

Assessing the impact of network data on the spatio-temporal spread of infectious diseases

Birgit Schrödle¹, Leonhard Held¹ and Håvard Rue²

¹ Division of Biostatistics, Institute for Social and Preventive Medicine, University of Zurich, Switzerland

² Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim

Networks of moving individuals like traded animals between farms represent a potential risk for the spatio-temporal spread of an infectious disease. To assess this relationship we propose two frameworks, namely parameter- and observation-driven models. We discuss both approaches in the context of uni- and multivariate time series of counts with specific emphasis on the direct inclusion of network data. In contrast to observation-driven models, where previous cases are included directly, the disease incidence in a parameter-driven model is governed by a latent stochastic process. We present ready-to-use software based on integrated nested Laplace approximations for inference in parameter-driven models. The predictive performance of both formulations is assessed using proper scoring rules and a score regression approach. The impact of cattle trade on the spatio-temporal spread of Coxiellosis in Swiss cows, 2004-2009, is finally investigated.

Keywords: Infectious disease counts; INLA; Network data; Observation-driven; Parameter-driven; Spatio-temporal.

1 Introduction

Networks of moving individuals represent a potential risk of disease transmission. As an example, cattle trade serves as contact node between infected herds and the ease of transportation can result in the spread of a disease even over long distances (Fèvre et al., 2006). Bigras-Poulin et al. (2007) state that contact structures have received little attention from epidemiologists, although they are very important to understand disease transmission. There are well-known events which illustrate the risk associated with animal trade. Gilbert et al. (2005) quantify the strong association between the movements from infected areas and the breakdown of herds for bovine tuberculosis in Great Britain. Fèvre et al. (2006) report that the foot-and-mouth epidemic in 2001 was spread by animal movement from the North of England to France and the Netherlands. This demonstrates even the risk associated with large-scale movement of animals over national borders.

Observed trade patterns are often analyzed using network analysis, where farms are treated as nodes and the movements of animals are treated as links (Natale et al., 2009).

1 Introduction

Such networks consider the direction of trade and try to identify farms, which have many contacts and are, therefore, hot spots. Furthermore, it is of interest to identify predictors related to having a high number of contacts, e.g. seasonal variations (Nöremark et al., 2011) and dependencies on enterprise type and gender (White et al., 2010). Often, the distribution of the contacts among farms is skewed with many farms having only few contacts while others have many (Bigras-Poulin et al., 2007). Furthermore, medium and large farms have a significantly higher movement activity than small farms. Nöremark et al. (2011) assess the ingoing infection chain which counts the number of ingoing contacts through other farms. However, in such analyses the trade network is never directly related to the actually observed cases. To assess the disease dynamics based on the network, computer simulations can be conducted using a Markov chain model (Natale et al., 2009). Such simulations can give hints on the effects of targeted removal of nodes. However, strong assumptions on the network must be made. Various probabilistic models were proposed in the context of the above mentioned foot-and-mouth epidemic. Green et al. (2006) use simulations to analyze the relationships between cattle trade and the temporal and spatial patterns of pre-detection foot-and-mouth cases. Ferguson et al. (2001) establish a model to simulate the impact of movement restrictions and additional control strategies. Jewell et al. (2009) introduce a Bayesian framework for stochastic transmission models and apply it as well to the foot-and-mouth data.

The importance of travel networks of humans has also been recognized within the last few years. Hufnagel et al. (2004) combine stochastic local infection dynamics among individuals with stochastic transport in a worldwide network to describe the spread of the severe acute respiratory syndrome (SARS). Merler and Ajelli (2010) predict a rapid diffusion of pandemic influenza within Europe because of the high mobility of the population. Paul et al. (2008) assess the spread of influenza in the US by incorporating air traffic information following investigations by Brownstein et al. (2006).

Our aim in this paper is to formulate a spatio-temporal statistical model that can adapt to epidemic outbreaks and directly relate network data to the observed disease cases. In our case study assessing cattle trade, we want to use appropriate model choice criteria to assess, if the assumption of a cattle trade spread mechanism outperforms a purely adjacency-based, local mechanism. We focus on the analysis of so-called aggregated area-level data for I administrative spatial regions over time. The cattle trade network information is available as an unsymmetrical, possibly time-dependent, $I \times I$ matrix containing the absolute number of traded cattle for each pair of regions.

Two different approaches known from the analysis of time series data are proposed in the following. Cox (1981) states a difference between so-called observation- and parameter-driven models. While in observations-driven models past observations or covariates are included directly (e.g. Zeger and Qaqish, 1988; Li, 1994), the dependence between subsequent observations is modelled by a latent stochastic process in parameter-driven models. This approach seems justifiable, if the disease is mostly subclinical.

A parameter-driven model was proposed by Zeger (1988) in a maximum likelihood framework assuming a Poisson distribution for the observed counts. Unfortunately, the integrals contained in the marginal likelihood of such so-called generalized linear mixed models are intractable (Nelson and Leroux, 2006). Inference for similar models in a hierarchical Bayesian setting has been discussed by many authors (Hay and Pettitt, 2001; Gamerman, 1998; Frühwirth-Schnatter and Wagner, 2006). Mugglin et al. (2002) use a hierarchical approach to model influenza epidemic dynamics in time and space. They assess the influence of first and second order neighbours on the disease spread by a vector autoregressive process. Knorr-Held and Richardson (2003) propose a discrete-time model

1 Introduction

incorporating area-specific latent time indicators to distinguish endemic from hyperendemic periods within meningococcal disease cases from France. For point-referenced data, Brix and Diggle (2001) propose a hierarchical log-Gaussian Cox process to monitor possible changes in the space-time incidence pattern of gastrointestinal infections and to predict the variation in the latent intensity. Diggle et al. (2004) review both point-referenced and area-level hierarchical models for infectious disease counts observed in discrete time, and discuss their application within on-line surveillance systems.

Within this paper we propose a general, parameter-driven approach to assess different networks of spatio-temporal disease spread. The building block is a vector autoregressive model including appropriate weights. A major emphasis is on the use of integrated nested Laplace approximations (INLA), a recently proposed method for approximate Bayesian inference in latent Gaussian models (Rue et al., 2009). INLA gives very accurate estimates in short computational time. Its usage is straightforward thanks to an available R package. A toolbox for the implementation of so-called dynamic models within INLA is described in Ruiz-Cardenas et al. (2010). Dynamic or state-space models are a broad class of parametric models where both, parameter variation and available data information, are described in a probabilistic way. Many hierarchical models for uni- and multivariate time series considered so far in the literature fit into this framework and can be fitted directly with the INLA software (www.r-inla.org).

