



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Knorr-Held:

Bayesian Modelling of Inseparable Space-Time Variation in Disease Risk

Sonderforschungsbereich 386, Paper 147 (1999)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



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 \times 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 1968–88 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 \times 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, 2, 3] propose extensions of the purely spatial models by Clayton and Kaldor^[4] and Besag *et al.*^[5] to space \times 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.

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 nonparametric estimation of temporal trends in disease risk including time-changing effects of covariates. In particular, they do not assume linearity nor 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 nonparametrically. 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 time and space, 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 nonparametric analogue of the Bernardinelli *et al.* formulation.

The proposed modelling framework is outlined in Section 2. We also include some details on the implementation by Markov chain Monte Carlo (MCMC) and outline several modifications and extensions. Section 3 describes an analysis of a dataset on mortality from lung cancer among white males between 55 and 64 years, for 21 successive years in the 88 counties of Ohio. This is a subset of a dataset 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 county i ($i = 1, \dots, n$) and year t ($t = 1, \dots, T$). We assume that the number of cases or deaths y_{it} , during year t , has a binomial distribution with parameters n_{it} and 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 or do not display temporal structure a priori. Similarly, θ_i and ϕ_i represent unspecified features of county i that respectively do or do not display spatial structure.

The formulation (1) is completed by assigning prior distributions to the four blocks $\alpha = (\alpha_1, \dots, \alpha_T)'$, $\gamma = (\gamma_1, \dots, \gamma_T)'$, $\theta = (\theta_1, \dots, \theta_n)'$ and $\phi = (\phi_1, \dots, \phi_n)'$ of η_{it} . Each prior is assumed to be multivariate Gaussian with mean zero and precision matrix κK , where κ is an unknown scalar 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

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

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

$$\eta_{it} = \alpha_i + \gamma_t + \theta_i + \phi_i + \delta_{it}. \quad (2)$$

The parameter vector $\delta = (\delta_{11}, \dots, \delta_{nT})'$ is assumed to be Gaussian with precision matrix $\kappa_\delta K_\delta$. As for the main effects, κ_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 to specify 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, $2 \cdot 2 = 4$ combinations 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. We discuss now each type separately, ordered by the degree of prior dependence.

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:

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

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

Type II Interaction. If we combine the random walk main effect α with the unstructured block ϕ by Clayton's rule, then each $\delta_i = (\delta_{i1}, \dots, \delta_{iT})'$, $i = 1 \dots, n$, follows a random walk, independently of all other counties. The structure matrix K_δ has rank $n \cdot (T - 1)$ and the

prior for δ can be written as

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

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.

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

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

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

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 time and space and can no longer be factorized into independent blocks.

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

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

with independent contrasts $\delta_{it} - \delta_{jt} - \delta_{i,t-1} + \delta_{j,t-1}$. This distribution is invariant to level transformations of the type

$$\begin{aligned} \tilde{\delta}_{it} &= \delta_{it} + c_i, & t = 1, \dots, T, \\ \tilde{\delta}_{it} &= \delta_{it} + c_t, & i = 1, \dots, n, \end{aligned}$$

for any constants $c_1, \dots, c_n, c_1, \dots, c_T$ and is therefore highly rank-deficient (K_δ has rank $(n-1) \cdot (T-1)$).

The conditional distribution of a pixel δ_{it} , given all the others, which can be derived^[13] from $K_\delta = K_\alpha \otimes K_\theta$ has mean

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

and precision

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

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 $\delta_{j,t+1}$, $j \sim i$), i.e. 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 diseases with an infectious aetiology.

2.3 Hyperpriors

Already in model (1), hyperparameters κ_α , κ_γ , κ_θ and κ_ϕ , which determine the variation of each block, have to be estimated from the data. In addition, κ_δ has to be estimated in model (2). We assign to all such parameters proper gamma priors, say $\kappa \sim G(a, b)$, to avoid problems with improper hyperpriors. Gamma priors are computationally convenient as the full conditional of κ will again be gamma, for example κ_α has full conditional $\kappa_\alpha \sim G(a + \frac{1}{2}rg(K_\alpha), b + \frac{1}{2}\alpha'K_\alpha\alpha)$, 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.

