Hierarchical Spatio–Temporal Mapping of Disease Rates

Lance A. WALLER, Bradley P. CARLIN, Hong XIA, and Alan E. GELFAND

Maps of regional morbidity and mortality rates are useful tools in determining spatial patterns of disease. Combined with sociodemographic census information, they also permit assessment of environmental justice; that is, whether certain subgroups suffer disproportionately from certain diseases or other adverse effects of harmful environmental exposures. Bayes and empirical Bayes methods have proven useful in smoothing crude maps of disease risk, eliminating the instability of estimates in low-population areas while maintaining geographic resolution. In this article we extend existing hierarchical spatial models to account for temporal effects and spatio–temporal interactions. Fitting the resulting highly parameterized models requires careful implementation of Markov chain Monte Carlo (MCMC) methods, as well as novel techniques for model evaluation and selection. We illustrate our approach using a dataset of county-specific lung cancer rates in the state of Ohio during the period 1968–1988.

KEY WORDS: Environmental justice; Identifiability; Metropolis algorithm; Model selection.

1. INTRODUCTION

A growing concern in the storage and disposal of hazardous substances is the concept of "environmental justice" (also known as "environmental equity"); that is, whether exposures to and adverse outcomes from environmental hazards are not unduly experienced within various sociodemographic subgroups. Subpopulations of specific interest include historically disadvantaged groups such as low-income groups or certain racial and ethnic minorities. Other populations of interest are those that might be more sensitive to the health effects associated with environmental exposures, such as children and expectant mothers.

Environmental justice is becoming an important part of the federal environmental regulatory process and plays a role in the setting of priorities for federal intervention. The U.S. Environmental Protection Agency established the Office of Environmental Equity in November of 1992 (Sexton, Olden, and Johnson 1993). In February 1994, President Clinton signed the Environmental Justice Executive Order, which requires an assessment of environmental justice in regulation decisions made by every federal agency involved with environmental and public health.

Sound statistical methodology for assessing environmental justice (or detecting environmental "injustices") is clearly needed but has been slow to develop. The assessment of environmental justice is a complex problem with many interesting statistical components (Wagener and Williams 1993). The three primary components of environmental justice are (1) exposure assessment at any location within a geographic region, (2) accurate estimation of sociodemographic variables across the same geographic region, and (3) disease (or other outcome of interest) incidence in small subregions of the same region. We concern ourselves primarily with modeling associations between components (2) and (3). Other work focuses on associations between components (1) and (2) (see, e.g., Anderton, Anderson, Oakes, and Fraser 1994) and between components (1) and (3) (e.g., bioassay). A full assessment of environmental justice will require a melding of developments from each of these three areas.

The data associated with the three components of environmental justice often are available in different formats. Exposure data over large regions are obtained from either monitoring stations or reported releases such as the U.S. Environmental Protection Agency's Toxic Chemical Release Inventory (Stockwell, Sorenson, Eckert, and Carreras 1993). Due to confidentiality concerns, disease incidence data in the United States most often are available as summary counts or rates from geographic regions forming a partition of the study area. Census regions such as counties or census tracts are often used. Demographic data are available as regional summaries again due to confidentiality requirements. The different data types require different types of analysis, and the past work on the various components of environmental justice revolves around methods appropriate for the data at hand.

Several geostatistical methods have been proposed for analysis of the spatial pattern of disease. In these methods, disease risk is modeled as a continuous surface over the study area. Classical approaches model dependence among observations as part of the likelihood, whereas Bayesian approaches typically assume observations to be conditionally independent given the model parameters, and subsequently incorporate dependence at the second stage as part of the prior distribution on the parameters. Carrat and Valleron (1992) and Webster, Oliver, Muir, and Mann (1994) explored the application of spatial prediction (kriging) to modeling disease rates or risks as a continuous spatial variable. Carrat and Valleron (1992) provided interesting time-sequenced predicted surfaces showing the spread of an

Lance A. Waller is Assistant Professor, Bradley P. Carlin is Associate Professor, and Hong Xia is a graduate student and Research Assistant, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455. Alan E. Gelfand is Professor, Department of Statistics, University of Connecticut, Storrs, CT 06269. This research was supported in part by University of Minnesota Graduate School Grant-in-Aid of Research, Artistry and Scholarship #15196 (Waller), National Institute of Allergy and Infectious Diseases FIRST Award 1-R29-AI33466 (Carlin and Xia), and National Science Foundation grant DMS-9301316 (Gelfand). The authors thank Owen Devine for furnishing the Ohio lung cancer dataset, as well as for frequent helpful advice.

influenza-like disease across France (although the accuracy of these predictions may be questionable, because some key assumptions for ordinary kriging are violated with rate data). Handcock and Wallis (1994) presented a Bayesian formulation of space–time kriging models for meterological data using Gaussian random fields. Random covariance parameters allow varying spatial dependence over time. Sampson and Guttorp (1992) considered kriging in time and space where the covariance structure is nonstationary (i.e., dependent on location). They proposed methods for finding a (possibly higher-dimensional) deformation of the spatial domain within which the covariance is stationary (i.e., a function of distance alone). In contrast to the approach of Handcock and Wallis (1994), this method assumes a consistent spatial covariance structure across time.

Other approaches estimate rates and counts for administrative districts, usually producing a rate map. Suppose that the rate scale is divided into exclusive and exhaustive sets, with a distinct color or gray level attached to each set in some monotonic fashion. A disease map is created by drawing a map of the regions that reflects the resultant color assigned to each region. Collections of maps for various diseases and/or geographic regions are often published together in a single atlas. (See Walter and Birnie 1991 for a comparison of 49 disease atlases published during the last 20 years.)

Crude rates (particularly for rare events such as cancers) are particularly unstable for regions with only a few thousand individuals at risk. United States census tracts typically contain 1,000–4,000 individuals. Bayesian and empirical Bayesian methods similar to those proposed for small-area analysis provide mechanisms for stabilizing or smoothing regional estimates of rates.

Clayton and Kaldor (1987) and Manton et al. (1989) outlined empirical Bayes approaches that account for spatial similarities among nearby rates. Devine, Halloran, and Louis (1994), Devine and Louis (1994), Devine, Louis, and Halloran (1994), and Ghosh (1992) considered constrained empirical Bayes approaches which guard against overshrinkage of the raw rates towards their grand mean. Cressie and Chan (1989) presented an exploratory spatial analysis of regional counts of sudden infant death syndrome (SIDS). The regional aspect of the data drives their analysis, which is based on conditionally autoregressive model specifications for data from a spatial lattice proposed by Besag (1974). Clayton and Bernardinelli (1992) reviewed Bayesian methods for modeling regional disease rates. Besag, York, and Mollié (1991) began with the Clayton and Kaldor (1987) approach and extended it to a model for separating spatial effects from overall heterogeneity in the rates. Although their parameterization is convenient for interpretation, it creates several implementation challenges (see Sec. 4 herein). Bernardinelli and Montomoli (1992) compared empirical and hierarchical Bayes methods, the latter implemented via Markov chain Monte Carlo (MCMC) methods. Breslow and Clayton (1993) placed the disease mapping problem within the larger framework of generalized linear mixed models and provide approximation schemes for inference. Most recently, Bernardinelli, Pascutto, Best, and Gilks (1997) used a spatial ecological regression model to account for imprecisely observed covariates.

Our contribution lies in extending the spatial models described by Besag et al. (1991) to accommodate general temporal effects, as well as space–time interactions. We give a hierarchical framework for modeling regional disease rates over space and time in Section 2. Our work also extends that of Bernardinelli, Clayton et al. (1995), who proposed a particular multiplicative form for space–time interaction. In Section 3 we provide a novel method for model comparison and selection using a predictive criterion. We present computational issues related to these models in Section 4, giving special attention to parameter identifiability and the associated convergence of our MCMC algorithms. In Section 5 we present an illustrative example using annual county-specific lung cancer rates by sex and race in Ohio for the period 1968–1988. Finally, we summarize our findings and discuss future implementation and interpretation issues in Section 6.