Another class of models, which is an additive mixture of parameter- and observation-driven components, has been proposed in Held, Höhle and Hofmann (2005), denoted by H^3 in the following. Here, the process of infection is modelled by a so-called epidemic part by directly including past observations in the region of interest and/or other regions to describe local epidemics (Paul et al., 2008). Additionally, the linear predictor contains an endemic part, which can consist of temporal and seasonal trends. This model perspective is motivated from a branching process with immigration. An important and very attractive advantage of this formulation is that for simple settings maximum likelihood inference can easily be realized by generic optimization routines, for example the function `optim()` in R (R Development Core Team, 2005). Relevant datasets and algorithms are available in the R package `surveillance`. For multivariate data random intercepts for each region can be included (Paul and Held, 2011); appropriate software is also freely available at <https://r-forge.r-project.org/projects/surveillance>. The evaluation of one-step-ahead predictive forecasts is an important issue for the proposed problem, since it is of major interest to predict the future incidence associated with cattle trade. An evaluation of probabilistic one-step-ahead predictions derived from the parameter-driven and H^3 model can be done using proper scoring rules (Gneiting and Raftery, 2007; Czado et al., 2009). In this paper we chose the squared error, the logarithmic and the ranked probability score, since they can easily be calculated from the INLA output for the parameter-driven and for the H^3 model (Paul and Held, 2011). Furthermore, to assess the calibration of the predictive distributions we apply a score regression approach based on the Dawid-Sebastiani score (Held et al., 2010).

For illustration, both methodologies are first applied to a univariate time series of weekly *Salmonella agona* counts in the UK, 1990-1995. This example is used to clearly state the different properties of observation- and parameter-driven models. The actual data of interest are Coxiellosis cases in Swiss cows from 2004-2009. They are available for 184 Swiss regions and the Principality of Liechtenstein. Coxiellosis is a widespread infectious, bacterial disease which is mostly subclinical. The spread happens from animal to animal by airborne infection. The bacterium can cause an abortion even in a late phase of the pregnancy and is mostly detected by a mandatory screening test due to

such an event. Furthermore, Coxiellosis is a zoonosis and can cause, e.g., Q-fever in humans. We will attempt to assess the spatio-temporal spread of the disease caused by cattle trade. Since it is recently mandatory in Switzerland to notify cattle trade, trade data are available for all Swiss regions for 2009. Cattle trade data for earlier years are not available, but experts stated that the trade pattern has not changed for at least the observed time period.

The rest of this paper is organized as follows: Section 2 provides an introduction to the INLA methodology and to proper scoring rules as building blocks for approximate Bayesian inference in parameter-driven models and model choice. Section 3 discusses univariate *Salmonella agona* counts as an illustrative example. Section 4 deals with multivariate time series models, their extension and implementation within INLA. Furthermore, the proposed models are applied to the Swiss Coxiellosis data. We close with a discussion of the findings in Section 5. The explicit R code is given in the Web Appendix A.

2 INLA, computation and model choice

Integrated nested Laplace approximations are a recently proposed method for approximate Bayesian inference in latent Gaussian models (Rue et al., 2009). The major advantage of INLA in comparison to the widely used MCMC algorithms is that it gives very precise estimates in short computational time. The posterior marginals are approximated directly by INLA and no Monte Carlo samples have to be drawn. The usage of MCMC algorithms for generalized dynamic models such as proposed for time series data is only feasible using data augmentation and requires complex sampling schemes to guarantee efficiency (Gamerman, 1998; Frühwirth-Schnatter et al., 2009). The inclusion of random effects for multivariate spatio-temporal data might require an elaborate re-parametrization of the model (Chib et al., 1998). In contrast, INLA can be run within a user-friendly R environment (R Development Core Team, 2005) using generic functions; all code is freely available from www.r-inla.org and can easily be installed. All analyses within this paper were run using the INLA version built on February 3rd, 2011. The R code for all examples in Sections 3.2 and 4.2 is provided in Web Appendix A. A detailed description of the INLA methodology is not given here, since this was done by various authors and for different applications. See, for example, Schrödle et al. (2011) in the context of spatio-temporal disease mapping models and Riebler et al. (2010) for correlated multivariate age-period-cohort models. It is also shown in both papers that there is excellent agreement between INLA and MCMC results.

A strength of INLA is that measures for model choice can easily be derived from the INLA output. For time series the predictive performance with respect to one-step-ahead predictions is of major interest. One-step-ahead predictions are obtained by successively leaving out the last observation, re-fitting the model and predicting the next count. The performance of probabilistic predictions can be evaluated using proper scoring rules (Gneiting and Raftery, 2007). Scoring rules assign a numerical score based on the predictive distribution and the actually observed count to each prediction and facilitate model comparison and selection. They address sharpness - the concentration of the predictive distribution - as well as calibration - the statistical consistency between the predicted and the later observed probability distributions (Paul and Held, 2011). The concept is elaborated for count data by Czado et al. (2009). Prominent examples of such scores are the squared error score (SES), the logarithmic score (logS) and the ranked probability score (RPS) (Paul and Held, 2011). Each score has specific properties and Czado et al.

3 Univariate time series

(2009) recommend to investigate several scores in applications. The scores used in this paper are defined as

$$\begin{aligned} \text{SES}(P, y) &= (y - \mu_P)^2 \\ \text{logS}(P, y) &= -\log(P(Y = y)) \\ \text{RPS}(P, y) &= \sum_{k=0}^{\infty} (P(Y \leq k) - \mathbf{1}(y \leq k))^2. \end{aligned}$$

Here, P is the predictive probability distribution, μ_P is its first moment and y the truly observed value. If scores are calculated for several predictions (e.g. for several timepoints or regions), the mean score is computed and compared. The smaller the mean score, the better the predictive performance of the model. A formal comparison of the mean scores of two competing models can be conducted by a Monte Carlo permutation test for paired individual scores (Paul and Held, 2011). The proposed scores are used to compare parameter-driven (PM) and H^3 models in Sections 3.2 and 4.2.

