STATISTICS IN MEDICINE
Statist. Med. 2000; **19**:2555–2567

Bayesian modelling of inseparable space-time variation in disease risk

Leonhard Knorr-Held*,†

Institute of Statistics, Ludwig-Maximilians-University Munich, Ludwigstr. 33, 80539 Munich, Germany

SUMMARY

This paper proposes a unified framework for a Bayesian analysis of incidence or mortality data in space and time. We introduce four different types of prior distributions for space×time interaction in extension of a model with only main effects. Each type implies a certain degree of prior dependence for the interaction parameters, and corresponds to the product of one of the two spatial with one of the two temporal main effects. The methodology is illustrated by an analysis of Ohio lung cancer data 1968–1988 via Markov chain Monte Carlo simulation. We compare the fit and the complexity of several models with different types of interaction by means of quantities related to the posterior deviance. Our results confirm an epidemiological hypothesis about the temporal development of the association between urbanization and risk factors for cancer.

1. INTRODUCTION

There has been much recent interest in the analysis of disease rates over space and time. The problem with such data is that the number of cases and the corresponding population at risk in any single unit of space × time are too small to produce a reliable estimate of the underlying disease risk without 'borrowing strength' from neighbouring cells. The goal here could be described as one of smoothing, in which both spatial and non-spatial considerations may arise, and spatio-temporal interactions may become an important feature.

Most of the Bayesian methods [1–3] propose extensions of the purely spatial models by Clayton and Kaldor [4] and Besag *et al.* [5] to space × time time data. Bernardinelli *et al.* [1] suggest a model in which both area-specific intercept and temporal trend are modelled as random effects, representing area-specific deviations from an overall risk profile. This formulation already allows for spatio-temporal interactions where temporal trends in disease risk may be different for different spatial locations and may even have spatial structure in itself. However, all temporal trends are assumed to be linear, which is a restrictive assumption.

*Correspondence to: Leonard Knorr-Held, Department of Epidemiology and Public Health, Imperial College School of Medicine, Norfolk Place, London W2 1PG, U.K.
†E-mail: l.knorr-held@ic.ac.uk

Contract/grant sponsor: German Science Foundation; contract/grant number: SFB 386

Waller *et al.* [2] use a nested model, where the hierarchical specification by Besag *et al.* is applied to each time point separately. The model does not have spatial main effects and therefore allows that the spatial patterns at each time point are completely different. There is less emphasis on modelling the temporal development in disease risk, as time is treated as essentially exchangeable.

The paper by Knorr-Held and Besag [3] proposes a model that combines the spatial model by Besag *et al*. with *dynamic models* [6, 7]. Such dynamic models allow for a non-parametric estimation of temporal trends in disease risk including time-changing effects of covariates. In particular, they do not assume linearity or stationarity and can be seen as the temporal analogue of the spatially structured component [8] in the Besag *et al.* model. Hence, both the temporal and the spatial risk profile are estimated non-parametrically. However, the model combines temporal and spatial main effects additively and does not allow for space $\times$ time interactions.

The present paper tries to fill this gap focusing on the case where the disease variation cannot be separated into temporal and spatial (main) effects and spatio-temporal interactions become an important feature. For simplicity we assume that the data consists of single observations on the number of persons under risk and the number of cases or deaths for each pixel in space$\times$time. We start with the (slightly modified) Knorr-Held and Besag model, separable in space and time, from which four interaction types arise naturally as the product of one of the two spatial with one of the two temporal main effects, based on a suggestion by Clayton [9]. These four types of space$\times$time interaction imply different prior assumptions about the interrelationship between interaction parameters, ranging from complete independence to complete dependence. Two of the corresponding models (combining additively each type of interaction with the main effects) can be seen as the non-parametric analogue of the Bernardinelli *et al.* formulation.

The proposed modelling framework is outlined in Section 2. We also include details on the implementation by Markov chain Monte Carlo (MCMC) and outline some modifications and extensions. Section 3 describes an analysis of a data set on mortality from lung cancer among white males between 55 and 64 years, for 21 successive years in the 88 countries of Ohio. This is a subset of a data set analysed previously in the literature [2, 3]. We have implemented all four models corresponding to the four types of interactions as well as a model with only main effects. We use the posterior deviance [10] for comparing the fit and the complexity of the different models. Deviance residuals are used for model diagnostics. The results confirm an epidemiological hypothesis [11] that the correlation between urbanization and risk factors for cancer decreases in time. Section 4 gives a short general discussion and outlines possible extensions of the model to data, which are further stratified by age.

2. BAYESIAN MODELS FOR SPACE-TIME VARIATION

2.1. The main effect model

Let n_{it} denote the number of persons at risk in country i ($i = 1, \ldots, n$) and year t ($t = 1, \ldots, T$). We assume that the number of cases or deaths y_{it}, for country i during year t, has a binomial distribution with parameters n_{it} and binomial probability π_{it}, and that the likelihood for the entire data is the corresponding product of binomial terms. In some contexts, a Poisson approximation to the binomial might be appropriate, in particular when the data are given as age-standardized rates. We follow a standard path in modelling π_{it} with a logit link to the binomial and start with a model where the linear predictor η_{it} decomposes additively into time- and space-dependent effects

[3]. More specifically, we assume that the log-odds

$$\eta_{it} = \ln\{\pi_{it}/(1-\pi_{it})\}$$

has the decomposition

$$\eta_{it} = \mu + \alpha_t + \gamma_t + \theta_i + \phi_i \tag{1}$$

where μ is an overall risk level and α_t and γ_t are temporal effects, representing unspecified features of year t that, respectively, do and do not display temporal structure *a priori*. Similarly, θ_i and ϕ_i represent unspecified features of country i that, respectively, do and do not display spatial structure.

