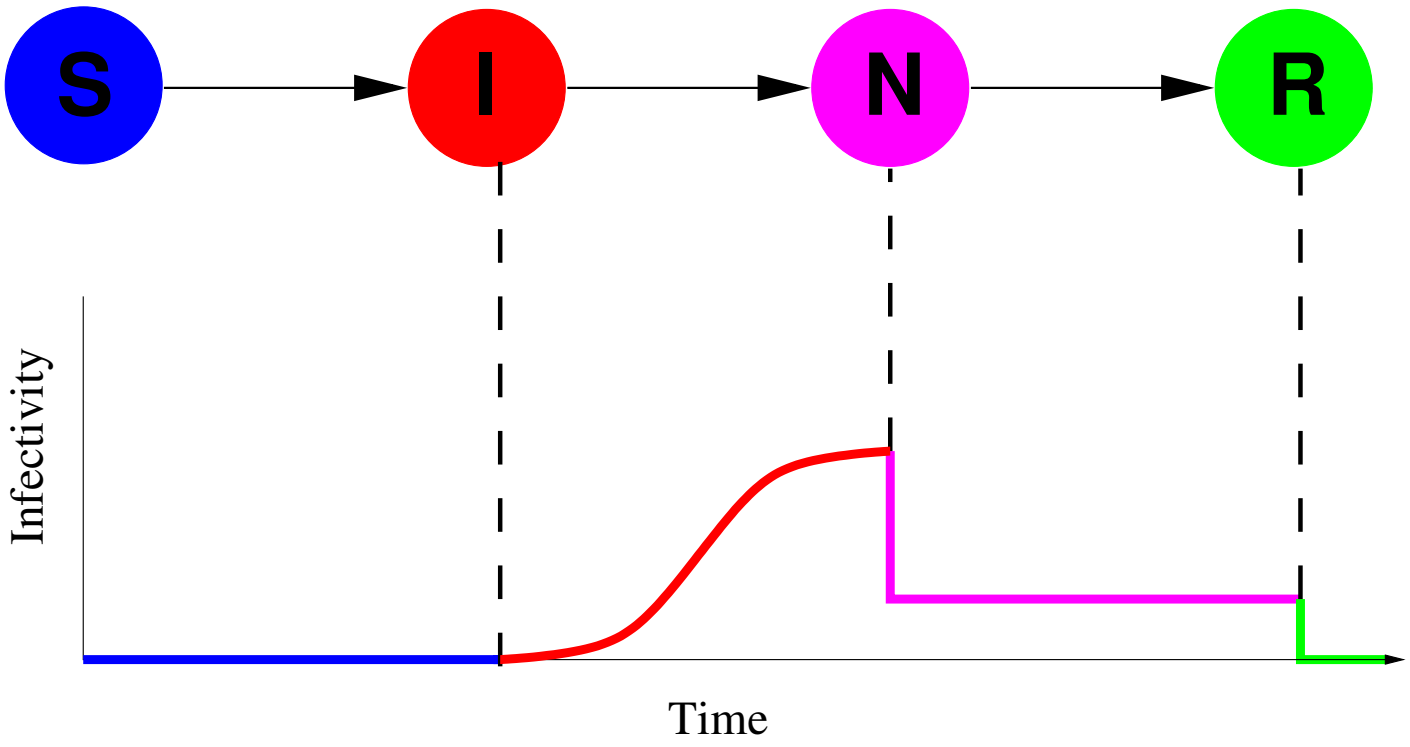
The Approach

1. Define a model for disease transmission in the *population*
2. Take prior opinion and field data - make inference on transmission and removal parameters
3. Use the results with forward simulation to make fully quantitative predictions



The Model

The basic assumption



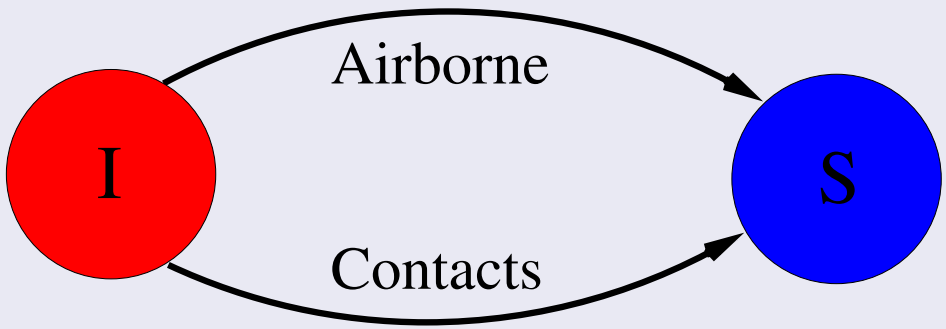
- Continuous time stochastic mechanistic model
- Individual = farm

The Model

Population structure

- Historical modelling treats all infectious contacts as equally likely: homogeneous mixing and often deterministic
  - eg. Anderson and May 1992
- In practice, populations are heterogeneously mixing and highly stochastic
  - Contact networks
  - Spatial proximity

Transmission



Outline

The Problem

- Missing data!
  - Infection times are not directly observed
  - What about occult infections?
- We can write down a statistical likelihood function for the model conditional on the infection times
- We cannot explicitly write a likelihood function if we do not observe infection times
  - Require an expectation over all possible infection times and occult status

A solution....

