Bayesian data assimilation for vector borne disease response: *Theileria orientalis* (Ikeda) in NZ cattle.

Chris Jewell

CHICAS
Lancaster University
c.jewell@lancaster.ac.nz

16th April, 2015

Lancaster Medical School | Lancaster University

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

CHICAS combining health information, computation and statistics

1. Motivation
2. Modelling
3. Inference and forecasting
4. T. orientalis Ikeda prediction
5. Roadmap

CHICAS combining health information, computation and statistics

Outline

1 Motivation

2 Modelling

3 Inference and forecasting

4 T. orientalis Ikeda prediction

5 Roadmap

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?
- Enterobacteriaciae?
- Casts doubt on food safety

CHICAS combining health information, computation and statistics

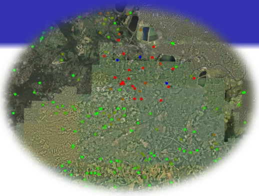
Epidemic Control

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

Epidemic control

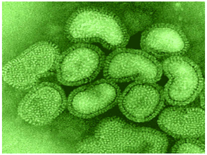


Forecasts of disease spread inform control:

- Provide real-time predictions of risk
  - Understanding the determinants of transmission
  - How many occult (undetected) infections are there?
  - Who is likely to be infected next?
- Explore a full range of possible outcomes given the current state of the epidemic
  - Movement restriction
  - Vaccines
  - Culls

Emerging infectious diseases

- Broad definition - Brown, C (2004) *Rev. sci. tech. Off. int. Epiz* **23**:435
  - a known agent appearing in a new geographic area
  - a known agent or its close relative occurring in a hitherto unsusceptible species
  - a previously unknown agent detected for the first time.
- Our only experience of how they behave is the current epidemic
- Model parameters may be uncertain



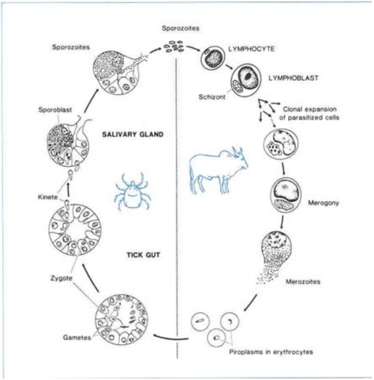
Theileria orientalis Ikeda

NZ Ministry for Primary Industries contract, Massey EpiCentre

- Protozoal vector borne disease
  - Host: Cattle
  - Vector: Tick *H. longicornis*

- Endemic, but...
- August 2012 new virulent subspecies Ikeda

- Case morbidity $< 35\%$
- Mortality $\approx 1\%$
- Cost NZ$25k per farm

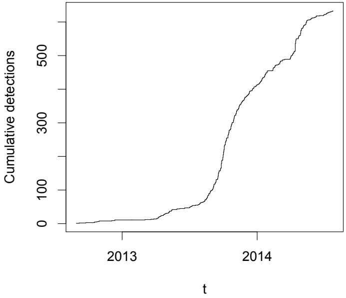


Theileria in NZ cattle

Questions

- What are the main determinants of transmission?
  - Environmental (spatial) spread
  - Spread via NAIT network
  - Importance of tick presence
- Where might any undetected infections be?
- How fast will the epidemic spread?

Theileria in NZ cattle



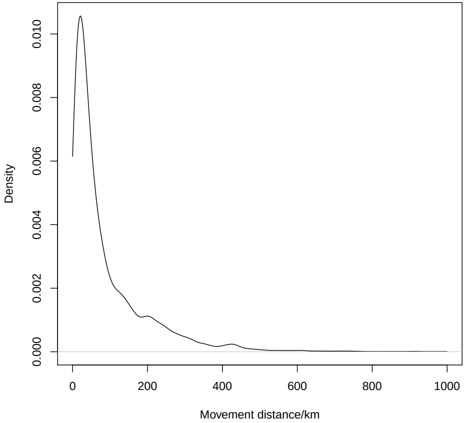
Epidemic at 1st Aug 2014

- 655 Case detections
- 136837 Farm locations
- Dairy/non-dairy
- NAIT cattle movement network frequencies

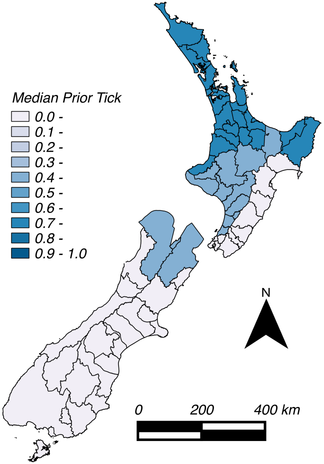
Databases: AgriBase, FarmsOnLine, NAIT

National Animal Identification and Tracing (NAIT)