The formulation (1) is completed by assigning prior distributions to the various components of η_{it}. For μ we choose a flat non-informative prior. Each of the four blocks $\alpha = (\alpha_1, \ldots, \alpha_T)'$, $\gamma = (\gamma_1, \ldots, \gamma_T)'$, $\theta = (\theta_1, \ldots, \theta_n)'$ and $\phi = (\phi_1, \ldots, \phi_n)'$ is assumed to be multivariate Gaussian with mean zero and precision matrix λK, where λ is an unknown scalar to be estimated from the data and K is a known *structure matrix* [9]. The structure matrix K will be different for each block in order to describe different assumptions about the prior interrelationship between parameters within each block.

For α, we adopt a prior in which effects for neighbouring time points tend to be alike. The simplest of such dynamic models is the random walk with independent Gaussian increments

$$p(\alpha|\lambda_\alpha) \propto \exp\left(-\frac{\lambda_\alpha}{2}\sum_{t=2}^{T}(\alpha_t - \alpha_{t-1})^2\right)$$

which has structure matrix

$$K_\alpha = \begin{bmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \vdots & \vdots & \vdots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{bmatrix}$$

see, for example, Clayton [9]. This reference also describes a possible alternative, the random walk of second order, which should be preferred, if one is interested in *predicting* future disease rates. For γ, we assume exchangeability of the components by taking $K_\gamma = I$, the identity matrix.

For the spatially structured block θ, we choose a simple Gaussian intrinsic autoregression; see, for example, Besag *et al.* [5]. Thus, the structure matrix K_θ has non-diagonal elements $k_{ij} = -1$ for geographically contiguous counties $i \sim j$ and diagonal entries k_{ii} equal to the number of counties, say m_i, that are geographically contiguous to county i. All other elements in K_θ are zero. The prior for θ can be written as

$$p(\theta|\lambda_\theta) \propto \exp\left(-\frac{\lambda_\theta}{2}\sum_{i\sim j}(\theta_i - \theta_j)^2\right).$$

This Markov random field prior is the spatial analogue of the random walk. It can be extended by introducing weights in the prior formulation [5]. Finally, unstructured spatial heterogeneity is

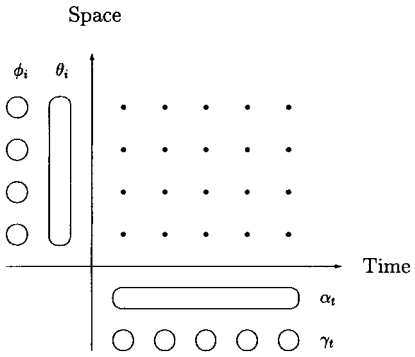

Figure 1. Symbolic representation of the main effect model. Circles represent prior independence, ovals represent prior dependence. Observations in time × space are indicated by small dots.

accounted for by taking $K_\phi = I$. A symbolic representation of the main effects model is given in Figure 1.

2.2. *Prior specification for interaction*

The above formulation, separable in space and time, requires appropriate expansion in the presence of space×time interactions. Formally we add interaction parameters δ_{it}, $i = 1, \ldots, n$, $t = 1, \ldots, T$, to (1):

$$\eta_{it} = \mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it}. \tag{2}$$

The parameter vector $\delta = (\delta_{11}, \ldots, \delta_{nT})'$ is assumed to be Gaussian with precision matrix $\lambda_\delta K_\delta$. As for the main effects, λ_δ is an unknown scalar and K_δ is a prespecified structure matrix. Note that model (2) reduces to (1) if all $\delta_{it} = 0$, hence δ captures only the variation that cannot be explained by the main effects.

Clayton [9] suggests specifying K_δ as the Kronecker product of the structure matrices of those main effects, which are assumed to interact. This rationale can be seen as the Bayesian analogue of modelling interactions by tensor products in a spline regression framework [12]. In our formulation, $2 \times 2 = 4$ combinations of spatio-temporal interaction are possible depending on which of the two temporal effects is assumed to interact with which of the two spatial effects. These four types of interactions imply different prior interrelationships between the δ_{it}, as illustrated in Figure 2. Now we discuss each type separately, ordered by the degree of prior dependence.

2.2.1. Type I *interaction* If the two unstructured main effects γ and ϕ are expected to interact, Clayton's rule gives $K_\delta = K_\gamma \otimes K_\phi = I \otimes I = I$, so all interaction parameters δ_{it} are *a priori* independent:

$$P(\delta|\lambda_\delta) \propto \exp\left(-\frac{\lambda_\delta}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}(\delta_{it})^2\right).$$

They can be thought of as unobserved covariates for each pixel (i, t), that do not have any structure in space × time.

