Chapter 16

Bayesian Disease Mapping for Public Health

Andrew Lawson[*,1] and Duncan Lee[†]
[*]*Medical University of South Carolina, Charleston, SC, United States*
[†]*School of Mathematics and Statistics, University of Glasgow, Glasgow, United Kingdom*
[1]*Corresponding author: e-mail: lawsonab@musc.edu*

ABSTRACT
Disease risk varies in space and time due to variation in many factors, including environmental exposures such as air pollution, and the prevalence of lifestyle behaviors such as smoking. Quantifying and explaining this spatio-temporal variation in disease risk is vital to improving public health, as for example it allows risk factors to be highlighted to the general public, as well as allowing informed decisions to be made about the future allocation of health resources. Research in this area is collectively known as *disease mapping* and has a variety of aims including estimating the spatio-temporal pattern in disease risk, explaining the variation in the risk by covariate factors, identifying high-risk subregions (clusters) for focused action, and surveillance of disease outbreaks. The data used in this field are typically counts of the numbers of disease cases in a set of nonoverlapping areal units, and in the simplest case these data are available for a single time period. However, recent work has extended this simple case by analyzing disease count data for multiple time periods (space–time modeling) or for multiple diseases for a single time period (multivariate modeling). This chapter provides an in-depth review of the disease mapping field, focusing on the four key areas listed above in spatial, spatio-temporal, and multivariate disease domains.

Keywords: Bayesian hierarchical models, Cluster detection, Disease mapping, Multivariate modeling, Space–time modeling

1 INTRODUCTION

Disease mapping concerns the analysis of the spatial distribution of disease. Usually, its focus is on the statistical modeling of disease outcomes when inference about disease risk is required. To this end, it can be considered that there are four main areas of focus: relative risk estimation, disease clustering, ecological analysis, and surveillance. The most developed part of the field is

Handbook of Statistics, Vol. 36. https://doi.org/10.1016/bs.host.2017.05.001

relative risk estimation, and a large range of models and associated software are now available. In the next section the basic models found in this area are described. Disease clustering can be defined as the study of the occurrence of "unusual" aggregations of disease and is an area of particular concern for public health. Public Health departments can be tasked to investigate clusters of disease on a regular basis. Ecological studies are those where predictor variables thought to be relevant in the explanation of disease variation are linked to small area health outcomes. By small area we mean any observation frame or unit which has a geo-reference or location. This could be a postal zone (zip code), census tract, census block, county, state, and municipality. The unit is usually not related to the health outcome but is arbitrary and could be defined for administrative or political purposes. The focus of ecological analysis is the relationship between a health outcome and predictors, at these small area levels, so that counts of disease are the outcome measure. Finally, surveillance is defined to be the activity of monitoring disease over time to assess whether changes to disease behavior have occurred. Prospective surveillance is carried out in (near) real time and aims to detect aberrations in disease incidence streams.

A model-based approach to the analysis of disease maps is beneficial. Statistical models allow the inclusion of a variety of features that capture the variation of disease risk. The paradigm that is mostly used in this context is Bayesian hierarchical modeling. Fundamentally, this approach to the analysis of disease variation allows the use of a data model (likelihood) for the observed health outcome, while specifying additional features of the data within hierarchical parameter structures. Hence, hierarchical modeling is fundamental to the Bayesian paradigm. Spatial health data can arise as collections of residential addresses of cases and their date/time of diagnosis. Often these (case event) data are aggregated in space and time to form counts of disease. Here, we focus on the commonly found count format. Case event data are discussed in Lawson (2013) and Lawson et al. (2016). For spatial health data observed as counts within small areas, a common model assumption is that the counts are independently distributed with a Poisson data model. Once this model is specified, then any spatial correlation in the counts can be specified within prior distributions which define the model hierarchy. In this way conventional-independent data likelihoods can be assumed for spatially correlated health outcomes. In the next section a detailed description of these models is provided in a purely spatial setting, before extensions to the spatio-temporal and multivariate disease domains are discussed in Sections 3 and 4, respectively. Section 5 provides an overview of the available software for fitting disease mapping models, while Sections 6 and 7 focus on clustering and boundary detection, respectively. Section 8 focuses on ecological regression, which are followed by sections on surveillance (Section 9), and survival analysis (Section 10), respectively. Section 11 provides an example of space–time disease mapping, and future work is outline in Section 12.

2 SPATIAL MODELING

2.1 Data and Overall Model

The simplest disease mapping setting is to have purely spatial data relating to I nonoverlapping small areas. Let $\mathbf{Y} = (Y_1, \ldots, Y_I)$ denote a vector of counts of the total numbers of disease cases from the populations living within each of the I small areas. These counts will depend on the overall size and demographic structure of the populations living within each area, and this is accounted for by computing the expected numbers of disease cases $\mathbf{e} = (e_1, \ldots, e_I)$ using indirect standardization. Specifically, the population in each area is split into strata based on age and sex (e.g., females 0–5, females 6–15), and $e_i = \sum_{\text{strata } r} N_{ir} q_r$, where N_{ir} is the number of people in area i and strata r, and q_r is the region-wide rate of disease for strata r. Based on these data the standardized mortality ratio (if the disease outcome in question is death) is an exploratory measure of disease risk and is given by $\text{SMR}_i = Y_i/e_i$. Here an SMR of 1.1 corresponds to a 10% increased risk compared with the region-wide average, while an SMR of 0.8 corresponds to a 20% reduced risk.

The SMR is an unstable estimate of disease risk, especially if the disease in question is rare or the population at risk is small (in which case e_i is small), which has led researchers to instead estimate risk using a model-based approach. One component of such a model is covariate risk factors that are thought to impact on disease risk. Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_I)$ denote a matrix of p such covariate factors (including a column of ones for the intercept term), where $\mathbf{x}_i = (1, x_{i2}, \ldots, x_{ip})$. Then a simple model for disease risk is given by

$$\begin{aligned} Y_i &\sim \text{Poisson}(e_i\theta_i) \quad i = 1, \ldots, I, \\ \ln(\theta_i) &= \mathbf{x}_i^\top \boldsymbol{\beta}, \end{aligned} \tag{1}$$

where a Poisson log-linear generalized linear model is appropriate because the disease data are nonnegative counts of the total numbers of disease cases. Here $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is a vector of unknown regression parameters, while θ_i represents disease risk (relative to e_i), and has the same scale of interpretation as the SMR. However, this model is often insufficient for modeling spatial disease data, because the set of available covariates $\mathbf{X}$ does not include all important risk factors. Thus if those important factors that are either unmeasured or unknown are spatially structured, then the disease counts will contain unmeasured spatial autocorrelation even after adjusting for the known covariates $\mathbf{X}$. This unmeasured confounding violates the independence assumption inherent in model (1), resulting in the use of the following Poisson log-linear generalized linear mixed model to represent these data.

$$\begin{aligned} Y_i &\sim \text{Poisson}(e_i\theta_i) \quad i = 1, \ldots, I, \\ \ln(\theta_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i. \end{aligned} \tag{2}$$

Here $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_I)$ is a vector of random effects, which are modeled as spatially autocorrelated to account for the influence of unmeasured confounders. A number of models have been proposed and a brief review is given below.

2.2 Random Effect Models

A number of different approaches have been proposed for modeling the unmeasured spatial autocorrelation, including geostatistical models (Biggeri et al., 2006), simultaneous autoregressive models (Kissling and Carl, 2008), and spline-based models (Ugarte et al., 2010). However, by far the most common approach is to use conditional autoregressive (CAR) models, which are a special case of a Gaussian Markov random field (GMRF). These models represent spatial closeness via an $I \times I$ neighborhood or adjacency matrix $\mathbf{W}$, where element w_{ir} defines whether areas (i, r) are spatially close. Typically a binary specification is used, so that $w_{ir} = 1$ if areas (i, r) are spatially close and $w_{ir} = 0$ otherwise. This specification leads to a sparse specification for $\mathbf{W}$, which makes the fitting of these models much more efficient than if $\mathbf{W}$ was a dense matrix. Commonly, border sharing is used to determine $\mathbf{W}$, so that $w_{ir} = 1$ if areas (i, r) share a common border, and $w_{ir} = 0$ otherwise. Given this neighborhood matrix, CAR models for a vector of random effects $\boldsymbol{\phi}$ are most often written as a set of univariate full conditional distributions, $f(\phi_i|\boldsymbol{\phi}_{-i})$, where $\boldsymbol{\phi}_{-i} = (\phi_1, \ldots, \phi_{i-1}, \phi_{i+1}, \ldots, \phi_I)$. However, this set of I conditional distributions is equivalent to the following multivariate Gaussian joint distribution

$$\boldsymbol{\phi} \sim \mathrm{N}(\mathbf{0}, \tau^2 \mathbf{Q}(\mathbf{W})^{-1}), \tag{3}$$

where $\mathbf{0}$ is an $I \times 1$ vector of zeros and $\mathbf{Q}(\mathbf{W})$ is an $I \times I$, potentially singular, precision matrix. The simplest CAR model is the *intrinsic* model (ICAR, Besag et al., 1991), which is given by

$$\phi_i|\boldsymbol{\phi}_{-i} \sim \mathrm{N}\left(\frac{\sum_{r=1}^{I} w_{ir}\phi_r}{\sum_{r=1}^{I} w_{ir}}, \frac{\tau^2}{\sum_{r=1}^{I} w_{ir}}\right), \tag{4}$$

and corresponds to the singular precision matrix $\mathbf{Q}(\mathbf{W}) = \mathrm{diag}(\mathbf{W1}) - \mathbf{W}$ in the above joint specification. This model captures spatial autocorrelation because the conditional expectation is the mean of the random effects in neighboring areas, while the conditional variance is inversely proportional to the number of neighboring areas. This latter specification makes sense if the data are spatially autocorrelated, because the more neighbors area i has with similar random effect values, then the more information and hence the less uncertainty there is about the value of ϕ_i. However, this model corresponds to an improper joint distribution for $\boldsymbol{\phi}$ with a singular precision matrix, and also only allows

for strong spatial correlation that can sometimes lead to oversmoothing. Therefore the *convolution* or *BYM* model was proposed by Besag et al. (1991), which augments the intrinsic model with a second set of spatially unstructured random effects.