- Construct a likelihood describing the continuous-time stochastic epidemic

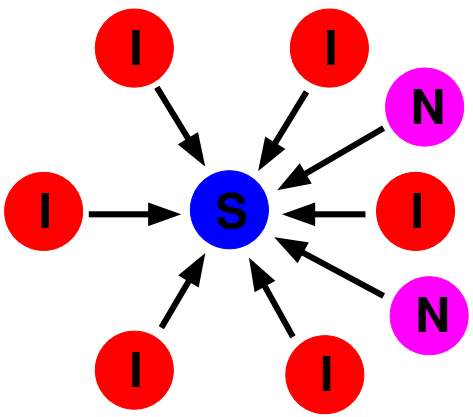
- Bayesian approach allows:
  1. Natural framework to include unobserved data by data augmentation MCMC methodology
     - Unobserved infection times
     - Occult infections
  2. Coherent inclusion of Prior information
     - Expert opinion
     - Previous disease outbreaks

- MCMC allows us to work in high dimensions

$$d = \dim(\boldsymbol{\beta}) + [\mathbf{I}] + [occults]$$

Infectious pressure – continuous time

- At any time t, susceptible j has infectious pressure exerted on it by
  - all infected farms i
  - "Background" - eg wildlife



In a small time interval Δt:

$$P(j \text{ infected}) \approx T_j \cdot \Delta t$$

$$T_j = \beta_0 + \sum_{i \in \{I_i < I_j < N_i\}} \beta_{ij}(t) + \sum_{i \in \{N_i < I_j < R_i\}} \beta^{\star}_{ij}$$

The Model
Infection times

- Infection times are not directly observed
- Require data imputation through data augmentation
- Notification time is observed
  - Assume a distribution for Infection to Notification time

I $\rightarrow$ N time

$$F_D(d) = \exp\left[-a \cdot \exp\left(b \cdot d\right) - 1\right] \qquad d \geq 0$$

The model
The Likelihood

- Likelihood uses a continuous time model:

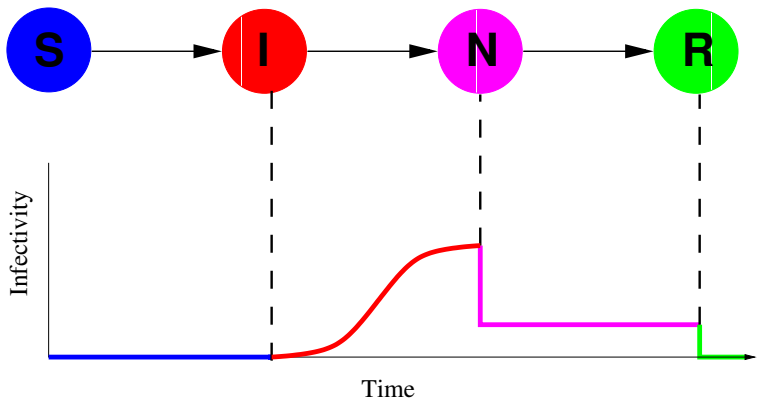
$$f(\mathbf{I}, \mathbf{N}, \mathbf{R}|\beta, a, b) = \prod_{l=1, l\neq k}^{[\mathbf{I}]} (T_l) \cdot exp\left[- \int_{I_k}^{T_{obs}} \left(\sum_{j=1}^{[\mathbf{S}]} T_j \right)_{t^-} dt \right] \times \prod_{j=1, j\neq k}^{[\mathbf{I}]} f_D(N_j - I_j)$$

- Order $[\mathbf{I}]^2 + [\mathbf{I}][\mathbf{S}]$ for update any β
- Order $[\mathbf{I}] + [\mathbf{S}]$ for updating an infection time

Statistical Inference

Inference on:

- d-dimensional vector of transmission parameters, β
- Unobserved data:
  - Infection times are never directly observed
  - Occult infections are individuals who are infected, but not yet detected (notified)



Prior distributions

- Gamma for rates ($\beta \geq 0$)
- Beta for probabilities ($0 \leq p \leq 1$)
- Uniform for occult status (0 or 1)
- Wide range of choices for infection times (or infectious period)

- Priors chosen to agree with expert opinion and previous knowledge of epidemics

MCMC scheme

Repeat the following steps

1. Model parameters β
   1. Propose $\tilde{\beta}$ en-bloc using MVN$(\beta^{(q)}, \Sigma)$
   2. Calculate likelihood
   3. Calculate acceptance probability, and accept or reject $\tilde{\beta}$ accordingly.
2. Infection times - repeat s times:
   1. Propose a move, add, or delete
   2. Update likelihood
   3. Calculate acceptance probability, and accept or reject $I_i^{(q+1)}$ accordingly

Parallel regions in red

Outline

Foot and Mouth 2007

- 2007 outbreak in Surrey comprising 9 'known' infected premises
- 21 farms slaughtered as Dangerous Contacts
  - Identified by Contact Tracing
- Can we identify undetected infected premises statistically?