2. STATISTICAL MODELS

2.1 Overall Structure

As noted in Section 1, disease incidence data are generally available as summary counts or rates for a formally defined region such as a county, district, or census tract. In fact, we assume that within a region, counts or rates are observed for subgroups of the population defined by sociodemographic variables relevant for investigations of environmental equity, such as gender, race, age, and ethnicity. In our Section 5 example we have four subgroups, arising through classification by gender (male or female) and race (white or nonwhite). We further assume that for each subgroup within each region, counts or rates are collected within regular time intervals; for example, annually, as in our example. In general notation an observation can be denoted by y_{ilt}, where $i = 1, \ldots, I$ indexes the regions, $l = 1, \ldots, L$ indexes the subgroups, and $t = 1, \ldots, T$ indexes the time periods. In application, subgroups are defined through factor levels, so that if we used, say, three factors, subscript l would be replaced by jkm.

In the sequel we assume that y_{ilt} is a count arising from an associated relative risk ψ_{ilt}. Hence Bayesian modeling involves two stages: We must specify a likelihood model for the vector of observed counts $\mathbf{y}$ given the vector of relative risks $\boldsymbol{\psi}$, and we must specify a prior model over the space of possible $\boldsymbol{\psi}$'s. Sophisticated MCMC computational algorithms yield a posterior for $\boldsymbol{\psi}$ given $\mathbf{y}$. The set of posterior means or medians of the ψ_{ilt} is then used to create a disease map, as described in Section 1.

Typically, the objective of the prior specification is to stabilize rate estimates by *smoothing* the crude map, although environmental equity concerns imply interest in *explaining* the rates. The crude map arises from the likelihood model alone using, for each (i, l, t), estimates (usually the maximum likelihood estimate [MLE]) of ψ_{ilt} based only on y_{ilt}. Such crude maps often feature large outlying relative risks in sparsely populated regions, so that the map is visually

dominated by the rates with the highest uncertainty. This likelihood approach also fails to account for the anticipated similarity of relative risks in nearby or adjacent regions. Hence, as we shall elaborate, an appropriate prior for ψ will incorporate exchangeability and/or spatial assumptions, enabling the customary Bayesian "borrowing of strength" to achieve the desired smoothing.

The likelihood model assumes that, given ψ, the counts y_{ilt} are conditionally independent Poisson variables. The Poisson serves as an approximation to a binomial distribution, say $y_{ilt} \sim \text{Bin}(n_{ilt}, p_{ilt})$, where n_{ilt} is the known number of individuals at risk in county i within subgroup l at time t and p_{ilt} is the associated disease rate. Often the counts are sufficiently large so that y_{ilt} or perhaps $\sqrt{y_{ilt}}$ (or the Freeman–Tukey transform, $\sqrt{y_{ilt}} + \sqrt{y_{ilt}+1}$) can be assumed to approximately follow a normal density. In our case, however, partitioning into subgroups results in many small values of y_{ilt} (including several 0's), so we confine ourselves to the Poisson model; that is, $y_{ilt} \sim \text{Po}(n_{ilt}p_{ilt})$.

We reparameterize to relative risk by writing $E_{ilt}\psi_{ilt} = n_{ilt}p_{ilt}$, where E_{ilt} is the expected count in region i for subgroup l at time t. In particular, if p^* is an overall disease rate, then $\psi_{ilt} = p_{ilt}/p^*$ and $E_{ilt} = n_{ilt}p^*$. The ψ's are said to be *externally standardized* if p^* is obtained from another data source, such as a standard reference table (see, e.g., Bernardinelli and Montomoli 1992). The ψ's are said to be *internally standardized* if p^* is obtained from the given dataset; for example, if $p^* = \sum_{ilt} y_{ilt} / \sum_{ilt} n_{ilt}$. In our data example we lack an appropriate standard reference table and so rely on the latter approach. Under external standardizing, a product Poisson likelihood arises, whereas under internal standardizing the joint distribution of the y_{ilt} is multinomial. However, because likelihood inference is unaffected by whether or not we condition on $\sum_{ilt} y_{ilt}$ (see, e.g., Agresti 1990, pp. 455–456), it is common practice to retain the product Poisson likelihood.

The interesting aspect in the modeling involves the prior specification for ψ_{ilt} or, equivalently, for $\mu_{ilt} \equiv \log \psi_{ilt}$. Models for μ_{ilt} can incorporate a variety of *main effect* and *interaction* terms. Consider the main effect for subgroup membership. In our example, with two levels for sex and two levels for race, it takes the form of a sex effect plus a race effect plus a sex–race interaction effect. In general, we can write this main effect as $\varepsilon_l = \mathbf{x}_l^T\boldsymbol{\beta}$ for an appropriate design vector $\mathbf{x}_l$ and parameter vector $\boldsymbol{\beta}$. In the absence of prior information we would take a vague, possibly flat, prior for $\boldsymbol{\beta}$. In many applications it would be sensible to assume additivity and write $\mu_{ilt} = \varepsilon_l + \varepsilon_{it}$. For instance, in our example this form expresses the belief that sex and race effects would not be affected by region and year. This is the only form that we have investigated thus far, although more complicated structures could of course be considered. Hence it remains to elaborate ε_{it}.

2.2 Temporal Modeling

As part of ε_{it}, consider a main effect for time δ_t. Although we might specify δ_t as a parametric function (e.g., linear or quadratic), we prefer a qualitative form that lets the data reveal any presence of trend via the set of posterior densities for the δ_t. For instance, a plot of the posterior means or medians against time would be informative. But what sort of prior shall we assign to the set $\{\delta_t\}$? A simple choice used in Section 5 is again a flat prior. An alternative is an autoregressive prior with say, AR(1) structure; that is, $\delta_t|\delta_{t-1} \sim N(\gamma_t + \rho(\delta_{t-1} - \gamma_{t-1}), \sigma_\delta^2)$. We might take $\gamma_t = \gamma t$, so that γ, ρ, and σ_δ^2 are hyperparameters.

2.3 Spatial Modeling

The main effect for regions, say η_i, offers many possibilities. For instance, we might have regional covariates collected in a vector $\mathbf{z}_i$ that contribute a component $h(\mathbf{z}_i)$. Typically, h would be a specified parametric function $h(\mathbf{z}_i;\boldsymbol{\omega})$; for example, $\mathbf{z}_i^T\boldsymbol{\omega}$ with a flat prior for $\boldsymbol{\omega}$. In addition to or in the absence of covariate information, we might include regional random effects θ_i to capture heterogeneity among the regions (Bernardinelli and Montomoli 1992; Clayton and Bernardinelli 1992). If so, because i indexes the regions arbitrarily, an exchangeable prior for the θ_i's is appropriate; that is, $\theta_i \overset{\text{iid}}{\sim} N(\kappa_\theta, 1/\tau)$. We then add a flat hyperprior for κ_θ but require a proper hyperprior (typically a gamma distribution) or τ. Formally, with a usual improper prior such as $\tau^{(d-3)/2}$, $d = 0, 1, 2$, the resultant joint posterior need not be proper (Hobert and Casella 1996). Pragmatically, an informative prior for τ ensures well-behaved MCMC model fitting; see Section 4 for further discussion. Note that the inclusion of heterogeneity effects yields association among the y's across regions.

