Bayesian modelling of space-time interactions on the Lexis diagram

Volker Schmid and Leonhard Held*
Ludwig-Maximilians-Universität München
Germany

First version: July 2003

Abstract

We propose a full model-based framework for a statistical analysis of incidence or mortality count data stratified by age, period and space, with specific inclusion of additional cohort effects. The setup will be fully Bayesian based on a series of Gaussian Markov random field priors for each of the components. Additional space-time interactions will be either modelled as space-period or space-cohort effects. Statistical inference is based on efficient algorithms to block update Gaussian Markov random fields, which have recently been proposed in the literature. We illustrate our approach in an analysis of stomach cancer data in West Germany.

Key words: Block updating; disease mapping; hierarchical models; Markov chain Monte Carlo; Markov random field models; age-period-cohort model; space-time interaction

*formerly Knorr-Held. Address for correspondence: Leonhard Held, Department of Statistics, Ludwig-Maximilians-Universität München, 80539 Munich, Ludwigstr. 33, Germany. Email: `leo@stat.uni-muenchen.de`

1 Introduction

Model-based statistical analyses of vital rates on the Lexis diagram has received much interest over the last two decades. Much of the early work has been done within a likelihood framework, see Holford (1983) and Clayton and Schifflers (1987a, b). The most general of the approaches proposed in these papers is the so-called age-period-cohort model. Later Bayesian approaches have been suggested within a hierarchical model, which has the advantage that they can incorporate additional parameters for unstructured variation, in order to account for overdispersion. Also they can impose a temporal structure on the age, period and cohort parameters which improves estimation properties and facilitates prediction of future rates (Berzuini, Clayton and Bernardinelli, 1993, Besag *et al.,* 1995, Knorr-Held and Rainer, 2001).

For purely spatial data, vital rates are typically analysed within the Bayesian framework assuming so-called Markov random field priors, see Besag *et al.* (1991) and Clayton and Bernardinelli (1992). Such models overcome problems with heterogeneity in the underlying population counts and provide a spatially smoothed version of the crude rates, if there is evidence for a spatial pattern. Alternatively or additionally, unstructured (spatial) variation can be incorporated as well.

For a full model-based space-time analysis of vital rates only preliminary suggestions have been made so far. Bernardinelli *et al.* (1995) suggest a model which already incorporates space-time interaction, assuming linear time trends. They work with age-standardized rates, which allows to ignore age- and cohort effects in the analysis. However, the uncertainty in the corresponding estimates is neglected in such a two-step analysis; a further practical problem is the choice of the reference rates for standardization. Assuncao *et al.* (2001) recently suggested an extension of the Bernardinelli *et al.* formulation with quadratic instead of linear time trends.

Knorr-Held (2000) suggested four different types of interaction for space-time data, which allow for a nonparametric estimation of the temporal trends, not only of the spatial pattern. However, again, the model does not incorporate any age or cohort effects. A different route has been taken by Knorr-Held and Besag (1998) who incorporate (time-changing) age effects, which allows in principle for estimation of any age-period interaction, including cohort effects. However, their model assumes that space and time are separable, hence does not include any space-time interaction.

Recently the proposal made in Knorr-Held (2000) has been used by Lagazio *et al.* (2001) in an analysis of lung cancer rates in Tuscany. They found the importance of modelling interactions of cohort parameters with space, which gave a better fit to the data considered than interactions of period with space. Again, not a full model was presented but the data analysed was standardized by age in advance.

The goal of this paper is to provide a framework for a full model-based analysis of data stratified by age, period and space, with specific inclusion of cohort effects. The setup will be fully Bayesian based on Gaussian Markov random field priors, which has the random walk priors used for the temporal parameters as a special case (Knorr-Held, 2000, Fahrmeir and Lang, 2001). The models we are proposing involve a large number of parameters, so it is vital to use efficient algorithms for inference via MCMC.

A full model-based analyis was also presented by Lagazio *et al.* (2003) again for an analysis of lung cancer rates in Tuscany. However, they used a single-site MCMC algorithm, which may be disadvantgeous. First, in the case of slow mixing the parameter estimates may be misleading (Knorr-Held and Rue, 2002). Second, sum-to-zero constraints can only be incorporated through ad-hoc re-centering approaches (i.e. sample each parameter from the full conditional *without* the constraint and re-center the parameters subsequently to have mean zero). In contrast, we use algorithms proposed by Rue (2001) for block updating of each Markov random field. This will lead in better mixing of the algorithms. Furthermore, these algorithms have the additional advantage that identifiability constraints on the latent paramter can be explicitly incorporated in the prior.

For comparison of several models we use the deviance information criteria proposed by Spiegelhalter *et al.* (2002) to assess model fit and complexity. We illustrate our modelling framework in an analysis of stomach cancer rates for males in West Germany.

The data were given by the "Deutsches Krebsforschungsinstitut", see Becker and Wahrendorf (1997). A descriptive analysis shows decreasing mortality rates from 1976 to 1990. The rates are especially high in Bavaria (south-eastern part of Germany), but with a strong downward trend. Boeing *et al.* (1991) found several risk factors for this mortality pattern as the preservation of meats and sausages by smoking, a high consumption of processed meat products and a large number of private wells unconnected to the public water supply in this area. Most of these factors were present in Bavaria in earlier days, but they became less important now. It is therefore of particular interest to investigate if there is evidence for space-time interactions in these data. Furthermore we will determine if a period-space or a cohort-space interaction model is more appropriate for these data. Finally we will show, how the models can be used to predict future mortality rates.

The paper is organized as follows. Section 2 outlines the general modelling framework and also gives details about implementation issues, Section 3 then describes an application to the dataset mentioned above. We end with some final comments.

2 Model

Let y_{ijl} and n_{ijl} denote the number of disease cases and the number of persons under risk respectively in age group $i = 1, \ldots, I$, period $j = 1, \ldots, J$ and area $l = 1, \ldots, L$. The cohort index $k = 1, \ldots, K$ can be derived from i and j, depending on the resolution of the age and period effects (Knorr-Held and Rainer, 2001). For example, for data stratified by the same grid, $k = I - i + j$. In the application considered in Section 3, $I = 13$, $J = 15$, $K = 75$ and $L = 30$.