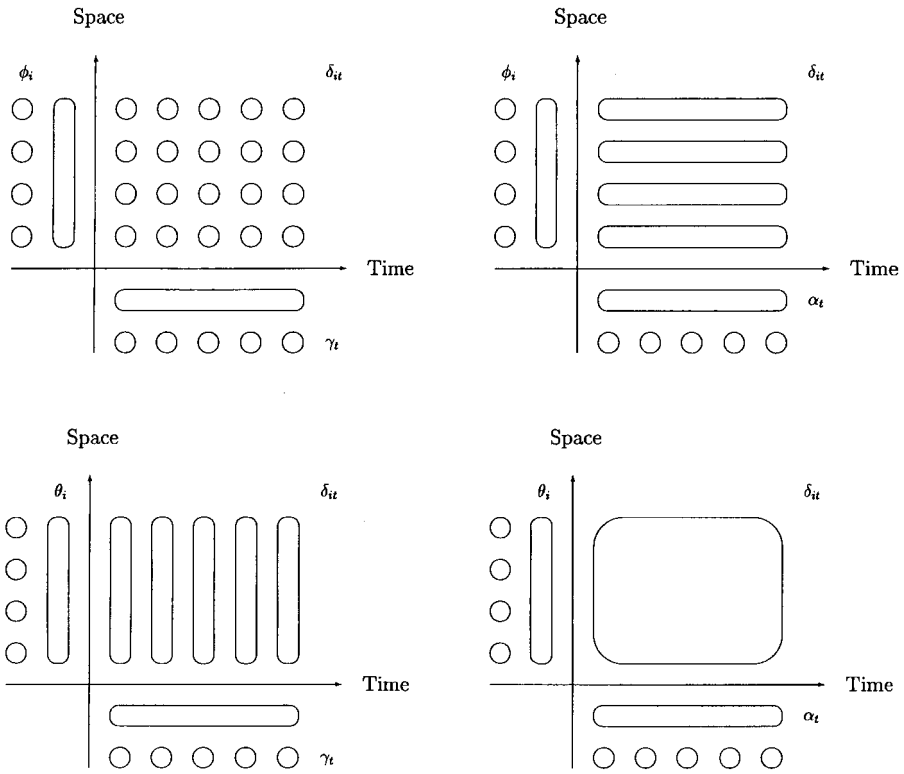

Figure 2. Symbolic representation of the four possible types of interactions. Circles represent prior independence, ovals represent prior dependence.

2.2.2. Type II *interaction* If we combine the random walk main effect α with the unstructured block ϕ by Clayton's rule, then each $\delta_i = (\delta_{i1}, \ldots, \delta_{iT})'$, $i = 1, \ldots, n$, follows a random walk, independently of all other counties. The structure matrix K_δ has rank $n(T-1)$ and the prior for δ can be written as

$$P(\delta|\lambda_\delta) \propto \exp\left(-\frac{\lambda_\delta}{2}\sum_{i=1}^{n}\sum_{t=2}^{T}(\delta_{it} - \delta_{i,t-1})^2\right).$$

Model (2) with δ of type II will be suitable, if temporal trends are different from county to county, but do not have any structure in space. It is the non-parametric analogue of one of the interaction models proposed in Bernardinelli *et al.* [1] where temporal trends are assumed to be linear.

2.2.3. Type III *interaction* If we assume that the main effects γ and θ interact, then each $\delta_t = (\delta_{1t}, \ldots, \delta_{nt})'$, $t = 1, \ldots, T$, follows an (independent) intrinsic autoregression. The rank of K_δ is now $(n-1)T$ and the prior on δ can be written as

$$P(\delta|\lambda_\delta) \propto \exp\left(-\frac{\lambda_\delta}{2}\sum_{t=1}^{T}\sum_{i\sim j}(\delta_{it} - \delta_{jt})^2\right).$$

Such a specification will be reasonable, if spatial trends are different from time point to time point, without any temporal structure.

2.2.4. Type IV *Interaction* From a theoretical point of view, the most interesting form of interaction arises as the product of the two dependent main effects, the random walk α and the intrinsic autoregression θ. Now δ is completely dependent over space and time and can no longer be factorized into independent blocks.

It can be shown that the prior for δ can be written as

$$P(\delta|\lambda_\delta) \propto \exp\left(-\frac{\lambda_\delta}{2}\sum_{t=2}^{T}\sum_{i\sim j}(\delta_{it}-\delta_{jt}-\delta_{i,t-1}+\delta_{j,t-1})^2\right)$$

with independent contrasts $\delta_{it}-\delta_{jt}-\delta_{i,t-1}+\delta_{j,t-1}$. Note that $K_\delta = K_\alpha \otimes K_\theta$ has only rank $(n-1)\,(T-1)$.

The conditional distribution of a pixel δ_{it}, given all the others, which can be derived [13] from K_δ has mean

$$\mu_{it} = \begin{cases} \delta_{i,t+1} + \frac{1}{m_i}\sum_{j\sim 1}\delta_{jt} - \frac{1}{m_i}\sum_{j\sim i}\delta_{j,t+1} & t=1 \\ \delta_{i,t-1} + \frac{1}{m_i}\sum_{j\sim 1}\delta_{jt} - \frac{1}{m_i}\sum_{j\sim i}\delta_{j,t-1} & t=T \\ \frac{1}{2}(\delta_{i,t-1}+\delta_{i,t+1}) + \frac{1}{m_i}\sum_{j\sim i}\delta_{jt} - \frac{1}{2m_i}\sum_{j\sim i}(\delta_{j,t-1}+\delta_{j,t+1}) & t=2,\ldots,T-1 \end{cases}$$

and precision

$$\tau_{it} = \begin{cases} m_i\lambda_\delta & t=1 \quad \text{or } t=T \\ 2m_i\lambda_\delta & t=2,\ldots,T-1 \end{cases}.$$