Viewed as geographic locations, the regions suggest the possibility of spatial modeling (Bernardinelli and Montomoli 1992; Besag et al. 1991; Clayton and Bernardinelli 1992; Clayton and Kaldor 1987; Cressie and Chan 1989). As with exchangeability, such spatial modeling introduces association across regions. We should thus consider whether it is appropriate for geographic proximity to be reflected in correlated counts. If so, we denote the spatial effect for region i by ϕ_i. In the spirit of Clayton and Kaldor (1987) and Cressie and Chan (1989), who extended Besag (1974), we model the ϕ_i using a conditional autoregressive (CAR) model. That is, we assume that the conditional density of $\phi_i|\phi_{j\neq i}$ is proportional to $\exp[-(\lambda/2)(a_i\phi_i - \sum_{j\neq i} w_{ij}\phi_j)^2]$, where $w_{ij} \geq 0$ is a weight reflecting the influence of ϕ_j on the expectation of ϕ_i and $a_i > 0$ is a "sample size" associated with region i. From Besag (1974), the joint density of the vector of spatial effects ϕ is proportional to $\exp(-(\lambda/2)\boldsymbol{\phi}^T\mathbf{B}\boldsymbol{\phi})$, where $\mathbf{B}_{ii} = a_i$ and $\mathbf{B}_{ij} = -a_i w_{ij}$. It is thus a proper multivariate normal density with mean $\mathbf{0}$ and covariance matrix $\mathbf{B}^{-1}$, provided that $\mathbf{B}$ is symmetric and positive definite. Symmetry requires $a_i w_{ij} = a_j w_{ji}$. A proper gamma prior is typically assumed for λ.

Two special cases have been discussed in the literature. One, exemplified by Cressie and Chan (1989), Devine, Halloran, and Lewis (1994), Devine and Louis (1994), and Devine, Louis, and Halloran (1994), requires a matrix of interregion distances d_{ij}. Then w_{ij} is set equal to $g(d_{ij})$ for a suitable decreasing function g. Cressie and Chan (1989) chose g based on the estimated variogram of the obser-

vations. Because $w_{ij} = w_{ji}$, symmetry of $\mathbf{B}$ requires a_i constant. In fact, $a_i = g(0)$, which without loss of generality we can set equal to 1. A second approach (as in, e.g., Besag et al. 1991) defines a set ∂_i of *neighbors* of region i. Such neighbors can be defined as regions contiguous to region i, or perhaps as regions within a prescribed distance of region i. Let n_i be the number of neighbors of region i and let $w_{ij} = 1/n_i$ if $j \in \partial_i$ and 0 otherwise. Then if $a_i = n_i$, $\mathbf{B}$ is symmetric. It is easy to establish that $\phi_i|\phi_{j\neq i} \sim N(\bar{\phi}_i, 1/(\lambda n_i))$, where $\bar{\phi}_i = n_i^{-1}\sum_{j\in\partial_i}\phi_j$. It is also clear that $\mathbf{B}$ is singular, because the sum of all its rows or columns is $\mathbf{0}$, so the joint density is improper. In Section 5 we confine ourselves to pairwise difference specification based on contiguity, as this requires no inputs other than the regional map, and refer to such specifications as CAR(λ) priors. The notions of interregion distance and neighbor sets can be combined following Cressie and Chan (1989) by setting $w_{ij} = g(d_{ij})$ for $j \in \partial_i$. The specific choice of g is sometimes based on parametric variograms; for our data example, a choice resulting in weights inversely proportional to distance is described briefly in Section 6.

We have chosen to model the spatial correlation in the prior specification using, in Besag's (1974) terminology, an auto-Gaussian model. One could also envision introducing spatial correlation into the likelihood (i.e., directly amongst the y_{ilt}), resulting in Besag's auto-Poisson structure (see also Ferrándiz, López, Llopis, Morales, and Tejerizo 1995). However, Cressie (1993) pointed out that this model does *not* yield a Poisson structure marginally. Moreover, Cressie and Chan (1989) observed the intractability of this model working merely with y_i's; for a set of y_{ilt}'s, model fitting becomes infeasible.

A general form for the regional main effects is thus $\eta_i = h(\mathbf{z}_i;\boldsymbol{\omega}) + \theta_i + \phi_i$, yielding a likelihood in which θ_i and ϕ_i are not identifiable. However, the prior specification allows the possibility of separating these effects, because the data can inform about each of them; see Section 4 for more discussion. A related issue is posterior integrability. One simple condition is that any effect in μ_{ilt} that is free of i and has a flat prior must have a fixed *baseline* level, such as 0. For example, μ_{ilt} cannot include an intercept. If δ_t terms are included, then we must include a constraint (say, $\delta_1 = 0$ or $\sum_t \delta_t = 0$), for otherwise we can add a constant to each ϕ_i and subtract it from each level of the foregoing effect without affecting the likelihood times prior, and hence the posterior. Similarly, if both ϕ_i and θ_i are in the model, then we must set $\kappa_\theta = 0$. Also note that θ_i and ϕ_i may be viewed as surrogates for unobserved regional covariates. That is, with additional regional explanatory variables, they might not be needed in the model. From a different perspective, for models using the general form for η_i, there may be strong collinearity between $\mathbf{z}_i$ and say ϕ_i, which again makes the ϕ_i difficult to identify and hence the models difficult to fit.

2.4 Spatio–Temporal Interaction

To this point μ_{ilt} consists of the main effects ε_l, δ_t, and η_i. We now turn to interactions, although as mentioned earlier the only ones that we consider are spatio-temporal. In general, defining a space–time component is thorny, because it is not clear how to reconcile the different scales. For the disease mapping problem, the only spatio-temporal analysis we are aware of is that of Bernardinelli, Clayton et al. (1995), which assumed the multiplicative form $(\nu_i^\theta + \nu_i^\phi)t$ for the heterogeneity by time and clustering by time effects. We again prefer a qualitative form and propose a *nested* definition, wherein heterogeneity effects and spatial effects are nested within time. In this way we can examine the evolution of heterogeneity and spatial patterns over time. We write these effects as $\theta_i^{(t)}$ and $\phi_i^{(t)}$; that is, $\theta_i^{(t)}$ is a random effect for the ith region in year t. Conditional exchangeability given t would lead us to the prior $\theta_i^{(t)} \overset{\text{iid}}{\sim} N(\kappa_\theta^{(t)}, 1/\tau_t)$. Similarly, we view $\phi_i^{(t)}$ as a spatial effect for the ith region in year t. Again we adopt a CAR prior but now given t as well. Hence we assume that $\phi_i^{(t)}|\phi_{j\neq i}^{(t)}$ has density proportional to $\exp[-(\lambda_t/2)(a_i\phi_i^{(t)} - \sum_{j\neq i} w_{ij}\phi_j^{(t)})^2]$. Again, we assume proper gamma priors for the τ_t and the λ_t. Such definitions for $\theta_i^{(t)}$ and $\phi_i^{(t)}$ preclude the inclusion of main effects θ_i and ϕ_i in the model. Also, if $\phi_i^{(t)}$ appears in the model, then we cannot include δ_t terms that lack an informative prior, because for any fixed t the model could not identify both. Similarly, if both $\theta_i^{(t)}$ and $\phi_i^{(t)}$ appear in the model, then we must set $\kappa_\theta^{(t)} = 0$.

In summary, the most general model that we envision for μ_{ilt} takes the form

$$\mu_{ilt} = \mathbf{x}_l^T\boldsymbol{\beta} + \mathbf{z}_i^T\boldsymbol{\omega} + \theta_i^{(t)} + \phi_i^{(t)}. \tag{1}$$

In practice, (1) might well include more effects than we need. In Section 5 we investigate more parsimonious reduced models.

3. MODEL SELECTION

For model (1) and its reduced versions, an important matter is how to choose among such models. Customary model choice procedures are not applicable. For example, Bayes factors are not interpretable with flat priors on $\boldsymbol{\beta}$ (and perhaps the δ_t) or with CAR priors for the spatial effects. Cross-validatory predictive selection schemes (Gelfand, Dey, and Chang 1992) are are also unavailable, again due to the presence of heterogeneity and clustering model parameters individually identified only by the prior. In any case, neither of these two options is feasible here, given the high dimensionality of our models. Penalized likelihood criteria (e.g., the Akaike information criterion [AIC] and the Bayesian information criterion [BIC]) seem attractive, but the penalty functions are motivated by asymptotic arguments. Because model dimension increases with sample size for models of interest, no extant asymptotic calculations are appropriate.