2.4 Computational Issues

We used Markov chain Monte Carlo to sample from the posterior distribution implied by the above formulation, applying univariate Metropolis steps^[14] 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%. 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.^[15, 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 omit μ and recentre either the α_t or the θ_i after each iteration cycle to mean zero, so that the overall risk level μ will be absorbed by the other block. We have centred θ both in model (1) and (2). For Type II, III and IV interactions, additional identifiability constraints have to be imposed with the δ_{it} recentred either column-wise, row-wise or both, the latter in an iterative loop.

2.5 Modifications and Extensions

Several modifications and extensions of the specification (2) are possible. For example, the modified linear predictor

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

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 to omit 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 dataset.

3. Application to Ohio lung cancer data

For illustration, we have analysed a dataset on mortality from lung cancer among white males between 55 and 64 years, 1968-1988, in the 88 counties of Ohio. Five different model specification 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 1, 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*^[16, 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 2500 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 1 gives the corresponding deviance summaries (median, mean, interquartile range and standard deviation).

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 Type II, III and IV.

For a diagnostic analysis, we have also calculated the posterior distribution of the *deviance residual*^[16]

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

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 Type 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 ($\hat{\pi} = .0026$) to -5.68 ($\hat{\pi} = .0034$).

We now provide a more detailed look at the results of the Type II model. Distinct *decreasing* temporal trends of interaction parameters were found for some highly urbanized counties such as Hamilton and Cuyahoga. In fact, for both counties, the interaction parameters were significantly different from zero for *all* time points $t = 1, \dots, 21$, based on 80% *simultaneous* credible regions, calculated with the method described in Besag *et al.*^[17] In contrast, pronounced *increasing* trends were found only for rural counties such as Clermont and Marion (simultaneously significant based on 75% and 85% credible regions, respectively). Figure 4 displays for the years 1968, 1975, 1982 and 1988 the spatial distribution of the estimated relative risk

$$\text{ARR}_{it} = \exp(u_i + v_i + \delta_{it}), \tag{4}$$

adjusted for the overall temporal trend α_t . Generally, the spatial pattern does not change much over the years but the heterogeneity of the estimates seems to increase. Note that 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 $\text{logit}(y_{it}/n_{it})$ are indicated by dots. It can be seen that the Type III model gives estimates, hardly distinguishable from the main effects model. Similar holds for Type I, especially for less populated Clermont county. Type II model estimates are different displaying the above mentioned trends of interaction parameters. 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 Type IV models, temporal trends of interaction parameters borrow strength from neighbouring counties. Hence, the decreasing trend in Hamilton county causes the estimated increase in Clermont county 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.^[18] 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.^[18, 3] The correlation is constantly decreasing from 0.15 (1968) to essentially zero (1986 and later), see Figure 6. Our findings therefore confirm a hypothesis by Greenberg,^[11] that the correlation between urbanization and risk factors for cancer *decreases* in time. However, note that the heterogeneity in relative risk among counties is increasing over the years. We therefore conclude that new latent risk factors, not associated with urbanization, become more and more important.

4. Discussion

We have proposed several formulations for the analysis of spatio-temporal disease data in the presence of interactions and have illustrated how the posterior deviance and deviance residuals can be used for model comparison. 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. Furthermore, simultaneous credible regions for interaction parameters are useful in identifying regions, which do not follow the overall time trend. In our application, we have found an interesting association between temporal trends of interaction parameters and urbanization, which might deserve further epidemiological research.

We have concentrated on the situation, where there is only one observation for each pixel in time \times space. Suppose now, the data are further stratified by age, which is rather common in descriptive epidemiology. A combination of Bayesian age-period-cohort models^[19, 17] with Bayesian spatial models^[5] might 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, and 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.