Hence, the type IV interaction prior is a Markov random field, where not only (first-order) temporal ($\delta_{i,t-1}$ and/or $\delta_{i,t+1}$) and spatial (δ_{jt}, $j\sim i$) neighbours enter in the full conditional for δ_{it}, but also second-order neighbours ($\delta_{j,t-1}$ and/or $\delta_{j,t+1}$, $j\sim i$), that is, spatial neighbours of temporal neighbours or, equivalently, temporal neighbours of spatial neighbours. This prior 'borrows strength' from spatial neighbours as it assumes that the temporal trend in county i (in terms of first differences) is similar to the average trend in neighbouring counties. Equivalently, one could also emphasize spatial trends here, as such a model 'borrows strength' from neighbouring time points ($t-1$ and/or $t+1$), assuming the spatial pattern in year t to be similar. This can be best seen from the conditional mean μ_{it}, which satisfies both

$$\mu_{it} - \bar{\delta}_{i\sim} = \bar{\delta}_{\sim t} - \bar{\delta}_{\sim\sim} \quad \text{and} \quad \mu_{it} - \bar{\delta}_{\sim t} = \bar{\delta}_{i\sim} - \bar{\delta}_{\sim\sim}.$$

Here $\bar{\delta}_{i\sim}$ is the mean of the neighbours in time, $\bar{\delta}_{\sim t}$ is the mean of the neighbours in space, and $\bar{\delta}_{\sim\sim}$ is the mean of the second-order neighbours.

Such a prior model will be suitable if temporal trends are different from county to county, but are more likely to be similar for adjacent counties. For example, it may be considered for non-infectious diseases where unobserved risk factors do have spatio-temporal structure, such as factors which can be attributed to air pollution from a specific source. Furthermore, such a prior might also be useful for modelling the spatio-temporal spread of diseases with an infectious aetiology. Model (2) with δ of type IV can be seen as the non-parametric analogue of the other interaction

model proposed in Bernardinelli *et al.* [1] in which estimates of (linear) temporal trends borrow strength from trends in adjacent counties.

2.3. Hyperpriors

Already in model (1), hyperparameters λ_α, λ_γ, λ_θ and λ_ϕ, which determine the variation of each block, have to be estimated from the data. In addition, λ_δ has to be estimated in model (2). We assign to all such parameters proper gamma priors, say $\lambda \sim \mathrm{G}(a, b)$ with mean a/b and variance a/b^2, to avoid problems with improper hyperpriors. Gamma priors are computationally convenient as the full conditional of λ will again be gamma, for example, λ_α has full conditional $\lambda_\alpha \sim \mathrm{G}(a + \frac{1}{2}rg(K_\alpha), b + \frac{1}{2}\alpha' K_\alpha \alpha)$, where $rg(K_\alpha)$ denotes the rank of K_α. In our application, highly dispersed gamma hyperpriors are chosen for all blocks with values $a = 1$ and $b = 0.01$. In a second run, we studied sensitivity and changed the values to $a = b = 0.01$. However, the fit of all the models got slightly worse (maintaining the order in median posterior deviance) and autocorrelations of the parameter samples increased considerably. For more details on the choice of hyperpriors in disease mapping see Bernardinelli *et al.* [14] and the discussion in Best *et al.* [15].

2.4. Computational issues

Statistical inference has been carried our using `C++` routines developed by the author. We used Markov chain Monte Carlo to sample from the posterior distribution implied by the above formulations, applying univariate Metropolis steps [16] for each parameter whereas hyperparameters were updated with samples from their full conditionals. The number of parameters in interaction models is extremely high, so tuning of the Metropolis steps was done in an automatic fashion. Specifically, the spread of each Metropolis proposal was fixed so that the corresponding acceptance rate of each parameter was around 40 per cent. An alternative to univariate Metropolis updating is block sampling based on conditional prior proposals, suggested in Knorr-Held [13]. This approach is especially useful if parameters are highly correlated in the posterior and has been successfully applied in related models [17, 7]. However, univariate Metropolis sampling is easier to implement and was sufficient in terms of convergence and mixing properties of the algorithm in the application reported here. In fact, the 2 500 samples we have stored have been virtually independent, as we have chosen extremely long run lengths (2 500 000 iterations plus burn-in) for each analysis.

Already the main effects model imposes an identifiability problem, because the overall level can be absorbed by both α and θ. A simple remedy is to recentre both α and θ after each iteration cycle to mean zero, or to omit μ and recenter either α or θ, so that the overall risk level μ will be absorbed by the other block. For type II, III and IV interactions, additional identifiability constraints have to be imposed with the δ_{it} recentred either row-wise (compare Figure 2), column-wise or both, the latter in an iterative loop.