To handle such nonregular models, we propose working in predictive space, extending methods of Laud and Ibrahim (1995). Here asymptotics are not needed, and intuitively appealing penalties emerge. The basic distribution that we

work with is

$$f(\mathbf{y}_{\text{new}}|\mathbf{y}_{\text{obs}}) = \int f(\mathbf{y}_{\text{new}}|\boldsymbol{\zeta}) f(\boldsymbol{\zeta}|\mathbf{y}_{\text{obs}})\, d\boldsymbol{\zeta}, \tag{2}$$

where $\boldsymbol{\zeta}$ denotes the collection of model parameters and $\mathbf{y}_{\text{new}}$ is viewed as a replicate of the observed data vector $\mathbf{y}_{\text{obs}}$. For model M_i, we revise the notation in (2) to

$$f(\mathbf{y}_{\text{new}}|\mathbf{y}_{\text{obs}}, M_i) = \int f(\mathbf{y}_{\text{new}}|\boldsymbol{\zeta}^{(i)}, M_i) f(\boldsymbol{\zeta}^{(i)}|\mathbf{y}_{\text{obs}}, M_i)\, d\boldsymbol{\zeta}^{(i)}.$$

The model selection criterion identifies a discrepancy function $d(\mathbf{y}_{\text{new}}, \mathbf{y}_{\text{obs}})$, computes

$$E[d(\mathbf{y}_{\text{new}}, \mathbf{y}_{\text{obs}})|\mathbf{y}_{\text{obs}}, M_i], \tag{3}$$

and selects the model that minimizes (3). A choice for Gaussian likelihoods, suggested by Laud and Ibrahim (1995, p. 250), is

$$d(\mathbf{y}_{\text{new}}, \mathbf{y}_{\text{obs}}) = (\mathbf{y}_{\text{new}} - \mathbf{y}_{\text{obs}})^T (\mathbf{y}_{\text{new}} - \mathbf{y}_{\text{obs}}). \tag{4}$$

For a non-Gaussian generalized linear mixed model, we may prefer to replace (4) by the corresponding deviance criterion. That is, with a Poisson likelihood, we would set

$$d(\mathbf{y}_{\text{new}}, \mathbf{y}_{\text{obs}}) = 2\sum_l \{y_{l,\text{obs}} \log(y_{l,\text{obs}}/y_{l,\text{new}}) - (y_{l,\text{obs}} - y_{l,\text{new}})\}, \tag{5}$$

where l indexes the components of $\mathbf{y}$. As before, our model selection criterion is based on $E[d(\mathbf{y}_{\text{new}}, \mathbf{y}_{\text{obs}})|\mathbf{y}_{\text{obs}}, M_i]$. Straightforward calculation shows that for (5) the lth term in the summation is strictly convex in $y_{l,\text{new}}$ if $y_{l,\text{obs}} > 0$. In fact, to avoid problems with zero counts, we correct (5) to

$$\tilde{d}(\mathbf{y}_{\text{new}}, \mathbf{y}_{\text{obs}}) = 2\sum_l \left\{ \left(y_{l,\text{obs}} + \frac{1}{2}\right) \log\left(\frac{y_{l,\text{obs}} + \frac{1}{2}}{y_{l,\text{new}} + \frac{1}{2}}\right) - (y_{l,\text{obs}} - y_{l,\text{new}}) \right\}.$$

Suppose that we write

$$\begin{aligned} E[\tilde{d}(\mathbf{y}_{\text{new}}, \mathbf{y}_{\text{obs}})|\mathbf{y}_{\text{obs}}, M_i] &= \tilde{d}(E[\mathbf{y}_{\text{new}}|\mathbf{y}_{\text{obs}}, M_i], \mathbf{y}_{\text{obs}}) \\ &\quad + E[\tilde{d}(\mathbf{y}_{\text{new}}, \mathbf{y}_{\text{obs}})|\mathbf{y}_{\text{obs}}, M_i] \\ &\quad - \tilde{d}(E[\mathbf{y}_{\text{new}}|\mathbf{y}_{\text{obs}}, M_i], \mathbf{y}_{\text{obs}}). \end{aligned} \tag{6}$$

The terms in (6) have attractive interpretations. The left side is the expected predictive deviance (EPD) for model M_i. The first term on the right side is essentially the likelihood ratio statistic with the MLE for $\boldsymbol{\mu}$ replaced by $E(\mathbf{y}_{\text{new}}|\mathbf{y}_{\text{obs}}, M_i)$. Using Jensen's inequality, the second term minus the third term is strictly positive and is the penalty associated with M_i. This difference becomes

$$2\sum_l \left(y_{l,\text{obs}} + \frac{1}{2}\right) \left\{ \log E\left[y_{l,\text{new}} + \frac{1}{2}\middle|\mathbf{y}_{\text{obs}}\right] - E\left[\log\left(y_{l,\text{new}} + \frac{1}{2}\right)\middle|\mathbf{y}_{\text{obs}}\right] \right\}. \tag{7}$$

Again, each term in the summation is positive. Using a second-order Taylor series expansion, we may show that (7) is approximately

$$\sum_l \frac{y_{l,\text{obs}} + \frac{1}{2}}{[E(y_{l,\text{new}} + \frac{1}{2}|\mathbf{y}_{\text{obs}})]^2} \cdot \text{Var}(y_{l,\text{new}}|\mathbf{y}_{\text{obs}}).$$

Hence (7) is a weighted predictive variability penalty, a natural choice in that if M_i is too large, predictive variances will become inflated. Computations of (6) requires only calculation of $E(y_{l,\text{new}}|\mathbf{y}_{\text{obs}}, M_i)$ and $E[\log(y_{l,\text{new}} + 1/2)|\mathbf{y}_{\text{obs}}, M_i]$. Such predictive expectations are routinely obtained as Monte Carlo integrations under the computational approach outlined in the next section.

4. COMPUTATIONAL APPROACH AND CONVERGENCE ISSUES

It is obvious from the high dimension and complexity of the models discussed in the previous section that some form of MCMC algorithm will be needed to obtain estimates of the posterior and predictive quantities of interest. However, the Gibbs sampler (Gelfand and Smith 1990) is not well suited to this problem, because, with the exception of the λ_t and τ_t, no full conditional distributions for the model parameters are standard. As such, we are instead drawn to the Metropolis algorithm (Chib and Greenberg 1995; Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953) as a way of obtaining the necessary samples. We began with a purely univariate version of this algorithm, wherein associated with each parameter was a univariate normal candidate density centered at the current value of this parameter and having some variance σ^2, chosen to provide a Metropolis acceptance ratio between 25% and 50% (see Gelman, Roberts, and Gilks 1996).

This collection of univariate Metropolis and Gibbs updating steps converges fairly slowly—no surprise in view of the between-parameter posterior correlations that we would expect in our model. Updating parameters in multivariate blocks (Liu, Wong, and Kong 1994) is one way to account for these correlations; another is through elementary transformations (Gelfand, Sahu, and Carlin 1995). The latter approach is particularly useful here and is motivated by the fact that the likelihood by itself cannot inform about the spatial heterogeneity and clustering parameters individually, only their sum. For simplicity, we illustrate in the case of a single period of observation ignoring subgroups. We make a linear transformation of variables from $(\boldsymbol{\theta}, \boldsymbol{\phi})$ to $(\boldsymbol{\theta}, \boldsymbol{\eta})$, where $\eta_i = \theta_i + \phi_i$, $i = 1, \ldots, I$. Writing the posterior on the old scale as $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y}) \propto L(\boldsymbol{\theta} + \boldsymbol{\phi}; \mathbf{y}) p(\boldsymbol{\theta}) p(\boldsymbol{\phi})$, the posterior on the new scale is thus $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\mathbf{y}) \propto L(\boldsymbol{\eta}; \mathbf{y}) p(\boldsymbol{\theta}) p(\boldsymbol{\eta} - \boldsymbol{\theta})$. Along with our conditional independence assumptions, this implies the full con-

ditional distributions

$$p(\eta_i|\eta_{j\neq i}, \boldsymbol{\theta}, \mathbf{y}) \propto L(\eta_i; y_i)p(\eta_i - \theta_i|\{\eta_j - \theta_j\}_{j\neq i}) \quad (8)$$

and

$$p(\theta_i|\theta_{j\neq i}, \boldsymbol{\eta}, \mathbf{y}) \propto p(\theta_i)p(\eta_i - \theta_i|\{\eta_j - \theta_j\}_{j\neq i}) \quad (9)$$

for $i = 1, \ldots, I$. Because the likelihood informs about η_i directly in (8), overall sampler convergence is improved. We also gain the serendipitous side benefit of a closed-form (normal) full conditional in (9), because the algebraically awkward Poisson likelihood component is no longer present.