- Mandatory tracing of cattle and deer
- Began 2012
- 517328 recorded movements
- 0.0003% connectivity
- Represented as sparse matrix



Theileria in NZ cattle
Tick habitat



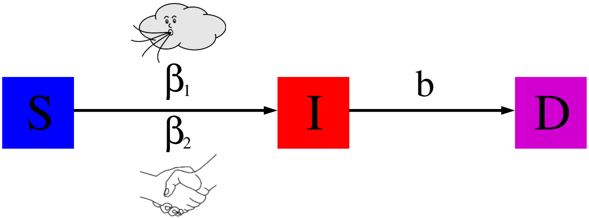
- Samples tested from BVD surveillance
- All *T. orientalis* spp.
- Aggregated at TLA level
- Prior risk: Alan Heath

Outline

1. Motivation
2. Modelling
3. Inference and forecasting
4. T. orientalis Ikeda prediction
5. Roadmap

SID models & population heterogeneity

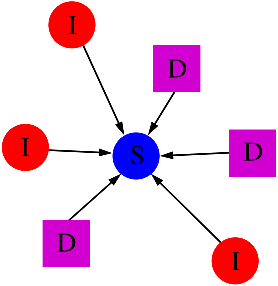
1. Individual herds infect each other
    - Spatial rate β_1
    - NAIT network β_2
2. Once infected, herds are infectious forever
    - Detected at rate b



$$\beta_1 = ?, \beta_2 = ?, b = ?$$

Infection process

- At any time t, susceptible j has infectious pressure exerted on it by
  - all infected or notified farms $i \in \mathcal{I}(t), \mathcal{D}(t)$
  - "Background" (not explicitly modelled)



In a small interval $\Delta t = (t, t + \delta]$:

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

$$\lambda_j(t) = \epsilon + \sum_{i \in \mathcal{I}(t)} \lambda_{ij}(t) + \sum_{i \in \mathcal{N}(t)} \lambda_{ij}^{\star}$$

Transmission model

SID model

$$\lambda_{ij}(t) = s(t;\boldsymbol{\alpha},\nu)\zeta^{\kappa_j} p_{k(j)} \left[\beta_1 K(i,j;\delta) + \beta_2 c_{ij}\right], \quad i \in \mathcal{I}, j \in \mathcal{S}$$
$$\lambda^{\star}_{ij}(t) = \beta_{ij}(t), \quad i \in \mathcal{N}, j \in \mathcal{S}$$

$$K(i,j;\delta) = \frac{\delta}{(\delta^2 + ||x_i - x_j||^2)^{1.2}}$$

$p_{k(j)}$ = tick occurence in TLA k

ζ = effect of dairy cf. beef

$$D - I \sim Gamma(4, b)$$

Unknowns
$\boldsymbol{\theta} = \{\beta_1, \beta_2, \delta, \boldsymbol{p}, \zeta, \boldsymbol{\alpha}, \nu, b\}$

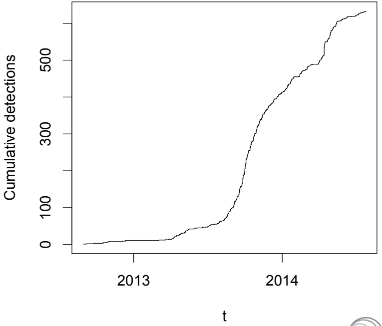
CHICAS combining health information, computation and statistics

Seasonality

- Biannual peak incidence – autumn/spring
- Due to vector ecology

- Vector seasonality
  - Smooth? (common sense!)
  - On/off threshold? (literature*)

- Candidates:
  - Periodic piecewise cubic spline
  - Periodic square wave



*e.g. Stafford KC (1994), Ogden *et al.* (2004)

Tick model – dynamics

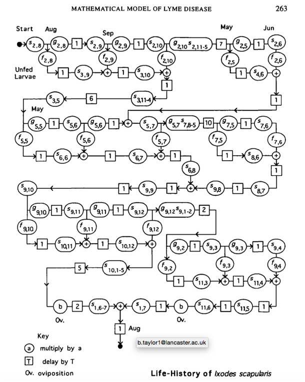
IMA Journal of Mathematics Applied in Medicine and Biology (1999) **16**, 261–296

A mathematical model of the ecology of Lyme disease

TRAVIS C. PORCO

Community Health Epidemiology and Disease Control, 25 Van Ness Avenue, Suite 710, San Francisco Department of Public Health, San Francisco, California 94102, USA

[Received 14 February 1994 and in revised form 17 June 1998]



Tick model – proxy sampling

- BVD samples collected during 2013
- Tested for *Theileria orientalis* spp.

- Implies Binomial sampling model, TLAs $k = 1, \ldots, m$

$$x_k \sim \text{Binomial}(n_k, p_k)$$

n_k herds sampled, x_k *Theileria* +ve, $p_k \propto$ tick occurrence

- Independent sampling $\rightarrow$ joint likelihood
- Robust to test Sensitivity and Specificity.