To compare the calibration of two predictive distributions, a score regression approach based on the Dawid-Sebastiani score (DSS) can be used (Held et al., 2010). The DSS is based on the first two moments of the predictive distribution (μ_P, σ_P^2). Held et al. (2010) state that if these two moments match the corresponding moments of the data-generating distribution (ideal forecast), the expectation of the DSS depends only on the log-predictive standard deviation. The DSS is defined as

$$\text{DSS}(P, y) = \frac{1}{2} \cdot (\log(\sigma_P^2) + ((y - \mu_P)/\sigma_P)^2).$$

The expectation is $\frac{1}{2} + \log(\sigma_P)$ for an ideal forecast. The smallest possible DSS is $\log(\sigma_P)$. Hence, we regress the individual scores on the logarithm of σ_P by a standard linear regression model

$$\text{DSS}_i = \kappa + \tau \cdot \log(\sigma_{P,i}) + \epsilon_i.$$

For an ideal forecast is $\kappa = \kappa_0 = \frac{1}{2}$ and $\tau = \tau_0 = 1$. To assess the null hypothesis $H_0 : \kappa = \kappa_0$ and $\tau = \tau_0$ we can construct a χ^2 -test with 2 degrees of freedom using an appropriate test statistic (Held et al., 2010).

3 Univariate time series

3.1 Statistical modelling

In parameter-driven models it is assumed that an unobserved autocorrelated mechanism drives the infection process (Zeger, 1988). Consider a model

$$Y_t \sim \text{Po}(\exp(\eta_t)) \quad \text{with } t = 1, \dots, T$$

for a univariate time series of counts, which is formulated in a hierarchical Bayesian framework. The disease risk is modelled by the linear predictor η_t :

$$\begin{aligned} \text{Stage 1: } \eta_t &= \alpha + \zeta_t \\ \text{Stage 2: } \zeta_t &= \lambda \cdot \zeta_{t-1} + \epsilon_t. \end{aligned} \tag{1}$$

This is a so-called dynamic model, where Stage 2 describes the evolution of $\zeta = (\zeta_1, \dots, \zeta_T)^T$. The vector ζ forms an autoregressive process of first order (AR(1)), where $\zeta_1 \sim N(0, \sigma_\zeta^2/(1 - \lambda^2))$. The errors $\epsilon = (\epsilon_1, \dots, \epsilon_T)^T$ are assumed to be iid normally distributed with variance σ_ϵ^2 . The model is stationary, if $|\lambda| < 1$. The parameter α on Stage 1 represents an

3 Univariate time series

Table 1: Mean squared error score ($\overline{\text{SES}}$), mean logarithmic score ($\overline{\log S}$) and mean ranked probability score ($\overline{\text{RPS}}$) for all PM and H^3 models for the Salmonella agona data. The scores are obtained by averaging the scores of one-step-ahead predictions for the last 100 observations. A negative binomial distribution was assumed for the H^3 models.

Model	Autoregressive term	Seasonal term	$\overline{\text{SES}}$		$\overline{\log S}$		$\overline{\text{RPS}}$	
			PM	H^3	PM	H^3	PM	H^3
1	λ	—	4.105	4.249	2.025	2.059	1.116	1.148
2	—	$\sin(\omega t)/\cos(\omega t)$	4.558	4.550	2.173	2.115	1.167	1.183
3	λ	$\sin(\omega t)/\cos(\omega t)$	4.012	4.084	2.011	2.045	1.106	1.126

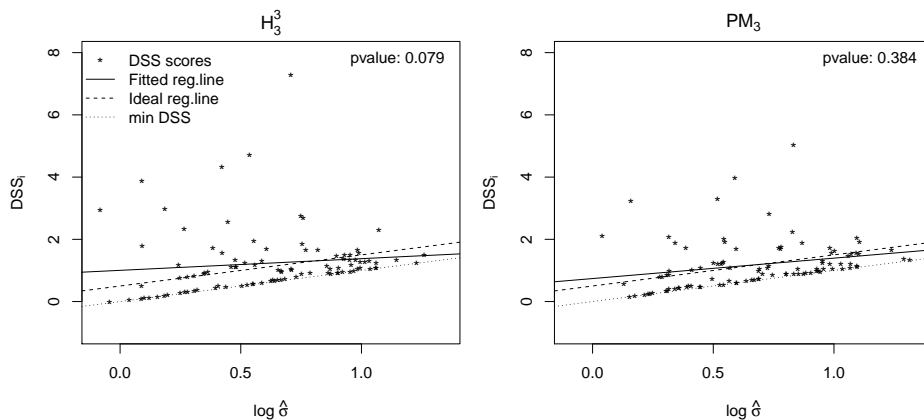


Figure 1: Individual DSS scores, fitted and ideal regression line and minimum DSS score are shown for models H_3^3 (left) and PM_3 (right) based on 100 one-step-ahead forecasts. Additionally, the p -value for a test of miscalibration of the predictive distribution is displayed.

intercept. In a full Bayesian setting hyperpriors for σ_ζ^2 and λ have to be defined. For convenience, the parameter λ is transformed to

$$\lambda^* = \text{logit}((\lambda + 1)/2)$$

using Fisher's z -transformation; λ^* then takes values over the whole real line.

Time series of infectious disease counts often show a seasonal pattern which can be accounted for by expanding model (1) on Stage 1 to

$$\eta_t = \alpha + \zeta_t + \beta \cdot t + \sum_{s=1}^S (\gamma_s \cdot \sin(\omega_s t) + \delta_s \cdot \cos(\omega_s t)). \quad (2)$$

Here, S is the number of harmonics to include and ω_s are the respective Fourier frequencies, for example $\omega_s = 2s\pi/52$ for weekly data.

Several variants of this model will be compared to the H^3 models introduced by Held et al. (2005). Here, the idea is to include the number of cases y_{t-1} in the past directly in the linear predictor to adjust for local epidemics. It is assumed that the counts are Poisson or negative binomially distributed with mean

$$\mu_t = \lambda \cdot y_{t-1} + \exp(\eta_t).$$

The term η_t can be defined equivalently to Stage 1 of the parameter-driven models (1) or (2), but without ζ . This formulation is motivated by a branching process model with immigration and stationary for $0 < \lambda < 1$ (Guttorp, 1995). Given stationarity, the endemic incidence is persistent with a stable temporal, perhaps seasonal pattern. The epidemic incidence will break out occasionally and eventually burn out (Held et al., 2005).

3.2 Salmonella agona in the UK, 1990-1995

As an illustrative example we consider weekly reported cases of Salmonella agona in the UK, 1990-1995. Different models with regard to the autoregressive and seasonal components were fitted, see Table 1. For all seasonal models of type (2) the number of harmonics $S = 1$ was sufficient.