We assume a binomial observation model for y_{ijl} given n_{ijl} with an unknown disease probability π_{ijl}. We decompose the log-odds $\eta_{ijl} = \log(\pi_{ijl}/(1 - \pi_{ijl}))$ of these probabilities additively into (a) main effects for age, period, cohort and space, (b) interaction between period and space, or cohort and space, and (c) parameters, describing additional unstructured heterogeneity in each cell (i, j, l).

More specifically we assume in the most complex formulation that

$$\eta_{ijl} = \mu + \theta_i + \varphi_j + \psi_k + \xi_l + \left[\begin{array}{c} \delta_{jl} \\ \delta_{kl} \end{array} \right] + z_{ijl} \qquad (1)$$

where μ is an intercept term, θ_i is the age, φ_j the period, ψ_k the cohort and ξ_l the spatial effect. Parameters for space-time interaction are denoted by δ_{jl} or δ_{kl} (either period or cohort with space; the brackets in equation (1) indicate that only one of the two options enter in the formulation) and z_{ijl} denotes parameters for additional unstructured heterogeneity. These parameters will adjust for residual overdispersion after adjusting for the main and interaction effects.

To achieve identifiability of the parameters, we have to include sum to zero restrictions on all main effects:

$$\sum_i \theta_i = \sum_j \phi_j = \sum_k \psi_k = \sum_l \xi_l = 0.$$

However, the age, the period and the cohort effect are still not identifiable, because any linear transformation of the type

$$\theta_i \to \theta_i + c \cdot i, \quad \phi_j \to \phi_j - c \cdot j, \quad \psi_k \to \psi_k + c \cdot k, \quad \mu \to \mu - c \cdot I \qquad (2)$$

with arbitrary $c \in R$ leaves the log-odds η_{ijl} unchanged (e.g. Clayton and Schifflers, 1987b). We will comment later on this issue.

2.1 Prior assumptions

For the intercept term μ we use a flat prior, that is:

$$p(\mu) \propto \text{const.}$$

The age, period and cohort effects are modeled with Gaussian random walk priors. For example a random walk of first order (RW1) for $\boldsymbol{\theta}$ is:

$$p(\theta_1) \propto \text{const.},$$

$$\theta_i | \theta_{i-1}, \kappa \sim \text{N}(\theta_{i-1}, \kappa^{-1}) \qquad \text{for } 2 \leq i \leq I$$

where κ is a precision parameter. This prior can also be written for the vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)^T$:

$$p(\boldsymbol{\theta}|\kappa) \propto \kappa^{rg(\mathbf{K}_\theta)/2} \exp\left(-\frac{\kappa}{2}\boldsymbol{\theta}^T \mathbf{K}_\theta \boldsymbol{\theta}\right)$$

where the *structure matrix* (Clayton, 1996) $\mathbf{K}_\theta$ has the form

$$\mathbf{K}_\theta = \begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \dots & \dots & \dots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}$$

for a random walk of first order and

$$\mathbf{K}_\theta = \begin{pmatrix} 1 & -2 & 1 & & & & & & \\ -2 & 5 & -4 & 1 & & & & & \\ 1 & -4 & 6 & -4 & 1 & & & & \\ & 1 & -4 & 6 & -4 & 1 & & & \\ & & \dots & \dots & \dots & \dots & \dots & & \\ & & & & 1 & -4 & 6 & -4 & 1 \\ & & & & & 1 & -4 & 5 & -2 \\ & & & & & & 1 & -2 & 1 \end{pmatrix}$$

for a random walk of second order (RW2).

Similarly, the priors for the period and cohort effects have the form

$$\begin{aligned} p(\boldsymbol{\phi}|\lambda) &\propto \lambda^{rg(\mathbf{K}_\phi)/2} \exp\left(-\frac{\lambda}{2}\boldsymbol{\phi}^T \mathbf{K}_\phi \boldsymbol{\phi}\right) \text{ and} \\ p(\boldsymbol{\psi}|\nu) &\propto \nu^{rg(\mathbf{K}_\psi)/2} \exp\left(-\frac{\nu}{2}\boldsymbol{\psi}^T \mathbf{K}_\psi \boldsymbol{\psi}\right). \end{aligned}$$

In general, the precision matrix $\boldsymbol{K}$ has a rank-deficiency of 1 (or 2) for the RW1 (or RW2) model. Thus, for example, for the age effects $\boldsymbol{\theta}$, the rank of $\mathbf{K}_\theta$ is $I-1$ and $I-2$ respectively. Hence all these priors are improper.

As pointed out by Knorr-Held and Rainer (2001, page 112), a RW1 model order imposes a further *stochastic constraint* on the age, period and cohort parameters; such a model will prefer *a priori*, among all possible values for $c \in R$ in (2) the one where the quadratic first differences (weighted with the corresponding precision parameters) are minimal. This allows us to visually examine the (non-linear) trends present in age, period and cohort parameters.

For the spatial effect we use a Markov random field prior (e.g. Besag *et al.,* 1991). Indeed this prior is similar to a random walk prior:

$$p(\boldsymbol{\xi}|\omega) \propto \omega^{rg(\mathbf{K}_\xi)} \exp\left(-\frac{\omega}{2}\boldsymbol{\xi}^T\mathbf{K}_\xi\boldsymbol{\xi}\right)$$

where $\mathbf{K}_\xi$ is now determined through the neighbouring structure of the districts. With $\mathbf{K}_\xi = (k_{lm})_{l,m\in\{1,\ldots,L\}}$ the off-diagonal elements k_{lm} are -1 for geographically contiguous districts $l \sim m$ and zero for all other non-diagonal entries. The diagonal elements k_{ll} are equal to the number of districts contiguous to district l. Usually the districts cannot be split up into two or even more completely separated pieces, the matrix $\mathbf{K}_\xi$ has then rank $rg(\mathbf{K}_\xi) = L - 1$.