$$\begin{aligned}\phi_i &= \phi_i^{(1)} + \phi_i^{(2)},\\ \phi_i^{(1)}|\boldsymbol{\phi}_{-i}^{(1)} &\sim \mathrm{N}\left(\frac{\sum_{r=1}^{I} w_{ir}\phi_r^{(1)}}{\sum_{r=1}^{I} w_{ir}}, \frac{\tau^2}{\sum_{r=1}^{I} w_{ir}}\right),\\ \phi_i^{(2)} &\sim \mathrm{N}(0,\sigma^2).\end{aligned} \tag{5}$$

This model represents the random effects $\boldsymbol{\phi}$ with a convolution of spatially autocorrelated and spatially unstructured effects, which are modeled by the intrinsic CAR model and a zero-mean Gaussian shrinkage model, respectively. This is the most commonly used CAR model in the literature and can induce varying levels of spatial autocorrelation by varying the amount of variation in each of the two components. However, these individual components $(\phi_i^{(1)}, \phi_i^{(2)})$ are not identifiable from the data, as only their sum can reliably be estimated. Therefore two further alternatives have been proposed, which each have a single set of random effects but introduce a spatial dependence parameter ρ to allow for varying levels of spatial autocorrelation. The first was proposed by Stern and Cressie (1999) and is given by

$$\phi_i|\boldsymbol{\phi}_{-i} \sim \mathrm{N}\left(\frac{\rho\sum_{r=1}^{I} w_{ir}\phi_r}{\sum_{r=1}^{I} w_{ir}}, \frac{\tau^2}{\sum_{r=1}^{I} w_{ir}}\right), \tag{6}$$

and corresponds to a joint distribution with precision matrix $\mathbf{Q}(\mathbf{W}, \rho) = \mathrm{diag}(\mathbf{W1}) - \rho\mathbf{W}$. Here $\rho = 1$ simplifies to the intrinsic model while $\rho = 0$ corresponds to independence, the latter being the case as then the conditional expectation does not depend on the random effects in other areas. One downside of this model is that if $\rho = 0$ then the conditional variance still depends on the number of neighboring areas, even though there is no spatial autocorrelation in the random effects. Therefore an alternative was proposed by Leroux et al. (2000) which is given by

$$\phi_i|\boldsymbol{\phi}_{-i} \sim \mathrm{N}\left(\frac{\rho\sum_{r=1}^{I} w_{ir}\phi_r}{\rho\sum_{r=1}^{I} w_{ir} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{r=1}^{I} w_{ir} + 1 - \rho}\right), \tag{7}$$

and corresponds to a joint distribution with precision matrix $\mathbf{Q}(\mathbf{W}, \rho) = \rho[\text{diag}(\mathbf{W1}) - \mathbf{W}] + (1 - \rho)\mathbf{I}$, where $\mathbf{I}$ is an identity matrix. In this model $\rho = 1$ simplifies to the intrinsic model, while $\rho = 0$ corresponds to independence with mean zero and a constant variance. More recently, Riebler et al. (2016) proposed an alternative CAR model to the above that accounts for scaling, and a number of other extensions are discussed in the remainder of this chapter. Inference for this model is typically undertaken in a Bayesian setting, and a summary of the inferential approaches and available software is given in Section 5.

2.3 Choice of Priors

The choice of prior distributions within Bayesian models is an important concern. Often prior distributions are chosen to provide only limited information about the parameters, and noninformative or flat/vague priors are chosen. On the other hand, informative prior distributions are useful when particular effects are to be ensured. In the case of regression models (Poisson or logistic) where predictors are to be included within the model, it is commonplace to assume an independent zero-mean Gaussian distribution with a variance $\tau^2_{\beta_r}$ for a regression parameter β_r, that is, $\beta_r \sim N(0, \tau^2_{\beta_r})$. Variance parameters for random effects τ^2 are often assigned apparently weakly informative conjugate inverse-gamma priors, that is $\tau^2 \sim$ Inverse Gamma(a, b), where common specifications are $a = 1.0$, $b = 0.0001$, or $a = 2.0$, $b = 1.0$. Finally, spatial dependence parameters such as ρ are typically assigned a uniform prior on the unit interval.

2.4 Goodness of Fit and Variable Selection

Goodness of fit criteria vary depending on the properties of the criteria and the nature of the model. In conventional generalized linear modeling with fixed effects, the deviance is an important measure. Usually this measure of model adequacy compares a fitted model with parameters $\widehat{\boldsymbol{\theta}}_{fit}$ to a saturated model with parameters $\widehat{\boldsymbol{\theta}}_{sat}$. It is based on the difference between the log-likelihood of the data under either model:

$$D = -2[l(\mathbf{y}|\widehat{\boldsymbol{\theta}}_{fit}) - l(\mathbf{y}|\widehat{\boldsymbol{\theta}}_{sat})],$$

where the saturated model has a single parameter per observation. Often a relative measure of fit is used so that the change in deviance between model 1 (with $\widehat{\boldsymbol{\theta}}_1$) and model 2 (with $\widehat{\boldsymbol{\theta}}_2$) is used:

$$\Delta D = -2[l(\mathbf{y}|\widehat{\boldsymbol{\theta}}_1) - l(\mathbf{y}|\widehat{\boldsymbol{\theta}}_2)].$$

Hence the saturated likelihood cancels in this relative comparison. The deviance is used in goodness-of-fit measures in Bayesian modeling, but usually

without reference to a saturated model. One disadvantage of using the deviance directly is that it does not allow for the degree of parameterization in the model: a model can be made to more closely approximate the data by increasing the number of parameters. Hence attempts have been made to penalize model complexity. Model parsimony should result in the use of such a penalized model comparison metric. Such metrics are widely used for fixed effect models, and in the Bayesian models considered here the most popular criterion is the *Deviance information criterion (DIC)* proposed by Spiegelhalter et al. (2002). Given a parameter vector $\boldsymbol{\theta}$, and now defining the deviance as $D(\boldsymbol{\theta}) = -2l(\mathbf{y}|\hat{\boldsymbol{\theta}})$, minus twice the log-likelihood, the DIC is defined as

$$\text{DIC} = 2E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] - D(E[\boldsymbol{\theta}|\mathbf{y}]).$$

Note that the DIC is based on a comparison of the posterior average deviance $E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})]$ and the deviance evaluated at the posterior mean $D(E[\boldsymbol{\theta}|\mathbf{y}])$. Given a Markov chain Monte Carlo (MCMC) algorithm with G samples $\{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(G)}\}$, then the two quantities are computed as: $E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] = \frac{1}{G}\sum_{i=1}^{G} D(\boldsymbol{\theta}^{(i)})$ and $D(E[\boldsymbol{\theta}|\mathbf{y}]) = D\left(\frac{1}{G}\sum_{i=1}^{G}\boldsymbol{\theta}^{(i)}\right)$. The effective number of parameters (pD) is estimated as $pD = E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] - D(E[\boldsymbol{\theta}|\mathbf{y}])$, and then an alternative algebraic formulation is DIC $= E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] + pD$. Unfortunately in some situations the pD can be negative (as it can happen that $2E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] < D(E[\boldsymbol{\theta}|\mathbf{y}])$). Instability in pD can lead to problems in the use of this DIC. For example, mixture models, or more simply, models with multiple modes can "trick" the pD estimate because the overdispersion in such models (when the components are not correctly estimated) leads to a negative DIC. However, it is also true that inappropriate choice of hyper-parameters for variances of parameters in hierarchical models can lead to this problem, as can nonlinear transformations (such as changing from a Gaussian model to a log normal model). In such cases it is sometimes safer to compute the effective number of parameters as half the posterior variance of the deviance (Gelman et al., 2004), which is straightforward to compute from the G MCMC samples. Other measures based on predictive ability of the model can be computed, such as mean square predictive error (MSPE) (Gelfand and Ghosh, 1998) or the log pseudo-marginal likelihood (PML) (Chen et al., 2000), and more recently the Watanabe–Akaike information criterion (WAIC, Gelman et al., 2014) has been proposed as an improved alternative to DIC.

3 SPACE–TIME MODELING

A wealth of research has extended the spatial models outlined above to the spatio-temporal domain, because areal unit disease data are typically available for $j = 1, \ldots, J$ consecutive nonoverlapping time periods. We note that for some data sets consecutive time periods overlap, requiring more advanced modeling techniques such as data augmentation (see, for an example, Lee and

Lawson, 2016). A wide range of spatio-temporal models have been developed fuelled by the free availability of spatio-temporal areal unit data, and the choice of model to use depends on the public health question being answered. For example, models have been developed to: (i) estimate the overall temporal trend and spatial pattern in disease risk (Knorr-Held, 2000; Ugarte et al., 2010), as well as area-specific temporal trends (Bernardinelli et al., 1995; MacNab and Dean, 2001); (ii) estimate the effects of covariates on disease risk (Greven et al., 2011; Wheeler et al., 2015); (iii) identify clusters of areas that exhibit similar disease risks (Lee and Lawson, 2016) or temporal trends (Lawson et al., 2012; Anderson et al., 2016); (iv) quantify any changes in spatial variation and health inequalities over time (Napier et al., 2016); (v) identify step changes in disease risk between geographically adjacent areal units (Rushworth et al., 2017); and (vi) identify the emergence and spread of epidemic diseases (Mugglin et al., 2002). In what follows we outline key models in the field and describe their uses, but stress that different models will be appropriate depending on the goal of the analysis. In all models the data likelihood for area i and time period j is $Y_{ij} \sim \text{Poisson}(e_{ij}\theta_{ij})$, where (Y_{ij}, e_{ij}) again denote the observed and expected numbers of disease cases. In what follows we do not include covariates for notational simplicity when describing the space–time latent structures, but note that covariates are straightforward to include by the addition of $\mathbf{x}_{ij}^{\top}\boldsymbol{\beta}$ to each linear predictor described below.

