



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2010

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

Schrödle, B ; Held, L ; Paul, M

Abstract: Two approaches to the analysis of registry data for bovine diseases with regard to the relationship between disease incidence and cattle trade are proposed. Firstly, a parameter-driven spatio-temporal disease mapping model formulated in a hierarchical Bayesian framework is used. Various cattle movement parameters, e.g. the number and proportion of in-movements from infected regions, can be included as potential covariates. Within this context problems of such an endogenous covariate are discussed. Since a purely parameter-driven approach is often not adequate to depict local epidemics, a so-called observationdriven infectious disease model is proposed as a second possibility. It includes an autoregressive part for counts in the region of interest in the past. Additionally, the sum of previous cases in other regions weighted by cattle movements is added to assess the spread of the disease by trading. Both models are applied to cases of Coxiellosis in Switzerland, 2005 to 2009.

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: <https://doi.org/10.5167/uzh-38476>
Conference or Workshop Item

Originally published at:

Schrödle, B; Held, L; Paul, M (2010). Analyzing veterinary surveillance data: Approaches to model the relationship between disease incidence and cattle trade. In: 25th International Workshop on Statistical Modelling, Glasgow, 5 July 2010 - 9 July 2010, 1-6.

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

Birgit Schrödle¹, Leonhard Held¹ and Michaela Paul¹

¹ University of Zurich, Institute for Social and Preventive Medicine, Hirschengraben 84, 8001 Zurich, Switzerland, Email:birgit.schroedle@ifspm.uzh.ch

Abstract: Two approaches to the analysis of registry data for bovine diseases with regard to the relationship between disease incidence and cattle trade are proposed. Firstly, a parameter-driven spatio-temporal disease mapping model formulated in a hierarchical Bayesian framework is used. Various cattle movement parameters, e.g. the number and proportion of in-movements from infected regions, can be included as potential covariates. Within this context problems of such an endogenous covariate are discussed. Since a purely parameter-driven approach is often not adequate to depict local epidemics, a so-called observation-driven infectious disease model is proposed as a second possibility. It includes an autoregressive part for counts in the region of interest in the past. Additionally, the sum of previous cases in other regions weighted by cattle movements is added to assess the spread of the disease by trading. Both models are applied to cases of Coxiellosis in Switzerland, 2005 to 2009.

Keywords: Cattle trade; Spatio-temporal disease mapping; Infectious disease; INLA; Observation-driven

1 Introduction

The spread of a bovine disease can take place over short distances between adjacent or nearby farms borne by wind or insects, which typically results in local clustering of cases. However, disease dispersal also takes place over long distances caused by trade of infectious animals (Gilbert et al., 2005). Hence, the inclusion of cattle trade in an analysis of veterinary surveillance data might give hints towards the association of animal movement and disease presence.

As a first approach disease counts can be analyzed using a disease mapping model that considers spatial, temporal, and spatio-temporal trends (Knorr-Held, 2000; Schrödle and Held, 2009). Additionally, various cattle movement parameters can be included in the model using ecological regression (Clayton et al., 1993). Bayesian inference is conducted using integrated nested Laplace approximations (INLA), which was recently proposed in Rue et al. (2009).

As an alternative to this purely parameter-driven approach the spread of a bovine disease by cattle trade can be modelled within a likelihood-based infectious disease framework (Paul et al., 2008). Here, an autoregressive term for past counts in the region of interest and the sum of past cases in other regions weighted by cattle movements are part of the model formulation. The advantage of this so-called observation-driven model is that it is able to describe local epidemics.

Coxiellosis in cattle is an infectious, bacterial disease among ruminant animals, which can be spread by airborne infection. It can be the reason for an abortion, even in a late phase of the pregnancy. Cases of Coxiellosis were reported to the Swiss Federal Veterinary Office (BVET) between 2005 and 2009 and are available aggregated for 184 Swiss regions and the Principality of Liechtenstein on a yearly basis. The number of herds m_i in each region i is known.

Since 2008 it has been mandatory for Swiss stock-keepers to notify the BVET of all cattle movements. As the spatial pattern of movements is similar for 2008 and 2009 the considered movement parameters are assumed to be consistent from year to year.