The priors for the interaction term were specified following a rationale originally proposed in Clayton (1996). The idea is to use the direct product of the precision matrices of the main effects as the precision matrix for the interaction effect. Here we will only consider interaction priors which include some form of temporal dependence. Using the terminology used in Knorr-Held (2000), these are called Type II and Type IV interaction. We will further distinguish if the interaction is between period and space or between cohort and space parameters.

The Type II model can be seen as independent district specific random walks. Let $\boldsymbol{\delta}_l$ denote the vector $(\delta_{1l}, \ldots, \delta_{Jl})^T$ for space-period interaction or $\boldsymbol{\delta}_l = (\delta_{1l}, \ldots, \delta_{Kl})^T$ for space-cohort interaction. Then the prior can be written as:

$$p(\boldsymbol{\delta}|\rho) \propto \exp\left(-\frac{\rho}{2}\sum_{l=1}^{L}\boldsymbol{\delta}_l^T\begin{bmatrix}\mathbf{K}_\phi\\ \mathbf{K}_\psi\end{bmatrix}\boldsymbol{\delta}_l\right).$$

As above the brackets indicate, that only one option enters in the formulation. The matrices $\mathbf{K}_\phi$ and $\mathbf{K}_\psi$ are the same as above and can be specified for a random walk of first or second order.

The Type II interaction model intrinsically assumes that $\boldsymbol{\delta}_l$, $l = 1, \ldots, L$ are independent across districts. In the Type IV interaction model, temporal trends are assumed to be similar in neighbouring districts.

In Knorr-Held (2000) the Type IV interaction model for $\boldsymbol{\delta}$ is given for the random walk of first order. The formulation can easily be modified to a random walk of second order by replacing first by second differences. This corresponds to a modified precision matrix for $\mathbf{K}_\phi$.

For space-period interaction, with $\boldsymbol{\delta} = (\delta_{11}, \ldots, \delta_{1L}, \delta_{21}, \ldots, \delta_{JL})^T$, the prior can be written as

$$
\begin{aligned}
P(\boldsymbol{\delta}|\rho) &\propto \rho^{(L-1)(T-2)/2} && (3)\\
&\times \exp\left(-\frac{\rho}{2}\sum_{l\sim m}\sum_{j=3}^{J}((\delta_{jl}-2\delta_{j-1,l}+\delta_{j-2,l})-(\delta_{jm}-2\delta_{j-1,m}+\delta_{j-2,m}))^2\right)\\
&\propto \rho^{(L-1)(T-2)/2}\exp\left(-\frac{\rho}{2}\left(\boldsymbol{\delta}^T(\mathbf{K}_\phi\otimes\mathbf{K}_\xi)\boldsymbol{\delta}\right)\right). && (4)
\end{aligned}
$$

This prior is a Gaussian Markov random field prior, where not only the spatial neighbours and the first and second temporal neighbours, but also the temporal neighbours of the spatial neighbours enter in the conditional distribution of δ_{jl}.

The density (4) is invariant to the addition of any arbitrary constants at any time j or at any district m. We therefore have to introduce $J+L-1$ additional restrictions, to make $\boldsymbol{\delta}$ identifiable. For example, we may use

$$
\sum_{l=1}^{L}\delta_{jl}=0 \quad \text{for } 1\le j\le J \quad \text{and} \quad \sum_{j=1}^{J}\delta_{jl}=0 \quad \text{for } 1\le l\le L-1.
$$

It can easily be seen that this implies

$$
\sum_{j=1}^{J}\delta_{jL}=0.
$$

Thus the row sums are all zero, that is the interaction is centred at zero at each period and the column sums are all zero, too, so the interaction is centred at zero in each district. A similar model with restrictions can be used for space-cohort interaction.

For the term z_{ijl} we use a white noise prior: $z_{ijl}\sim \mathrm{N}(0,\epsilon^{-1})$. For the hyperparameters $\lambda,\kappa,\nu,\omega,\rho$ and ϵ we assume proper gamma priors $\mathrm{G}(a,b)$, with $a=1$ and $b=0.05$.

2.2 MCMC simulation

Following Besag *et al.* (1995) we do not directly update z_{ijl} but reparametrize the model and use the linear predictor η_{ijl} as an unknown parameter rather than z_{ijl}, see equation (1). This has the advantage that the full conditionals of all effects mentioned above are Gaussian and hence Gibbs steps can be used for updating. The linear predictor can be sampled by independent Metropolis-Hastings steps.

The full conditional of the main effects have all a similar form, for example for the age effect $\boldsymbol{\theta}$ we get

$$p(\boldsymbol{\theta}|..) \propto \exp\left(-\frac{\lambda}{2}\left(\boldsymbol{\theta}^T\mathbf{K}_\theta\boldsymbol{\theta}\right)\right)\cdot\exp\left(-\frac{\epsilon}{2}\left(\boldsymbol{\theta}^T\mathbf{I}\boldsymbol{\theta}\right)+\epsilon\left(\boldsymbol{\beta}^T\boldsymbol{\theta}\right)\right) \tag{5}$$

$$\propto \exp\left(-\frac{1}{2}(\boldsymbol{\theta}^T A_{\boldsymbol{\theta}}\boldsymbol{\theta})+(\boldsymbol{\beta}^T\boldsymbol{\theta})\right) \tag{6}$$

where $\boldsymbol{\beta}=(\beta_1,\ldots\beta_I)^T$ is a $I\times 1$ vector with

$$\beta_i=\sum_{j,l}\left(\eta_{ijl}-\mu-\phi_j-\psi_k-\xi_l-\begin{bmatrix}\delta_{jl}\\ \delta_{jk}\end{bmatrix}\right)$$

and $A_{\boldsymbol{\theta}}=\lambda\mathbf{K}_\theta+\epsilon\mathbf{I}$. The full conditional $p(\boldsymbol{\theta}|..)$ is therefore multivariate normal with mean $A_{\boldsymbol{\theta}}^{-1}\boldsymbol{\beta}$ and precision matrix $A_{\boldsymbol{\theta}}$. One can easily see that $A_{\boldsymbol{\theta}}$ is a band matrix of the same band-width as the prior precision $\mathbf{K}_\theta$. Similary, the full conditionals for the period, cohort and space parameters are also multivariate normal with a band width of the precision matrix equal to the band width of the corresponding prior precision matrix.

