Analyzing veterinary surveillance data: Approaches to model the relationship between disease incidence and cattle trade

Birgit Schrödle, Leonhard Held & Michaela Paul

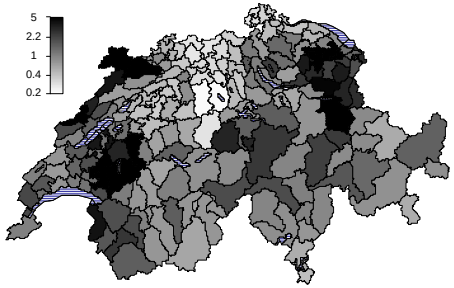
Biostatistics Unit
Institute of Social and Preventive Medicine
University of Zurich

Glasgow, July 7th, 2010

Overview

Swiss registry data

- Registry data are a tool for passive surveillance.
- Here: Reported cases of Coxiellosis in cattle
  → Time: 2005 - 2009 (5 years)
  → Space: 184 regions of Switzerland and the Principality of Liechtenstein
- Relative spatial risk, estimated by a spatio-temporal model:



Cattle trade as a risk factor

- Spread of a bovine disease over short distances by wind or insects → local clustering
- Gilbert *et al.* (2005, Nature):
  Dispersal over long distances by trade of infectious animals
- Swiss cattle trade database:
  Since 2008 it is mandatory for stock-keepers to notify all cattle movements
- Assumption: Movement/trade pattern is consistent over time
  → use data from 2009

→ Association of animal trade and disease presence?

Two approaches

- Question: Is there an between association of animal movement and disease presence?
- Approach 1: Spatio-temporal disease mapping
  - hierarchical Bayesian model (→ INLA, Rue *et al.* (2009, JRSSB))
  - include cattle trade quantities as covariates
  - parameter-driven approach
  - Knorr-Held (2000, Stat Med)
- Approach 2: Infectious disease model
  - likelihood framework (→ numerical optimization)
  - include cattle trade as a weight for spatial spread
  - observation-driven approach
  - Paul *et al.* (2008, Stat Med)

Hierarchical Bayesian model - parameter-driven

- Data Y: Number Y_{it} of reported cases in region i at time t
- Y_{it} are Poisson distributed given a rate λ_{it}:

$$Y_{it}|\lambda_{it} \sim \text{Poisson}(\lambda_{it})$$

- Linear predictor:

$$\log(\lambda_{it}) = \eta_{it} = \log(m_i) + \xi + \nu_i + \psi_i + \gamma_t + \phi_t + \delta_{it} \quad (1)$$

  - m_i, ξ: Number of units in region i, intercept
  - ν_i, ψ_i: Spatial components (i.i.d. normally distr., ICAR)
  - γ_t, ϕ_t: Temporal components (i.i.d. normally distr., RW1)
  - δ_{it}: Space-time interaction term (IGMRF)
- Expand model (1) to account for covariates x_{it}:

$$\eta_{it} = (1) + \beta \cdot x_{it}$$

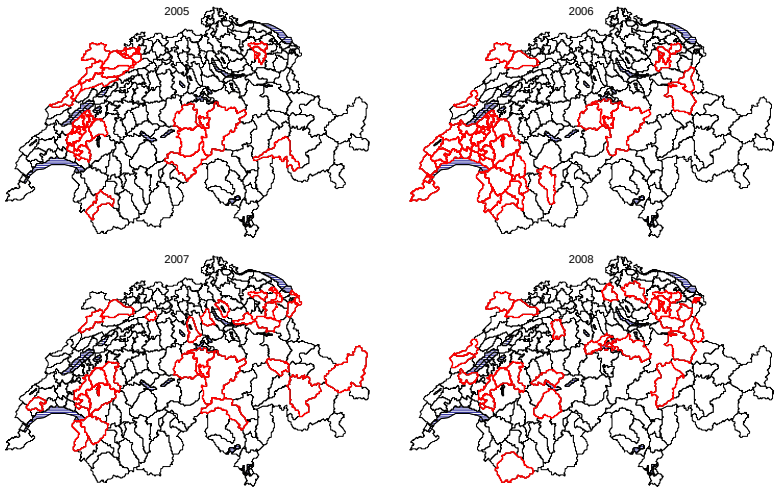
→ ecological regression

Cattle trade parameters

1. Total number of in-movements (TOT) per region, constant over time
2. Absolute number (A) and proportion (P) of in-movements from regions with elevated risk, time-varying