2 Spatio-temporal disease mapping

For modelling spatio-temporal disease counts $y_{it} \sim \text{Po}(\exp(\eta_{it}))$ a nonparametric hierarchical Bayesian setting as proposed in Knorr-Held (2000) is used. The respective linear predictor can be written as

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

where ξ is an intercept, ν_i and ψ_i are spatially unstructured and structured effects and γ_t and ϕ_t are temporal main effects, specified as an i.i.d. term and a random walk of first order, respectively. The term δ_{it} accounts for spatio-temporal interaction and can be specified assuming four different types of interaction between time and space (Knorr-Held, 2000). To account for covariates x_{it} as introduced in the following paragraph (1) can be extended to

$$\eta_{it} = \log(m_i) + \xi + \nu_i + \psi_i + \gamma_t + \phi_t + \delta_{it} + \beta \cdot x_{it}. \quad (2)$$

The total number of in-movements (model acronym: TOT) and the absolute number (A) and proportion (P) of in-movements from regions with elevated risk are considered as potentially associated with disease presence. Since we assume that the movement pattern does not change from year to year the total number of in-movements is constant over time. The movements from infected regions are time-varying and defined using a two-stage process: A separate spatial disease mapping model is fitted for each year, including only ν_i and ψ_i from (1) (Besag et al., 1991). Regions are indicated

as infected when exceeding two different thresholds, namely an estimated relative risk larger than 2 and 3, respectively. At the second stage the models are fitted using a time lag of one and two years, respectively, to detect the incubation period of the disease (TL1 and TL2). As the time-varying covariate is derived using previous observations it is a so-called endogenous or feedback variable. This issue will be discussed briefly in Section 4.

All models are fit using integrated nested Laplace approximations (INLA). This approach for approximate Bayesian inference was recently proposed by Rue et al. (2009) as an alternative to Markov chain Monte Carlo mechanisms. The advantage of INLA is that it runs in remarkably fast computational time and returns accurate parameter estimates for a wide range of models. Additionally, INLA computes the deviance information criterion (DIC) as tool for Bayesian model choice. All analyses in this paper were conducted using the R INLA package build on the 1st of February 2010, version 1.668.

3 Infectious disease model

A purely parameter-driven model as proposed in Section 2 might not be able to describe localized epidemics which can often be found in veterinary disease surveillance data (Held et al., 2005). Hence, a so-called observation-driven model is built including the number of cases $y_{i,t-1}$ in the past. In its simplest formulation the observations y_{it} are Poisson distributed with mean

$$\mu_{it} = \lambda \cdot y_{i,t-1} + m_i \cdot \exp(\alpha) \quad (3)$$

and $\lambda > 0$ (Held et al., 2005). The parameter α accounts for all residual variation. Cases at times $t - k$, $k > 1$, could be considered as well.

As an addition, the sum of counts in all other regions j weighted by a factor w_{ji} can be added to model the spatial spread of the disease over time. In Paul et al. (2008) the respective mean is specified as

$$\mu_{it} = \lambda \cdot y_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot y_{j,t-1} + m_i \cdot \exp(\alpha) \quad (4)$$

with $\lambda, \rho > 0$. To assess the association between cattle movement and disease presence the square root of the absolute number of cattle movements (CM) between regions j and i are used as weights w_{ji} in this application. Other weights w_{ji} can also be considered (Paul et al., 2008). Here, models with $w_{ji} = 1$ for all j and for all $j \sim i$, respectively, are fit as alternatives. The term $j \sim i$ denotes all regions j which are neighbours of region i . The parameter α in (4) can be split into an intercept and a linear time trend

$$\mu_{it} = \lambda \cdot y_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot y_{j,t-1} + m_i \cdot \exp(\alpha + \zeta \cdot t). \quad (5)$$

TABLE 1. Spatio-temporal disease mapping (see Section 2): The DIC and the posterior mean of the respective cattle trade parameter along with its 95%-credible interval are shown for each model. For the model without covariate (1) the DIC is shown only.

		RR > 2		RR > 3	
		DIC	$\hat{\beta}$	DIC	$\hat{\beta}$
TL1	A	939.3	0.017 [0.008; 0.027]	942.7	0.012 [0.004; 0.020]
	P	947.0	1.29 [0.39; 2.18]	950.0	0.96 [-0.02; 1.94]
TL2	A	945.2	0.009 [-0.000; 0.018]	951.0	0.007 [-0.003; 0.016]
	P	948.0	0.83 [-0.038; 1.69]	951.1	0.85 [-0.20; 1.90]
TOT			955.8	0.005 [-0.006; 0.017]	
(1)				954.6	

Extensions for a region-specific random effect are also possible (Paul et al., 2009), but the computation of the results might suffer from numerical problems if the number of regions is large. Hence, they are not considered here.