For the spatial effect we use a clever trick described in Rue (2001) and reorder the indices of the districts, so that $\mathbf{K}_\xi$ and consequently the precision matrix of the full conditional is a band matrix with minimal band-width m (see Rue, 2001 for details). In our application $m=10$.

Using the band structure of the precision matrix, we can efficently sample from (6) using the algorithms described Rue (2001). The idea of these algorithms is to use Cholesky decomposition of the precision matrix, which is very fast for band matrices of small band-width. As mentioned above we need to sample conditional on the sum to zero restrictions. This imposes no further complication, as also in this case we can use an algorithm based on an efficient Cholesky decomposition of the precision matrix, see Rue (2001) for further details.

Turning to the interaction models, the full conditional precision of the Type II model is only 1 and 2 depending on the chosen model. In Type IV models, the band width is usually larger. Using the reordered districts, the band-width of $\mathbf{K}_\phi\otimes\mathbf{K}_\xi$ is $L+m$ for a random walk of first order and $2\cdot L+m$ for a random walk of second order. Typically the band-width of the precision matrix can not be further reduced by another reordering.

The full conditionals of the hyper parameters are all independent Gamma distributed and can be sampled in Gibbs steps. Finally, the full conditional of the linear predictors

$$p(\eta_{ijl}|..) \propto \frac{\exp(\eta_{ijl}y_{ijl})}{(1+\exp(\eta_{ijl}))^{n_{ijl}}} \exp\left(-\frac{\epsilon}{2}\left(\eta_{ijl}-\mu-\theta_i-\phi_j-\psi_k-\xi_l-\left[\begin{array}{c}\delta_{jl}\\ \delta_{kl}\end{array}\right]\right)^2\right),$$

is not a standard distribution. Using a Taylor approximation we approximate this with a normal distribution, which can be used as a proposal in a Metropolis-Hastings algorithm. With this approach we get acceptance rates of 90%-99%.

2.3 Predictions

An important aspect of our formulation is that it allows for the prediction of future mortality rates while allowing for space-time interaction. Indeed, using a random walk prior, we can easily extend our model (1) to predict future incidence rates. The predicted log-odds of the mortality at period $J+1$ is

$$\eta_{i,J+1,l} = \mu + \theta_i + \varphi_{J+1} + \psi_k + \xi_l + \left[\begin{array}{c}\delta_{J+1,l}\\ \delta_{kl}\end{array}\right] + z_{i,J+1,l}. \tag{7}$$

The period effect φ_{J+1} can be sampled by extending the random walk prior to the future, that is $\varphi_{J+1} \sim \mathrm{N}(\varphi_J, \lambda^{-1})$ for first order and $\varphi_{J+1} \sim \mathrm{N}(2\cdot\varphi_J - \varphi_{J-1}, \lambda^{-1})$ for second order. The cohort effect ψ_{K+1} can be generated similarly.

In the model with Type II interactions $\delta_{J+1,l}$ can be sampled the same way by $\delta_{J+1,l} \sim \mathrm{N}(\delta_{Jl}, \omega^{-1})$ in each period $l = 1, \ldots, L$ and analogous for the model with cohort interactions $\delta_{K+1,l}$.

The Type IV interactions can only be generated by including the restriction $\sum_{l=1}^{L} \delta_{jl} = 0$ for each period $j = J+1, J+2, \ldots$ to be predicted. We propose a *sequential* algorithm for this rather challenging task which is described in detail in Appendix A.

3 Application

3.1 Model selection

In order to measure the goodness of fit of our models, we use the posterior deviance

$$D = -2\cdot\sum_{ijl}\left(l(\pi_{ijl}) - l(\hat{\pi}_{ijl})\right),$$

where $l(y_{ijl})$ is the log likelihood. A low deviance indicates a good fit of the estimated mortality rates to the data.

As the deviance decreases with the number of parameters, we use the deviance information criteria DIC introduced by Spiegelhalter *et al.* (2002) to compare the different models. The DIC is defined as

$$DIC = (\bar{D}) + p_D$$

with p_D the effective number of parameters. This effective number of parameters penalizes the model complexity so that the models can be compared by the DIC. It is calculated by

$$p_D = (\bar{D}) - D\left(\hat{\pi}_{ijl}\right).$$

with

$$D\left(\hat{\pi}_{ijl}\right) = -2 \cdot \sum_{ijl} \left[l(\pi_{ijl}) - l(\hat{\pi}_{ijl})\right].$$

Table 1 shows deviance, effective number of parameters and DIC of all mentioned models.

	interaction		deviance	p_D	DIC
RW 1	none		5623.1	39.9	5663.0
	Type II	period	5374.0	196.4	**5570.4**
	Type II	cohort	5235.5	1573.5	6809.0
	Type IV	period	5424.9	269.5	5694.4
	Type IV	cohort	5289.5	1483.0	6772.5
RW 2	none		5656.8	19.4	5676.2
	Type II	period	5406.2	180.0	**5586.2**
	Type II	cohort	5309.7	1202.7	6512.4
	Type IV	period	5461.8	266.5	5728.3
	Type IV	cohort	5358.1	1312.4	6670.5

Table 1: deviance, effective number of parameters and DIC of all models

The results are very similar for both the RW1 and RW2 prior. The deviance is the lowest for the Type II interactions and the highest for the no interaction model. As mentioned above, a higher number of nominal parameters leads naturally to a good fit. However, the fit of the Type IV interaction models is not so good, because it imposes more dependence structure on the interaction parameter than the Type II model.