Hyperpriors for the autoregressive and variance components must be defined for the parameter-driven models. We chose a $N(0,5)$ prior for the z -transformed λ^* (Paul et al., 2010). Note that in the INLA setting the hyperprior with regard to the variance of an AR(1) process is imposed on the unconditional variance $\sigma_\zeta^2/(1 - \lambda^2)$. We chose a highly dispersed $IGa(0.1,0.001)$ distribution (Frühwirth-Schnatter and Wagner, 2006). A non-informative $N(0,10^3)$ distribution is used as hyperprior for the intercept α and all fixed effects as an INLA default.

To assess the predictive performance of all models, mean scores of one-step-ahead predictions for the last 100 observations were calculated. Note that the PM models are more complex due to their hierarchical structure. If we count the absolute number of (time-independent) parameters, the PM models have one more than the H^3 models (assuming a Poisson distribution), namely the smoothing variance σ_ζ^2 . Hence, to make way for a fair comparison, the H^3 models were fitted assuming a negative binomial distribution for the response, which contains an additional overdispersion parameter. Nevertheless, the absolute number of parameters does not reflect the true complexity of a hierarchical Bayesian model. An appropriate measure is the effective number of parameters p_D proposed by Spiegelhalter et al. (2002). It is calculated using the deviance of the model and depends on the data. For model PM_3 it grows from around 40 (if the model is fitted to the first 100 observations only) to around 65 (if the model is fitted to all 312 observations). Although the absolute and effective number of parameters cannot be directly compared, this is evidence that the hierarchical model is more complex.

The best model in terms of \overline{SES} , $\overline{\log S}$ and \overline{RPS} is model PM_3 . Model H_3^3 performs best among the H^3 models, but all scores are larger than for the respective PM model. Hence, in this example the predictive performance of the PM models is slightly better. However, note that the observational distribution (i.e. negative binomial distribution) is used as predictive distribution for the H^3 models by plugging in the obtained estimates. This approach ignores parameter uncertainty (Paul and Held, 2011). For the INLA results, the full posterior predictive probability distribution for each forecast is approximated directly when re-running the model. As a consequence, the plug-in predictive distribution might be more narrow than the directly approximated predictive INLA distribution. To assess the impact of these two strategies on the calibration of the predictive distribution, we use a score regression approach based on the DSS. The obtained individual DSS scores for 100 one-step-ahead forecasts and the fitted and ideal regression line are shown in Figure 1 for models H_3^3 and PM_3 . The estimated regression coefficients are $\hat{\kappa} = 1.00$ and $\hat{\tau} = 0.38$ (H_3^3) and $\hat{\kappa} = 0.74$ and $\hat{\tau} = 0.66$ (PM_3), respectively. The p -values for the test of miscalibration are 0.079 (H_3^3) and 0.384 (PM_3). Hence, calibration is better for the parameter-driven model, but the test detects no sig-

4 Multivariate time series

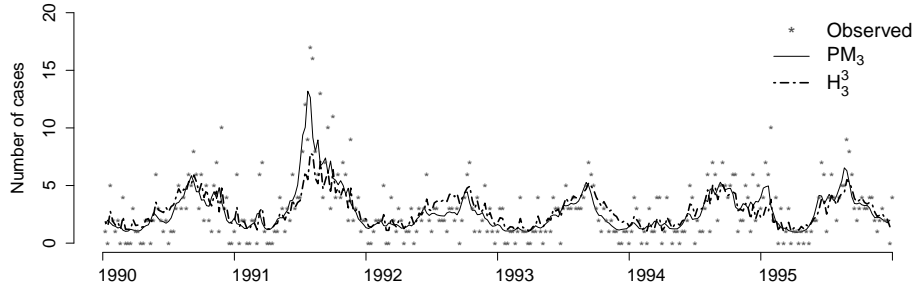


Figure 2: Observed and fitted number of *Salmonella agona* cases. Fitted number of cases are obtained from the best H^3 model (H_3^3 , dashed and dotted line) and the respective parameter-driven model (PM_3 , solid line).

nificant miscalibration of the predictive distribution for both models.

Figure 2 shows the observed and fitted values for models H_3^3 and PM_3 . Apparently, the latent model is more able to adjust for the epidemic peaks. Furthermore, the resulting curve for the parameter-driven model is more smooth.

4 Multivariate time series

4.1 Statistical modelling

Multivariate time series of counts can arise in different contexts, e.g. if time series are available for different age groups, pathogens or geographic regions. We concentrate on spatio-temporal data, but the concepts can also be used in different contexts. Assume that data in region i at time t are available and distributed as

$$Y_{it} \sim \text{Po}(m_{it} \cdot \exp(\eta_{it})) \quad \text{with } t = 1, \dots, T \text{ and } i = 1, \dots, I. \quad (3)$$

The quantity m_{it} is an offset that adjusts for possibly different numbers of exposed individuals in each region and time. A multivariate dynamic model can be written in analogy to model (1) using two stages

$$\begin{aligned} \text{Stage 1: } \eta_{it} &= \alpha + \zeta_{it} \\ \text{Stage 2: } \zeta_{it} &= \lambda \cdot \zeta_{i,t-1} + \epsilon_{it}. \end{aligned} \quad (4)$$

Here, identical replicates of the AR(1) process $\zeta_i = (\zeta_{i,1}, \dots, \zeta_{i,T})^T$ for each region i govern the latent evolution. Such a process is often called a Gaussian multivariate or vector autoregressive process and can alternatively be written as

$$\zeta_t = \Omega \cdot \zeta_{t-1} + \epsilon_t, \quad (5)$$

where $\zeta_t = (\zeta_{1,t}, \dots, \zeta_{I,t})^T$ and $\epsilon_t = (\epsilon_{1,t}, \dots, \epsilon_{I,t})^T$ (Harvey, 1981, 1989; Lütkepohl, 2005). The autoregressive coefficient matrix Ω has dimension $I \times I$. In model (4) the matrix Ω is simply diagonal with entries equal to λ .