Foot and Mouth 2007

Data and priors

- Covariate Data
  - Location - OS Grid coords
  - Number of cattle and sheep on farm (treat pigs as sheep)
- Priors
  - Based on posteriors of Kypraios 2007

- Epidemiological Data

Premises	N	R
IP1	0	2
IP2	4	5
IP3a	40	42
IP3b	40	42
IP4	41	41.5
IP5	45	45.5
IP6	49	50
IP7	52	52.5
IP8	58	58.5

Foot and Mouth 2007
Transmission model

- Use the model of Keeling et al (2001)
  - Spatial location (x,y)
  - Number of cattle
  - Number of sheep

Transmission model

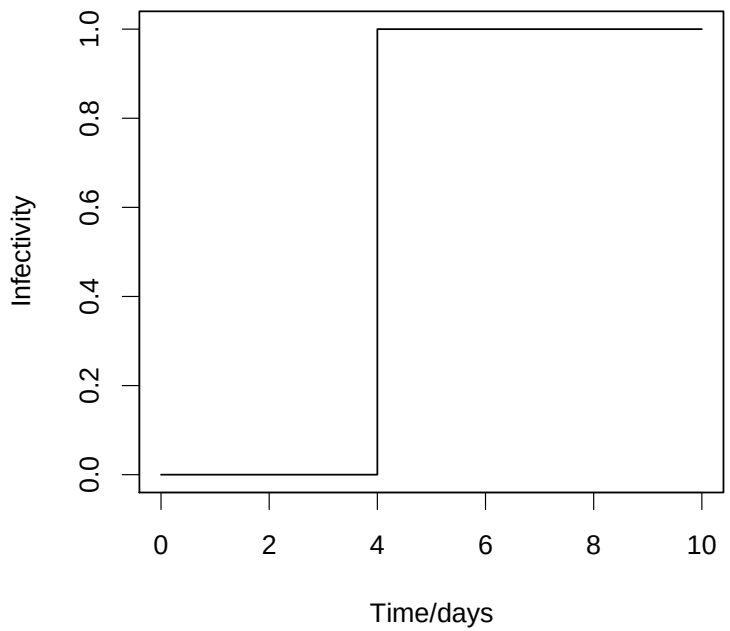
$$\beta_{ij} = \left(\beta_1 n_i^{c\psi} + n_i^{s\psi}\right)\left(\beta_2 n_j^{c\psi} + n_j^{s\psi}\right) \cdot \beta_3 \cdot \frac{\delta^2}{\rho_{ij}^2 + \delta^2} \quad i \in \mathbf{I}, j \in \mathbf{S}$$

$$\beta_{ij}^{\star} = \left(\beta_1 n_i^{c\psi} + n_i^{s\psi}\right)\left(\beta_2 n_j^{c\psi} + n_j^{s\psi}\right) \cdot \beta_4 \cdot \frac{\delta^2}{\rho_{ij}^2 + \delta^2} \quad i \in \mathbf{N}, j \in \mathbf{S}$$

Foot and Mouth 2007

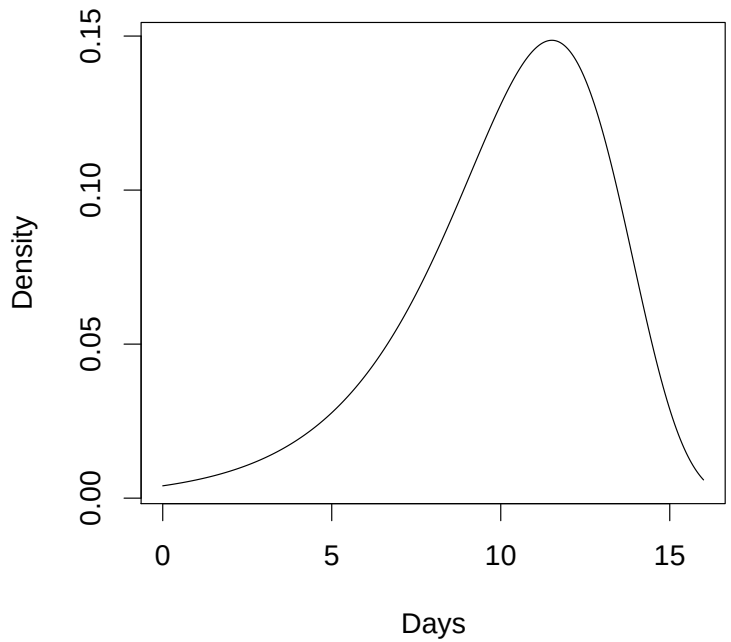
Infectivity and Infectious period functions