The first space–time disease risk model was proposed by Bernardinelli et al. (1995), who estimated separate linear time trends for each areal unit using the model

$$\ln(\theta_{ij}) = \beta_0 + \alpha_i + (\lambda + \gamma_i)j, \tag{8}$$

where θ_{ij} is the risk of disease in the i-th areal unit and j-th time period. The i-th areal unit has a linear temporal trend in disease risk with intercept $\beta_0 + \alpha_i$ and slope $\lambda + \gamma_i$, and nearby areas are modeled as having similar intercepts and slopes by assigning CAR priors such as (4) to both $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_I)$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_I)$. This model was extended by MacNab and Dean (2001), who replaced the linear trends with area-specific natural cubic splines with the same number of knots. The corresponding knot coefficients were smoothed spatially via the CAR prior (7), leading to geographically neighboring areas exhibiting similar nonlinear trends. The original linear trend model (8) was extended by Anderson et al. (2016) by additionally clustering areas into groups based on sharing similar intercept and slope parameters.

Probably the most commonly used model is that proposed by Knorr-Held (2000), which uses an ANOVA-style decomposition of the variation in disease risk into spatial and temporal main effects and a space–time interaction. This model allows one to estimate the average spatial and temporal trends in disease risk and is given by

$$\ln(\theta_{ij}) = \beta_0 + \alpha_i + \lambda_j + \gamma_{ij}, \tag{9}$$

where both the overall spatial (α_i) and the overall temporal (λ_j) terms are modeled by BYM CAR priors (5) with spatial (**W**) and temporal (**D**) neighborhood matrices. Here **D** is a binary $J \times J$ matrix whose (j, j')th element equals 1 if $|j - j'| = 1$ and zero otherwise. Finally, four different priors are proposed for the space–time interactions γ_{ij}: *Type 1*, spatially and temporally independent interactions; *Type 2*, spatially independent and temporally autocorrelated interactions; *Type 3*, patially autocorrelated and temporally independent interactions; and *Type 4*, spatially and temporally autocorrelated interactions. These interaction structures are induced via the $IJ \times IJ$ spatio-temporal precision matrix $\mathbf{Q}(\mathbf{W}, \mathbf{D}) = \mathbf{Q}(\mathbf{W}) \otimes \mathbf{Q}(\mathbf{D})$, the Kronecker product of spatial and temporal precision matrices. Here $\mathbf{Q}(\mathbf{W}) = \mathbf{I}$, the $I \times I$ identity matrix for spatial independence, while $\mathbf{Q}(\mathbf{W}) = \text{diag}(\mathbf{W1}) - \mathbf{W}$ corresponding to the intrinsic CAR model (4) for spatial autocorrelation. Similarly $\mathbf{Q}(\mathbf{D}) = \mathbf{I}$ or $\mathbf{Q}(\mathbf{D}) = \text{diag}(\mathbf{D1}) - \mathbf{D}$ for temporal independence and autocorrelation, respectively.

The other commonly used spatio-temporal model replaces the three sets of spatial, temporal, and interaction random effects with a single set of spatio-temporally autocorrelated random effects modeled by a multivariate autoregressive process. Variations of this model have been proposed by Ugarte et al. (2012) and Rushworth et al. (2014), and it is an appropriate model if one wishes to estimate how the spatial risk surface has evolved over time. The model by Rushworth et al. (2014) has the form

$$\begin{aligned} \ln(\theta_{ij}) &= \beta_0 + \phi_{ij}, \\ \boldsymbol{\phi}_j | \boldsymbol{\phi}_{j-1} &\sim \text{N}\left(\gamma \boldsymbol{\phi}_{j-1}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho)^{-1}\right) \quad j = 2, \ldots, J, \\ \boldsymbol{\phi}_1 &\sim \text{N}\left(\mathbf{0}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho)^{-1}\right). \end{aligned} \tag{10}$$

Here $\boldsymbol{\phi}_j = (\phi_{1j}, \ldots, \phi_{Ij})$ is the vector of random effects at time j, which is allowed to change over time via a first-order autoregressive process with temporal autocorrelation parameter γ. The spatial dependence is induced via the precision matrix $\mathbf{Q}(\mathbf{W}, \rho) = \rho(\text{diag}(\mathbf{W1}) - \mathbf{W}) + (1 - \rho)\mathbf{I}$, which, when included in the joint distribution (3), corresponds to the CAR model (7) proposed by Leroux et al. (2000).

The above models have been described without covariate adjustment, as in this case the random effects represent the spatio-temporal variation in (log) disease risk. Covariates can easily be included in any of the above models by replacing the intercept term β_0 with $\mathbf{x}_{ij}^{\top}\boldsymbol{\beta}$, but now the interpretation changes as the random effects relate to the residual variation in disease risk after covariate adjustment. Thus if identifying spatio-temporal patterns, clusters, and changes in disease risk is the goal of the analysis, then covariates should not be included and the random effects are the primary parameters of interest. However, if the goal of the analysis is to estimate the effect of a

risk factor on disease, then covariates (the risk factor and confounders) should be included in the model and the random effects are nuisance parameters to remove the residual spatio-temporal autocorrelation rather than being of direct interest. In this latter case the latter model (10) could be viewed as preferable to (8) and (9), as it does not make the restrictive assumption of linearity in the temporal trends as (8) does and is more parsimonious than (9).

Spatio-temporal modeling is a burgeoning research field, and a large number of models have been proposed for specific purposes, such as epidemic models that allow for step changes in risk over time (Mugglin et al., 2002), and varying-variance models for identifying changes in the magnitude of health inequalities (spatial variation in disease risk) over time (Napier et al., 2016). Other spatio-temporal model development has been undertaken in the fields of cluster identification, boundary detection, and ecological regression, all of which are discussed in later sections of this chapter.

4 MULTIVARIATE MODELING

Often it is appropriate to consider the analysis of the geo-referenced distribution of more than one disease. For example, the focus may be on a group of diseases with similar etiology in an epidemiological study. Another example, in the context of public Health, could be the examination of the general health status of a region (possibly following a cluster alarm signal). In the latter case a range of disease types might be considered to find out if any show signs of unusual risk variation. In some cases, the focus is on the spatial distribution of the vector of disease risks, while in other cases the diseases are to be contrasted, or correlation between their spatial distributions are to be considered. In this chapter the focus will be on relative risk estimation and modeling of risk both in terms of correlation between diseases and in terms of comparison.

4.1 Likelihood Models

The data consist of counts of $l = 1, \ldots, L$, diseases in each of the $i = 1, \ldots, I$ small-areas, resulting in a matrix of $I \times L$ observations. The disease data for area i and disease l are (Y_{il}, e_{il}), which are, respectively, the observed and expected numbers of disease cases. Various approaches can be adopted to modeling these data depending on the focus. First, by conditioning on the total count within the small area, $Y_{T_i} = \sum_{l=1}^{L} Y_{il}$, it is possible to consider the multinomial distribution for the count probability vector $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iL})$. On the other hand, it is also possible to examine the unconditional distribution of the counts assuming conditional independence and a Poisson count distribution. In the first case, assume that

$$\mathbf{Y}_i \sim \text{Multinomial}(\mathbf{p}_i | Y_{T_i}),$$

where $\mathbf{p}_i = (p_{i1}, \ldots, p_{iL})$. Due to the constraint, the probability vector is defined to be $0 < p_{il} < 1$, and $\sum_{l=1}^{L} p_{il} = 1 \; \forall i$. The log-likelihood is then considered to be

$$l(\mathbf{Y}|\mathbf{p}) = \sum_{i=1}^{I}\sum_{l=1}^{L} Y_{il} \ln(p_{il}).$$

To model the probabilities, it is useful to assume that they arise from a normalization such as:

$$p_{il} = \frac{\lambda_{il}}{\sum_{k=1}^{L} \lambda_{ik}}$$

It is convenient to assume that rate terms consist of a log-linear function of covariates or random effects. A typical general example could be

$$\lambda_{il} = e_{il}\theta_{il},$$
$$\theta_{il} = \exp(\alpha_l + \mathbf{x}_i^{\top}\boldsymbol{\beta} + \phi_{il}),$$

where α_l is a disease-specific intercept and ϕ_{il} is a disease and area-specific random effects, possibly modeled by the BYM CAR model outlined earlier. Clearly by normalization, the conditioning on the total disease count in each area yields relative inference concerning the disease distribution. An example of the application of such a model was for three diseases at the county-level in Georgia, where the goal was a description of the relative risks of the three diseases. Alternative formulations of multivariate risk can be envisaged. Shared component models have been extended to multiple diseases by Held et al. (2005). In their formulation a Poisson likelihood is assumed:

$$Y_{il} \sim \text{Poisson}(e_{il}\theta_{il}),$$
$$\ln(\theta_{il}) \sim \text{N}(\alpha_l + \sum_{k=1}^{L} \delta_{kl}\phi_{ki}, \tau_l^2),$$

where the weights $\sum_{l=1}^{L} \delta_{kl} = 0$ and can be modeled by a multivariate log-normal distribution. Once more than two diseases are examined, however, the interpretation of a shared component is more difficult.

4.2 Multivariate Spatial Correlation and MCAR Models

4.2.1 Multivariate Gaussian Models

In general, once multiple diseases are admitted into an analysis there is a need to consider relations between the diseases. This can be done in variety of ways. A basic approach to this is to consider cross-correlation between the diseases. There is a considerable literature on the specification of cross-correlation models for Gaussian processes (see, e.g., Banerjee et al., 2004).