Interestingly the estimated effective number of parameters for the Type IV interaction models is higher than for the Type II interactions. This is because in the latter models in many regions the median of the interaction parameter is zero for mostly all periods whereas for the former, due to the underlying Markov random field more parameters are estimated non-zero.

The models with cohort interactions have a much higher number of effective parameters without decreasing the deviance. This leads to the conclusion, that the model with period-district interactions is more appropriate for these data.

3.2 Results

We first show the estimated main effects of the best model in terms of the DIC criterion, the model with Type II period-district interactions. As the age, period and cohort effects are not identifiable in the RW2 model (see Knorr-Held and Rainer, 2001), we only show the result for the model based on the random walk of first order.

Figures 1 to 3 displays median and 90% credible intervals of the age, period and cohort effects. One should keep in mind that the patterns can be transformed through linear transformations of type (2) without any change in the likelihood. Therefore only non-linear trends can be interpreted. While the period effect appears to be roughly linear, the cohort effect has an interesting change-point around birth cohort 1940. After 1940 the effect appear to be linear. However, these estimates are based on few data and should therefore not be overinterpreted.

Figure 4 shows the spatial effect. The highest values are, as expected, in the south east of Germany in the dictricts Upper Palatinate and Lower Bavaria.

Figures 5 displays the period-district interactions for the district of Lower Bavaria. In all the models, the interactions show a decreasing trend. A similar trend can be seen in the other districts in this area, which indicated that - after adjusting for the overall period trend - there is a decreasing mortality rate in the south-east of Germany.

The model based on cohort-district rather than period-district interactionsproduces very unstable estimates (Figure 6). It seems that the estimated global variance of this interaction term is rather high. Consequently, the effects in the last birth cohorts, which are based on only few cases, show high variability. As

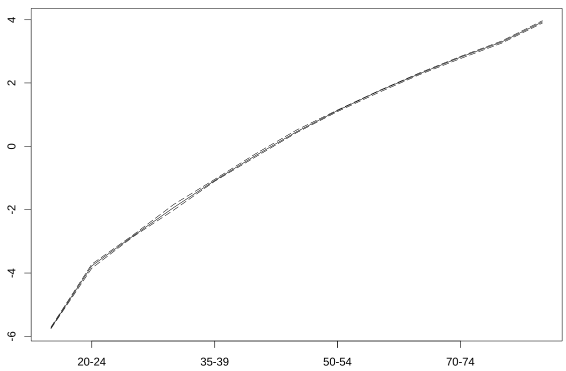

Figure 1: Posterior median within 90% credible intervals of the age effects.

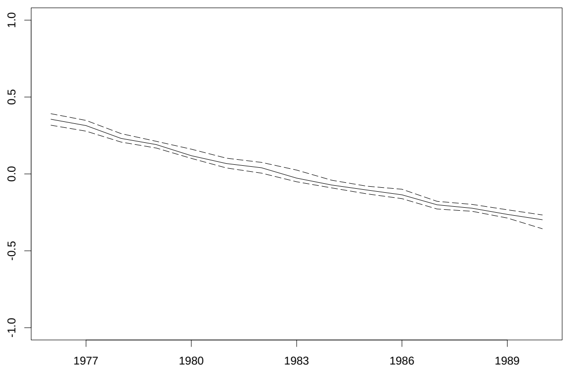

Figure 2: Posterior median within 90% credible intervals of the period effects.

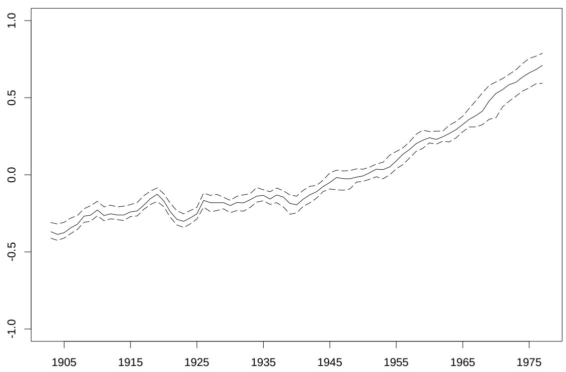

Figure 3: Posterior median within 90% credible intervals of the cohort effects.

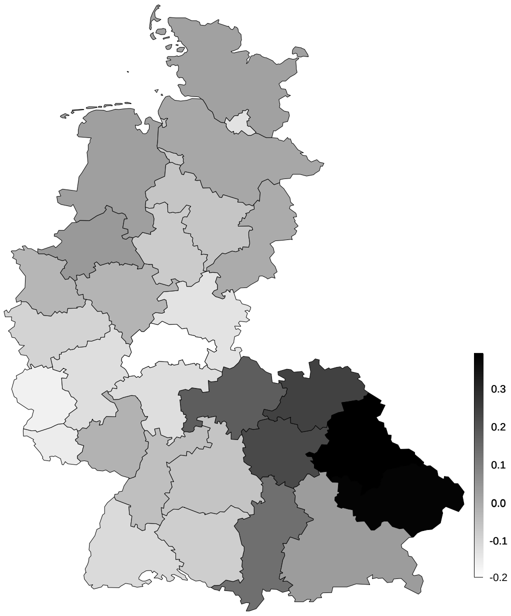

Figure 4: Posterior median of the spatial effects.