In order to model the latent spatial spread of the disease, a weighted sum of the past

4 Multivariate time series

states in other regions j than the region of interest (i) can be included in model (4) on Stage 2:

$$\begin{aligned} \text{Stage 1: } \eta_{it} &= \alpha + \zeta_{it} \\ \text{Stage 2: } \zeta_{it} &= \lambda \cdot \zeta_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot \zeta_{j,t-1} + \epsilon_{it}. \end{aligned} \quad (6)$$

Different choices for the network and, hence, relevant regions j and the respective weights w_{ji} are possible, see also Mugglin et al. (2002). As a first idea all neighbouring regions $i \sim j$ are considered in the ρ term to describe a purely adjacency-driven, local disease spread; in this case $w_{ji} = 1$ for all $i \sim j$ and 0 otherwise. In Section 4.2 the number of traded cattle from region j to region i (and appropriate transformations of this quantity) are used as weights w_{ji} in order to assess the spatio-temporal spread of a cow disease by cattle trade. Model (6) can also be formulated in the form (5), but now Ω also has off-diagonal elements $\rho \cdot w_{ji}$. For the neighbourhood weights the matrix Ω is symmetric. However, it is not symmetric for the cattle trade weights, since the number of cattle traded from region j to region i can be different from the number of cattle traded from region i to region j .

We assume throughout that the cattle trade weights are constant over time, since no time-dependent weights are available for the case study. Since cattle trade also takes place within a region (w_{ii}), we can include this quantity as a fixed weight into the first component on Stage 2, resulting in $\lambda \cdot w_{ii} \cdot \zeta_{i,t-1}$. However, this did not improve the models considered in the case study and is therefore neglected.

One can show that a vector autoregressive AR(1) process is stationary, if all eigenvalues of Ω are less than one in absolute value (Lütkepohl, 2005). However, in practice the region of parameter stationarity cannot be easily derived from this condition (Sun and Ni, 2004). In analogy to the univariate problem, Mugglin et al. (2002) restrict all parameters in the autoregressive coefficient matrix to be in a range of $[-1; 1]$ and assign a Gaussian prior, but, apparently, this transformation does not guarantee stationarity of the multivariate AR(1). For this condition to hold the time series in all regions have to be jointly stationary (Harvey, 1981). Sun and Ni (2004) argue that in the case of no strong prior information the use of reference or noninformative priors is desirable. However, finding suitable reference priors for vector-autoregressive models is almost intractable. For our case study in Section 4.2 we conducted a sensitivity analysis using both Gaussian priors defined on a Fisher z -transformed and unrestricted parameter space.

If the data vary seasonally, a term similar to (2) could be included on Stage 1 of, e.g., model (4). For spatio-temporal data it is also possible to include random components for each region i on Stage 1 of models (4) and (6), if the disease spread is more of an endemic than epidemic nature. These $\boldsymbol{\nu} = (\nu_1, \dots, \nu_I)^T$ random components can be iid normally distributed with mean 0 and variance σ_{ν}^2 . As an alternative an intrinsic conditionally autoregressive (ICAR) effect $\boldsymbol{\psi} = (\psi_1, \dots, \psi_I)^T$ can be assumed, which takes into account the incidence in neighbouring regions (Rue and Held, 2005).

The dynamic model (4) can be fitted in INLA using already implemented standard model specifications, see Web Appendix A. For inference of a system like (6) within INLA an augmented model must be coded, where the column of actual observations is merged with a column of pseudo-observations. These pseudo-observations are 0's obtained from equating the evolution equation on Stage 2 to zero. Hence, the resulting response matrix

contains two columns, namely

$$\begin{bmatrix} y_{11} & \text{NA} \\ \vdots & \vdots \\ y_{I,T} & \text{NA} \\ \text{---} & \text{---} \\ \text{NA} & 0 \\ \vdots & \vdots \\ \text{NA} & 0 \end{bmatrix}$$

where lines $(1, \dots, IT)$ correspond to the actual observations and lines $(IT + 1, \dots, 2IT)$ correspond to the pseudo-observations. Different likelihoods are assumed for the observed counts and pseudo-observations (here: Poisson and Gaussian distribution). The R code for the case study in Section 4.2 is given in Web Appendix A. In the augmented model the parameters λ and ρ are treated as so-called scaling factors for the respective realizations of the process ζ . A detailed description of the provided toolbox and several similar examples are given in Ruiz-Cardenas et al. (2010) and on <http://www.r-inla.org/models/tools>.

The idea of the proposed models is very similar to H^3 models for multivariate time series (Paul and Held, 2011). In the easiest case, the counts y_{it} are Poisson or negative binomially distributed with mean

$$\mu_{it} = \lambda \cdot y_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot y_{j,t-1} + m_{it} \cdot \exp(\eta_{it}). \quad (7)$$

An inclusion of weights w_{ii} is possible, but currently not implemented in the available software for H^3 models. Analog to the univariate case this model can be written as a multivariate branching process with immigration

$$\boldsymbol{\mu}_t = \Lambda \cdot \mathbf{y}_{t-1} + \mathbf{m}_t \cdot \exp(\boldsymbol{\eta}_t)$$

with suitable defined column vectors $\boldsymbol{\mu}_t$, \mathbf{y}_{t-1} , \mathbf{m}_t and $\boldsymbol{\eta}_t$. The matrix Λ contains the λ 's as diagonal and the $\rho \cdot w_{ji}$'s as off-diagonal elements. If the largest eigenvalue of Λ is smaller than unity, the process is ergodic (Held et al., 2005). In analogy to the parameter-driven model the term $\boldsymbol{\eta}_t$ can contain temporal and seasonal components. Furthermore, Paul and Held (2011) discuss the inclusion of regional random effects of iid (ν) or ICAR (ψ) type into $\boldsymbol{\eta}_t$. In their framework the autoregressive parameters λ and ρ could also be modelled as random effects and vary from region to region. However, this is not possible in the INLA setting.

4.2 Coxiellosis in Swiss cows, 2004-2009

The methodology is applied to data on Coxiellosis incidence on Swiss farms. As noted before, the data are available from 2004 to 2009 for 184 Swiss regions and the Principality of Liechtenstein. A herd is denoted a case, if at least one diseased animal was detected. The number of herds per region is included as an offset m_i in equation (3). Coxiellosis incidence is quite low, with a mean number of 47 cases per year for the observed time period. Hence, the data were yearly aggregated. Such a coarse aggregation can be justified by the fact that the date of confirmation of a case by a laboratory, which is available for our case study, does not directly coincide with the date of an infection. Often the animals have been infected some time before, but the disease is not detected until an abortion takes place. Hence, reporting delay and underreporting are likely to

4 Multivariate time series

Table 2: Total number of pairs of regions and number of pairs with no cattle trade. Furthermore, summary statistics for pairs of regions with cattle trade are displayed.