Let $\mathbf{Y}$ denote the vector of disease counts for all areas, ordered so that the first I observations are for all spatial units for disease 1, and so on. Then for a multivariate Gaussian process, a common assumption would be that

$$\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{A}_{\mathbf{Y}})$$

where $\boldsymbol{\mu}$ is an $I \times L$ mean vector and $\mathbf{A}_{\mathbf{Y}}$ is a $IL \times IL$ covariance matrix. It is convenient to consider a block representation of $\mathbf{A}_{\mathbf{Y}}$ which stresses the cross-covariances

$$\mathbf{A}_{\mathbf{Y}} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & . & \mathbf{A}_{1L} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & . & . \\ . & . & . & . \\ \mathbf{A}_{L1} & & & \mathbf{A}_{LL} \end{pmatrix}.$$

Here each of the diagonal block matrices is internal covariances across space, whereas the off-diagonal block matrices define the cross-correlations between components.

Various models can be assumed for the overall covariance structure of a set of Gaussian fields. Often simple assumptions are made to allow for computation. Banerjee et al. (2004) discuss various examples of separable models and asymmetric cases (mainly for simple situations where each field is measured on the same grid). They also extend the analysis by considering the linear model for coregionalization (LMC), which specifies that a multivariate process is a linear function of *iid* spatial processes with zero mean, variance 1 and a weakly stationary and isotropic spatial covariance function $\rho(h)$ for distance h. More generally separate covariance functions $\rho_l(h)$ can be assumed for each disease, so that the cross-covariance is defined as $\mathbf{A}_{ll'} = \sum_{j=1}^{L} \rho_j(||\mathbf{s} - \mathbf{s}'||) T_j$ for locations $\mathbf{s}$ and $\mathbf{s}'$, where T_j is the covariance matrix for the j-th component. An alternative, computationally attractive, conditional specification was also proposed by Royle and Berliner (1999).

While in general full multivariate cross-correlation models could be employed for modeling continuous multivariate spatial processes, their implementation is not straightforward and in particular their computational demands often force the consideration of simpler formulations. Note that within a disease mapping context these models could form joint prior distributions for spatial random effects (rather than models for observed Gaussian fields), especially for case event models where continuous spatial effects are naturally favored. Hence for the i-th case event we might be interested in the vector of intensities:

$$\boldsymbol{\lambda}(\mathbf{s}_i|\boldsymbol{\psi}) = \exp(\boldsymbol{\Delta}_i + \mathbf{V}_i)$$

where $\boldsymbol{\Delta}_i$ includes fixed and uncorrelated random effects and $\mathbf{V}_i$ is part of a multivariate spatial Gaussian process. For count data this might take the form, for the i-th small area with area (denoted a_i) as:

$$\boldsymbol{\theta}_i = \exp(\boldsymbol{\Delta}_i + \boldsymbol{\mu}_i)$$
$$\boldsymbol{\mu}_i = \int_{a_i} \mathbf{V}(u)du$$

Often for count data and approximately for case event data a Markov random field (MRF) specification is adopted at least for simplicity of implementation. In the next section these multivariate CAR models are discussed.

4.2.2 MCAR Models

The MCAR model of Gelfand and Vounatsou (2003) specifies that the $I \times L$ matrix of random effects $\boldsymbol{\phi}$ in the model

$$Y_{il} \sim \text{Poisson}(e_{il}\theta_{il} = \exp[\mathbf{x}_i^\top \boldsymbol{\beta}_l + \phi_{il}])$$

is defined with a constraint that the spatial effects separate into nonspatial and spatially structured effects:

$$\boldsymbol{\phi} \sim \text{N}(\mathbf{0}, \mathbf{H}_1),$$

where $\mathbf{H}_1 = [\boldsymbol{\Lambda} \otimes (\mathbf{D} - \alpha\mathbf{W})]^{-1}$ with $\otimes$ denoting Kronecker product, and $\mathbf{D}$ is a $I \times I$ diagonal matrix with elements which are the number of neighbors of the i-th region and $\mathbf{W}$ is an adjacency matrix. Here $\boldsymbol{\Lambda}$ is a $L \times L$ positive definite matrix of nonspatial precisions, defining the relation between diseases and α is a common spatial autocorrelation parameter. This is denoted as the MCAR(α, $\boldsymbol{\Lambda}$) model. This model can be extended to allow for separate autocorrelation (smoothing) for each disease:

$$\boldsymbol{\phi} \sim \text{N}(\mathbf{0}, \mathbf{H}_2),$$

where $\mathbf{H}_2 = [\mathbf{Q}(\boldsymbol{\Lambda} \otimes \mathbf{I}_{I\times I})\mathbf{Q}^\top]^{-1}$ and $\mathbf{Q} = diag(\mathbf{R}_1, \ldots, \mathbf{R}_L)$ and $\mathbf{R}_l = chol(\mathbf{D} - \alpha_l\mathbf{W})$, $l = 1, \ldots, L$, where $chol()$ denotes the Cholesky decomposition. This has been termed the MCAR($\boldsymbol{\alpha}$, $\boldsymbol{\Lambda}$). Extensions and variants to these models have been proposed by Kim et al. (2001) and Jin et al. (2005). Restriction to the conditional ordering of the effects in the GMCAR model of Jin et al. (2005) has led to a different approach.

5 SOFTWARE

The growth in research in spatial and spatio-temporal areal unit modeling has exploded since the seminal work of Besag et al. (1991), and in addition to disease mapping includes fields as diverse as agriculture (Besag and Higdon, 1999), ecology (Brewer and Nolan, 2007), education (Wall, 2004), image analysis (Molina et al., 1999), and social science (Leckie and Goldstein, 2014). The growth in this field has been aided by increasing amounts of freely available data at the areal unit level, such as the Surveillance, Epidemiology, and End Results (SEER, https://seer.cancer.gov) cancer registry, and small-area statistics repositories such as Statistics.gov.scot (http://statistics.gov.scot) in

Scotland and Neighbourhood Statistics (http://www.neighbourhood.statistics.gov.uk/dissemination/) in England. Early model development in this field was implemented in `Fortran` or `C++` and was not made available to a general audience (for example, Knorr-Held, 2000). However, both the wealth of methodological advances and the desire from applied researchers to fit spatial and spatio-temporal models have led to a recent growth in freely available software for fitting these models.

When fitting these models the first decision a researcher has to make is whether to use the frequentist or Bayesian paradigms. Frequentist implementations of these models typically use penalized quasi-likelihood algorithms (PQL, Breslow and Clayton, 1993), while Bayesian implementations use either MCMC (see Robert and Casella, 2010) simulation or integrated nested Laplace Approximations (INLA, Rue et al., 2009). The majority of the software implements the Bayesian paradigm, with inference based on MCMC simulation. The software generally falls into one of two types, general purpose software where the user can develop their own models with a degree of programming, and specialized software that can fit a limited set of models with one-line function calls. The advantage of the former is its flexibility and generality, but the latter is much easier to use due to not requiring any programming to fit models.

5.1 General Purpose Software

The most predominant piece of software used to fit spatial models is the Bayesian inference using Gibbs sampling (`BUGS`, Lunn et al., 2000, http://www.mrc-bsu.cam.ac.uk/software/bugs/) software environment, which is a general purpose programming environment for fitting models in a Bayesian setting using MCMC simulation. The software includes both `WinBUGS` and `OpenBUGS` versions and has inbuilt functionality to fit the intrinsic and BYM CAR models (4) and (5) via the function `car.normal()`, as well as the proper CAR model (6) via the function `car.proper()`. Additionally, it has the inbuilt function `mv.car()` for fitting a simple MCAR model to multivariate disease data. Other model types can be programmed by the user, such as autoregressive time series processes, making BUGS one of the most flexible spatio-temporal modeling tools available. However, this flexibility means that the user has to write the model they wish to fit out syntactically, which is not straightforward for nonexperts. Additionally, when fitting spatial models BUGS requires the neighborhood information contained in $\mathbf{W}$ to be given in three different forms, making spatial models lengthy to set up.

The second piece of general purpose software is the `R` package `INLA` (Rue et al., 2009, http://www.r-inla.org), which fits models in a Bayesian setting using INLA. The software is designed to fit models where the linear predictor includes, among other things, a latent GMRF, of which a CAR model is a special case, and is thus suited to fitting spatio-temporal areal unit models. For purely spatial data it can fit models (4) through (7) using the functions `besag()`, `besagproper()` and `generic1()`, while temporal models can be

fitted using autoregressive-type models such as `ar1()` and `ar()`. Thus spatio-temporal models can be built by combining these elements together. In common with BUGS it can fit a wide range of models, but again requires a degree of specialist knowledge in how to combine the models to implement them. It also does not use simulation-based inference for posterior estimation of parameters, so is much faster than BUGS, but does provide the user with simulated draws from the joint posterior distribution from which quantities of interest can be computed. In addition, it does not currently provide multivariate models for spatial correlation such as MCAR or MVCAR.

The other general purpose modeling tool that can fit spatio-temporal areal unit models is `BayesX` (Belitz et al., 2012, http://www.statistik.lmu.de/~bayesx/bayesx.html), which can fit the intrinsic and BYM models, as well as temporally autocorrelated autoregressive models. The software can implement these models in both frequentist and Bayesian frameworks and focuses on the general class of structured additive regression models.

5.2 Specialized Spatial Modeling Software

One of the most comprehensive `R` packages designed to fit spatial CAR-type models is `CARBayes` (Lee, 2013), which is available at https://cran.r-project.org. The package fits models with binomial, Gaussian and Poisson data likelihoods in a Bayesian setting using MCMC simulation, where the linear predictor can depend on covariates, an offset and a set of random effects. The latter can be modeled with a number of different CAR priors, including common specifications such as (4) and (7) using the function `S.CARleroux()`, and (5) using the function `S.CARbym()`. The software can also fit more complex models, such as the boundary detection model proposed by Lee and Mitchell (2012) and the localized smoothing model proposed by Lee and Sarran (2015). The package can also fit a simple MCAR model to multivariate disease data based on the CAR priors (4) and (7) using the function `MVS.CARleroux()`.