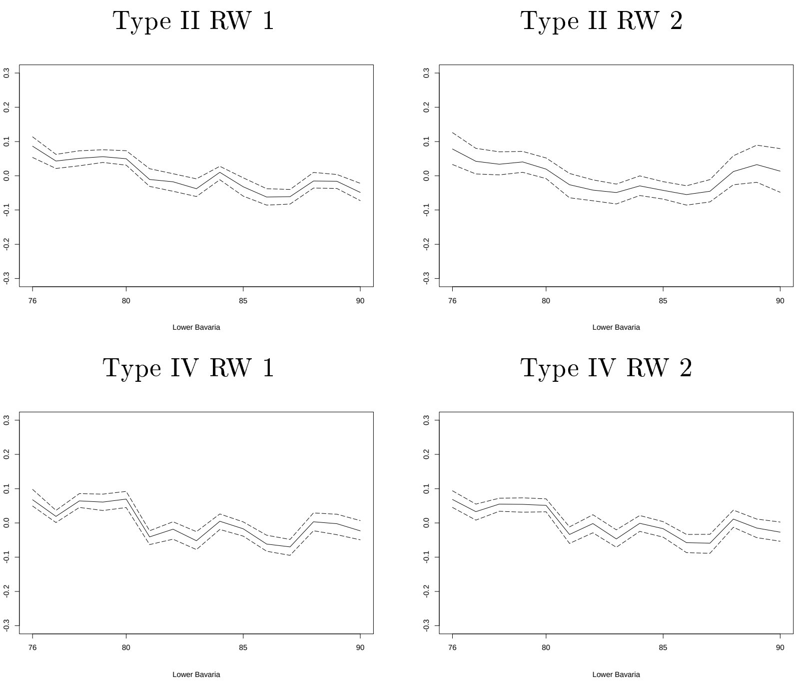

Figure 5: Posterior median within 90% credible intervals of period-district interactions in Lower Bavaria.

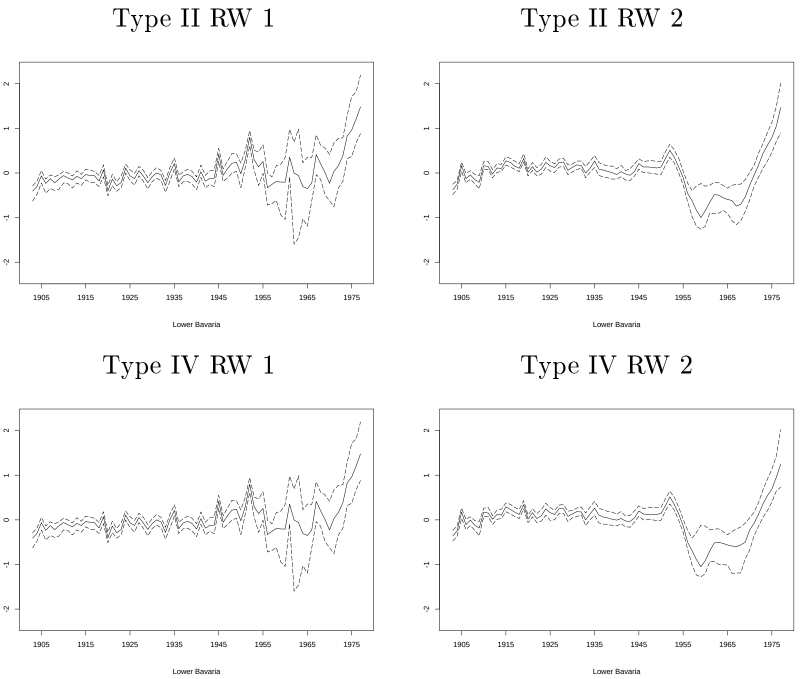

Figure 6: Posterior median within 90% credible intervals of Cohort-district interactions in Lower Bavaria.

one can expect, the effects in the RW2 model are smoother then in the RW1 model.

For a further analysis to investigate the predictive quality of the different models, we fitted the models to the first ten observation years, obmitting the last five years, and then made predictions for those last five years. As the cohort-district interactions appear to be not appropriate for the data, we only use the models with period-district interactions here. Figure 7 shows the predicted median number of cases within 90%-credible intervals (lines) and the actually observed data (x) for Lower Bavaria. Note that we have aggregated the counts in this district over all age groups.

The models without interactions produce very narrow credible intervals and therefore do not make good predictions. The models with interactions seem to produce more sensible prediction intervals. The median of the predicted cases is closer to the actually data and the 90% credible intervals contain most of the data points.

For are more detailed analysis, Table 2 gives the percentage of observed values that lie within credibility intervals of a given level. Note that this table is based on the original grid, i.e. predictions have been compared with the actual observed data for each of the $13 \times 15 \times 30 = 5,850$ age group times period times district combinations.

The first thing to note is that the model without interaction terms has poor coverage percentages. The model with interaction come closer to the actual credibility levels, in particular those based on the RW2 prior. But still they are all smaller, with the only exception of Type IV interactions with the RW2 prior, where 62% of the data are contained in the 50% credible intervals. This indicates that this model might overfit the data. Therefore it seems that the Type II RW2 model has the best prediction qualities: It comes closest to the actual credible levels, while not overestimating them.

Credibility level	no interaction		Type II		Type IV	
	RW1	RW2	RW1	RW2	RW1	RW2
(%)	(%)	(%)	(%)	(%)	(%)	(%)
50	11	41	22	50	42	62
80	20	55	39	62	56	68
95	33	62	50	66	64	71

Table 2: The percentage of observed cases that lie within predictive credible intervals

4 Discussion

The specification of model for space-time interaction, while adjusting for age, period and cohort effects is a challenging task. In this paper we have proposed various formulations which take into account interaction effects between period and space or cohort and space. In our application, one of these models provide a better fit to the data, even after adjusting for the additional complexity of the formulations. Further confirmation of the superiority of this model was obtained through a study of the empirical coverage percentages of out-of-sample predictions.