2.5. Modifications and extensions

Several modifications and extensions of the specification (2) are possible. In particular, it is not always necessary to allow for both structured and unstructured type of variation, both in space and time. For example, the modified linear predictor

$$\eta_{it} = \mu + \alpha_t + \theta_i + \phi_i + \delta_{it} \tag{3}$$

might be useful if δ is of type II or IV. Such a model will often be reasonable in practice as temporal trends are typically strong for most diseases, so that the unstructured temporal block γ can be neglected. In general, however, we recommend omitting only those main effects that are not assumed to interact. For illustration, consider model (3) with δ of type II. This model implies that, for each region, both the level and the temporal trend in disease risk is estimated by globally borrowing strength from the other regions. This structure would be destroyed if one of the main effects, α or ϕ, is omitted. For example, without α, temporal trends would be estimated completely separately and the δ parameters can no longer be interpreted as interaction parameters.

An extension of our formulation (2) is to include more than one type of interaction but the model will become rather crude. Waller *et al.* [2] include both type I and type III interaction in a different formulation without main effects. They report, however, that this model turned out to be inferior to simpler specifications in an analysis of the Ohio lung cancer data set.

3. APPLICATION TO OHIO LUNG CANCER DATA

For illustration, we have analysed a data set on mortality from lung cancer among white males between 55 and 64 years, 1968–1988, in the 88 counties of Ohio. Five different model specifications have been implemented: model (1) without any form of space $\times$ time interaction and model (2) with one of the four interaction priors.

In Figure 3 and Table I we report the posterior distribution of the deviance for comparing the fit and the complexity of each model. More specifically, we have calculated the *saturated deviance* (Reference [18], p. 34) [10]

$$D = 2\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it}\log\left(\frac{y_{it}}{n_{it}\pi_{it}}\right) + (n_{it}-y_{it})\log\left(\frac{n_{it}-y_{it}}{n_{it}(1-\pi_{it})}\right)\right)$$

as a functional of unknown parameters. Based on 2 500 samples, the left panel in Figure 3 gives the empirical distribution of D for each model. Smaller values of D indicate a better fitting model. Furthermore, we propose to roughly assess the complexity of the model by the *variation* of the posterior deviance. Table I gives the corresponding deviance summaries (median, mean, interquartile range IQR and standard deviation STD). Note that Spiegelhalter *et al.* [10] propose an alternative measure of model complexity, called p_D, which can be combined with the posterior mean of D to a single model selection criterion, called DIC.

In terms of median (or mean) posterior deviance, the type II interaction model gives the best model fit, followed by type I. Type III is the worst fitting interaction model, not much better than the model without any interaction parameters. Concerning model complexity, not surprisingly, the main effects model has the smallest deviance variation. For the interaction models, the deviance variation seems to be inversely related to the degree of prior dependence for interaction parameters. Indeed, the type I model has the highest deviance variation, followed by types II, III and IV. In passing we note that the ordering of the models in terms of deviance variation has essentially been maintained by the alternative measure of model complexity p_D.

For a diagnostic analysis, we have also calculated the posterior distribution of the *deviance residual* (Reference [18], p. 39)

$$d_{it} = \sqrt{2}\left(y_{it}\log\left(\frac{y_{it}}{n_{it}\pi_{it}}\right) + (n_{it}-y_{it})\log\left(\frac{n_{it}-y_{it}}{n_{it}(1-\pi_{it})}\right)\right)^{\frac{1}{2}} \times \operatorname{sign}(y_{it}-n_{it}\pi_{it})$$