Outline

1. Motivation
2. Modelling
3. Inference and forecasting
4. T. orientalis Ikeda prediction
5. Roadmap

Approach to inference

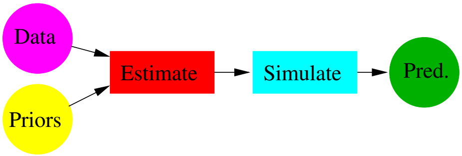
Bayesian approach for risk forecasting

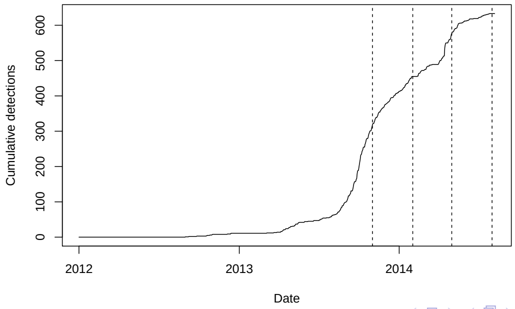
$$P(\theta|X) = \frac{L(X|\theta)f(\theta)}{\int_\Theta L(X|\theta)f(\theta)d\theta}$$

- Coherent inclusion of Prior information
- Likelihood assimilates *all* data
- Posterior encodes uncertainty $\rightarrow$ Predictive distribution
- Complicated integral $\rightarrow$ MCMC
  - Unobserved infection times
  - Occult infections

Approach to real time forecasting

At any point during an epidemic:





Approach to inference

- Bayesian approach for risk forecasting

$$
\begin{aligned}
L(\boldsymbol{\theta}|\boldsymbol{I},\boldsymbol{D}) &= \sum_{j:I_\kappa<I_j<T_{obs}} \left(\lambda_j(I_j^-)\right) - \exp\left[\int_{I_\kappa}^{T_{obs}} \left(\sum_{j\in\mathcal{P},j\neq\kappa} \lambda_j(t)\right) \mathrm{d}t\right] \\
&\times \prod_{k=1}^{m} p_k^{x_k}(1-p_k)^{n_k-x_k} \\
&\times \sum_{j:D_j\leq T_{obs}} \left(f_D\left(D_j - I_j\right)\right) + \sum_{j:D_j>T_{obs}} \left(1 - F_D\left(T_{obs} - I_j\right)\right)
\end{aligned}
$$

- Jump algorithm integrates over dimension of $\boldsymbol{I}$
- GPU as a likelihood coprocessor
  - NVIDIA CUDA software libraries
  - Parallelise likelihood within MCMC.

MCMC algorithm

Repeat the following steps

1. Model parameters (adaptive logRWMH)
   1. Update $\{\beta_1, \beta_2, \zeta, \epsilon, \delta\}$, $\boldsymbol{p}$, $\boldsymbol{\alpha}$, ν
   2. Update b
   3. Update b with $U_i = bD_i \sim \text{Gamma}(4, 1)$ (non-centred)
2. Infection times - repeat z times:
   Equal probability
   1. Move $I_s \rightarrow I_x^\star$
   2. Add $\{\boldsymbol{I} + s\} \rightarrow \boldsymbol{I}^\star$
   3. Delete $\{\boldsymbol{I} - s\} \rightarrow \boldsymbol{I}^\star$