Spatio-temporal modeling can be implemented in the sister `R` package `CARBayesST` (Lee et al., 2016), which can fit a range of spatio-temporal models including those commonly used and given by (8), (9), and (10), albeit with (9) utilizing the CAR priors (4) and (7) rather than (5). It can also fit a range of more complex spatio-temporal models, such as those proposed by Lee and Lawson (2016), Napier et al. (2016), and Rushworth et al. (2017). In common with the `BUGS` software both `CARBayes` and `CARBayesST` provide the user with a summary output about the MCMC samples, including posterior medians, 95% credible intervals, the convergence diagnostic proposed by Geweke (1992) and the effective number of independent samples generated. An example using the `CARBayesST` software is given in the next section of this chapter.

A number of other `R` packages exist for fitting specialized spatial and spatio-temporal areal unit models. Species distribution models can be fitted using the `hSDM` package (Vieilledent et al., 2014), while zero-inflated binomial

and Poisson models can be fitted using `Spatcounts` package (Schabenberger, 2009). Simultaneous autoregressive (SAR) models can be fitted within the `spdep` package (Bivand and Piras, 2015), while CAR models can be fitted in a frequentist setting via PQL using the `hglm` package (Ronnegard et al., 2010). In a spatio-temporal setting the `plm` (Croissant and Millo, 2008) and `splm` (Millo and Piras, 2012) packages model panel data, while the `surveillance` package (Michael Höhle and Paul, 2015) models epidemic data. Finally, multivariate spatio-temporal modeling is in its infancy, and thus no specific software currently exists for this type of data.

6 CLUSTER IDENTIFICATION

Cluster detection refers to the identification of an area or a small group of spatially contiguous areas that exhibit an elevated risk of disease compared to the surrounding areas. The identification of such high (and low) risk clusters is of key interest to public health practitioners, as it enables them to target interventions, such as an educational campaign about the risk factors of the disease in question, at areas at the greatest need. Additionally, the identification of high-risk clusters aids the search for, hitherto unknown, etiological risk factors that may affect the disease. However, cluster identification is an inherently difficult task, as one it attempting to partition the unknown continuously varying risk surface into two (or more) groups, such as high-risk and baseline. However, for the reasons outlined above it is a very popular research area, and a number of different methodologies have been proposed.

One of the most popular approaches is scan statistics via the SaTScan software (Kulldorff et al., 2005; http://www.satscan.org). This approach uses a hypothesis testing framework and identifies a small number of potential clusters and tests the null hypothesis that their risks are the same as those in the remaining areas. Different shaped spatial or spatio-temporal clusters can be identified, but they do not provide an estimate of the entire risk surface. A critique of their limitations is provided by Wakefield and Kim (2013), who also propose a Bayesian alternative and software for implementation (the `R` package `SpatialEpi`).

The simplest approach to cluster detection within the framework of the general model (2) is to use a posterior exceedance probability (PEP) as discussed by Richardson et al. (2004), which for area i is given by

$$v_i = \mathbb{P}(\theta_i > 1|\mathbf{y}), \tag{11}$$

the posterior probability of having a risk greater than 1 given the data $\mathbf{y}$. Area i can be classified in a high-risk cluster if $v_i > C$, for a given threshold C that is likely to be problem dependent. An alternative approach is to fit model (2) to the data and applies a clustering algorithm or model to the estimated risk surface in a postestimation step. One could choose two clusters, baseline and high-risk, and use the algorithm to identify areas in the latter group.

An example of this general approach is described in Charras-Garrido et al. (2013), who applied a Gaussian mixture model to the risk surface. However, this approach does not guarantee spatially contiguous clusters, so the authors used an additional postprocessing step to obtain such contiguity. Numerous clustering methodologies exist to carry out this postestimation step, ranging from algorithmic approaches such as *K-means* to fully model-based approaches such as mixture models. A more general criticism of approaches based on model (2) fits a spatial smoothing model to the data, thus creating a spatially smoothed risk surface. This smooth risk surface is then analyzed to identify clusters that have boundaries (discontinuities) separating areas within and outside the cluster. Thus one is attempting to identify changes in risk between neighboring areas (one inside and one outside a cluster) that have had their risks smoothed toward each other, thus smoothing away the very variation one is trying to identify.

Therefore a number of researchers including Gangnon and Clayton (2000), Knorr-Held and Raßer (2000), Green and Richardson (2002), Charras-Garrido et al. (2012), and Forbes et al. (2013) have proposed partition models, where the risk surface is partitioned into a small number of different risk classes. The data likelihood for the general model (2) is then altered to the form:

$$Y_i \sim \text{Poisson}(e_i\theta_{z_i}) \quad i=1,\ldots,I, \tag{12}$$

where z_i is the allocation variable for area i taking values $z_i = 1, 2, \ldots, q$, denoting the q possible risk levels. Here the risk levels are denoted by $(\theta_1, \ldots, \theta_q)$ and are typically ordered so that $\theta_r < \theta_{r+1}$ to ensure parameter identifiability. The allocation variables $\mathbf{z} = (z_1, \ldots, z_I)$ are typically modeled by a spatially autocorrelated discrete Markov model, and Charras-Garrido et al. (2012) use the hidden Markov model given by

$$\mathbb{P}(\mathbf{Z}=\mathbf{z}|\alpha,\gamma) \propto \exp\left(-\sum_{i=1}^{I}\varphi_1(z_i|\alpha)+\gamma\sum_{i=1}^{I}\sum_{r=1}^{I}w_{ir}\varphi_2(z_i,z_r)\right). \tag{13}$$

Here (φ_1, φ_2) are potential functions of orders 1 and 2, and, respectively, control the proportions of the different risk classes and the spatial autocorrelation in the risk classes. The latter could be the Potts model, but in their work (Charras-Garrido et al., 2012) consider L1 and L2 norm MRF potential functions given by

$$\varphi_2(z_i,z_r)=1-\frac{(z_i-z_r)^2}{q-1} \qquad \varphi_2(z_i,z_r)=1-\frac{|z_i-z_r|}{q-1}, \tag{14}$$

which induce spatial smoothness in $\mathbf{Z}$. An EM algorithm is used to fit the model, and full details are provided with the paper.

Alternatively, a two-stage approach was proposed by Anderson et al. (2014), who created a set of possible cluster structures in stage 1 by applying a hierarchical agglomerative clustering algorithm to data on disease risk from

an earlier time period to that being studied. The second stage then fits a model similar to (1) to the study data with cluster fixed effects for each cluster structure, and the best model is chosen by the DIC (Spiegelhalter et al., 2002). Finally, cluster identification has been extended to the spatio-temporal domain by Lee and Lawson (2016) using a partition-type approach similar in spirit to those outlined above, while Lawson et al. (2012) and Anderson et al. (2016) have extended the clustering paradigm by clustering areas based on temporally evolving latent structures and linear trends, respectively.

7 BOUNDARY DETECTION (WOMBLING)

A related area is that of boundary detection, which is also known as *Wombling* following the seminal work of Womble (1951). Boundaries occur where there are large step-changes, called *difference boundaries*, in disease risk between geographically adjacent areal units. Thus two areas that are separated by a difference boundary can be studied to identify differences in possibly unknown etiological factors that may affect disease risk. A simple approach to identifying such boundaries was proposed by Lu and Carlin (2005), who proposed a model similar to (2), and then computed

$$\hat{\Delta}_{ir} = \frac{1}{G}\sum_{g=1}^{G}|\theta_i^{(g)} - \theta_r^{(g)}|, \tag{15}$$

the posterior mean absolute difference in risk between neighboring areas (i, r), where $\{\theta_i^{(1)}, \ldots, \theta_i^{(G)}\}$ are G samples from the posterior distribution of risk in area i. They then identified difference boundaries between pairs of areas with the largest $\hat{\Delta}_{ir}$ values, and the top 20% and 50% of values were chosen to visually display on a map. However, the choice of the top 20% or 50% is arbitrary, and in common with the critique of clustering methods above, identifying differences between neighboring areas in a spatially smoothed risk surface is somewhat counter intuitive. This observation hints at a more fundamental problem often ignored when modeling spatial areal unit data, namely that given these difference boundaries exist, one should not naively use a CAR prior that smooths spatially over these boundaries. For example, multivariate Gaussian theory shows that the partial autocorrelation between (ϕ_i, ϕ_r) given all the remaining random effects $\boldsymbol{\phi}_{-ir}$ from the intrinsic CAR model (4) is given by:

$$\text{Corr}[\phi_i, \phi_r | \boldsymbol{\phi}_{-ir}] = \frac{w_{ir}}{\sqrt{\left(\sum_{l=1}^{n} w_{il}\right)\left(\sum_{l=1}^{n} w_{rl}\right)}}. \tag{16}$$

Thus all pairs of geographically close areal units that are defined to be neighbors via $\mathbf{W}$ will have partially autocorrelated random effects (ϕ_i, ϕ_r),

even if their true risks are very different and a difference boundary exists between them. Thus fitting this model will smooth away these difference boundaries, leading to poorer estimation of the risk surface. Therefore a number of approaches have extended CAR models to allow for localized (or adaptive) smoothing, by modeling the elements of $\mathbf{W}$ corresponding to geographically adjacent areal units as random quantities to be estimated rather than being fixed equal to one. In this setting if w_{ir} is estimated as greater than zero then (ϕ_i, ϕ_r) are modeled as partially autocorrelated and are smoothed toward each other, where as if w_{ir} is estimated as zero then (ϕ_i, ϕ_r) are modeled as conditionally independent and no such spatial smoothing is enforced. In this latter case a *difference boundary* is said to exist between areas (i, r), giving a natural framework for undertaking *Wombling*. The first example of this enhanced *Wombling* framework was Lu et al. (2007), who modeled w_{ir} by the logistic regression model

$$\begin{aligned} w_{ir} &\sim \text{Bernoulli}(p_{ir}) \\ \ln\left(\frac{p_{ir}}{1-p_{ir}}\right) &= \alpha_0 + \alpha_1 z_{ir}. \end{aligned} \tag{17}$$