The presence in our model of parameters unidentified by the likelihood has repercussions not only for reparameterization and MCMC convergence acceleration, but also for convergence monitoring and diagnosis. As pointed out by Besag, Green, Higdon, and Mengerson (1995), samplers operating on overparameterized spaces (i.e., having a subset of parameters identified by neither the likelihood nor the prior) are perfectly legitimate, provided that their samples are used only to summarize the posterior distributions of identifiable functions of the parameters. There is no notion of "convergence" for unidentified parameters, so their presence in the sampling order will have no negative effect on the convergence of the functions of interest.

Our model is more difficult, however, as it features many *weakly* identified parameters (i.e., identified only by a vague prior). Convergence will eventually obtain for these parameters, but very slowly, causing a correspondingly slow convergence rate for the parameters of interest. Moreover, this slow convergence may not be apparent from the sample paths of the model's well-identified parameters. Carlin and Louis (1996, p. 203) discussed this behavior in the case of the simple two-parameter likelihood model $Y|\theta, \phi \sim N(\theta + \phi, 1)$. As a result, in our subsequent data analyses we monitor a representative subset of *all* the parameters present (e.g., the $\theta_i^{(t)}$ and $\phi_i^{(t)}$ corresponding to a few counties and years). We also use simple reparameterizations of the form leading to (8) and (9) to speed overall convergence.

5. EXAMPLE: SPATIO-TEMPORAL MODELING OF LUNG CANCER MORTALITY

5.1 Description of Dataset and Basic Model

As an illustration, we fit the model described in Section 2 to data originally analyzed by Devine (1992, chap. 4). Here y_{ijkt} is the number of lung cancer deaths in county i during year t for gender j and race k in the state of Ohio, and n_{ijkt} is the corresponding exposed population count. These data were originally taken from a public use data tape (Centers for Disease Control 1988) containing age-specific death counts by underlying cause and population estimates for every county in the United States. Our subset of lung cancer data is recorded for $J = 2$ genders (male and female) and $K = 2$ races (white and nonwhite) for each of the $I = 88$ Ohio counties over an observation period of $T = 21$ years (1968–1988 inclusive), yielding a total of 7,392 observations. We obtain internally standardized expected death counts as $E_{ijkt} = n_{ijkt}\bar{y}$, where $\bar{y} = (\sum_{ijkt} y_{ijkt} / \sum_{ijkt} n_{ijkt})$, the average statewide death rate over the entire observation period.

The study area includes the U.S. Department of Energy Fernald Materials Processing Center. The Fernald facility recycles depleted uranium fuel from U.S. Department of Energy and Department of Defense nuclear facilities. The Fernald plant is located in the southwest corner of Ohio, approximately 25 miles northwest of Cincinnati. The recycling process creates a large amount of uranium dust. Peak production occurred between 1951 and the early 1960's, and some radioactive dust may have been released into the air due to inadequate filtration and ventilation systems. Lung cancer is of interest because inhalation is the primary route of the putative exposure and lung cancer is the most prevalent form of cancer potentially associated with exposure to uranium. We use reported rates for the years 1968–1988 to allow an appropriate (10 to 20-year) temporal lag between exposure and disease development.

We apply spatio-temporal models to extend the spatial-only modeling of Devine (1992). The temporal component is of interest to explore changes in rates over a relatively long period. Demographic issues (related to environmental justice) are of interest because of possible variation in residential exposures for various population subgroups. In addition, the demographic profile of plant workers and nearby residents most likely evolved over the time period of interest.

Devine (1992) and Devine, Louis, and Halloran (1994) applied Gaussian spatial models using a distance matrix to the average lung cancer rates for white males over the 21-year period. Xia, Carlin, and Waller (1997) fit Poisson likelihood models incorporating the sex and race covariates to each year separately, thus providing a rough estimate of the change in heterogeneity and clustering present in the data over time, and motivating our spatio-temporal analysis. We begin by fitting a version of model (1),

$$\mu_{ijkt} = s_j\alpha + r_k\beta + s_j r_k \xi + \theta_i^{(t)} + \phi_i^{(t)}, \quad (10)$$

where we adopt the gender and race scores

$$s_j = \begin{cases} 0 & \text{if male} \\ 1 & \text{if female} \end{cases}$$

and

$$r_k = \begin{cases} 0 & \text{if white} \\ 1 & \text{if nonwhite} \end{cases}.$$

Letting $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \ldots, \theta_I^{(t)})'$, $\boldsymbol{\phi}^{(t)} = (\phi_1^{(t)}, \ldots, \phi_I^{(t)})'$, and denoting the I-dimensional identity matrix by $\mathbf{I}$, we adopt the prior structure

$$\boldsymbol{\theta}^{(t)}|\tau_t \overset{\text{ind}}{\sim} N\left(\mathbf{0}, \frac{1}{\tau_t}\mathbf{I}\right)$$

and

$$\boldsymbol{\phi}^{(t)}|\lambda_t \overset{\text{ind}}{\sim} \text{CAR}(\lambda_t), \qquad t = 1, \ldots, T, \quad (11)$$

so that heterogeneity and clustering may vary over time. Note that the sociodemographic covariates (gender and race) do not interact with time or location.

To complete the model specification, we require prior distributions for α, β, ξ, the τ_t, and the λ_t. Because α, β, and ξ are identified by the likelihood, we may use a flat prior on these three parameters. Next, for the priors on the τ_t and λ_t we used conjugate, conditionally iid gamma(a, b) and gamma(c, d) priors. As the discussions in Section 2.3 and Section 4 revealed, some precision is required to facilitate implementation of an MCMC algorithm in this setting. On the other hand, too much precision risks likelihood–prior disagreement. To help settle this matter, we fit a spatial-only (reduced) version of model (10) to the data from the middle year in our set (1978, $t = 11$), using vague priors for λ and τ with both mean and standard deviation equal to 100 ($a = c = 1$, $b = d = 100$). The resulting posterior .025, .50, and .975 quantiles for λ and τ were (4.0, 7.4, 13.9) and (46.8, 107.4, 313.8). As such, in fitting our full spatio-temporal model (10) we retain $a = 1$, $b = 100$ for the prior on τ, but reset $c = 1$, $d = 7$ (i.e., prior mean and standard deviation equal to 7). Although these priors are still quite vague, the fact that we have used a small portion of our data to help determine them does give our approach a slight empirical Bayes flavor. Still, our specification is consistent with the advice of Bernardinelli, Clayton, and Montomoli (1995), who suggested that the heterogeneity parameters have prior standard deviation roughly $\frac{7}{10}$ of that assigned to the clustering parameters. Recasting this advice in terms of prior precisions and the adjacency structure or our CAR prior for the $\phi_i^{(t)}$, we have $\lambda \approx \tau/(2\bar{m})$, where $\bar{m}$ is the average number of counties adjacent to a randomly selected county (about 5–6 for Ohio).

5.2 Model Fitting

Using the cycle of univariate Metropolis and Gibbs steps described in Section 4, we ran five parallel, initially overdispersed MCMC chains for 500 iterations, a task that took roughly 20 minutes on a Sparc 10 workstation. Graphical monitoring of the chains for a representative subset of the parameters, along with sample autocorrelations and Gelman and Rubin (1992) diagnostics, indicated an acceptable degree of convergence by around the 100th iteration. Using the final 400 iterations from all five chains, we obtained the 95% posterior credible sets $(-1.10, -1.06)$, $(0, .05)$, and $(-.27, -.17)$ for α, β, and ξ. The corresponding point estimates are translated into the fitted relative risks for the four subgroups in Table 1. It is interesting that the fitted sex–race interaction ξ reverses the slight advantage that white men hold over nonwhite men, making nonwhite females the healthiest subgroup, with a relative risk nearly four times smaller than either of the male groups. Many Ohio counties have very small nonwhite populations, so this result could be an artifact of our inability to model covariate–region interactions.