References

- [1] Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M. ‘Bayesian analysis of space–time variation in disease risk’, *Statistics in Medicine*, **14**, 2433–2443 (1995).
- [2] Waller, L. A., Carlin, B. P., Xia, H. and Gelfand, A. E. ‘Hierarchical spatio–temporal mapping of disease rates’, *Journal of the American Statistical Association*, **92**, 607–617 (1997).
- [3] Knorr-Held, L. and Besag, J. ‘Modelling Risk from a Disease in Time and Space’, *Statistics in Medicine* **17**, 2045–2060 (1998).
- [4] Clayton, D. G. and Kaldor, J. ‘Empirical Bayes estimates of age–standardized relative risks for use in disease mapping’, *Biometrics*, **43**, 671–681 (1987).
- [5] Besag, J. E., York, J. C. and Mollié, A. ‘Bayesian image restoration with two applications in spatial statistics’ (with discussion), *Annals of the Institute of Statistical Mathematics*, **43**, 1–59 (1991).
- [6] West, M. and Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models* (2nd Edition). New York: Springer Verlag.
- [7] Fahrmeir, L. and Knorr-Held, L. ‘Dynamic and semiparametric models’, In: M. Schimek (ed.), *Smoothing and Regression: Approaches, Computation and Application*, New York: John Wiley & Sons, to appear, 1999.
- [8] Clayton, D. G. and Bernardinelli, L. ‘Bayesian methods for mapping disease risks’, In: J. Cuzick and P. Elliot (eds), *Small Area Studies in Geographical and Environmental Epidemiology*, Oxford University Press, pp. 205–220 (1992).

- [9] Clayton, D. G. ‘Generalized linear mixed models’, *in*: Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall, 1996, 275–301.
- [10] Spiegelhalter, D. J., Best, N. G. and Carlin, B. P. ‘Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models’, Technical Report, MRC Biostatistics Unit, Cambridge, UK (1998).
- [11] Greenberg, M. R. *Urbanization and Cancer Mortality*, New York: Oxford University Press, 1983.
- [12] Stone, C. J., Hansen, M., Kooperberg, C. and Troung, Y. K. ‘Polynomial splines and their tensor products in extended linear modeling’ (with discussion), *Annals of Statistics*, **25**, 1371–1470 (1997).
- [13] Knorr-Held, L. ‘Conditional Prior Proposals in Dynamic Models’, *Scandinavian Journal of Statistics* **26**, 129–144 (1999).
- [14] Smith, A. F. M. and Roberts, G. O. ‘Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods’ (with discussion), *Journal of the Royal Statistical Society B*, **55**, 3–23 (1993).
- [15] Fahrmeir, L. and Knorr-Held, L. ‘Dynamic discrete time duration models: estimation via Markov chain Monte Carlo’, *Sociological Methodology*, **27**, 417–452 (1997).
- [16] McCullagh, P. and Nelder, J. A. *Generalized Linear Models* (2nd Edition), London: Chapman & Hall, 1989.
- [17] Besag, J. E., Green, P. J., Higdon, D. M. and Mengersen, K. L. ‘Bayesian computation and stochastic systems’ (with discussion), *Statistical Science*, **10**, 3–66 (1995).
- [18] Kafadar, K. and Tukey, J. W. ‘U.S. cancer rates: a simple adjustment for urbanization’, *International Statistical Review*, **61**, 257–281 (1993).

- [19] Berzuini, C. and Clayton, D. ‘Bayesian analysis of survival on multiple time scales’, *Statistics in Medicine*, **13**, 823–838 (1994).

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

Table 1: Deviance summaries for Ohio lung cancer data.

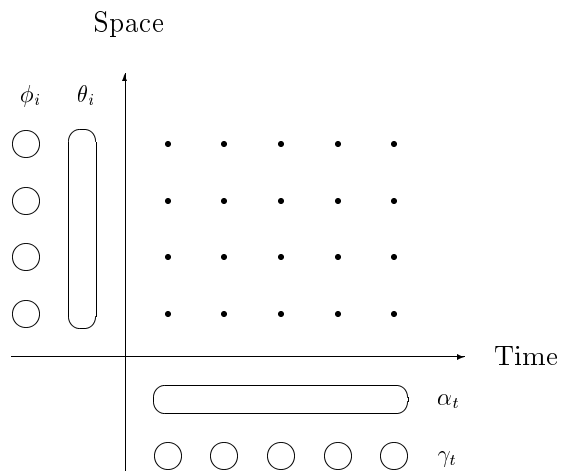


Figure 1: Symbolic representation of the main effect model. Circles represent prior independence, ovals represent prior dependence. Observations in time \times space are indicated by small dots.

