Predictive assessment of a non-linear random effects model for space-time surveillance data

Michaela Paul, Leonhard Held
Department of Biostatistics, Institute of Social and Preventive Medicine,
University of Zurich

Abstract: Notification data collected by national surveillance systems are typically available as weekly time series of counts of confirmed new cases, stratified e.g. by geographic areas. This work outlines the statistical modeling framework in Paul and Held (2011) for the analysis of such data. Inherent (spatio-)temporal dependencies are incorporated via an observation-driven formulation. Using region-specific and possibly spatially correlated random effects, we are able to address heterogeneous incidence levels. Inference is based on penalized likelihood methodology for mixed models. The predictive performance of models is assessed using probabilistic one-step-ahead predictions and proper scoring rules.

Keywords: Time series of counts, infectious diseases, proper scoring rules.

1 Introduction

Notification data on infectious diseases typically consist of counts of confirmed new infections, which are observed in defined geographical areas at regular time intervals. Retrospective surveillance aims to identify outbreaks and (spatio-)temporal patterns through statistical modeling. Motivated by a branching process with immigration, Held et al. (2005) propose to decompose the mean incidence additively into three components: an *autoregressive*, a *neighbor-driven* and an *endemic* component. The first two components represent an autoregression on past counts in the same and in other regions, respectively, and should capture occasional outbreaks and dependencies across regions. The third component parametrically models regular trends and seasonal variation, e.g. by a sine-cosine formulation. Overdispersion can be allowed for by replacing the Poisson with a negative binomial distribution.

In the case of spatially correlated time series, the assumption of equal disease transmission or incidence levels across all regions is questionable. For instance, transmission might be influenced by age, vaccination status, or environmental conditions. Such factors could be incorporated into the model as covariates if suitable information is available. As an alternative, Paul and Held (2011) suggest to include regional random effects to allow for heterogeneity across regions. The predictive quality of the models is then investigated using one-step-ahead predictions and proper scoring rules (Gneiting and Raftery, 2007).

2 Methods

2.1 Modeling framework

Let y_{rt} denote the number of cases of a specific disease in region $r = 1, \ldots, R$ at time $t = 1, \ldots, T$. The counts are assumed to be Poisson or negative binomially distributed with conditional mean

$$\mu_{rt} = \lambda_r y_{r,t-1} + \phi_r \sum_{q \neq r} w_{qr} y_{q,t-1} + e_{rt}\nu_{rt}\,, \tag{1}$$

where $\lambda_r, \phi_r, \nu_{rt} > 0$ are unknown quantities, w_{qr} are suitably chosen known weights and e_{rt} corresponds to an offset (e.g. population numbers). A simple choice for the weights is $w_{qr} = 1$ if units q and r are adjacent and 0 otherwise.

The three unknown quantities are further decomposed additively on the log-scale and specified for example as

$$\log(\lambda_{rt}) = \alpha_0 + a_r \tag{2}$$

$$\log(\phi_{rt}) = \beta_0 + b_r \tag{3}$$

$$\log(\nu_{rt}) = \gamma_0 + c_r + \gamma_1 \sin(2\pi/52\, t) + \gamma_2 \cos(2\pi/52\, t) \tag{4}$$

where $\alpha_0, \beta_0, \gamma_0$ are intercepts, a_r, b_r, c_r are regional random effects, and the terms in curly brackets in (4) define the model seasonal variation. In applications, each of the three components may be suitably modified or omitted.

The stacked vector of all random effects is assumed to follow a normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. For instance, one may choose $\mathbf{\Sigma} = \mathbf{\Omega} \otimes \mathbf{I}$, where $\mathbf{\Omega}$ is an unknown 3×3 covariance matrix, and $\mathbf{I}$ is the $R \times R$ identity matrix. This formulation correlates the random effects (a_r, b_r, and c_r) between components, and leaves the random effects within each component (e.g., $\mathbf{c} = (c_1, \ldots, c_R)^\top$) uncorrelated.

In hierarchical models for spatio-temporal data, it is often reasonable to assume spatially correlated random effects rather than independent and identically distributed (iid) ones. Therefore, one might also adopt an intrinsic conditional autoregressive (ICAR) model (Besag et al., 1991) for the incidence levels $\mathbf{c}$, say. As the associated precision matrix has a rank deficiency of one, we apply a transformation $\mathbf{c} = \gamma_0 + \mathbf{Z}\tilde{\mathbf{c}}$ and estimate a reduced set of $R - 1$ random effects, $\tilde{\mathbf{c}}$, that are iid Gaussians (see Paul and Held, 2011).