Here $z_{ir} = |z_i - z_r|$ is a dissimilarity metric, a measure of the difference in an important covariate between areas (i, r). Thus the larger the dissimilarity metric the more likely it is that a difference boundary exists between areas (i, r). This approach was extended by Ma et al. (2010) to additionally include a set of random effects in the linear predictor modeled by a second stage CAR prior, while Ma and Carlin (2007) replaced the logistic regression model with an Ising model. Following concerns from Li et al. (2011) over the parsimony in these models Lee and Mitchell (2012) removed the logistic regression component and proposed the following similar but more parsimonious alternative

$$w_{ir} = \begin{cases} 1 & \text{if } \exp(-\alpha z_{ir}) \geq 0.5 \text{ and } i \sim r \\ 0 & \text{otherwise} \end{cases}, \tag{18}$$

where $i \sim r$ means areas (i, r) are spatial neighbors. This model only has a single additional parameter α and can be applied using the `CARBayes` package.

An alternative algorithmic approach was proposed by Lee and Mitchell (2013), which iterated between: (i) estimating the model parameters conditional on the current fixed $\mathbf{W}$ matrix; and (ii) deterministically updating $\mathbf{W}$ based on the current posterior distribution of the other model parameters. A deterministic update of $\mathbf{W}$ was used so that the algorithm did not suffer from the parameter identifiability problems outlined above. The algorithm terminates when $\mathbf{W}$ does not change from one iteration to the next, and extensive simulations showed this normally occurred before five iterations of the algorithm. In contrast, Rushworth et al. (2017) extended the random $\mathbf{W}$ framework to the spatio-temporal domain and used the temporal replication in the data to better estimate the elements of $\mathbf{W}$ corresponding to geographically adjacent

areal units. Each element was modeled on the unit interval via a logit transformation, and the values were smoothed globally by a nonspatial shrinkage prior. A second level CAR model was also considered, but did not perform as well as difference boundaries are not typically spatially clustered. Finally, modeling $\mathbf{W}$ is not the only way to achieve adaptive spatial smoothing in areal unit data, as Brewer and Nolan (2007) allowed the variances to vary spatially, yielding smaller variances and hence more smoothing in some regions compared to others. Alternatively, Lawson and Clark (2002) augmented the BYM CAR prior with an additional set of random effects with a prior based on the nonsmooth L1-norm penalty. They extended the linear predictor in model (2) to

$$\ln(\theta_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \zeta_i \phi_i^{(1)} + (1-\zeta_i)\delta_i + \phi_i^{(2)}.$$

Here $(\phi_i^{(1)} + \phi_i^{(2)})$ are modeled by the BYM prior (5). The spatially autocorrelated variation is partitioned into smooth variation modeled by $\phi_i^{(1)}$ and jumps modeled by δ_i. The latter is assigned a L1 Laplace like penalty prior given by

$$f(\boldsymbol{\delta}|\sigma^2) \propto \frac{1}{\sqrt{\sigma^2}} \exp\left(\frac{1}{\sigma^2}\sum_{i=1}^{I}\sum_{r=1}^{I} w_{ir}|\delta_i - \delta_r|\right). \tag{19}$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_I)$. Finally, ζ_i is the mixture weight assigned to the spatially autocorrelated components.

8 ECOLOGICAL REGRESSION IN PUBLIC HEALTH

The majority of this chapter has focused on estimating spatio-temporal patterns in disease risk, but another important area in public health is ecological regression, which involves estimating the health impact of both good (e.g., green space) and bad (e.g., air pollution) exposures. The results from such studies are prone to *ecological bias*, which occurs when the estimated population-level association is assumed to relate to individuals. A full discussion of ecological bias can be found in Wakefield and Salway (2001). Despite these problems ecological regression is an active research area, with examples focusing on the effects of air pollution (Greven et al., 2011), greenspace (Wheeler et al., 2015) and socio-economic deprivation (e.g., Mackenbach et al., 1997).

In addition to ecological bias, a number of other problems need to be addressed when conducting such studies, including (i) allowing for uncertainty in the exposure when estimating its health effect; and (ii) the potential for collinearity between spatially smooth exposures and the random effects included in the model to account for unmeasured residual spatial autocorrelation in the disease data. Model development for the former is in its infancy,

and early solutions are outlined in Blangiardo et al. (2016) and Lee et al. (2017). The potential for collinearity between spatially smooth covariates and spatially smooth random effects was first noticed by Clayton et al. (1993), and has since resulted in detailed analyzes by Reich et al. (2006) and Paciorek (2010). A number of general solutions have been outlined to this problem, including retaining the CAR prior for the random effects and increasing its flexibility by treating $\mathbf{W}$ as a random quantity, using the models outlined in Section 7. Modeling $\mathbf{W}$ in this way allows the random effects to be more flexible and capture areas of spatial smoothness as well as distinct step changes (difference boundaries) in disease risk, which should result in less correlation between the random effects and a globally spatially smooth covariate (see, for an example, Lee et al., 2014).

An alternative approach was proposed by Lawson et al. (2012), who suggested estimating the covariate effects and the residual spatio-temporal autocorrelation separately in two steps. First they computed the residuals, on the linear predictor scale, from a simple Poisson log-linear model with no random effects, before modeling these residuals with a mixture model in stage 2. This mixture model allows complex spatio-temporal autocorrelation structures to be estimated, and these residual autocorrelation structures are then included in a final model as a fixed offset term. A third approach is orthogonal smoothing, which is where the random effects used to capture the residual spatial autocorrelation are forced to be orthogonal to the covariates, thus removing the collinearity between the two. The first approach in this vein was proposed by Reich et al. (2006), but they did not force the random effects to be spatially smooth in their model. Therefore their approach was extended by Hughes and Haran (2013), who replaced the random effects $\boldsymbol{\phi}$ with a set of spatially smooth basis functions that were orthogonal to the covariates. This model can be implemented by the `R` package `ngspatial`, for further details see Hughes and Haran (2013).

A special case of ecological analysis arises when environmental pollution exposure is thought to be sourced from fixed locations. This special case is sometimes known as putative source analysis or focussed clustering and has spawned a wide range of methods. Space limits our review of this important Public Health topic. However, recent reviews of this area can be found in Lawson (2013, chap. 7) and Lawson et al. (2016, chap. 14).

9 DISEASE MAP SURVEILLANCE

In many situations the retrospective analysis of disease maps is a major focus. In fact, the analysis described above has been applied to a retrospective collection of health outcomes. By *retrospective*, we mean that the events have already happened, i.e., we know the history of the process. There are situations where there is a need to consider how new data will be characterized and how well we can predict *new* events. In application to disease maps, we would be

focussed on the prediction of (say) counts of asthma in census tracts in 2018, when we only have counts up to 2017. This is in essence *prospective* analysis of disease outcomes and is the focus of this section.

9.1 Surveillance Concepts

The US Center for Disease Control (CDC) defines public health surveillance as:

> *the ongoing, systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those who need to know. The final link of the surveillance chain is the application of these data to prevention and control. A surveillance system includes a functional capacity for data collection, analysis, and dissemination linked to public health programs.*

This definition stresses that surveillance is essential in public health practice and that timeliness is a key component, as is dissemination to relevant parties. The ability to predict future disease outcomes is an important aspect of this activity. Timeliness can be in terms of minutes, hours, or days for diseases that spread quickly (highly infectious agents), or in terms of months or years for less infectious or noninfectious diseases. Following the September 11, 2001 World Trade Center attacks (9/11), there was raised awareness of the need for fast response to potential public health threats. This awareness has increased interest in being able to monitor and respond quickly to changes in disease behavior. These both lead to the need to have efficient computational algorithms that can quickly process health data, and the need to make inference as early as possible.

Traditional public health surveillance was originally designed as a passive system whereby cases of disease (usually infectious) were reported individually as they arise, usually by primary care physicians. This individual case reporting allowed more detailed analysis of each occurrence. Prospective surveillance of data on disease by national or regional health agencies has historically been limited by the fear of data trawling and a high false positive rate. However, since 9/11 there is now a great emphasis on public health agencies' ability to react quickly and to detect unusual disease events within large databases. In fact data mining of multiple disease data streams has become important, partly because of the unfocused targeting of potential bioterrorist threats (Azarian et al., 2009). That is, it is unlikely that the targeting of a particular disease will be known, and so a range of diseases must be considered. Added to this problem is the fact that rare diseases are often not well understood in their transmission dynamics or degree of risk (for example, anthrax inhalation has hardly been studied in humans (Brookmeyer, 2006; Buckeridge et al., 2006).

9.2 Syndromic Surveillance

Sosin (2003) noted that Syndromic surveillance can be defined as the

> *systematic and ongoing collection, analysis, and interpretation of data that precede diagnosis (e.g., laboratory test requests, emergency department chief complaint, ambulance response logs, prescription drug purchases, school or work absenteeism, as well as signs and symptoms recorded during acute care visits) and that can signal a sufficient probability of an outbreak to warrant public health investigation.*

A recent review of the area (Katz et al., 2011) stresses the range of definitions available, but also stresses the usefulness of syndromic surveillance within the greater public health community. Statistical challenges facing early detection of disease outbreaks has been reviewed recently by Shmeuli and Burkom (2010) and Fricker (2011). However, only limited discussion has been focussed on Bayesian approaches or spatial and spatiotemporal approaches. An early volume that does address spatial issues is Lawson and Kleinman (2005).