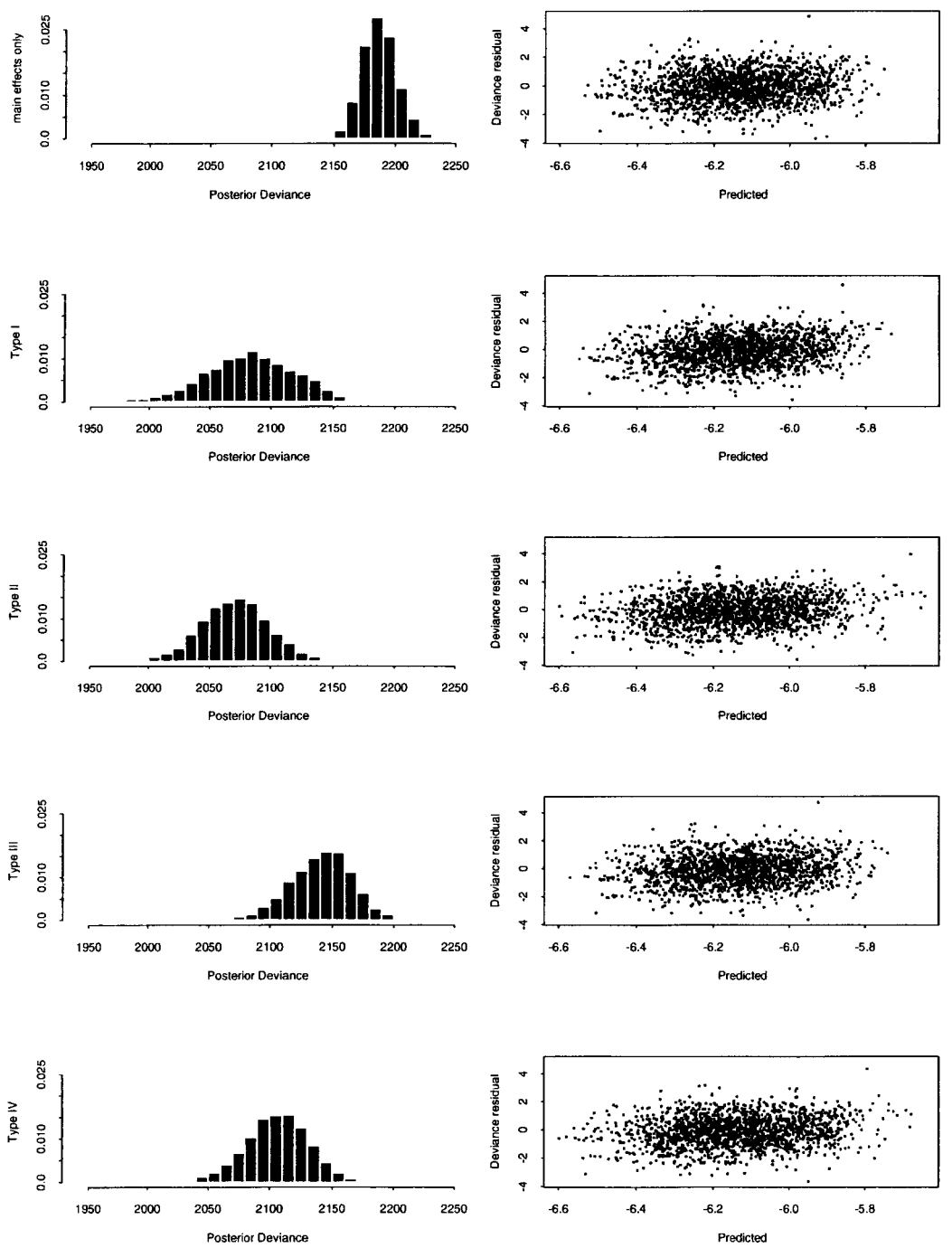

Figure 3. The posterior distribution of the deviance (left panel) and deviance residual versus predicted diagnostics plots (right panel) for the five different models.

Table I. Deviance summaries for Ohio lung cancer data.

Model	Median	Mean	IQR	STD
Main effects only	2187	2187	18.5	13.9
Type I interaction	2083	2082	48.6	35.9
Type II interaction	2071	2071	36.6	27.0
Type III interaction	2142	2141	32.8	24.8
Type IV interaction	2106	2106	32.8	24.7

for each observation (i, t). The right panel of Figure 3 gives the residual versus predicted diagnostics plots, where the posterior median $\hat{d}_{it}$ of the deviance residual is plotted against the posterior median $\hat{\eta}_{it}$ of the linear predictor. The plots for types I and III are nearly identical to the main effects model. The type IV and especially the type II model show more variation of the predicted values and less strong outliers in terms of deviance residuals. In particular, in the type II model, the deviance residual of the most extreme outlier is reduced from 4.86 in the main effects model to 3.94 with the corresponding predicted value increased from -5.95 ($\hat{\pi} = 0.0026$) to -5.68 ($\hat{\pi} = 0.0034$).

We now provide a more detailed look at the results of the type II model as this model gives the best fit with moderate posterior deviance variation and less strong outliers in terms of deviance residuals. Distinct *decreasing* temporal trends of interaction parameters were found for some highly urbanized counties such as Hamilton and Cuyahoga. In fact, for both counties, 80 per cent *simultaneous* credible regions for δ_{it}, $t = 1, \ldots, 21$, did not cover the zero line $\delta_{it} = 0$, $t = 1, \ldots, 21$, which corresponds to the case of no interaction. In contrast, pronounced *increasing* trends were found only for rural counties such as Clermont and Marion, where simultaneous credible regions on similar credible levels did also not cover the zero line. This indicates that the temporal trends in all four counties are substantially different from the global Ohio trend. All simultaneous credible regions have been calculated with the method described in Besag *et al.* [19].

For a closer examination of the temporal change of the estimated spatial risk profile, we define the *adjusted relative risk* by

$$\mathrm{ARR}_{it} = \exp(\theta_i + \phi_i + \delta_{it}) \tag{4}$$

which is automatically calibrated on a common base for the temporal main effects α_t and γ_t. Figure 4 displays the spatial distribution of the posterior median ARRs for the years 1968, 1975, 1982 and 1988. Generally, the spatial pattern does not change much over the years. However, some regions have interesting time trends, for example the two adjacent counties in the south-west corner (Hamilton and Clermont) where opposite trends in disease risk can be detected.

As a further illustration, Figure 5 gives estimated linear predictors $\hat{\eta}_{it}$ (posterior medians) for Hamilton and Clermont county. Each of the four interaction specifications is contrasted with estimates from the main effects model. The logit transformed rates $\mathrm{logit}(y_{it}/n_{it})$ are indicated by dots. It can be seen that the type I and type III model gives estimates hardly distinguishable from the main effects model. In contrast, the estimates from the type II and type IV model are substantially different from the main effects model. Note that type II and type IV estimates are very similar for Hamilton county, whereas for Clermont the estimated trend is less different from the main effects model for type IV interactions. This can be explained by the fact that, in the type IV model, temporal trends of interaction parameters borrow strength from neighbouring counties. Hence, the