The estimation of parameters involves integration of the likelihood with respect to the random effects which cannot be done analytically. Paul and Held (2011) suggest a penalized likelihood approach for inference, where variance components are treated as known when estimating the fixed and random effects. The variance components themselves are estimated through maximizing the approximated marginal likelihood obtained via a Laplace approximation.

2.2 Predictive model assessment

Model choice based on classical information criteria such as AIC is well explored and understood for models that correspond to fixed-effects likelihoods. However, their use can be problematic in the presence of random effects (Burnham and Anderson, 2002, p. 316). For model selection in time series models, the comparison of successive one-step-ahead predictions with the actually observed data is especially attractive. The often used mean squared error of several point predictions does not take prediction uncertainty into account. Instead, Gneiting and Raftery (2007) recommend the use of strictly proper scoring rules to evaluate probabilistic predictions in the form of a predictive distribution.

Strictly proper scoring rules simultaneously measure the sharpness and calibration of a prediction by assigning a numerical score based on a stated predictive distribution and the later observed actual value. The smaller the score, the better the predictive quality. Several proper scoring rules for count data are discussed by Czado et al. (2009). A popular scoring rule is the logarithmic score

$$\text{logS} = -\log(P(Y = y)) \tag{5}$$

which corresponds to the log predictive density at the observed value y. It is highly sensitive to extreme cases as it strongly penalizes low probability events. A more robust alternative is the ranked probability score

$$\text{RPS} = \sum_{k=0}^{\infty} \Big(P(Y \leq k) - 1(y \leq k) \Big)^2, \tag{6}$$

where 1 is the indicator function.

Typically, mean scores over a set of predictions are used to rank and compare different models informally or via tests such as a Monte Carlo permutation test for paired observations (see Paul and Held, 2011).

3 Case study

In a case study, Paul and Held (2001) applied the model to weekly influenza surveillance counts in 140 districts of Southern Germany for the years 2001–2008. Data were obtained from the SurvStat database of the Robert Koch Institute and analyzed using the functions implemented in the R package `surveillance` (Höhle, 2007). Exemplary R code to reproduce the analysis is given in the package vignette available at `https://r-forge.r-project.org/projects/surveillance/`.

The negative binomial model which yielded the lowest average logarithmic score, called 'B2', was specified by $\log(\lambda_{rt}) = \alpha_0$, $\log(\phi_{rt}) = \beta_0 + b_r$, and $\log(\nu_{rt}) = \gamma_0 + c_r + \gamma_1 t + \sum_{s=1}^{3} \gamma_{2s} \sin(2\pi s/52\, t) + \gamma_{2s+1} \cos(2\pi s/52\, t)$, where $(\mathbf{b}^\top, \mathbf{c}^\top)^\top \sim \text{N}(\mathbf{0}, \boldsymbol{\Omega} \otimes \mathbf{I})$ with $\boldsymbol{\Omega} = \begin{pmatrix} \sigma_b^2 & \rho\sigma_b\sigma_c \\ \rho\sigma_b\sigma_c & \sigma_c^2 \end{pmatrix}$. Here we consider a further model 'S', where the autoregressive

Model	$\overline{\text{logS}}$	$\overline{\text{RPS}}$
B2: with seasonal variation in (4)	0.5633	0.4363
S: with seasonal variation in (2) and (4)	0.5571	0.4224

Table 1: Average scores based on $140 \cdot 104$ one-step-ahead predictions.

component (2) additionally contains $S = 1$ seasonal terms. Average scores for this model, based on one-step-ahead predictions for years 2007–2008, can be found together with the scores for model B2 in Table 1.

4 Concluding remarks

The analysis showed that the predictive performance improves when the autoregressive parameter is also allowed to vary over time. In Paul and Held (2011), the inclusion of spatially correlated random incidence levels instead of iid ones did not substantially improve the predictive performance of a model which already incorporated spatio-temporal correlation via the neighbor-driven component.

References

Besag J., York J., Mollié A. (1991) Bayesian image restoration with two applications in spatial statistics, *Annals of the Institute of Statistical Mathematics*, 43, 1–20.

Burnham K. P., Anderson D. R. (2002) *Model Selection and Multimodel Inference. A Practical InformationTheoretic Approach*, Springer, New York.

Czado C., Gneiting T., Held L. (2009) Predictive model assessment for count data, *Biometrics*, 65, 1254–1261.

Gneiting T., Raftery A. E. (2007) Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association*, 102, 359–378.

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

Höhle M. (2007) Surveillance: an R package for the monitoring of infectious diseases, *Computational Statistics*, 22, 571–582.

Paul M., Held L. (2011) Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts, *Statistics in Medicine*, 30, 1118–1136.