Maximum likelihood inference is performed using iterative algorithms as described in Held et al. (2005) and Paul et al. (2008). The AIC is calculated for model choice.

4 Results

Regarding spatio-temporal disease mapping, model (1) was run without covariate for all four possible types of space-time interaction. The model including an interaction term of Type II had the lowest DIC and was chosen as basis model for the ecological regression including cattle trade quantities. All results are summarized in Table 1. The DIC is lowest for the model including the absolute number of in-movements from infected regions with relative risk larger than 2 and a time lag of one year. This is plausible considering the nature of the disease. Models including the absolute number of in-movements are generally preferred compared to models involving the proportion of cattle trade from infected regions. A positive association is obtained for all covariates. For three of the models with a time lag of one year the 95%-credible interval includes only positive values.

As noted in Section 2 the number of in-movements from infected regions is an endogenous covariate. If the pattern of the disease exhibits local clusters, the infected areas chosen by the two-stage process typically are a few groups of neighboring regions. As cattle trade is much larger between neighboring regions, the respective parameters might just explain parts of the local spatial clustering of cases in the data. One hint pointing in this direction is the fact that the estimated variance of the spatially structured effect ψ_i in

TABLE 2. Infectious disease model (see Section 3): The AIC and the estimated coefficients along with their standard errors are shown for each model.

	w_{ji}	AIC	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\zeta}$
(3)	—	1092.8	0.44 (0.05)		
(4)	1	1094.8	0.44 (0.05)	0.0000 (0.0000)	
(4)	1, if $j \sim i$	1062.7	0.43 (0.05)	0.0491 (0.0103)	
(4)	$\sqrt{\text{CM}}$	1082.0	0.44 (0.05)	0.0005 (0.0002)	
(5)	$\sqrt{\text{CM}}$	1083.5	0.44 (0.05)	0.0005 (0.0002)	0.08 (0.11)

(1) drops after inclusion of cattle movement in the model. Unfortunately, it cannot be quantified to what extent such confounding is present.

Results for the infectious disease models are shown in Table 2. With regard to AIC model (4) using $\sqrt{\text{CM}}$ as weights performs better than model (3). Hence, the autoregressive inclusion of counts from other regions weighted by cattle trade provides a better fit. The respective parameter estimate $\hat{\rho}$ is positive (0.0005) with a small standard error (0.0002). In contrast, the alternative model with $w_{ji} = 1$ for all j is not better than (3). If the counts in neighbouring regions $j \sim i$ are considered as additional explanatory variables the AIC is even smaller than for the model including cattle trade and the estimated coefficient is significantly positive (0.0491). Hence, a high local clustering of cases is present. For (5) a positive linear time trend is estimated (0.08), but it is not significantly different from zero.

5 Discussion

Two very different approaches were applied to data on Coxiellosis in cattle to assess the spread of the disease by cattle trade. In both cases model choice criteria and estimated coefficients indicate a positive association between animal movement and disease presence. Nevertheless, both approaches are not without problems. The disease mapping approach makes use of an endogenous covariate which might result in a confounding problem. With regard to the infectious disease model first steps in the direction of a mixture of the parameter- and observation-driven approach are taken when fitting (5). Nevertheless, it would be desirable to include spatial effects as well, especially a spatially structured effect to account for local clustering of the disease. Hence, it must be explored if both approaches could be combined by substituting α in (3) by (1) (without $(\log m_i)$). In this new setting parameter estimation might be possible in a Bayesian framework as algorithms used for maximum likelihood inference (Paul et al., 2008) will possibly suffer from numerical problems. Furthermore, the two approaches are not comparable at the moment as different model choice criteria are derived and different components are included.

Acknowledgments: Financial support by the Swiss Federal Veterinary Office (BVET) is gratefully acknowledged.

References

- Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1-59.
- Clayton, D., Bernardinelli, L. and Montomoli, C. (1993). Spatial correlation in ecological analysis. *International Journal of Epidemiology*, **22**, 1193-1202.
- 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**, 2555-2568.
- Paul, M., Held, L. and Toschke, A. (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine*, **27**, 6250-6267.
- Paul, M. and Held, L. (2009). Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Technical Report, University of Zurich*.
- Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of the Royal Statistical Society, Series B*, **71**, 319-392.
- Schrödle, B. and Held, L. (2009). Spatio-temporal disease mapping using INLA. *Technical Report, University of Zurich*.