CHICAS combining health information, computation and statistics

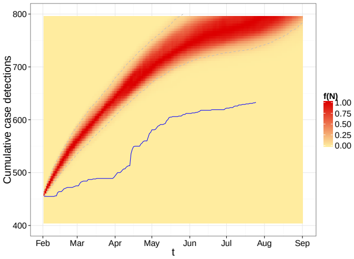
Outline

1. Motivation
2. Modelling
3. Inference and forecasting
4. T. orientalis Ikeda prediction
5. Roadmap

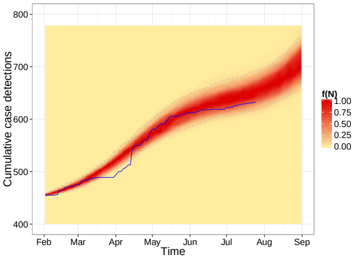
Theileria in NZ cattle

Prediction: in sample cumulative detections

Cubic Spline



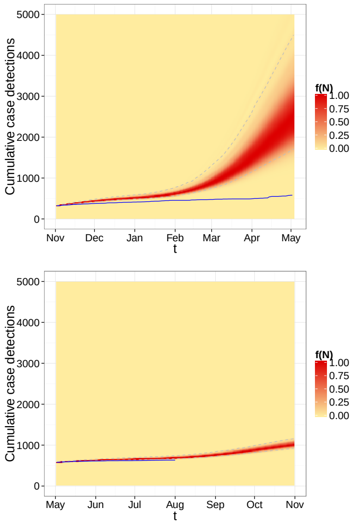
Square Wave



Analysis at 1st Aug 2014, simulate from 1st Feb 2014

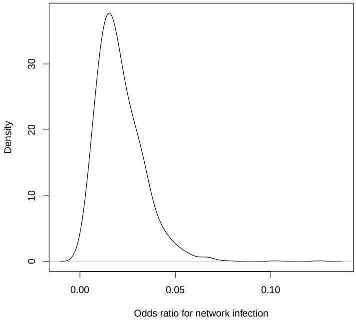
Theileria in NZ cattle

Prediction: out of sample cumulative detections – justification for square wave

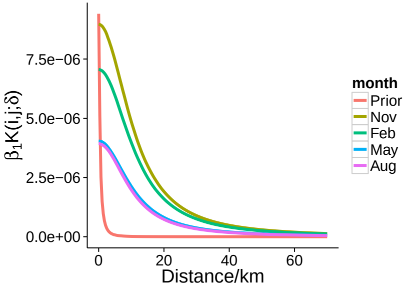


Theileria in NZ cattle
Parameter estimation

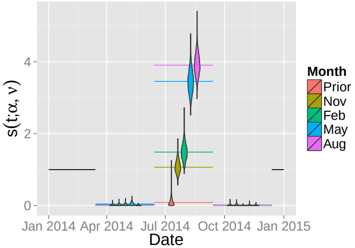
NAIT vs Environmental spread

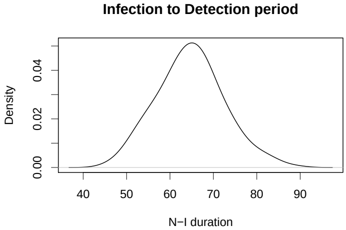


Distance Kernel



Seasonality and infectious period

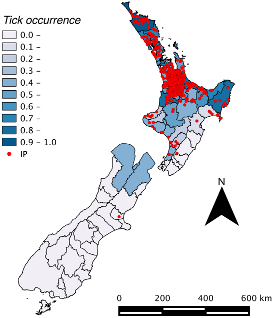


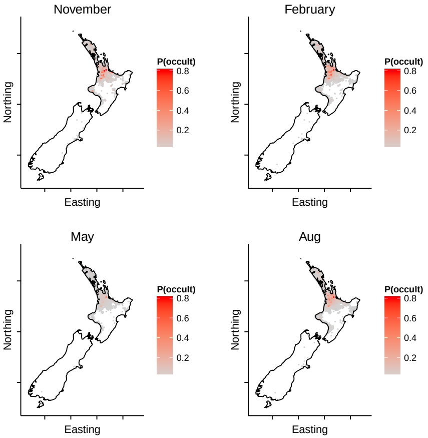


Caveat: non-identifiability between phase and infec-detect time!

Theileria in NZ cattle

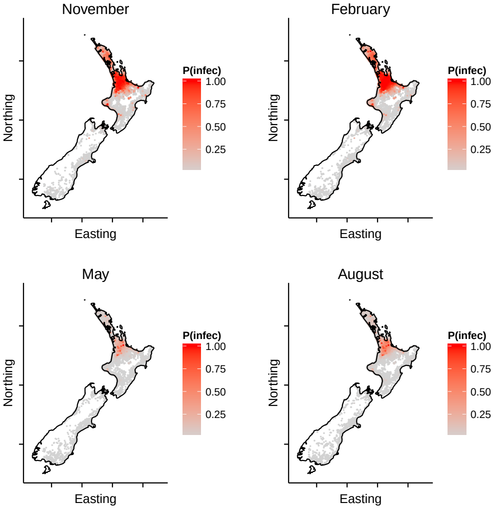
Tick occurrence probability and occults





Theileria in NZ cattle

Prediction: 6 month infection risk



Outline

1 Motivation

2 Modelling

3 Inference and forecasting

4 T. orientalis Ikeda prediction

5 Roadmap

Theileria in NZ cattle
Conclusions

- Environmental spread main driver
  - NAIT spread limited (currently)
  - Caveat – AgriBase, FOL, NAIT joining issues
- Epidemic appears to be slowing
  - Complacency vs. effective pop size.
- Modelling issues
  - Ticks
    - Seasonality – tie in with climate data (resolve $Cor(b, \nu)$)
    - e.g. MAXLIK and MAXENT, etc.
  - Endemic stability
    - Require SIDS model
    - Within-farm epidemic
- Lack of coherent livestock databases (CEBRA/MPI)

Thanks to...

- IFS
  - Richard Brown
- MPI
  - Daan Vink
  - Kevin Lawrence
  - Andy McFadden
  - Mary van Andel
- AsureQuality
  - Robert Sanson