# of pairs	# 0's	Min	1st Qu	Median	Mean	3rd Qu	Max
34039	17060	1	3	11	49	37	4768

Table 3: Mean squared error score ($\overline{\text{SES}}$), mean logarithmic score ($\overline{\log S}$) and mean ranked probability score ($\overline{\text{RPS}}$) for all PM and H^3 models for the Coxiellosis data. The scores are obtained by averaging over the scores of a one-step-ahead prediction for 2009 for all 185 regions.

Model index	w_{ji}	$\overline{\text{SES}}$		$\overline{\log S}$		$\overline{\text{RPS}}$	
		PM	H^3	PM	H^3	PM	H^3
1	—	0.611	0.645	0.583	0.624	0.239	0.257
2	$i \sim j$	0.505	0.603	0.547	0.593	0.218	0.246
3	CT_{ji}	0.494	0.527	0.549	0.590	0.214	0.236
4	CT_{ji}/n_j	0.518	0.540	0.554	0.583	0.218	0.234
5	$\sqrt{CT_{ji}}$	0.644	0.638	0.557	0.619	0.217	0.255
6	$\log(CT_{ji} + 1)$	0.750	0.645	0.575	0.624	0.230	0.257
7	$(CT_{ji}/n_j) \cdot b_j$	0.444	0.527	0.549	0.578	0.212	0.232

be present. To ensure that the temporal aggregation does not dilute any important facts, we re-ran all analyses discussed below using half-yearly and quarterly aggregated data, but the conclusions did not change. However, not all parameter-driven models did converge for half-yearly and quarterly data. As a consequence of the yearly aggregation, seasonality is not modelled for this case study.

A first question to be answered is, if any spatial spread of the disease along a network takes place. Hence, we compare models of type (4) (no spatial spread) with models of type (6) (spatial spread) using predictive scores. Furthermore, different spread mechanisms, namely adjacency-based weights (model index 2) and cattle trade weights, are compared.

Descriptive statistics for the entries of the cattle trade matrix can be found in Table 2. There is no cattle trade between about half of all possible pairs of regions. For those pairs of regions with cattle trade the quantity is very skewed. Hence, a transformation of the cattle trade counts (CT) might be useful. Ideally, a data-driven transformation $f(CT)$ is preferred, but such an algorithm is not yet available. So we compared different choices (see Table 3, model index 3 to 6) based on the one-step-ahead forecast scores, which are discussed below.

Note that hyperpriors for all variance and autoregressive components must be defined for the PM models. For model PM_1 , a z -transformed λ^* is used in INLA as a default. The three proposed hyperpriors in Paul et al. (2010) ($N(0,1.25)$, $N(0,2.5)$, $N(0,5)$) did not result in different posterior marginals and we chose an $N(0,5)$ for this analysis. In analogy to the univariate case a highly dispersed $\text{IGa}(0.1,0.001)$ prior is chosen for the unconditional variance of the $\text{AR}(1)$ process. For models of type (6), a sensitivity analysis concerning the hyperprior choice for λ and ρ was conducted. As discussed above, we

4 Multivariate time series

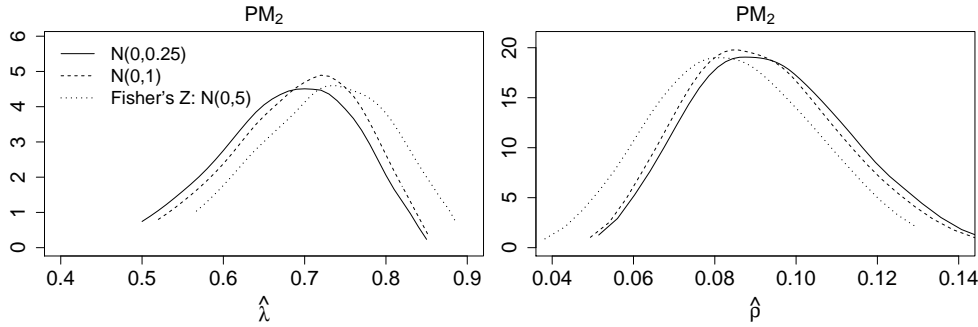


Figure 3: Posterior marginals of λ and ρ for model PM_2 (neighbourhood weights) for three different hyperpriors: an $N(0,0.25)$ (solid) and an $N(0,1)$ (dashed) prior for λ and ρ and an $N(0,5)$ prior for a Fisher's z -transformed λ^* and ρ^* (dotted).

use adjacency-based and cattle trade weights. To guarantee that the chosen hyperpriors mean the same for all weight schemes, the (transformed) cattle trade numbers were standardized in such a way that they are 1 on average. As first and second hyperprior option we chose a $N(0,0.25)$ and $N(0,1)$ prior for λ and ρ . As third to fifth option we chose the three above mentioned priors for a z -transformed λ^* and ρ^* . Small differences in the posterior marginals can be detected for all five hyperpriors. This is probably due to the sparseness of the data. As an example, the results for prior options 1 to 3 for model PM_2 are shown in Figure 3. Prior options 4 and 5 are omitted from the plot, since their results are very similar to those of option 3. The z -transformed priors resulted in slightly better predictive scores than the unrestricted priors. However, small differences in the hyperparameter estimates do not affect the predictive scores in such a way that the following model comparison is falsified. To guarantee a fair comparison with models of type (4), all following results are calculated using a z -transformed $N(0,5)$ prior.

Following Paul et al. (2008), we use the absolute (CT_{ji}) and relative number of traded cattle (CT_{ji}/n_j) as weights in model 3 and 4, respectively, see Table 3; here n_j denotes the mean number of cattle in region j . Since the cattle trade counts are skewed, a square root ($\sqrt{CT_{ji}}$) and log transformation ($\log(CT_{ji} + 1)$) might also be reasonable (models 5 and 6). Models 1 to 6 are fitted using a PM and H^3 approach. A negative binomial distribution was assumed for the H^3 models. To assess the predictive performance of a model, a multivariate one-step-ahead prediction for all regions for 2009 is made and the mean score over all 185 regions is computed. For the PM models it turns out that the best overall predictive performance taking into account all three scores is obtained using the absolute cattle trade counts as weights (model PM_3). Note that the absolute cattle trade counts also slightly outperform the adjacency-based weights. Among the H^3 models the relative cattle trade performs slightly better than the absolute CT counts for the logarithmic and ranked probability score. The respective scores of model H_4^3 exhibit a better predictive performance than for the neighbourhood weights (H_2^3). The models without spatial spread (PM_1 , H_1^3) are clearly outperformed.