The models have a very large nominal number of a priori dependent parameters. It is therefore advantageous to use block updating MCMC algorithms to avoid slow mixing of the Markov chain. Furthermore, blocking also allows for a proper incorporation of identifiability constraints, such as sum-to-zero constraints. However, further improvement is to be expected by joint updates of

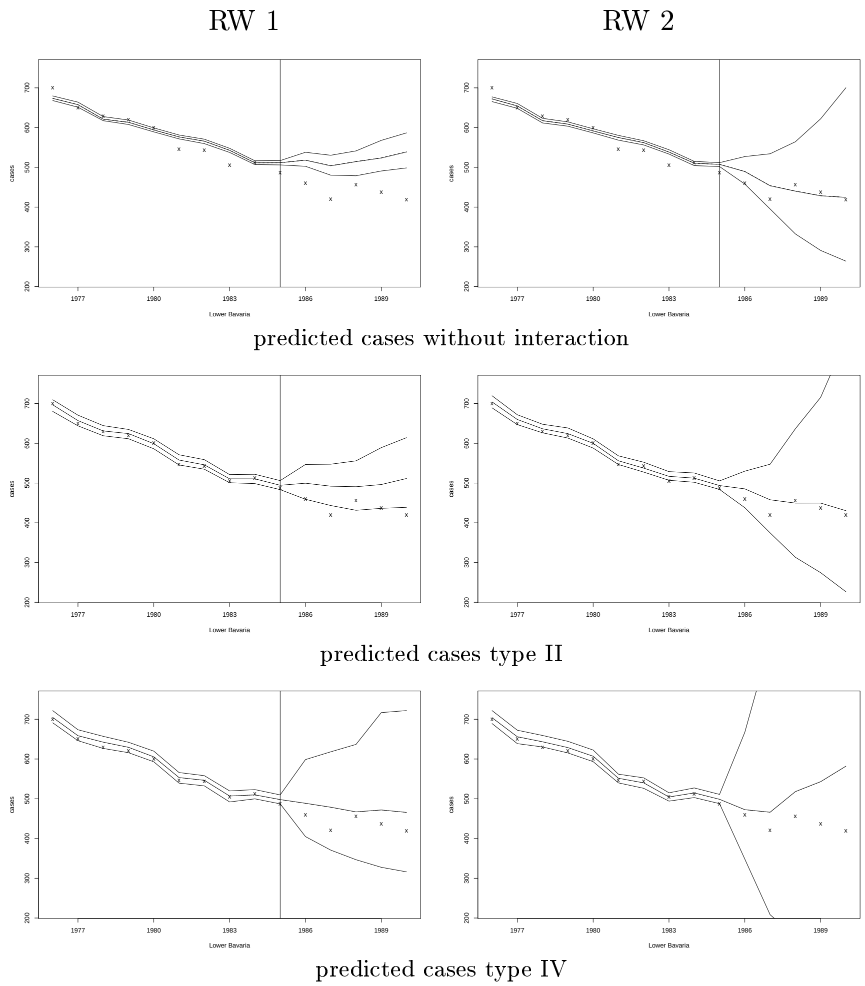

Figure 7: Observed and predicted number of cases within pointwise 90% credible intervals for Lower Bavaria based on the RW1 (left) and RW2 (right) model. Crosses represent the observed rates.

hyperparameters and parameters (Knorr-Held and Rue, 2002).

The most complex formulation allows for a dependence of interaction parameters across space and time, with additional sum-to-zero constraints for each area and time point. New methodological developments, as described in the Appendix, have been made in order to predict this formulation into the future.

A Prediction with Type IV interactions

The prior for the Type IV period-district interaction in the RW1 model is

$$p(\boldsymbol{\delta}|\rho) \propto \exp\left(-\frac{\rho}{2}\sum_{l\sim m}\sum_{j=2}^{j}((\delta_{jl}-\delta_{j-1,l})-(\delta_{jm}-\delta_{j-1,m}))^2\right).$$

Thus a natural approach to predict future values of $\boldsymbol{\delta}$ is based on a algorithm *sequentially* in time: We start by sampling $\boldsymbol{\delta}_{J+1}|\boldsymbol{\delta}_J,\rho$ where $\boldsymbol{\delta}_{J+1} = (\delta_{J+1,1},\ldots,\delta_{J+1,L})$:

$$\begin{aligned} p(\boldsymbol{\delta}_{J+1}|\boldsymbol{\delta}_J,\rho) &\propto \exp\left(-\frac{\rho}{2}\sum_{l\sim m}(\delta_{J+1,l}-\delta_{J+1,m}-\delta_{Jl}+\delta_{Jm})^2\right). && (8)\\ &\propto \exp\left(-\frac{1}{2}\boldsymbol{\delta}'_{J+1}(\rho\mathbf{K}_\xi)\boldsymbol{\delta}_{J+1}+\rho\boldsymbol{\delta}'_{J+1}\boldsymbol{\nu}\right) && (9)\end{aligned}$$

with $\boldsymbol{\nu} = \nu_1,\ldots,\nu_J$ and $\nu_i = \sum_{l\sim m}(\delta_{il}-\delta_{im})$. This is the density of a singular Gaussian Markov random field. Thus we can not sample from the density without further restrictions. As in Section 2.1 we will use a sum to zero restriction $\sum_l \delta_{J+1,l} = 0$. To obtain the restricted density, we use a Lemma by Box & Tiao (1973, p.419):

Lemma. Let $\mathbf{x},\mathbf{a}$ and $\mathbf{b}$ be $k\times 1$ vectors and $\mathbf{A}$ and $\mathbf{B}$ be two $k\times k$ positive semidefinite symetric matrices. Suppose the rank of the matrix $\mathbf{A}+\mathbf{B}$ is $q(<k)$. Then, subject to the constraints $\mathbf{G}\mathbf{x}=\mathbf{0}$,

$$\begin{aligned} (\mathbf{x}+\mathbf{a})'\mathbf{A}(\mathbf{x}-\mathbf{a})+(\mathbf{x}-\mathbf{b})'\mathbf{B}(\mathbf{x}-\mathbf{b}) &= (\mathbf{x}-\mathbf{c})'(\mathbf{A}+\mathbf{B}+\mathbf{M})(\mathbf{x}-\mathbf{c}) && (10)\\ &+ (\mathbf{a}-\mathbf{b})'\mathbf{A}(\mathbf{A}+\mathbf{B}+\mathbf{M})^{-1}\mathbf{B}(\mathbf{a}-\mathbf{b}) \end{aligned}$$

where $\mathbf{G}$ is any $(k-q)\times k$ matrix of rank $k-q$ such that the rows of $\mathbf{G}$ are linearly independent of the rows of $\mathbf{A}+\mathbf{B}, \mathbf{M}=\mathbf{G}'\mathbf{G}$ and

$$\mathbf{c} = (\mathbf{A}+\mathbf{B}+\mathbf{M})^{-1}(\mathbf{A}\mathbf{a}+\mathbf{B}\mathbf{b}).$$