Infectivity function



(fixed - equiv to SEINR model)

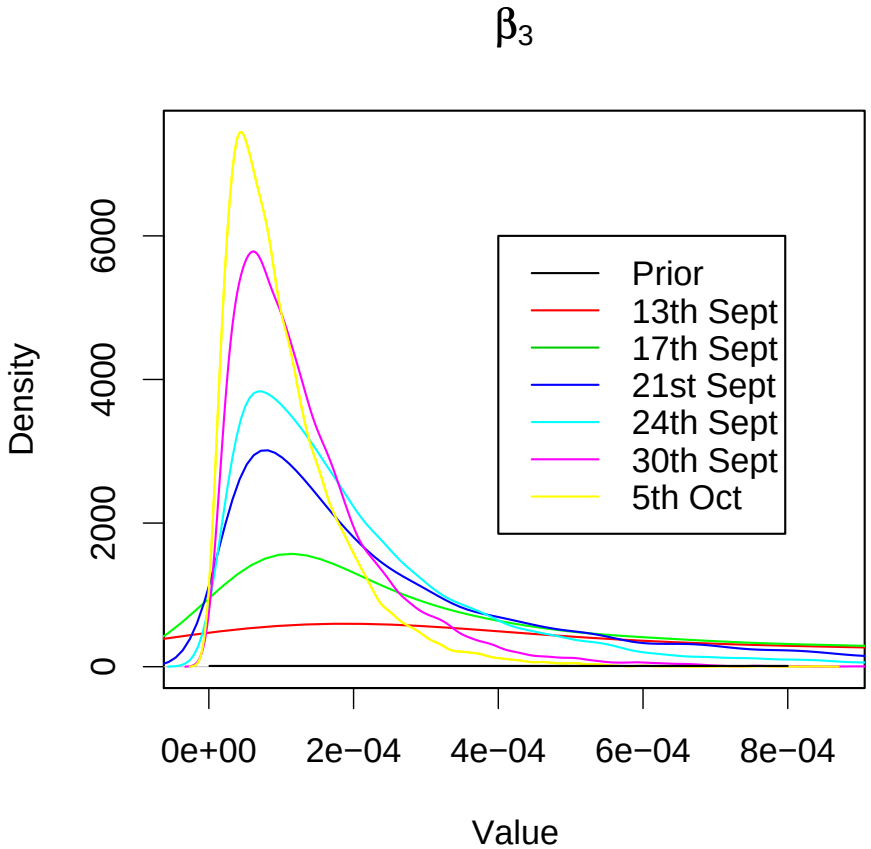
Infection to Notification time



(modified by data)