   Obtained within a two-stage process:
   - → Fit a spatial disease mapping model for each year separately.
   - → Indicate regions with elevated risk in the previous year using a threshold (e.g. RR>2) and extract covariates x_{it} accordingly.
   - → Fit a spatio-temporal ecological regression model.
3. DIC can be used for model comparison
4. Problem: Covariate is derived using previous observations $y_{i,t-1}$ → Endogenous covariate

Regions with elevated risk (RR>2), 2005-2008



Results for parameter-driven approach

M(1): $\eta_{it} = \log(m_i) + \xi + \nu_i + \psi_i + \gamma_t + \phi_t + \delta_{it} \qquad (+\beta \cdot x_{it})$

Model	DIC	$\hat{\beta}$ [95%-CI]
M(1)	857.7	–
TOT	857.1	0.010 [−0.001; 0.021]
RR>2 & A	855.8	0.019 [0.009; 0.028]
RR>2 & P	859.9	1.268 [0.369; 2.163]

Problem endogenous covariate:

- Chosen regions of elevated risk typically are a few spots of neighbouring regions
- Cattle trade between neighbouring regions is much larger → Confounding by location
- Hint in this direction: Variability of spatially structured component ψ_i drops after inclusion of covariate

Infectious disease models - observation-driven

- A purely parameter-driven model might not be able to describe localized epidemics (Held *et al.*, 2005, Stat Mod)
- Solution: Include the number of cases $y_{i,t-1}$ in the past
- Simplest model (IDM1):
  Observations y_{it} are Poisson distributed with mean

$$\mu_{it} = \lambda \cdot y_{i,t-1} + m_i \cdot \exp(\alpha) \quad \lambda > 0,$$

  where α accounts for all residual variation.

Infectious disease models... continued

- Model the spatial spread of the disease (IDM2):
  Add the sum of counts in other regions j, weighted by w_{ji}

$$\mu_{it} = \lambda \cdot y_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot y_{j,t-1} + m_i \cdot \exp(\alpha)$$

  with $\lambda, \rho > 0$.
  Various choices for weights w_{ji}:
  1. $w_{ji} = 1$ for all neighbouring regions $j \sim i$
  2. Square root of the absolute number of cattle movements ($\sqrt{\text{CM}}$) from region j to i
- Split α in an intercept and a linear time trend (IDM3):

$$\mu_{it} = \lambda \cdot y_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot y_{j,t-1} + m_i \cdot \exp(\alpha + \zeta \cdot t)$$

- AIC can be used for model comparison

Results for observation-driven approach

IDM1: $\mu_{it} = \lambda \cdot y_{i,t-1} + m_i \cdot \exp(\alpha)$

IDM2: $\mu_{it} = \lambda \cdot y_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot y_{j,t-1} + m_i \cdot \exp(\alpha)$

IDM3: $\mu_{it} = \lambda \cdot y_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot y_{j,t-1} + m_i \cdot \exp(\alpha + \zeta \cdot t)$

Model	w_{ji}	AIC	$\hat{\lambda}$ (s.e.)	$\hat{\rho}$ (s.e.)	$\hat{\zeta}$ (s.e.)
IDM1	–	1092.8	0.44 (0.05)		
IDM2	$1 \forall j$	1094.8	0.44 (0.05)	0.0000 (0.0000)	
IDM2	1, if $j \sim i$	1062.7	0.43 (0.05)	0.0491 (0.0103)	
IDM2	$\sqrt{\text{CM}}$	1082.0	0.44 (0.05)	0.0005 (0.0002)	
IDM3	$\sqrt{\text{CM}}$	1083.5	0.44 (0.05)	0.0005 (0.0002)	0.08 (0.11)

Discussion

- Model choice criteria and parameter estimates indicate a positive association between animal movement and disease presence for both approaches.
- Unfortunately, the two approaches are not comparable at the moment.
- Problems:
  → Approach 1 makes use of an endogenous covariate.
  → Approach 2 does not include, e.g., a random effect to account for local spatial clustering.
- Solution:
  Mixture of a parameter- and observation-driven model!
  → Might be possible within an hierarchical Bayesian framework...

Literature

Gilbert, M., Mitchell, A., Bourn, D., Mawdsley, J., Clifton-Hadley, R. and Wint, W. (2005). Cattle movement and bovine tuberculosis in Great Britain, *Nature* **435**: 491–496.

Held, L., Höhle, M. and Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts, *Statistical Modelling* **5**: 187–199.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk, *Statistics in Medicine* **19**: 2555–2567.

Paul, M., Held, L. and Toschke, A. (2008). Multivariate modelling of infectious disease surveillance data, *Statistics in Medicine* **27**: 6250–6267.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion), *Journal of the Royal Statistical Society: Series B* **71**: 319–392.