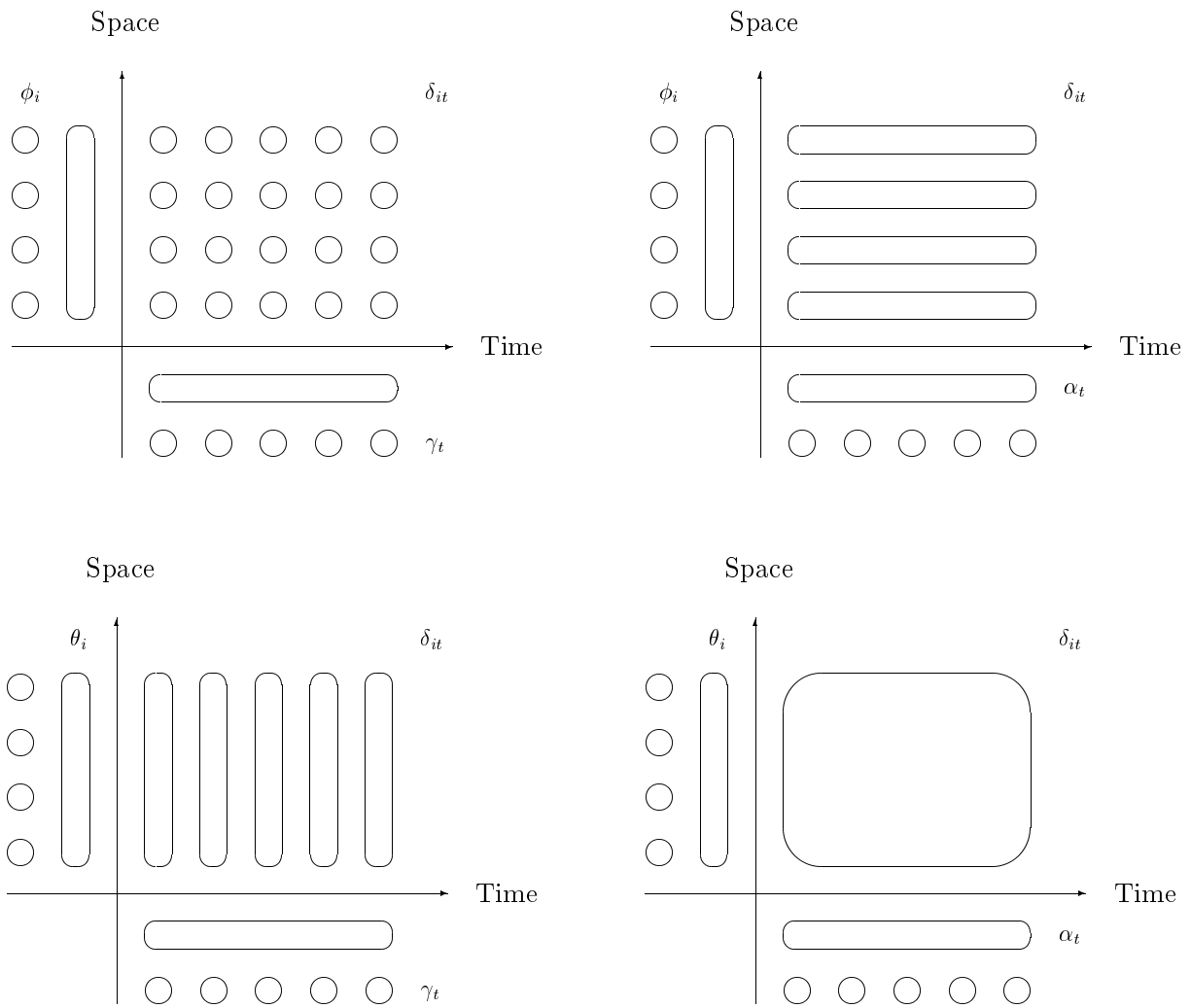


Figure 2: Symbolic representation of the four possible types of interactions. Circles represent prior independence, ovals represent prior dependence.