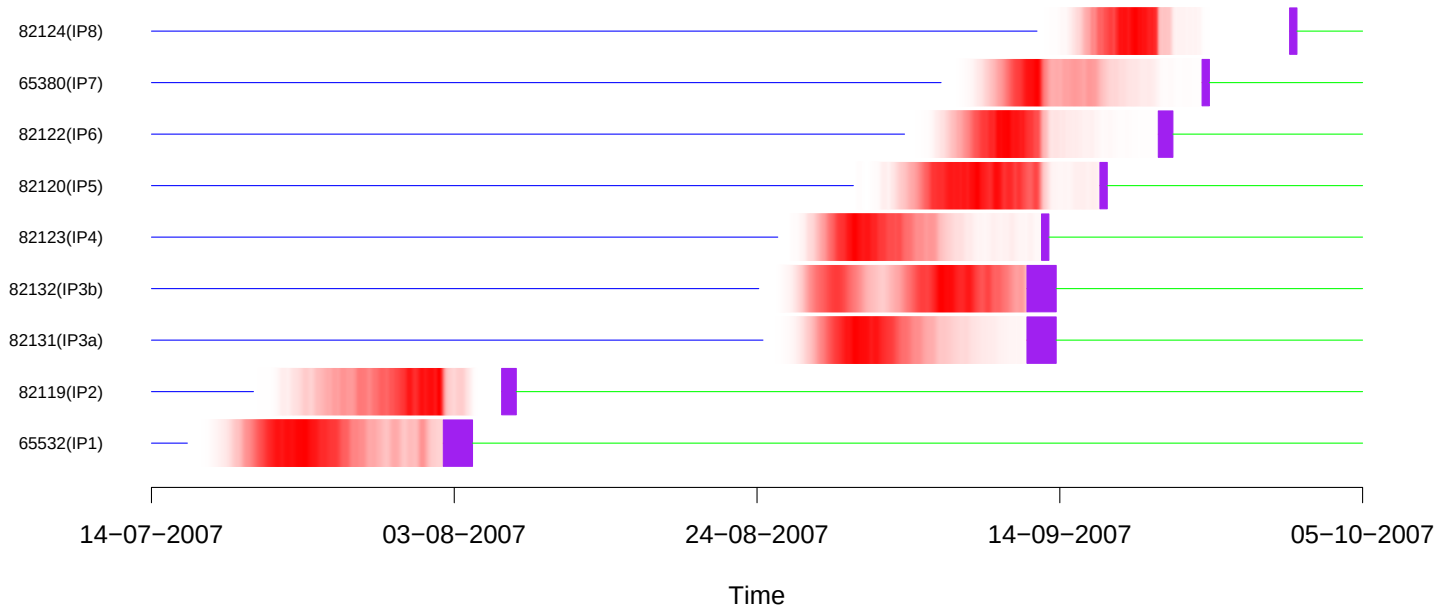
Parameter learning

As the epidemic progresses, parameters information grows



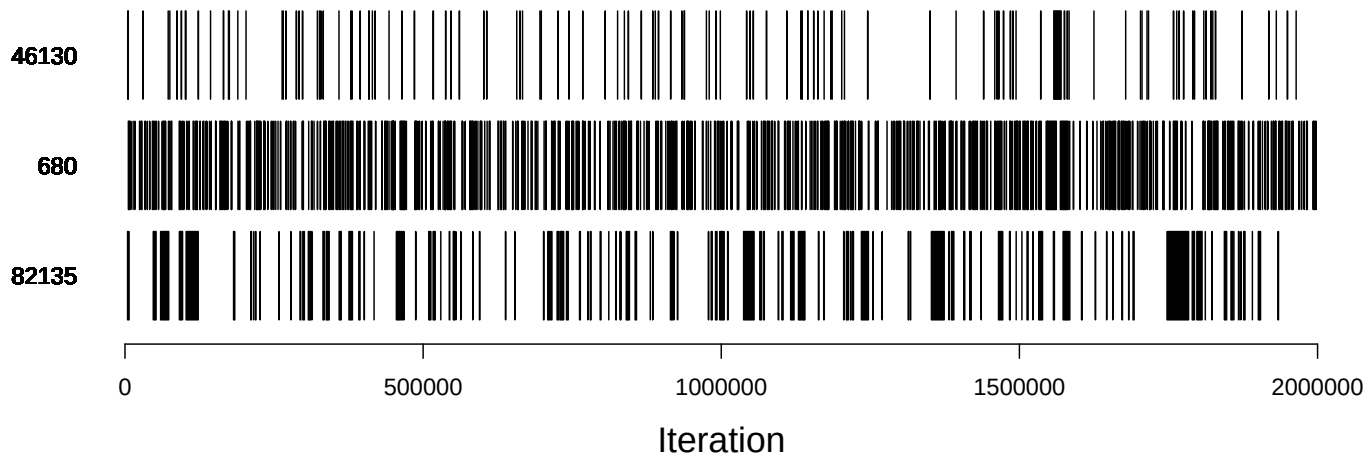
Infection Times

- Estimation of infection times



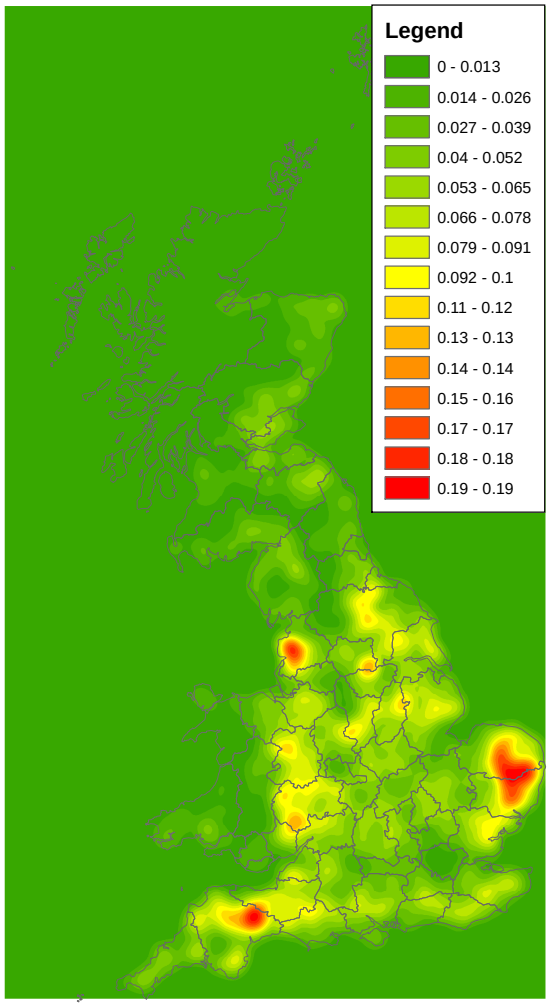
Occult infections

- Probability of being infected assigned to each *presumed* susceptible
- Direct consequence of incorporating occult infections into the analysis.