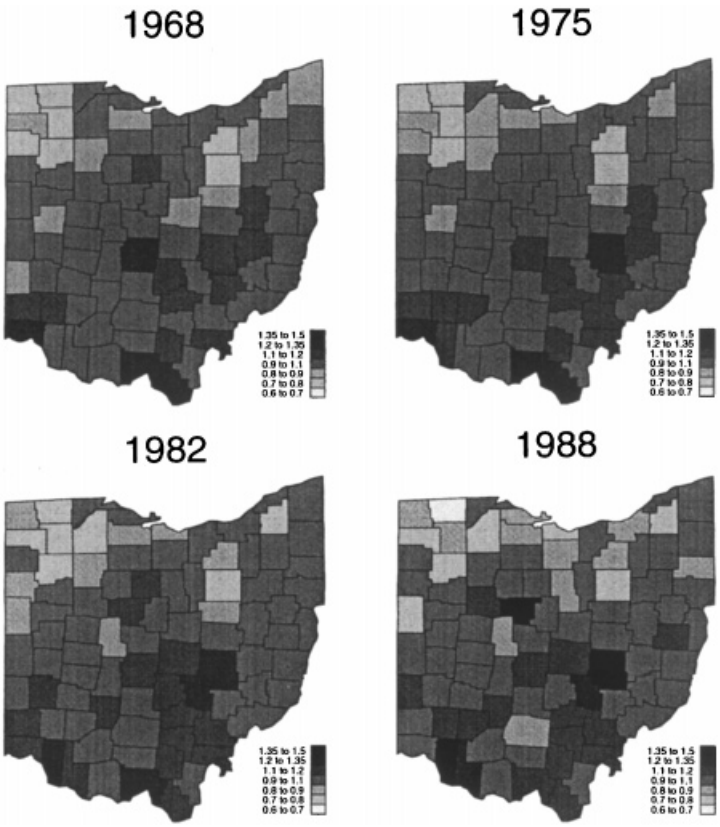

Figure 4. Adjusted relative risk for lung cancer in Ohio.

decreasing trend in Hamilton county causes the estimated increase in Clermont county, which is less populated than Hamilton, to be less pronounced.

Urbanization, as a surrogate for cigarette consumption and other risk factors associated with urban areas, is known to explain part of the spatial variation of lung cancer rates [20]. However, the temporal trends of urbanized and rural areas indicate a changing relationship between urbanization and lung cancer mortality. For each year t, we have therefore calculated the correlation of the adjusted log relative risk (the logarithm of (4)) from the type II model with a simple measure of urbanization, defined as the logarithm of the population size of the largest city in each county in 1970 [20, 3]. The correlation is constantly decreasing from 0.15 (1968) to essentially zero (1986 and later). Our findings therefore seem to confirm a hypothesis by Greenberg [11] that the correlation between urbanization and risk factors for cancer decreases in time. However, the variation in adjusted relative risk among counties in Figure 4 seems to be largest in 1988. In fact, the variance of the estimated log relative risks increases by 50 per cent from 1978 to 1988. We therefore conclude that new latent risk factors, not associated with urbanization, become more and more important.

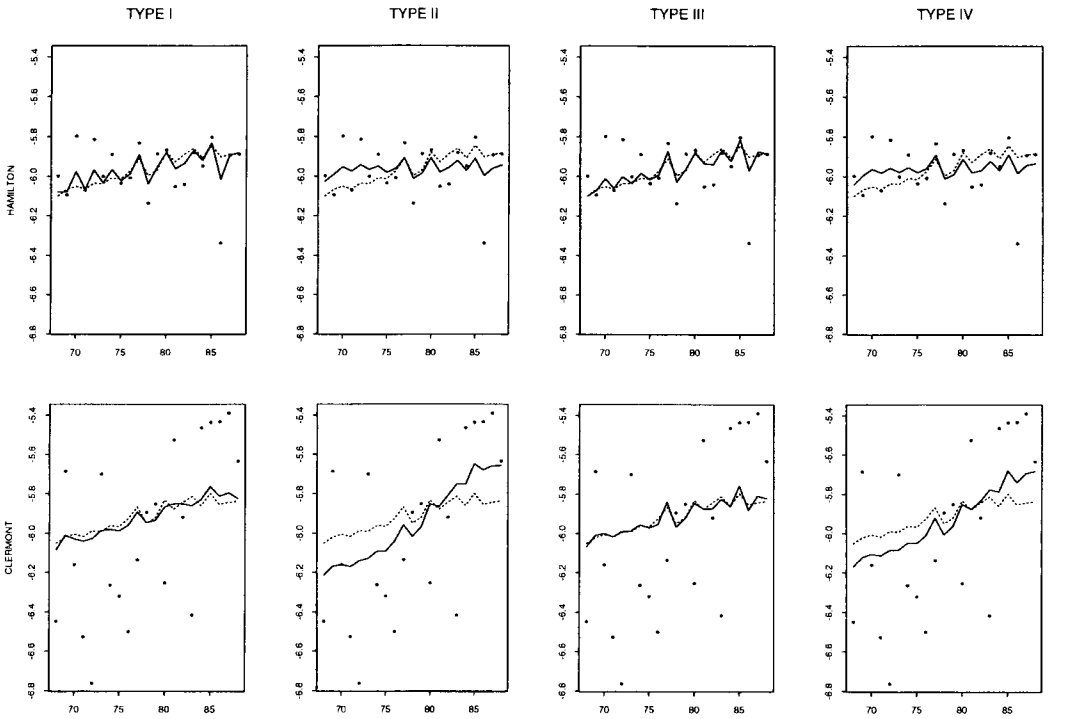

Figure 5. The temporal development of $\hat{\eta}_{it}$ for Hamilton (top row) and Clermont (bottom row) county. The estimates of each interaction type model (solid lines) are contrasted with estimates from the main effects model (dashed lines). The dots are the actually observed rates y_{it}/n_{it} on a logit scale.