An extension of the above discussed models is examined. The mean herd size per region might be a risk factor for the disease spread by cattle trade across regions. The larger a herd, the larger is the risk for at least one subclinically diseased animal. Furthermore, Nöremark et al. (2011) identified herd size as a significant predictor of increased cattle trade activity. In Switzerland, the mean herd size per region b_j varies between about 10 and 60 animals. Hence, we estimated a model with $(CT_{ji}/n_j) \cdot b_j$ as weight w_{ji} , to asses

the effect of the mean herd size in the origin trade region j multiplied by the relative cattle trade on the spatial spread of the disease (model PM_7). For models PM_7 and H_7^3 the scores even slightly improved in comparison to models PM_3 and H_4^3 .

The variation of the mean scores across different models is generally stronger for the SES than for the logS and RPS, particularly for the PM models. Note that the three scores have different properties. The logS and RPS depend on the full predictive distribution, while the SES depends only on its first moment. The large differences in mean SES between some of the models are caused by only a few observations with different predictive means, to which the mean score reacts severely. In contrast, the logS and RPS appear to be more robust, because the whole predictive distribution is taken into account. The RPS varies less than the logS, which confirms that it is more robust to slight changes in the predictive distribution (Gneiting and Raftery, 2007).

To finally answer the question, if the best transformed cattle trade weights outperform the local adjacency-based weights, we conducted a permutation test for each score and model type. However, the p -values of the permutation tests of PM_2 vs. PM_7 ($p(\overline{\text{SES}}) = 0.48$, $p(\overline{\text{logS}}) = 0.77$, $p(\overline{\text{RPS}}) = 0.45$) and H_2^3 vs. H_7^3 ($p(\overline{\text{SES}}) = 0.18$, $p(\overline{\text{logS}}) = 0.19$, $p(\overline{\text{RPS}}) = 0.06$) do not exhibit a significant difference. But still, the transformed cattle trade weights are at least as good as the adjacency-based weights for the PM and H^3 framework. The differences are larger for the H^3 models.

A further criterion as regards the comparison of the best cattle trade and adjacency-based weights are the p -values of the score regression test for calibration. The respective p -values for the PM and H_3 models are 0.518 (PM_2) vs. 0.738 (PM_7) and 0.424 (H_2^3) vs. 0.571 (H_7^3). Hence, the calibration of the predictive distribution is slightly better for the best cattle trade weights in both modelling frameworks.

In the following the predictive performance of the PM and H^3 models will be compared. All mean scores obtained for the PM models are smaller than for the respective H^3 model. A permutation test comparing the best PM and H^3 models PM_7 and H_7^3 gives the following p -values: $p(\overline{\text{SES}}) = 0.50$, $p(\overline{\text{logS}}) = 0.02$ and $p(\overline{\text{RPS}}) = 0.10$. Hence, as regards the logarithmic score there is some evidence that the PM model performs significantly better.

We finally discuss possible reasons for the differences in the mean scores of models PM_7 and H_7^3 . For a closer investigation, we looked at the mean scores stratified by the number of actually observed counts per region. For each score we observed the same pattern: There are no cases for 150 regions and the difference between the PM and H^3 scores is essentially 0. Hence, only the remaining 35 regions with 1 to 6 observed cases cause the differences in the mean scores. The larger the observed count, the larger is the difference between the PM and H^3 scores. Hence, the PM model seems to perform better in predicting higher counts.

A different behavior of the PM and H^3 models can be observed, when iid (ν) or ICAR random effects (ψ) are included on Stage 1 of model (6) or into η_{it} of model (7), respectively. For the PM models in our case study such iid or ICAR random effects are estimated to be zero and the obtained scores from the Table 3 do change only minimally. However, if one such random effect is included into an H^3 model and the weights w_{ji} are chosen as in, e.g., model H_7^3 , the parameters $\hat{\lambda}$ and $\hat{\rho}$ are estimated to be zero and all variability between the regions is explained by the spatial random effect ν or ψ . The predictive scores obtained for models with an additional random effect are higher than for models including a weight scheme but no random effects (e.g. for inclusion of ν : $\overline{\text{SES}} = 0.607$, $\overline{\text{logS}} = 0.603$ and $\overline{\text{RPS}} = 0.239$, cf. Table 3). This behavior might be due to the fact that the data are very sparse. Hence, there are only a few yearly transitions

5 Discussion

Table 4: The estimated parameters $\hat{\lambda}$ and $\hat{\rho}$ for models PM_2 , PM_7 , H_2^3 and H_7^3 for the Coxiellosis data.

Model	w_{ji}	$\hat{\lambda}$ (95%-CI)	$\hat{\rho}$ (95%-CI)
PM_2	$i \sim j$	0.73 (0.58;0.87)	0.0845 (0.0484;0.1229)
PM_7	$(CT_{ji}/n_j) \cdot b_j$	0.78 (0.67;0.83)	0.0048 (0.0028;0.0072)
H_2^3	$i \sim j$	0.38 (0.24;0.60)	0.0498 (0.0283;0.0876)
H_7^3	$(CT_{ji}/n_j) \cdot b_j$	0.36 (0.23;0.56)	0.0302 (0.0193;0.0470)

with observed cases for both successive years present in the data, which are needed to estimate the parameters λ and ρ in the H^3 framework. So, if random effects are included additionally, they dominate the estimation process.

For completeness the estimated autoregressive parameters for some of the competing models are displayed in Table 4. In the PM context the estimates $\hat{\lambda}$ are generally larger than for the H^3 models. Note that the $\hat{\rho}$'s cannot be compared directly, since the weights are of a different absolute size.

5 Discussion

It was shown in this paper, how parameter- and observations-driven models for time series of infectious disease counts can be applied to network data using ready-to-use software. An emphasis is on the usage of the INLA method for approximate Bayesian inference in parameter-driven models, which can be conducted using a toolbox for generalized dynamic models.

The proposed parameter-driven models can be compared to so-called H^3 models on the basis of proper scoring rules, which are a tool to evaluate the predictive performance in terms of one-step-ahead predictions. In both case studies the parameter-driven models give better mean scores. Furthermore, the PM models tend to give better scores, if the observed number of cases is relatively large. A possible reason for this is their greater flexibility due to the complex hierarchical model structure. Unfortunately, the difference in model complexity between both approaches can hardly be quantified. A disadvantage of the parameter-driven models discussed here is that the autoregressive parameters cannot be modelled as random effects and, hence, are the same for all regions in the study area. This restriction is not made by the H^3 models.