HPAI in UK Poultry

- Extract from Great Britain Poultry Register (May 2006)
- 8363 registered poulty premises after data cleaning
  - Production stock only (10 types)
- 3 contact networks identified
  - Feed lorries
  - Slaughterhouse lorries
  - Company association
- OS National Grid coordinates for each premises



The Model

Infection rate

Inter-farm transmission rate

$$\beta_{ij} = \eta_{sp,j}\left(\beta_1 C_{ij}^{FM} + \beta_2 C_{ij}^{SH} + \beta_3 C_{ij}^{CP} + \beta_4 e^{-\beta_6 \cdot \rho[i,j]}\right) \qquad i \in \mathbf{I}, j \in \mathbf{S}$$

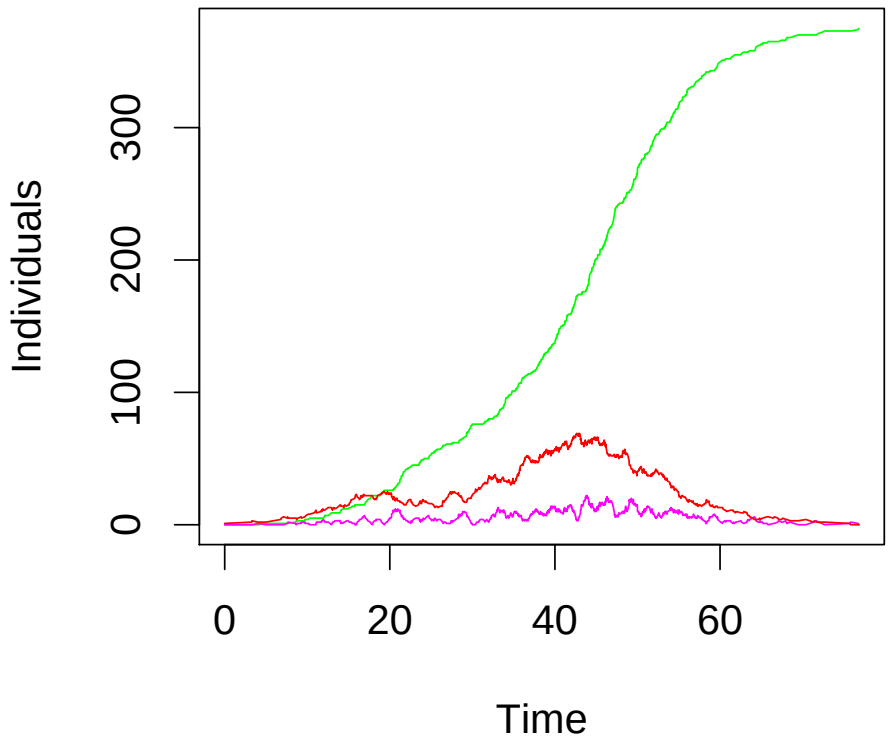
$$\beta_{ij}^{\star} = \eta_{sp,j}\left(\beta_5 e^{-\beta_6 \cdot \rho[i,j]}\right) \qquad i \in \mathbf{N}, j \in \mathbf{S}$$

- $\eta_{sp,j}$ = susceptibility of major species on farm j
- $\beta_1 C_{ij}^{FM}$ = feedmill infection rate
- $\beta_2 C_{ij}^{SH}$ = slaughterhouse infection rate
- $\beta_3 C_{ij}^{CP}$ = company infection rate
- $\beta_{\{4,5\}} e^{-\beta_6 \cdot \rho[i,j]}$ = spatial rate between farms $\rho[i,j]$km apart

Simulated epidemic

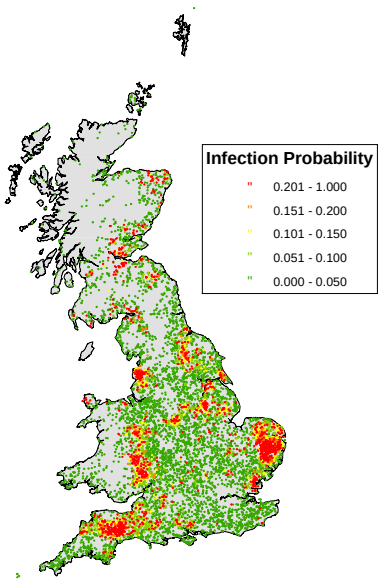
- No HPAI epidemic in the UK yet!
- Simulate epidemic on our dataset