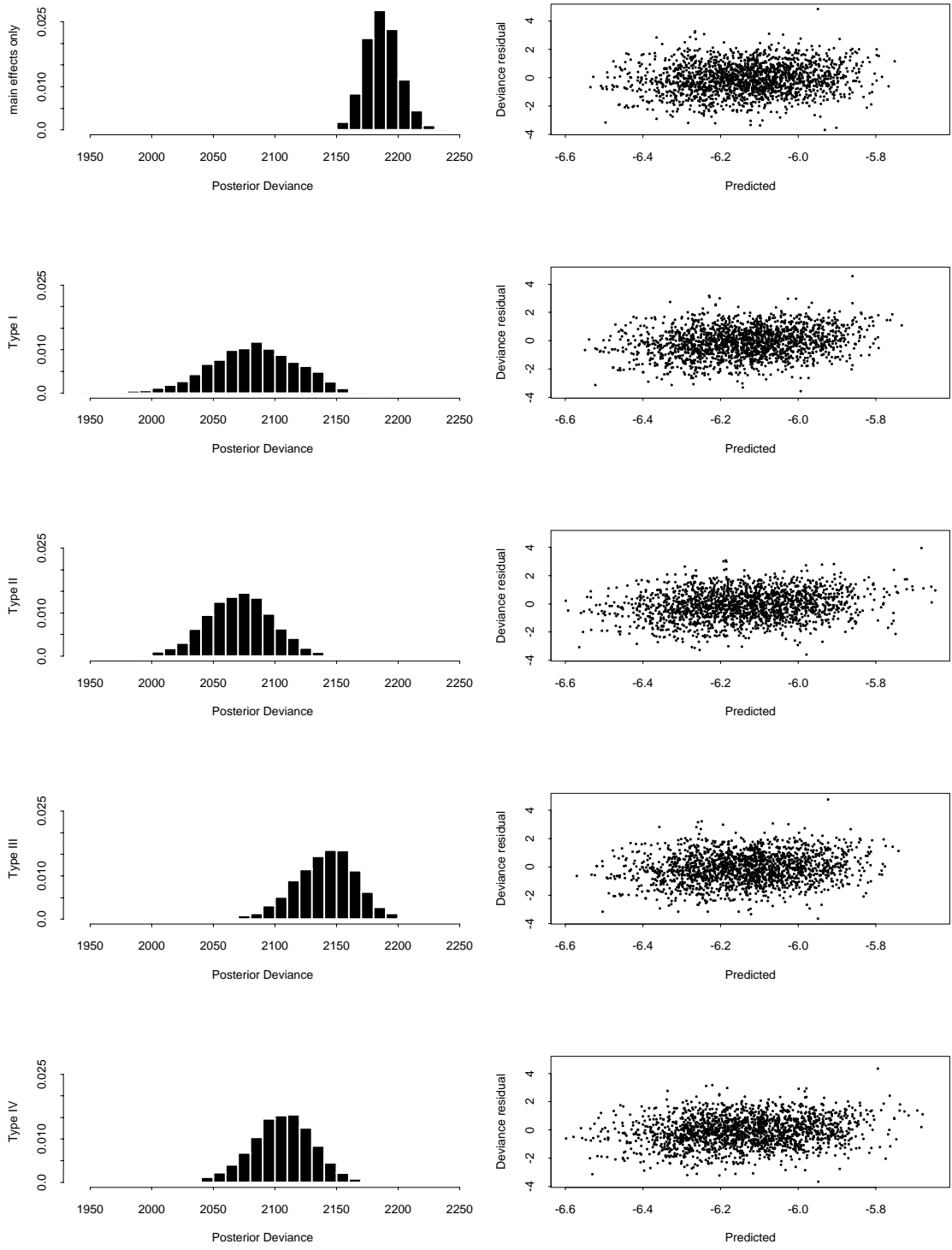
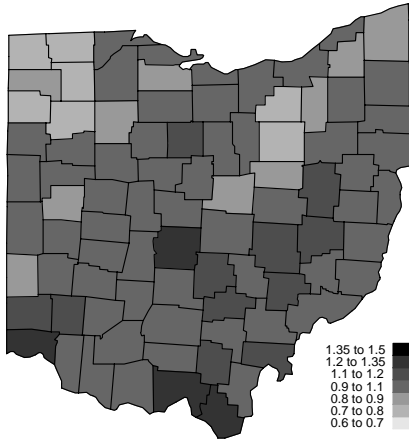
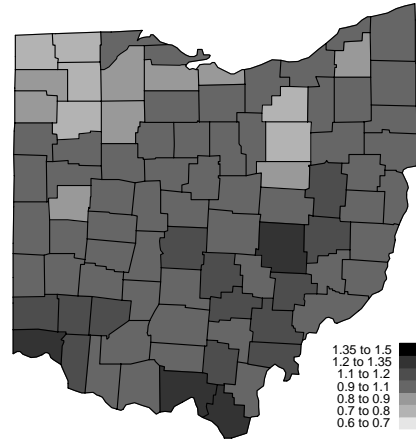