Table 1. Fitted Relative Risks for Four Sociodemographic Subgroups in the Ohio Lung Cancer Data

Demographic subgroup	Contribution to ε_{jk}	Fitted log-relative risk	Fitted relative risk
White males	0	0	1
White females	α	−1.08	.34
Nonwhite males	β	.02	1.02
Nonwhite females	$\alpha + \beta + \xi$	−1.28	.28

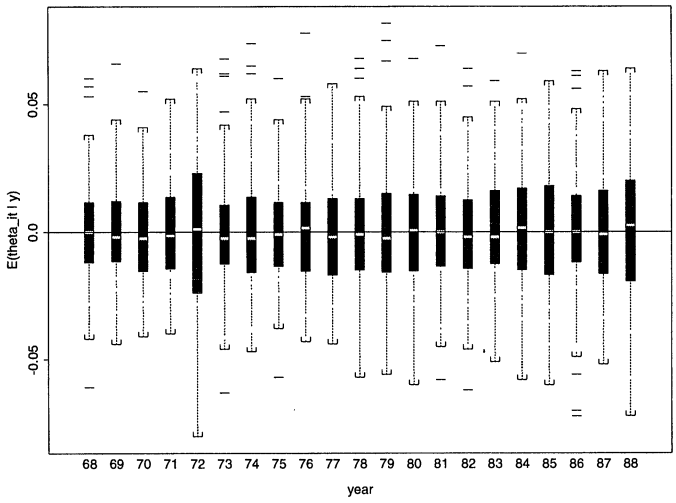

Figure 1. Boxplots of Estimated Posterior Means for $\theta_i^{(t)}$ Versus t, Full Model.

Turning to the spatio-temporal parameters, boxplots of the $I = 88$ estimated posterior means of the $\theta_i^{(t)}$ and $\phi_i^{(t)}$ for each year t are given in Figures 1 and 2. Figure 2 clearly shows the increasing trend in overall lung cancer mortality; the median increase of $.74 \approx \log 2$ suggests that the relative risk for a typical county has roughly doubled over the 21-year observation period. In contrast, Figure 1 shows the $E(\theta_i^{(t)}|\mathbf{y})$ distributions to be tightly centered near 0, suggesting no significant additional heterogeneity in the data beyond that explained by the CAR prior.

Figures 3 and 4 investigate this issue a bit further by checking for differential heterogeneity and clustering effects over time. For example, Figure 3 plots the estimated posterior medians for the clustering parameters λ_t versus t.

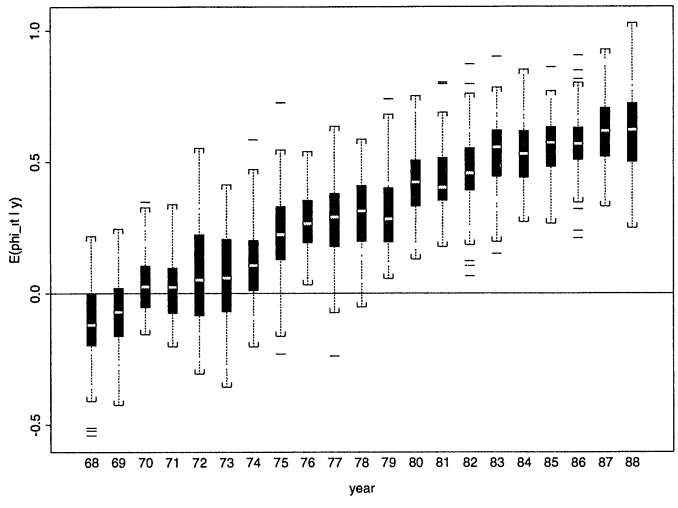

Figure 2. Boxplots of Estimated Posterior Means for $\phi_i^{(t)}$ Versus t, Full Model.

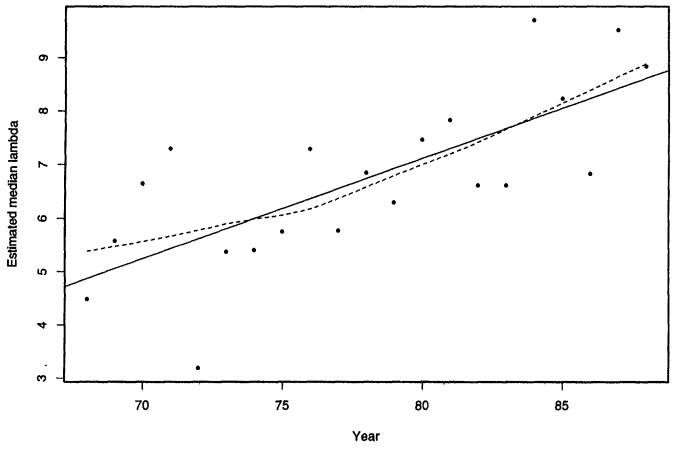

Figure 3. Estimated Posterior Medians for λ_t Versus t, Full Model. The solid line represents ordinary least squares; the dashed line, lowess smoothing.

A clear, almost linear increase is observed, suggesting that the spatial similarity of lung cancer cases is increasing over the 21-year time period. On the other hand, the posterior medians for τ_t plotted versus t in Figure 4 are all quite near the prior mean of 100 (again suggesting very little excess heterogeneity) and provide less indication of trend. What trend exists appears to be downward, suggesting that heterogeneity is also increasing over time. (Recall that τ_t is the precision in the mean 0 prior for $\theta_i^{(t)}$).

Because under our model the expected number of deaths for a given subgroup is county i during year t is $E_{ijkt}\exp(\mu_{ijkt})$, we have that the (internally standardized) expected death rate per 1,000 population is given by $1{,}000\bar{y}\exp(\mu_{ijkt})$. The first row of Figure 5 maps point estimates of these fitted rates for nonwhite females during the first (1968), middle (1978), and last (1988) years in our dataset. These estimates are obtained by plugging in the estimated posterior medians for the μ_{ijkt} parameters calculated from the output of the Gibbs sampler. The second row of the figure shows estimates of the variability in these rates (as measured by the interquartile range, or IQR) for the same subgroup during these three years.

Figure 5 reveals several interesting trends. Clearly, lung cancer death rates are increasing over time, as indicated by the gradual darkening of the counties in the figure's first row. But their variability is also increasing somewhat, as we would expect given our Poisson likelihood. This variability is smallest for high-population counties, such as those containing the cities of Cleveland (northern border, third from the right), Toledo (northern border, third from the left), and Cincinnati (southwestern corner). Lung cancer rates are high in these industrialized areas, but there is also a pattern of generally increasing rates as we move from west to east across the state for a given year. One possible explanation for this is a lower level of smoking among persons living in the predominantly agricultural west, as compared to those in the more mining- and manufacturing-oriented east. Finally, we note that although our conclusions of increasing clustering *and* increasing heterogeneity over time may have at first seemed contradictory, they are confirmed by the fitted death rate maps. We see increasing evidence of clustering among the high-rate counties, but with the higher rates increasing and the lower rates remaining low (i.e., increasing heterogeneity statewide). Again, because the higher rates tend to emerge in the poorer, more mountainous eastern counties, we have an indication of decreasing environmental equity over the last few decades.

Regarding the possible impact of the Fernald facility, Figure 6 shows the fitted lung cancer death rates per 1,000 population by year for white males in Hamilton County (which contains the facility), as well as in Butler, Warren, and Clermont Counties (which are adjacent to Hamilton). The statewide white male death rate is also plotted by year for comparison. (We focus on white male mortality because this is the demographic group most likely to have been affected at the facility during the potential exposure window in the 1950s.) We observe substantially elevated rates of lung cancer in Hamilton County. Rates in the adjacent counties are similar in magnitude to the statewide rate, but do seem to be increasing a bit more rapidly than the statewide rate. In addition, the adjacent county with the highest rates overall is Butler, which is immediately north of Hamilton and second-closest to Fernald.