With $\mathbf{B}=\mathbf{0}$ and $\mathbf{b}=\mathbf{0}$ we can rewrite (10) as

$$\begin{aligned} (\mathbf{x}-\mathbf{a})'\mathbf{A}(\mathbf{x}-\mathbf{a}) &= (\mathbf{x}-\mathbf{c})'(\mathbf{A}+\mathbf{M})(\mathbf{x}-\mathbf{c}) \\ \Leftrightarrow \mathbf{x}'\mathbf{A}\mathbf{x} - 2\mathbf{x}'\mathbf{A}\mathbf{a} + \mathbf{a}'\mathbf{A}\mathbf{a} &= \mathbf{x}'(\mathbf{A}+\mathbf{M})\mathbf{x} - 2(\mathbf{x}'(\mathbf{A}+\mathbf{M})\mathbf{c}) + \mathbf{c}'(\mathbf{A}+\mathbf{M})\mathbf{c} \end{aligned}$$

Applying this Lemma, we set $\mathbf{A}=\rho\mathbf{K}_\xi, \mathbf{a}=\boldsymbol{K}_\xi^-\boldsymbol{\nu}$ and $\mathbf{G}=\mathbf{1}'$ and hence the density of the restricted Markov random field is proportional to

$$p(\boldsymbol{\delta}_{J+1}|\rho,\boldsymbol{\delta}_J) \propto \exp\left(-\frac{1}{2}\boldsymbol{\delta}'_{J+1}(\rho\boldsymbol{K}_\xi+\mathbf{1}\mathbf{1}')\boldsymbol{\delta}_{J+1}+\rho\boldsymbol{\delta}'_{J+1}\boldsymbol{\nu}\right). \tag{11}$$

This is the density of a (proper!) multivariate normal distribution with mean $\rho(\rho\boldsymbol{K}_\xi+\mathbf{1}\mathbf{1}')^{-1}\boldsymbol{\nu}$ and covariance matrix $(\rho\boldsymbol{K}_\xi+\mathbf{1}\mathbf{1}')^{-1}$. We now use the algorithm described in Rue (2001) to get a sample from this distribution.

For the interactions with random walk of second order, the density for the distribution of $\boldsymbol{\delta}_{J+1}$ can be extended from (4). An analogous calculation leads to the same formulation of the restricted density as (11) but with $\nu_i = \sum_{l\sim m}((\delta_{Jl}-\delta_{Jm}) - 2(\delta_{J-1,l}-\delta_{J-1,m}) + (\delta_{J-2,l}-\delta_{J-2,m}))$.

After we have sampled $\boldsymbol{\delta}_{J+1}$ we can use the same algorithms to sample $\boldsymbol{\delta}_{J+2}$ and so on. The prediction for cohort-district interactions has been implemented in a similar way.

References

Assuncao, R. M., Reis, I. A. and O., C. D. L. (2001). Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Statistics in Medicine*, **20**, 2319-2335.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M. (1995). Bayesian analysis of space–time variation in disease risk. *Statistics in Medicine*, **14**, 2433–2443.

Becker, N. and Wahrendorf, J. (1997). Atlas of Cancer Mortality in the Federal Republic of Germany 1981-1990. Berlin: Springer Verlag.

Beruzini, C., Clayton, D. and Bernardinelli, L. (1993). Bayesian inference on the Lexis diagramm. *Bulletin of the International Statistical Institute*, **50**, 149-164.

Besag, J. E., Green, P. J., Higdon, D. M. and Mengersen, K. L. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10**, 3-66.

Besag, J. E., York, J. C. and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, **43**, 1-59.

Boeing, H., Frentzel-Beyme, R., Berger, M. *et al.* (1991). Case-control study on stomach cancer in Germany. *International Journal of Cancer*, **47**, 858-864.

Box, G. and Tiao, G. (1973): *Bayesian Inference in Statistical Analysis.* Reading: Addison-Wesley.

Clayton, D. G. (1996). Generalized linear mixed models. *in*: W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.), *Markov chain Monte Carlo in Practice*, pp. 275–301. London: Chapman & Hall.

Clayton, D. G. and Bernardinelli, L. (1992). Bayesian methods for mapping disease risks, *in*: J. Cuzick and P. Elliot (eds.), *Small Area Studies in Geographical and Environmental Epidemiology*, pp. 205-220. Oxford University Press, Oxford.

Clayton, D.G. and Schifflers, E. (1987a). Models for temporal variation in cancer rates. I: age-period and age-cohort models. *Statistics in Medicine*, **6**, 449-467.

Clayton, D.G. and Schifflers, E. (1987b). Models for temporal variation in cancer rates. II: age-period-cohorts models. *Statistics in Medicine*, **6**, 468-481.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society*, Series C (*Applied Statistics*), **50**, 201-220.

Holford, T.R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, **39**, 311-324.

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

Knorr–Held, L. and Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine*, **17**, 2045-2060.

Knorr-Held, L and Rainer, E. (2001). Projections of lung cancer mortality in West Germany: a case study in Baysian prediction. *Biostatistics*, **2**, 109-129.

Knorr–Held, L. and Rue, H. (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics*, **29**, 597-614.

Lagazio, C., Dreassi, E. and Biggeri, A. (2001). A hierarchical Bayesian model for space-time variation of disease risk. *Statistical Modelling*, **1**, 17-29.

Lagazio, C., Biggeri, A. and Dreassi, E. (2003). Age-period-cohort models and disease mapping. *Environmetrics*, **14**, 475-490.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society*, Series B, **63**, 325-338.

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**, 583-640