9.3 Process Control Ideas

In general it is possible to consider the monitoring of health outcomes as a form of process control, in that we want to consider action when some "unusual" disease event or event sequence occurs. In industrial process control usually an overall mean level for a process is specified, and limits are set around the mean level. Action is taken when the process strays too far from the mean level and exceeds the limits specified. Control of variability of a process can also be considered. It is possible to consider jointly variability and mean level and also to extend this to multivariate applications, where multiple processes are to be monitored. Extending these ideas to public health disease monitoring is possible but some important caveats must be considered:

- disease events are discrete and at the finest resolution level form a point process.
- disease events occur within a population that can vary in time both in size but also in characteristics.

In the case of aggregate counts within time periods, then it is common to assume a Poisson data level model with observed counts $\{Y_j\}$ for J time periods $\{t_j\}$ with associated expected counts $\{e_j\}$ and relative risks $\{\theta_j\}$. Hence, $Y_j \sim \text{Poisson}(e_j\theta_j)$ is the commonly used model. Often a log link to a linear predictor is assumed and then models similar to those for the intensity of point process can be defined, such as:

- linear trend: $\ln(\theta_j) = \alpha + \beta t_j$.
- linear trend with random effects: $\ln(\theta_j) = \alpha + \beta t_j + \phi_j$, where ϕ_j is a correlated or uncorrelated noise term.

Extensions to these models can include a variety of additional features such as seasonal or cyclic effects or specific forms of background and foreground clustering and also nonlinear or semiparametric forms. A very useful review of temporal methods, albeit in the context of infectious diseases, can be found in Unkel et al. (2012).

9.4 Single Disease Sequence

It is common to use the predictive distribution in a Bayesian model to compare the observed with predicted outcomes (see, e.g., Lawson et al., 2004, Vidal-Rodeiro and Lawson, 2006a). The use of residual-based methods has been criticized by Frisén and Sonesson (2005), as they do not consider any past evidence for changes in risk. Optimal surveillance (Frisen, 1992; Frisen and Mare, 1991) seeks to evaluate the historical backdrop of a risk change. Some alternative approaches to risk monitoring have been proposed (see, e.g., Grigg and spiegelhalter, 2007), such as Bayesian Hidden Markov models for temporal surveillance data by Strat and Carrat (1999), but they only apply the approach to retrospective analysis. Switching models where a latent binary variable have been proposed by Martinez-Beneito et al. (2008) and Al-Hadhrami and Lawson (2011).

9.5 Multiple Disease Sequences

Within a Bayesian formulation it is straightforward to extend temporal models into multiple time series. Define the vector of L disease outcomes at the j-th time period as $\boldsymbol{Y}_j = (Y_{j1}, \ldots, Y_{jL})$. At the data level a Poisson model could be appropriate, that is $Y_{jk} \sim \text{Poisson}(e_{jk}\theta_{jk})$. A log link can be assumed so that $\ln(\theta_{jk}) = \alpha_0 + \alpha_{1k} + \mathbf{x}_j^\top \boldsymbol{\beta}_k + \phi_{jk}$, which includes an overall (α_0) and disease-specific intercepts (α_{1k}), and ϕ_{jk} a set of time-dependent random effects. In this formulation it is possible to assume a correlation between parameters across diseases. The regression parameters $\boldsymbol{\beta}_k$ could be correlated across diseases via a multivariate Gaussian prior, and similar assumptions could be made for the correlation between ϕ_{jk}. An alternative, syndromic, view is that different diseases drive or presage other diseases, and so a conditional model might be favored. For example, with two diseases (a, b) it might be that:

$$\begin{aligned}\ln(\theta_{jb}) &= \alpha_0 + \alpha_{1b} + \mathbf{x}_j^\top \boldsymbol{\beta}_b + \phi_{jb},\\ \ln(\theta_{ja}|b) &= \alpha_0 + \alpha_{1a} + \mathbf{x}_j^\top \boldsymbol{\beta}_a + f(\theta_{jb}, \theta_{j-1,b}) + \phi_{ja}.\end{aligned}$$

Examples of multivariate modeling in retrospective analysis are found in Paul et al. (2008). However, there are few examples of prospective analysis in this area (Paul and Held, 2011).

9.6 Infectious Disease Surveillance

In the case of infectious diseases, it is usually important to consider mechanisms of transmission in the formulation. Hence some transmission dynamic is

usually included within models and transmission rates are estimated. A recent review of this area is given by Unkel et al. (2012). For influenza and meningococcal disease time series, Paul et al. (2008) examined likelihood models only, although Bayesian switching models have been proposed by Martinez-Beneito et al. (2008). Usually, infectious disease models have mechanistic dependencies on previous levels or counts of disease so that $E(Y_j) = f(Y_{j-1})$ or $E(Y_j) = f(Y_{j-1}, Y_{j-2}, Y_{j-3}\ldots)$ or specified as a function of the mean level: $E(Y_j) = f(\mu_j, \mu_{j-1}\ldots)$, where $\mu_j = e_j\theta_j$. Combinations of both forms have also been proposed. An additional consideration in infectious disease surveillance is whether an endemic component should be modeled. Some diseases have background incidence which occurs outside of epidemic periods, and so a two component model is sometimes recommended: with an endemic and epidemic part (Held et al., 2006). Very rare diseases usually do not have such an endemic component (e.g., plague), but more common diseases may require one (e.g., STDs). Influenza, for example, often displays limited background incidence outside of epidemic periods. Endemicity does depend on the temporal scale of the study as well. Longer time periods are more likely to find incidence of rare diseases.

9.7 Spatial and Spatiotemporal Surveillance

At a fundamental level, observation of disease in space and time usually consists of a location ($\mathbf{s}$) and a date of diagnosis (t) and over a study period a sequence of disease occurrences is found at different locations and times. This sequence: $\{(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2), (\mathbf{s}_3, t_3)\ldots\}$ forms a point process in space–time. In this situation, spatio-temporal point process models can be applied and there has been some application of these in a non-Bayesian setting (see, e.g., Lawson and Leimich, 2000, Diggle et al., 2004, Diggle, 2005, Rowlingson et al., 2005, Diggle, 2007). Aggregation of events into I fixed spatial and J temporal units, yields counts of disease: Y_{ij}, for which the usual Poisson disease model would be appropriate with risk θ_{ij}. Then a log-linear form $\ln(\theta_{ij}) = P_{ij}$ is typically used, where P_{ij} is a predictor of some form which could include covariates (such as syndromic variables) and also smooth (random) effects. Two major issues arise first, how should components appear in P_{ij} to correctly describe "normal" health behavior and "unusual" behavior? Second, how can prospective analysis be carried out within a Bayesian paradigm?

9.7.1 *Components of* P_{ij}

In conventional spatio-temporal modeling (usually for noninfectious diseases) a separable decomposition is often assumed: that is $P_{ij} = \alpha + S_i + T_j + ST_{ij}$. This decomposition consists of main effects in space (S_i), time (T_j) and an interaction effect (ST_{ij}). The question then arises: is this parameterization appropriate for prospective surveillance? For case events it can be assumed that S_i and T_j can be regarded as background and the space–time interaction

component (ST_{ij}) would be monitored. The questions then raised could be how identifiable are these components? and how can estimation proceed when differential risk changes occur? How can endemicity be accommodated?

9.7.2 Component Identification and Estimation Issues

Prospective surveillance requires that decisions be made about the health system as new observations arrive. Define the current disease count map at the j-th time as $\{Y_{ij}\}$, while historical data are denoted $\{Y_{i,\ j-1}\}$, $\{Y_{i,\ j-2}\}$, … For prospective analysis there are a variety of estimation issues: (1) should refitting of a model take place when new data arrives? (2) If refitting is not carried out then do we adjust the model when unusual events occur? (3) should there be temporal components estimated for the normal background disease levels? Measures that add surveillance detection to the posterior estimation have been developed by Corberán-Vallet and Lawson (2011), Corberán-Vallet and Lawson (2014), Rotejanaprasert et al. (2016), and Rotejanaprsasert and Lawson (2017).

9.7.2.1 Refitting

A fundamental feature of prospective analysis is that new data arrives and enlarges the data set available and also possibly the parameter set may increase. To accommodate this it has often been proposed that models should be refitted with new data (see, e.g., Vidal-Rodeiro and Lawson, 2006b, Paul and Held, 2011). Note that if new parameters appear then the model actually changes its structure each time data appears. However, refitting of models in prospective surveillance may lead to accommodation of data that is "unusual," as the model attempts to adapt to the new data (Lawson, 2004, 2005). Clearly some comparison of the new data with model "predictions" at the new time point should be considered (before any refitting is considered). Following this comparison, decisions about the state of the system should be made. This leads on to estimation considerations. With a set of spatial units not all units will signal at any given time. Hence, some areas will no longer be in a background state. If areas have signaled, then a decision must be made concerning whether to "freeze" the background effect for those spatial units. The inclusion of temporal effects in normal background can also lead to estimation problems.

9.7.2.2 Background and Endemicity

The issue of what should be included in the "normal" disease background level is a major issue. Note that endemicity, a term synonymous with prevalence, represents the usual behavior of disease in an area. The correct modeling of this endemicity or background is important. Using historical data in the estimation of background has advantages, not least of which is the ability to self-control area risk. However, there is a disadvantage in that a choice of

historical period that is stable could be misjudged. If the historical data includes unusual disease artifacts then false negatives could be found. Of course the use of expected counts could also be influenced by an inappropriate choice of reference population or period.