For the Coxiellosis data we found that models assuming a spatial spread between neighbouring regions or by cattle trade exhibit a better predictive performance than models using an autoregressive component within a region only. The spatial spread is modelled best including the relative number of traded cattle weighted by the mean herd size in the origin region of the traded cattle as spread network. Hence, the cattle trade slightly outperforms a purely local, adjacency-driven spread mechanism. The mean herd size in a region elevates the risk of a spatial spread of the disease by cattle trade.

For future work it is desirable to estimate a data-driven transformation of the cattle trade weights. Furthermore, the results might be more precise, if time-varying cattle trade data are available. As an additional issue, Nöremark et al. (2011) state that between-country comparisons or the inclusion of data from adjacent countries can be helpful to parametrize disease spread models. Unfortunately, there are often differences in the data structure between national registries.

Acknowledgements

Financial support by the Swiss Federal Veterinary Office (BVET) is gratefully acknowledged. Special thanks to my colleague Michaela Paul for sharing her expertise about H^3 models.

References

- Bigras-Poulin, M., Barfod, K., Mortensen, S., and Greiner, M. (2007). Relationship of trade patterns of the Danish swine industry animal movements network to potential disease spread. *Preventive Veterinary Medicine* **80**, 143–165.
- Brix, A. and Diggle, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society - Series B* **63**, 823–841.
- Brownstein, J. S., Wolfe, C. J., and Mandl, K. D. (2006). Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Medicine* **3**, e401.
- Chib, S., Greenberg, E., and Winkelmann, R. (1998). Posterior simulation and Bayes factors in panel count data models. *Journal of Econometrics* **86**, 33–54.
- Cox, D. R. (1981). Statistical analysis of time series: some recent developments (with discussion and reply). *Scandinavian Journal of Statistics* **8**, 93–115.
- Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics* **65**, 1254–1261.
- Diggle, P. J., Knorr-Held, L., Rowlingson, B., Su, T., Hawtin, P., and Bryant, T. N. (2004). On-line monitoring of public health surveillance data. In Brookmeyer, R. and Stroup, D., editors, *Monitoring the health of populations*. Oxford University Press.
- Ferguson, N. M., Donnelly, C. A., and Anderson, R. M. (2001). The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science* **292**, 1155–1160.
- Fèvre, E. M., de C. Bronsvoort, B. M., Hamilton, K. A., and Cleaveland, S. (2006). Animal movements and the spread of infectious diseases. *Trends in Microbiology* **14**, 125–131.
- Frühwirth-Schnatter, S., Frühwirth, R., Held, L., and Rue, H. (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing* **19**, 479–492.
- Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state-space modelling. *Biometrika* **93**, 827–841.
- Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika* **85**, 215–227.
- 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.

References

- Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**, 359–378.
- Green, D. M., Kiss, I. Z., and Kao, R. R. (2006). Modelling the initial spread of foot-and-mouth disease through animal movements. *Proceedings of the Royal Society B: Biological Sciences* **273**, 2729–2735.
- Guttorp, P. (1995). *Stochastic modeling of scientific data*. Chapman & Hall, London.
- Harvey, A. C. (1981). *Time series models*. Phillip Alan, Oxford.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. University Press, Cambridge.
- Hay, J. L. and Pettitt, A. N. (2001). Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics* **2**, 433–444.
- 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.
- Held, L., Rufibach, K., and Balabdaoui, F. (2010). A score regression approach to assess calibration of continuous probabilistic predictions. *Biometrics* **66**, 1295–1305.
- Hufnagel, L., Brockman, D., and Geisel, T. (2004). Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15124–15129.
- Jewell, C. P., Kypraios, T., Neal, P., and Roberts, G. O. (2009). Bayesian analysis for emerging infectious diseases. *Bayesian Analysis* **4**, 465–496.
- Knorr-Held, L. and Richardson, S. (2003). A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Applied Statistics* **5**, 169–183.
- Li, W. (1994). Time series models based on generalized linear models: some further results. *Biometrics* **50**, 506–511.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer, Berlin.
- Merler, S. and Ajelli, M. (2010). The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society B: Biological sciences* **277**, 557–565.
- Mugglin, A. S., Cressie, N., and Gemmell, I. (2002). Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Statistics in Medicine* **21**, 2703–2721.
- Natale, F., Armando, G., Savini, L., Palma, D., Possenti, L., Fiore, G., and Calstri, P. (2009). Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Preventive Veterinary Medicine* **92**, 341–350.
- Nelson, K. P. and Leroux, B. G. (2006). Statistical models for autocorrelated data. *Statistics in Medicine* **25**, 1413–1430.
- Nöremark, M., Håkansson, N., Sternberg Lewerin, S., Lindberg, A., and Jonsson, A. (2011). Network analysis of cattle and pig movements in Sweden: Measures relevant for disease control and risk based surveillance. *Preventive Veterinary Medicine* DOI: 10.106/j.preventmed.2010.12.009.

References

- Paul, M. and Held, L. (2011). Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Statistics in Medicine* DOI: 10.1002/sim.4177.
- Paul, M., Held, L., and Toschke, A. M. (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine* **27**, 6250–6267.
- Paul, M., Riebler, A., Bachmann, L. M., Rue, H., and Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Statistics in Medicine* **29**, 1325–1339.
- R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Riebler, A., Held, L., and Rue, H. (2010). Correlated multivariate age-period-cohort models. Technical report, University of Zurich, Switzerland.
- Rue, H. and Held, L. (2005). *Gaussian Markov random fields*. Chapman & Hall/CRC, London.
- 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.
- Ruiz-Cardenas, R., Krainski, E. T., and Rue, H. (2010). Fitting dynamic models using integrated nested Laplace approximations - INLA. Technical report, NTNU Trondheim, Norway.
- Schrödle, B., Held, L., Riebler, A., and Danuser, J. (2011). Using INLA for the evaluation of veterinary surveillance data from Switzerland: A case study. *Journal of the Royal Statistical Society - Series C* **60**, 261–279.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society - Series B* **64**, 583–639.
- Sun, D. and Ni, S. (2004). Bayesian analysis of vector-autoregressive models with non-informative priors. *Journal of Statistical Planning and Inference* **121**, 291–309.
- White, P., Frankena, K., O’Keeffe, J., More, S. J., and Martin, S. W. (2010). Predictors of the first between-herd animal movement for cattle born in 2002 in Ireland. *Preventive Veterinary Medicine* **97**, 264–269.
- Zeger, S. L. (1988). A regression model for time series counts. *Biometrika* **75**, 621–629.
- Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics* **44**, 1019–1031.