These results are consistent with those obtained by Devine (1992) using a Gaussian spatial-only model. However, they are difficult to interpret due to two confounding factors. First, Hamilton County is also home to Cincinnati, a large urban area. We have already seen in Figure 5 that elevated rates are often present in such areas. Second, our model does not incorporate information on smoking prevalence, the most important risk factor for lung cancer. Although we currently have no county- or year-specific information on these two factors, they conceivably could be accounted for using census information (say, using population density and the number of cigarettes sold per capita). We hope to obtain this important covariate information and report on the resulting model in a subsequent article.

5.3 Model Validation, Comparison, and Selection

The results of the previous section suggest that our full

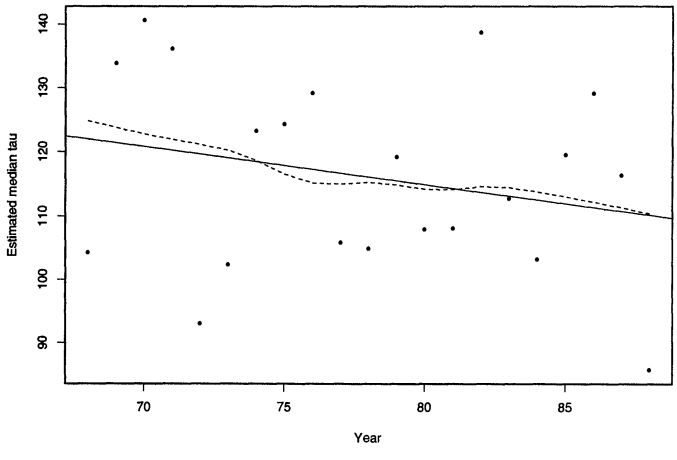

Figure 4. Estimated Posterior Medians for τ_t Versus t, Full Model. The solid line represents ordinary least squares; the dashed line, lowess smoothing.

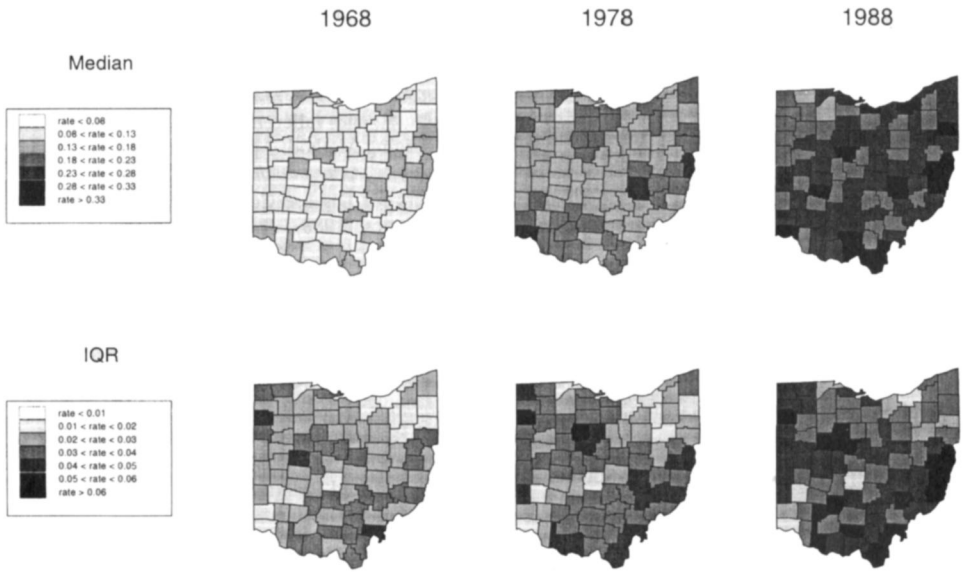

Figure 5. Fitted Lung Cancer Death Rates per 1,000 Population, Nonwhite Females.

model (10) is overfitting the data, especially with respect to heterogeneity over time. As such, we now contemplate several reduced models, judging their fit relative to the full model using the expected predictive deviance (EPD) score (6). These models and their scores are shown in Table 2, along with the partitioning of the EPD into its likelihood ratio statistic (LRS) and predictive variability penalty (PEN) components. The first row of the table shows the full model and its scores; the second and third rows show scores for the two models that eliminate the entire collection of spatial and heterogeneity effects. As we might have guessed, the latter change leads to no significant increase in the overall score, but the former change also seems to have only a small impact. (The EPD scores for these three models are essentially equivalent to the order of accuracy in our MCMC calculations.) Apparently, despite being centered around a common mean κ_θ a priori, the $\theta_i^{(t)}$ parameters are able to adapt their posterior levels to account for the increasing cancer rate over time.

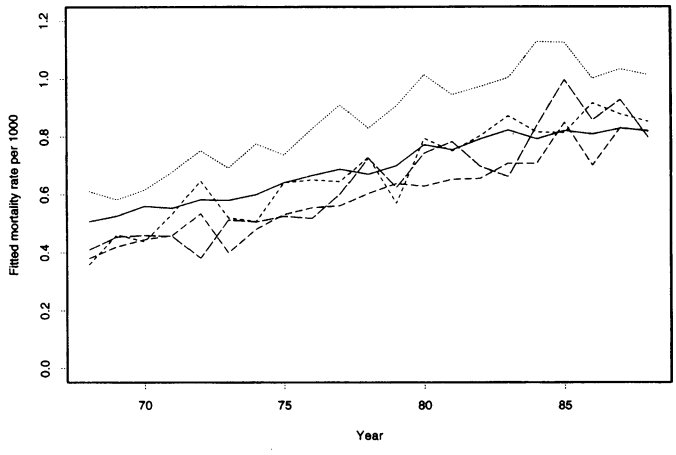

Figure 6. Estimated White Male Lung Cancer Death Rates per 1,000 by Year, Statewide Average and in Counties Near the Fernald Materials Processing Center. ———, all counties; · · ·, Hamilton (home of Fernald, Cincinnati); - - -, Butler (north of Hamilton); – – –, Warren (northeast of Hamilton); — — —, Clermont (southeast of Hamilton).

The model considered in the fourth row of Table 2 eliminates the excess heterogeneity and space–time interaction effects but still allows a general form for the overall time trend, subject to an identifiability constraint such as $\sum_t \delta_t = 0$. We implemented this constraint numerically at the end of each iteration by recentering the sampled δ_t values around 0 and also adjusting the ϕ_i values accordingly. The prior for the ϕ_i is a similar, simplified version of that in (11), namely $\phi_i \sim \text{CAR}(\lambda)$. Despite the smaller penalty associated with this more parsimonious model, the increase in the LRS leads to a larger (i.e., poorer) overall EPD score. This increase in score over the first three models is "significant" in the sense that it reoccurred in five other independent MCMC runs.

Finally, the last row of the table eliminates the spatial aspect of the model altogether and specifies a linear form for the temporal trend. The resulting model has no random effects and only five parameters in total. Again we see that the decrease in the penalty term fails to offset the reduction in quality of fit, resulting in a further significant increase in the overall EPD score.

6. CONCLUSIONS

We have presented and illustrated a method for hierarchical spatio-temporal analysis, with specific attention to the mapping of county-level disease rates over time. Our MCMC/Bayesian approach enables a previously inaccessible model of very high dimension, but this in turn necessitates great care in implementation to avoid the pitfalls of overmodeling and unidentifiability. We have also presented a new predictive criterion for model choice that is reminiscent of traditional penalized likelihood criteria but requires neither proper priors nor asymptotic arguments for its validity. In the specific context of our Ohio lung cancer dataset, our analysis confirms and extends the preliminary findings of Xia, Carlin, and Waller (1997) with respect to the in-

Table 2. Likelihood Ratio Statistics (LRS), Predictive Variability Penalties (PEN), and Overall Expected Predicted Deviance (EPD) Scores for the Spatio-temporal Ohio Lung Cancer Models

Model for μ_{ijkt}	LRS	PEN	EPD
$\varepsilon_{jk} + \theta_i^{(t)} + \phi_i^{(t)}$	5,374.87	5,806.07	11,180.94
$\varepsilon_{jk} + \theta_i^{(t)}$	4,941.32	6,017.06	10,958.39
$\varepsilon_{jk} + \phi_i^{(t)}$	5,180.25	5,808.73	10,988.98
$\varepsilon_{jk} + \delta_t + \phi_i$	7,274.70	5,725.19	12,999.89
$\varepsilon_{jk} + \gamma t$	9,444.20	4,756.31	14,200.50

crease in both heterogeneity and clustering over time. Those authors fit heterogeneity-only and clustering-only models separately to each of the 21 years of data, and thus were not able to "borrow strength" across years nor discover the lack of excess heterogeneity beyond that implied by the clustering process.