Time/days	Infections
0	1
14	10
25	61
50	290
76	375

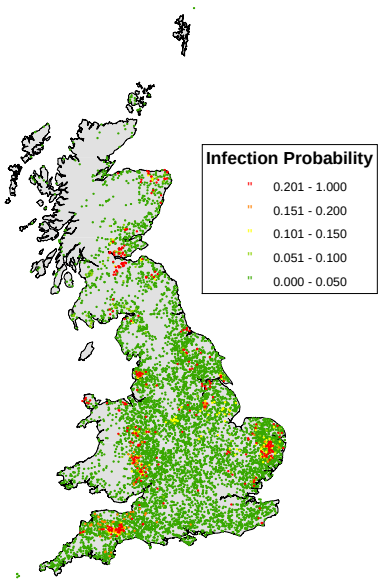


Risk Maps

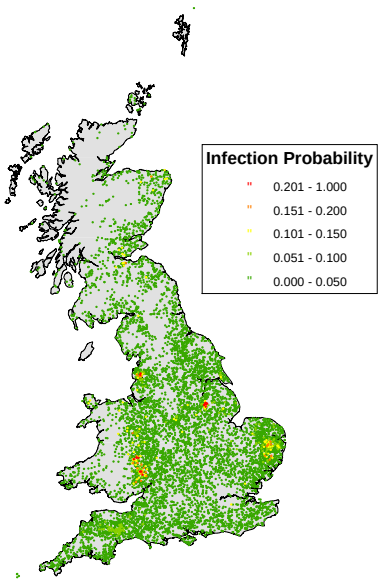
"The probability of farms becoming infected given the current situation"



14 days



25 days



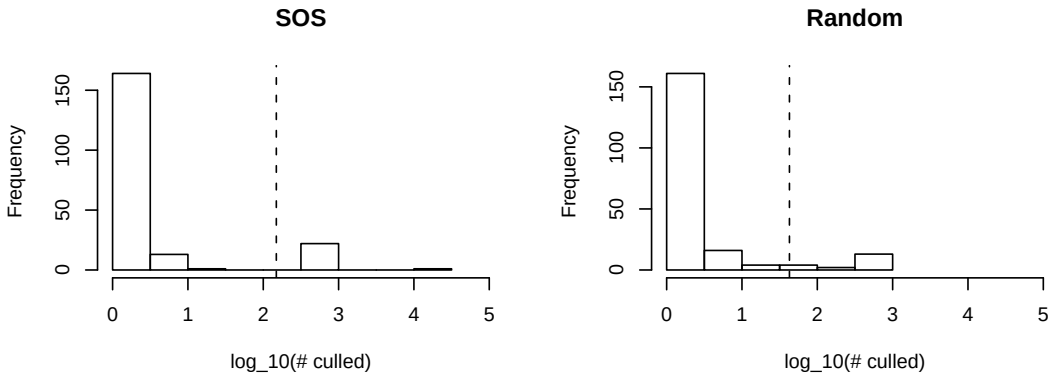
50 days

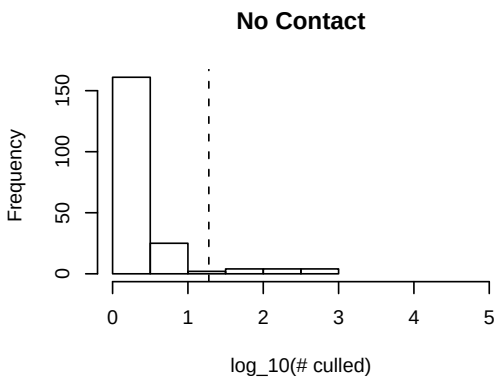
Table: Risk Maps

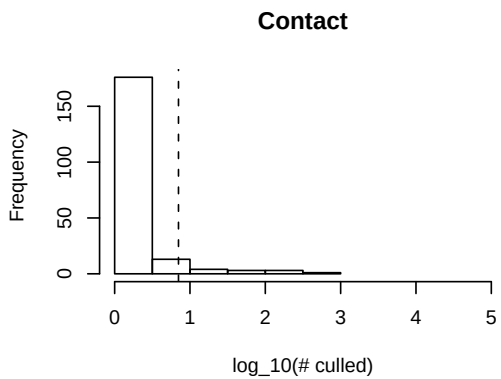
Bayesian guided surveillance

Speculative!

- Active surveillance scenarios: how should a limited active surveillance resource be targetted?
- If used: 15 farms surveyed per day, 10km radius of IPs
- Perfect on farm test, depopulated within 24h







Strategy	$\overline{\text{culled}}$
Reactive	74.9 (723)
Random	42.7 (666)
Bayes	18.7 (417)
Bayes CT	7.04 (186)

(Brackets: upper 95% HPD)

Conclusions

- We have constructed a robust flexible likelihood-based Bayesian approach for real-time parameter inference
- Solves the problem of censored data in epidemics
- In conjunction with forward simulation, this provides a powerful risk assessment resource for use during a disease epidemic in the UK
  - Bayesian predictive risk easily calculated using forward simulation.
- Evidence to suggest highly effective for optimising allocation of limited control resource.
- Such an approach can be easily adopted for other model structures as well.