Figure 3: The posterior distribution of the deviance (left panel) and deviance residual versus predicted diagnostics plots (right panel) for the five different models.

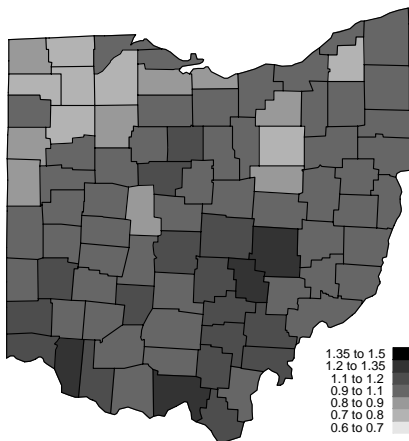
1968



1975



1982



1988

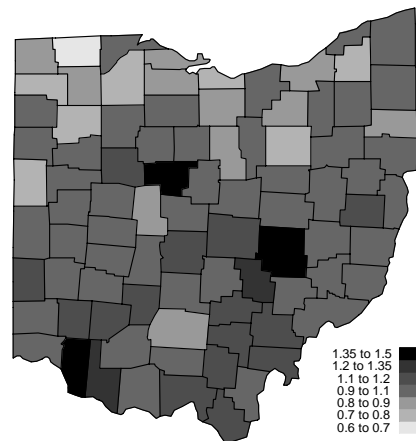


Figure 4: Adjusted relative risk for lung cancer in Ohio.

