Livestock and likelihoods: statistical risk forecasting for epidemic disease outbreaks

Chris Jewell<sup>1</sup>

Statistics and Bioinformatics Group Institute of Fundamental Sciences, Massey University

24 October 2012



Presenter: c.p.jewell@massey.ac.nz

# 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 various population characteristics in propagating the *current* epidemic?
- Measure uncertainty to better inform control policies



# A job for the statistician...?

• Results of forward simulation depend on model parameters

- To realise predictive power, require formal estimates for parameters given the model
- Historically:
  - Estimation of R<sub>0</sub> is relatively easy
    - \* Relevance to heterogeneous populations?
  - Difficult to estimate infection and removal rates together

References: Bailey (1975); Becker (1989)



#### Available data

#### Covariate data

Location, number and type of animals, contact networks





# The Approach

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



#### The Model Population structure

- $\bullet~\mbox{Farms progress}~\mbox{S} \rightarrow \mbox{I} \rightarrow \mbox{N} \rightarrow \mbox{R}$
- Populations are heterogeneously mixing and highly stochastic
  - Contact networks
  - Spatial proximity



# The Problem

- Missing data!
  - Infection times are not directly observed
  - What about occult infections?

 $f(I, N, R|\theta) = \prod_{\substack{j=1\\j\notin A\\j\notin A\\j\in A}}^{|I|} \left[ \epsilon + \sum_{i=1}^{|T_{i_j}|} S_{ij}(I_j - I_i) + \sum_{i=1}^{|T_{i_j}|} \right]$ pserved  $\times \prod_{\substack{j=1\\j\in A\\j\in A}}^{|I|} p_k^{1(C_{jkl_j}\neq\emptyset)}$  $\cdot \times \prod_{\substack{i=1\\j\in A}}^{|I|} \frac{|p|}{|I|} - p_i e^{[c_{jkl_j}\in I_j]}.$ 

- Statistical likelihood function for the model conditional on the infection times
- We cannot explicitly write a likelihood function for censored 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:
  - Coherent inclusion of Prior information
    - ★ Expert opinion
    - Previous disease outbreaks
  - 2 Measures uncertainty
    - \* avoids asymptotic assumptions



- Natural framework to include unobserved data by data augmentation MCMC methodology
  - ★ Unobserved infection times
  - Occult infections



# **Prior distributions**

- Gamma for rates ( $\beta \ge 0$ )
- Beta for probabilities ( $0 \le p \le 1$ )
- Wide range of choices for infection times (or infectious period)

- Priors chosen to agree with expert opinion and previous knowledge of epidemics
- Frameworks and software, eg. SHELF (O'Hagan *et al*), Elicitor (Kynn)



## 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 (no pigs!)

- Priors based on posterior of Kypraios 2007 (FMD2001)
- Case data by email

#### Transmission model

$$\beta_{ij} = \left( \mathbf{c}_i^{\psi} + \xi_s \mathbf{s}_i^{\psi} \right) \left( \mathbf{c}_j^{\psi} + \zeta_s \mathbf{s}_j^{\psi} \right) \cdot \beta_3 \cdot \frac{\delta}{\rho_{ii}^2 + \delta^2} \quad i \in \mathbf{I}, j \in \mathbf{S}$$



### Parameter learning

As the epidemic progresses, parameters information grows



Parameter value



#### **Infection Times**

#### Estimation of infection times



Time



Bayesian guided surveillance - HPAI example Jewell and Roberts (2012) *Biostatistics* 

- 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



# FMD2001 - Non-orthogonality



#### Conclusions

- 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.
- Currently building into a prototype disease management system.



### Road map

- Formal methods for model diagnostics PRIORITY!
- Assimilation of diverse data, eg pathogen sequencing, contact tracing
- Alternatives to standard MCMC, eg MWMH, GIMH, SMC
- HPC development of more effective parallel algorithms
  - parallel clusters, GPU computing
- Software:
  - ► R-package BERP: Bayesian inference for Epidemic Risk Prediction
  - Interspread, the next generation ??



## Acknowledgments

- Warwick
  - Matt Keeling
  - Mike Tildesley
  - Gareth Roberts
  - Laura Green
  - Judith Brown
- Nottingham
  - Theo Kypraios
- AHVLA
- BBSRC





### Sellke diagnostics



IDReC