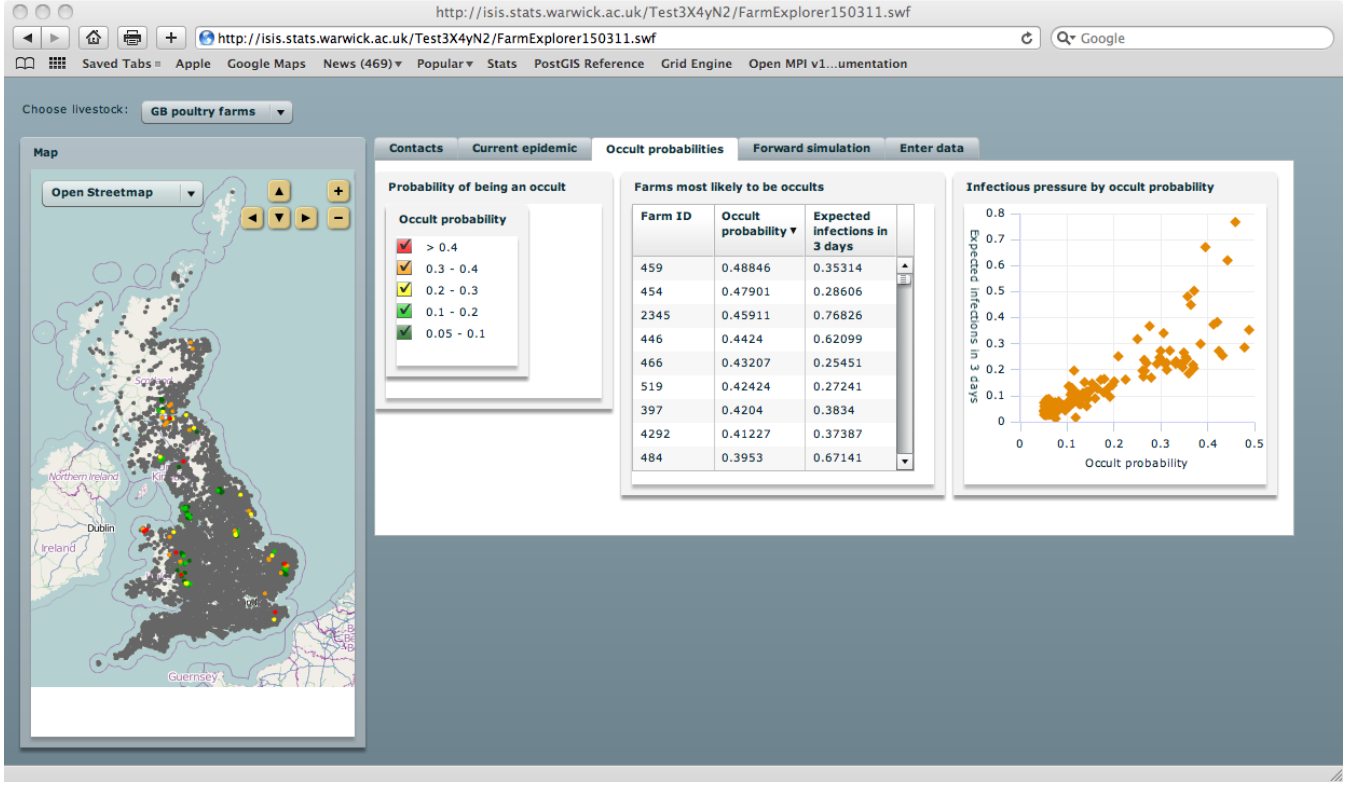
Outline

Software development

- Developer:
  - C++ API for rapid development of application-specific analysis of epidemics
  - Modular design separates model, data, parameters, and algorithm engines (simulation, MCMC etc)
  - Open source!
- User:
  - Frontend: Flash-based web application
  - Backend: Web server, relational database server, high-performance cluster
  - User concentrates on data and outputs
  - Statistician monitors usage, MCMC mixing, intervenes if necessary
  - Software updates automatic, controlled centrally

Frontend

- Web app written using Adobe Flex (open source)
- Concept stage - input/ideas/feature requests welcome!



Road map

- How does a farm's infectivity build during its infection?
- Methods for coping with uncertainty in covariate data
- Effective models for different data resolutions (Spatial and Spatio-temporal Epidemiology, in print)
- Formal methods for model diagnostics
  - Behaviour of DIC? Sellke construction + methods from survival analysis?
- HPC - development of more effective parallel algorithms
- R-package `BERP`: Bayesian inference for Epidemic Risk Prediction

Acknowledgments

- Warwick
  - Matt Keeling
  - Mike Tildesley
  - Gareth Roberts
- Nottingham
  - Theo Kypraios
- Massey University
  - Nigel French
  - Phan Minh
  - Mark Stevenson
- Defra
- BBSRC