4. DISCUSSION

We have proposed several formulations for the analysis of spatio-temporal disease data in the presence of interactions. Our framework is built in the spirit of classical interaction models, where main effects are combined with interaction parameters. One advantage of such an approach is that we are able to simplify the model if interaction turns out to be negligible. If not, we can examine the posterior distribution of interaction parameters in order to identify patterns of the disease variation, which cannot be attributed to the main effects. In our application, simultaneous credible intervals for interaction parameters have been useful to find those counties which do not follow the overall time trend. In this way we have discovered an interesting association between temporal trends of interaction parameters and urbanization, which might deserve further epidemiological research.

We have illustrated how the posterior deviance and deviance residuals can be used for screening the different formulations. The results may provide clues about the aetiology of the disease and give suggestions about the spatio-temporal structure of the underlying risk factors. Of course, the use of location and variation of the posterior deviance as a measure of model fit and complexity, respectively, is slightly *ad hoc*. Another interesting aspect is how to combine those two quantities into a single model selection criterion. Nevertheless, in our application the posterior deviance has given sensible results and turned out to be very helpful for model selection. In addition, deviance residuals have been useful for a more detailed look at the fit of individual observations.

We have concentrated on the situation where there is only one observation for each pixel in space $\times$ time. Suppose now the data are further stratified by age, which is rather common in descriptive epidemiology. A combination of Bayesian age–period–cohort models [21, 19] with Bayesian spatial models [5] might then be useful and Clayton's rule is a guideline for the specification of interaction priors. The author currently investigates models where cohort or period effects are allowed to interact with space (type II and type IV). However, there are lots of other possible formulations. Model selection criteria and model diagnostics, such as those we have used in our application, will be necessary in selecting an appropriate model from the many possible formulations.

ACKNOWLEDGEMENTS

This research was supported by the German Science Foundation (DFG), SFB 386. The author expresses thanks to Julian Besag for helpful discussions, which have initiated this work, to Markus Rieß, a former student, for realizing the potential of the Clayton approach in his diploma thesis on space–time modelling of disease rates, and to two referees for helpful comments.

REFERENCES

1. Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine* 1995; **14**:2433–2443.
2. Waller LA, Carlin BP, Xia H, Gelfand AE. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* 1997; **92**:607–617.
3. Knorr-Held L, Besag J. Modelling risk from a disease in time and space. *Statistics in Medicine* 1998; **17**:2045–2060.
4. Clayton DG, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987; **43**:671–681.
5. Besag JE, York JC, Mollié A. Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* 1991; **43**:1–59.
6. West M, Harrison PJ. *Bayesian Forecasting and Dynamic Models*, 2nd edn, Springer Verlag: New York, 1997.
7. Fahrmeir L, Knorr-Held L. Dynamic and semiparametric models. In *Smoothing and Regression: Approaches, Computation and Application*, Schimek M (ed.). Wiley: New York, 2000.
8. Clayton DG, Bernardinelli L. Bayesian methods for mapping disease risks. In Cuzick J, Elliot P (eds). *Small Area Studies in Geographical and Environmental Epidemiology*, Oxford University Press, pp. 205–220 (1992).
9. Clayton DG. Generalized linear mixed models. In *Markov Chain Monte Carlo in Practice*, Gilks WR, Richardson S, Spiegelhalter DJ. (eds). Chapman & Hall: London, 1996; 275–301.
10. Spiegelhalter DJ, Best NG, Carlin BP. Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. Technical Report, MRC Biostatistics Unit, Cambridge, U.K., 1998.
11. Greenberg MR. *Urbanization and Cancer Mortality*. Oxford University Press: New York, 1983.
12. Stone CJ, Hansen M, Kooperberg C, Troung YK. Polynomial splines and their products in extended linear modeling (with discussion). *Annals of statistics*, 1997; **25**:1371–1470.
13. Knorr-Held L. Conditional prior proposals in dynamic models. *Scandinavian Journal of Statistics* 1999; **26**:129–144.
14. Bernardinelli L, Clayton D, Montomoli C. Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine* 1995; **14**:2411–2431.
15. Best NG, Arnold RA, Thomas A, Waller LA, Conlon EM. Bayesian models for spatially correlated disease and exposure data (with discussion). In *Bayesian Statistics 6*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford University Press, 1999; 131–156.
16. Smith AFM, Roberts GO. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society*, *Series B* 1993; **55**:3–23.
17. Fahrmeir L, Knorr-Held L. Dynamic discrete time duration models: estimation via Markov chain Monte Carlo. *Sociological Methodology* 1997; **27**:417–452.
18. McCullagh P, Nelder JA. *Generalized Linear Models*, 2nd edn. Chapman & Hall: London, 1989.
19. Besag JE, Green PJ, Higdon DM, Mengersen KL. Bayesian computation and stochastic systems (with discussion). *Statistical Science* 1995; **10**:3–66.
20. Kafadar K, Tukey JW. U.S. cancer rates: a simple adjustment for urbanization. *International Statistical Review* 1993; **61**:257–281.
21. Berzuini C, Clayton D. Bayesian analysis of survival on multiple time scales. *Statistics in Medicine* 1994; **13**:823–838.