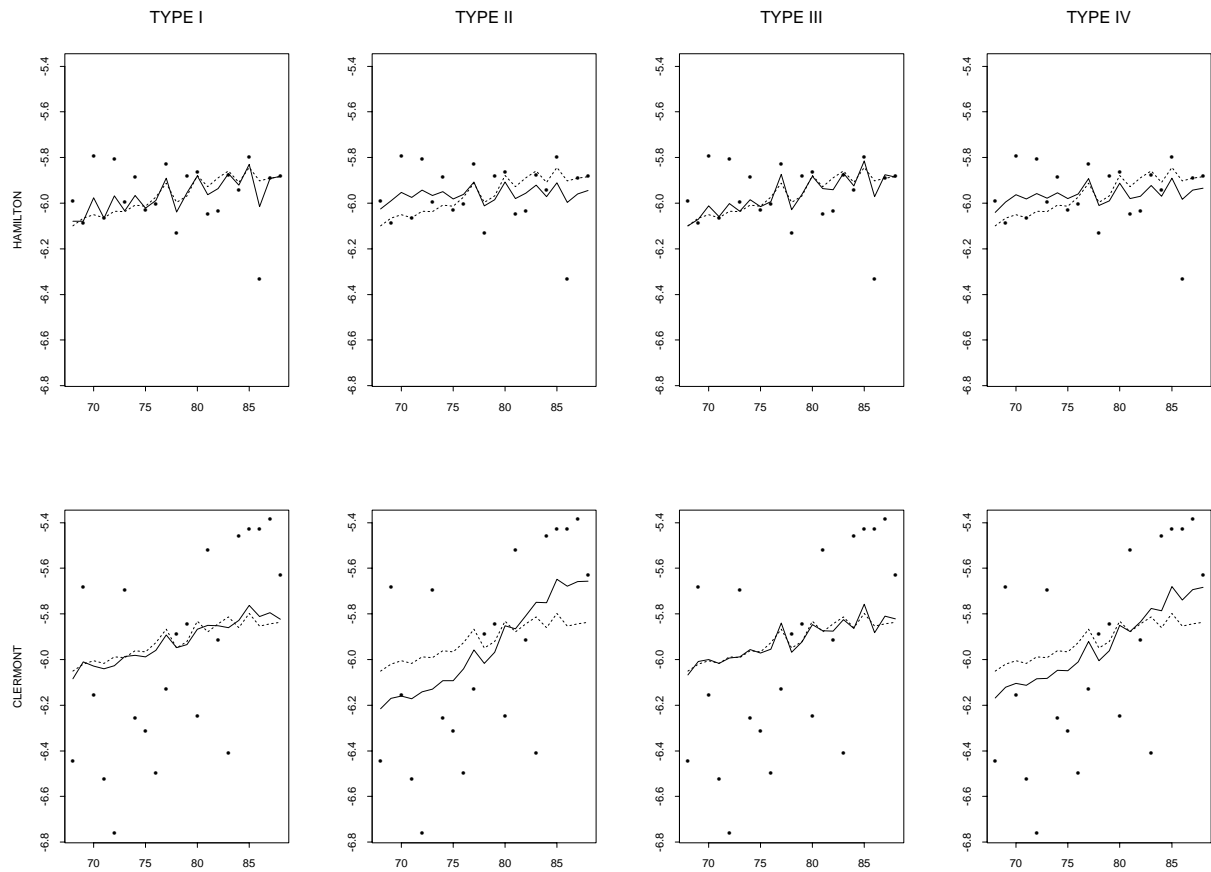


Figure 5: The temporal development of $\hat{\eta}_{it}$ for Hamilton and Clermont county. The estimates of each interaction type model (solid lines) are contrasted with estimates from the main effects model (dashed lines). The dots are the actually observed rates y_{it}/n_{it} on a logit scale.

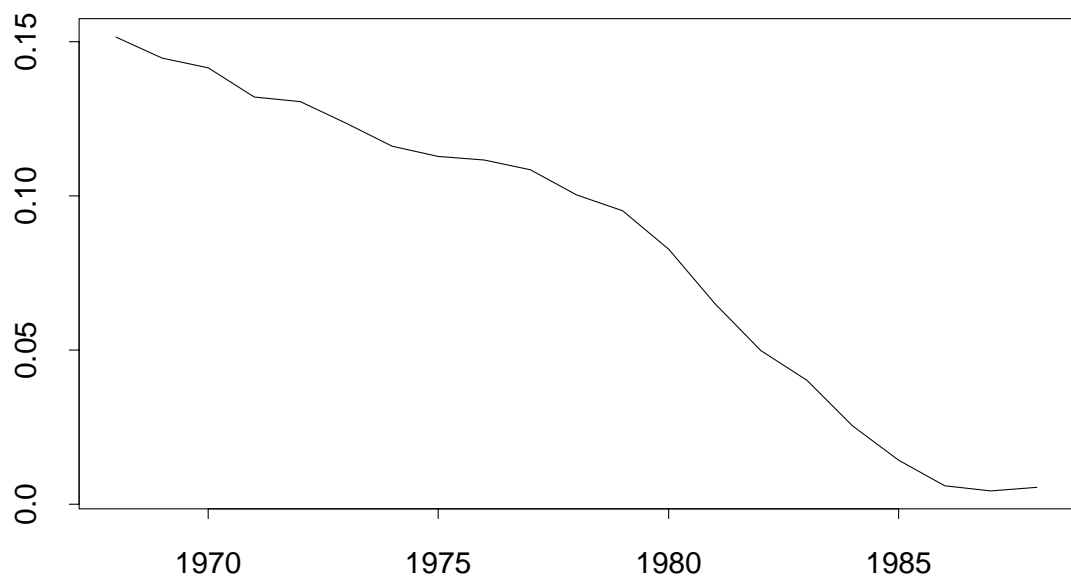


Figure 6: Time series plot of the correlation between the estimated adjusted relative risk and urbanization.