Many other models even more complicated than ours might well be considered. We experimented briefly with more sophisticated forms for the weights w_{ij} in the CAR prior for the clustering parameter, such as $w_{ij} = c/(c + d_{ij})$, where we took d_{ij} to be the distance in miles between the county seats of counties i and j. Although further investigation is warranted, the high level of aggregation present in our dataset seemed to preclude any significant benefit over our original, adjacency-based models. Other model enhancements, such as functional forms or perhaps an AR(1) structure for the δ_t and multiplicative forms for the spatio-temporal interaction, could also be considered. Finally, inclusion of the aforementioned covariate information on smoking status, population density, and socioeconomic status of the counties might also pay big dividends. We defer such examination to a companion work.

[Received August 1995. Revised June 1996.]

REFERENCES

Agresti, A. (1990), *Categorical Data Analysis*, New York: Wiley.

Anderton, D. L., Anderson, A. B., Oakes, J. M., and Fraser, M. R. (1994), "Environmental Equity: The Demographics of Dumping," *Demography*, 31, 229–248.

Bernardinelli, L., Clayton, D., and Montomoli, C. (1995), "Bayesian Estimates of Disease Maps: How Important are Priors?," *Statistics in Medicine*, 14, 2411–2431.

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.

Bernardinelli, L., and Montomoli, C. (1992), "Empirical Bayes Versus Fully Bayesian Analysis of Geographical Variation in Disease Risk," *Statistics in Medicine*, 11, 983–1007.

Bernardinelli, L., Pascutto, C., Best, N. G., and Gilks, W. R. (in press), "Disease Mapping With Errors in Covariates," *Statistics in Medicine*.

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 36, 192–236.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), "Bayesian Computation and Stochastic Systems" (with discussion), *Statistical Science*, 10, 3–66.

Besag, J., 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.

Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 9–25.

Carlin, B. P., and Louis, T. A. (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, London: Chapman and Hall.

Carrat, F., and Valleron, A.-J. (1992), "Epidemiologic Mapping Using the "Kriging Method: Application to an Influenza-Like Illness Epidemic in France," *American Journal of Epidemiology*, 135, 1293–1300.

Centers for Disease Control and Prevention, National Center for Health Statistics (1988), *Public Use Data Tape Documentation Compressed Mortality File, 1968–1985*, Hyattsvile, MD: U.S. Department of Health and Human Services.

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis–Hastings Algorithm," *The American Statistician*, 49, 327–335.

Clayton, D. G., and Bernardinelli, L. (1992), "Bayesian Methods for Mapping Disease Risk," in *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*. eds. P. Elliott, J. Cuzick, D. English, and R. Stern, Oxford, U.K.: Oxford University Press.

Clayton, D. G. and Kaldor, J. (1987), "Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping," *Biometrics*, 43, 671–681.

Cressie, N.A.C. (1993), *Statistics for Spatial Data* (2nd ed.), New York: Wiley.

Cressie, N.A.C., and Chan, N. H. (1989), "Spatial Modeling of Regional Variables," *Journal of the American Statistical Association*, 84, 393–401.

Devine, O. J. (1992), "Empirical Bayes and Constrained Empirical Bayes Methods for Estimating Incidence Rates in Spatially Aligned Areas," unpublished Ph.D. dissertation, Emory University, Division of Biostatistics.

Devine, O. J., Halloran, M. E., and Louis, T. A. (1994), "Empirical Bayes Methods for Stabilizing Incidence Rates Prior to Mapping," *Epidemiology*, 5, 622–630.

Devine, O. J., and Louis, T. A. (1994), "A Constrained Empirical Bayes Estimator for Incidence Rates in Areas With Small Populations," *Statistics in Medicine*, 13, 1119–1133.

Devine, O. J., Louis, T. A., and Halloran, M. E. (1994), "Empirical Bayes Estimators for Spatially Correlated Incidence Rates," *Environmetrics*, 5, 381–398.

Ferrándiz, J., López, A., Llopis, A., Morales, M., and Tejerizo, M. L. (1995), "Spatial Interation Between Neighbouring Counties: Cancer Mortality Data in Valencia (Spain)," *Biometrics*, 51, 665–678.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992), "Model Determination Using Predictive Distributions with Implementation via Sampling-Based Methods" (with discussion), in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds., Oxford, U.K.: Oxford University Press, pp. 147–167.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995), "Efficient Parametrizations for Normal Linear Mixed Models," *Biometrika*, 82, 479–488.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996), "Efficient Metropolis Jumping Rules," in *Bayesian Statistics 5*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith. Oxford, U.K.: Oxford University Press, pp. 599–607.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences" (with discussion), *Statistical Science*, 7, 457–511.

Ghosh, M. (1992), "Constrained Bayes Estimation With Applications," *Journal of the American Statistical Association*, 87, 533–540.

Handcock, M. S., and Wallis, J. R. (1994), "An Approach to Statistical Spatial–Temporal Modeling of Meteorological Fields" (with discussion), *Journal of the American Statistical Association*, 86, 368–390.

Hobert, J. P., and Casella, G. (1996), "The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models," *Journal of the American Statistical Association*, 91, 1461–1473.

Laud, P., and Ibrahim, J. (1995), "Predictive Model Selection," *Journal of the Royal Statistical Society*, Ser. B, 57, 247–262.

Liu, J., Wong, W. H., and Kong, A. (1994), "Covariance Structure of the Gibbs Sampler With Applications to the Comparisons of Estimators and Augmentation Schemes," *Biometrika*, 81, 27–40.

Manton, K. G., Woodbury, M. A., Stallard, E., Riggan, W. B., Creason,

J. P., and Pellom, A. C. (1989), "Empirical Bayes Procedures for Stabilizing Maps of U.S. Cancer Mortality Rates," *Journal of the American Statistical Association*, 84, 637–650.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087–1091.

Sacks, J., and Steinberg, L. J. (1994), "Environmental Equity: Statistical Issues, Report of a Forum," Technical Report 11, National Institute of Statistical Sciences.

Sampson, P. D., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87, 108–119.

Sexton, K. Olden, K., and Johnson, B. L. (1993), " 'Environmental Justice': The Central Role of Research in Establishing a Credible Scientific Foundation for Informed Decision Making," *Toxicology and Industrial Health*, 9, 685–727.

Stockwell, J. R., Sorensen, J. W., Eckert, J. W., and Carreras, E. M. (1993), "The U.S. EPA Geographic Information System for Mapping Environmental Releases of Toxic Chemical Release Inventory (TRI) Chemicals," *Risk Analysis*, 13, 155–164.

Wagener, D. K., and Williams, D. R. (1993), "Equity in Environmental Health: Data Collection and Interpretation Issues," *Toxicology and Industrial Health*, 9, 775–795.

Walter, S. D., and Birnie, S. E. (1991), "Mapping Mortality and Morbidity Patterns: An International Comparison," *International Journal of Epidemiology*, 20, 678–689.

Webster, R., Oliver, M. A., Muir, K. R., and Mann, J. R. (1994), "Kriging the Local Risk of a Rare Disease From a Register of Diagnoses," *Geographic Analysis*, 26, 168–185.

Xia, H., Carlin, B. P., and Waller, L. A. (in press), "Hierarchical Models for Mapping Ohio Lung Cancer Rates," *Environmetrics*.