10 SPATIAL SURVIVAL ANALYSIS

In many biostatistical applications there is a need to consider temporal variation. The commonest examples are often found in clinical or behavioral intervention trials where a state can be reached by a patient. The time at which the state is reached could be of primary interest. The end point could be a vital outcome such as death, or disease remission, cure, or cessation of a behavior. In all cases the time of the event is the important random variable. This is the typical scenario where *survival analysis* is employed. If the researcher wants (a) to explore factors affecting the study, and (b) to make sure that confounding is allowed for in the analysis, then the use of spatial information can help with both these tasks. Clayton et al. (1993) have stressed the usefulness of including spatial correlation terms in models to make allowance for confounder and ecological biases, and so there is reasonably strong arguments for always including contextual terms that have spatial structure in survival analysis.

In survival analysis the time to endpoint (T) is the random variable. Assume that a sample of individuals have associated endpoint times. Denote this sample of times as $\{t_i\}$ $i = 1, \ldots, m$. For now assume these are all observed exactly. In addition to an endpoint time a geo-reference is also available. The geo-reference could be an address location, in which case it is denoted as $\mathbf{s}_i$, or a contextual spatial effect, denoted as w_i. The contextual effect is simply a factor that may have spatial correlation so that areas close together have similar risks. In addition to these ingredients, each observation unit can have covariates associated and, for the i-th unit/person these are denoted by the vector $\mathbf{x}_i$. These covariates could be individual or contextual/ecological. Reviews of Bayesian survival methods can be found in Gustafson (1998) and Ibrahaim et al. (2000). For Bayesian proportional hazard modeling see Carlin and Hodges (1999). A review of spatial survival appears in Banerjee (2016).

10.1 Endpoint Distributions

Often a failure time or endpoint distribution is specified for the time to endpoint, and this is often chosen from distributions on the positive real line. Common choices are among the Weibull or extreme value, lognormal or gamma families. Hence for a parametric survival model, at the 1st level of the hierarchy, the data model consists of an endpoint distribution. For flexibility, we will assume a Weibull distribution in what follows. The probability of an endpoint at time t_i under a Weibull distribution is specified by

$$f_T(t_i) \equiv f(t_i) = \rho \mu_i t_i^{\rho-1} \exp(-\mu_i t_i^{\rho}). \tag{20}$$

The survival and hazard functions derived from this specification are

$$S(t_i) = 1 - \int_0^{t_i} f(u)du = \exp(-\mu_i t_i^{\rho}),$$
$$h(t_i) = f(t_i)/S(t_i) = \rho \mu_i t_i^{\rho-1}.$$

The parameterization emphasizes the modeling of a function of the mean of the distribution via μ_i. Note that this allows a straightforward interpretation of the model component for this distribution: covariates and contextual effects can be included within μ_i and the parameter ρ provides the shape of the distribution. Often modeling proceeds via the hazard function, rather than the density, and for the Weibull this decomposes into two components:

$$h(t_i) = h_0(t_i).h_1(t_i) = \rho t_i^{\rho-1} \mu_i.$$

Here $h_0(t_i)$ is regarded as a baseline hazard, while $h_1(t_i)$ is a nonbaseline component which is usually the focus of modeling. Usually it is assumed that a predictor term is linked to the parameter μ_i and each unit will have a different μ_i depending on covariates. A log-linear specification is often assumed. A nonspatial example could be

$$\ln(\mu_i) = \mathbf{x}_i^{\top} \boldsymbol{\beta} + \phi_i, \tag{21}$$

where $\mathbf{x}_i^{\top}$ is a row vector of fixed covariates, $\boldsymbol{\beta}$ the corresponding parameter vector and ϕ_i is a uncorrelated random effect. For the Weibull distribution, Carlin and Hodges (1999) suggested the formulation (21) in a nonspatial setting. This can be extended easily to include a spatial contextual effect. For example, we could specify

$$\begin{aligned} \ln(\mu_i) &= \mathbf{x}_i^{\top} \boldsymbol{\beta} + \phi_i, \\ \phi_i &= \phi_i^{(1)} + \phi_i^{(2)}, \end{aligned} \tag{22}$$

where ϕ_i contains a convolution of unit specific random effect terms similar to the (5). These effects could be individual unit level or they could be contextual. While the Weibull is a flexible distribution there are many alternatives, some of which do not impose the constraints of proportionality of hazard. A wider class of models is the accelerated failure time (AFT) models. These models replace the t within the survival and hazard functions with a modulated function of covariates: $t \exp(\mathbf{x}_i^{\top} \boldsymbol{\beta})$. This leads to a covariate acceleration /deceleration of risk. The Weibull is a special case of this general class. This lead to a linear model in the log of time:

$$\ln(T) = \alpha + \mathbf{x}_i^{\top} \boldsymbol{\beta} + \epsilon,$$

where ϵ is an error term and α is an intercept. This model could also be extended with the addition of random effects which are individual or contextual.

10.2 Censoring

Censoring is always an important issue in survival analysis, as it is often the case that times are not observed exactly. For parametric models this can be treated via a survival function term product in the likelihood. For example, for right censoring we can assume $L = \prod_u f(t_u) \prod_c S(t_c)$, where u denotes uncensored and c denotes right censored. Note that this likelihood simplifies if you assume a censoring indicator γ_i for the i-th observation, which takes 0 if it is censored and 1 if it is uncensored, that is: $L = \prod_{all\ t_i} h(t_i)^{\gamma_i} S(t_i)^{1-\gamma_i}$. Other likelihood forms can similarly be derived for alternative censoring mechanisms.

10.3 Random Effect Specification

As most survival data are observed at the individual unit level, there could be either individual covariates or random effects or contextual effects relating to the individual. For example, the age of an individual could be a personal covariate and the location coordinates of the individual's address could be regarded as personal covariates also. In addition, there could be an individual level random effect which allows for frailty amongst individuals. This could be correlated spatially or uncorrelated. If the residential address of the individual (unit) were known then this could be used directly. An uncorrelated frailty can easily be specified for each individual via $\phi_i \sim \mathrm{N}(0, \tau^2)$.

For a spatially correlated term, ϕ_i could be specified to have a full multivariate normal prior distribution with spatial correlation included within the covariance matrix, where the elements of the covariance are a function of distance between locations. Essentially this is a zero-mean spatial Gaussian process prior distribution. Note that it would also be possible to specify an intrinsic Gaussian prior distribution. As contextual effects, it is relatively straightforward to include spatial effects. For example, with cancer registry data, often individual outcomes have associated county, postal district, or zip code of residence. They often do not have address location for confidentiality reasons. Hence contextual spatial information is often available for these data. This consists of aggregated spatial groupings or factors. Henderson et al. (2002) modeled individual level frailty but resorted to modeling the spatial structure of leukemia survival data using district level (contextual) effects. In that work the covariance function was modeled using the power exponential and Matérn covariance functions. At this level of aggregation, it is natural to consider a CAR specification. Henderson et al. (2002) also compared their contextual models with fully specified covariance and CAR components and found that a Matérn covariance gave the best fitting model based on the DIC goodness-of-fit criterion. This was not found to be the case in other comparisons of Gaussian prior distributions (see, e.g., Best et al., 2005). Banerjee et al. (2003) first reported the use

of a spatially correlated frailty in the application to linked birth-death individual level infant mortality data for the US state of Minnesota. In their formulation a Weibull parametric model was assumed, and the hazard was assumed to be

$$h(t_{ij}|\mathbf{x}) = \rho t_{ij}^{\rho-1} \exp(\mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + \phi_i),$$

for the i-th subject in the j-th stratum.

10.4 General Hazard Model

Banerjee and Carlin (2003) proposed a relaxation of the Weibull model to allow a semiparametric formulation, whereby for subject j in the i-th county the model was:

$$h(t_{ij}|\mathbf{x}_{ij}) = h_{0i}(t_{ij}) \exp(\mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + \phi_i)$$

where h_{0i} denotes the county-specific baseline hazard. For an individual with censoring indicator γ_{ij} (0 if alive, and 1 if dead) the likelihood contribution is then

$$h(t_{ij}; \mathbf{x}_{ij})^{\gamma_{ij}} \exp\{-H_{0i}(t_{ij}) \exp(\mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + \phi_i)\}$$

where $H_{0i}(t_{ij}) = \int_0^{t_{ij}} h_{0i}(u)du$ is a county-specific cumulative baseline hazard, and the covariates are assumed to be not time dependent. The baseline hazard appears in this likelihood and so this must be estimated. Different approaches have been proposed for the estimation of this baseline. One approach assumes a Gamma process which is a function of a parametric cumulative hazard (see, e.g., Ibrahaim et al., 2000). Alternatives can be found in Banerjee and Carlin (2003) and Bastos and Gamerman (2006). Extensions of the above approaches have been proposed in a variety of more complex applications (see, for examples, Cooner et al., 2006, Zhou et al., 2008, Zhang and Lawson, 2011, Lawson et al., 2014, Onicescu et al., 2016).

11 EXAMPLE

We illustrate Bayesian disease mapping via a case study investigating the spatio-temporal trends in measles susceptibility in children in Glasgow, Scotland, between 1998 and 2014. The study arises from the controversy surrounding the now discredited link between the measles, mumps, and Rubella (MMR) vaccination and an increased risk of autism, following the work of Wakefield et al. (1998). The paper attracted media attention and resulted in MMR vaccination rates falling to 87% across Scotland by 2003. The paper linking the MMR vaccination and autism was partially retracted in 2004 and then fully discredited in 2010, as multiple epidemiological studies failed to find any such association (see Elliman and Bedford, 2007 for a review of the evidence). Nonsusceptibility to measles is based upon the receipt of one or two MMR vaccinations that each have a failure rate of 10%. Of interest here

is to estimate what impact the publication, media attention, and subsequent retraction of the Wakefield article have on susceptibility to measles in Glasgow between 1998 and 2014? Also of interest is to examine what the spatial inequality is in measles susceptibility over that time period, and to what extent it has changed over time.

The data come from the Scottish Immunisation and Recall System hosted by the Information and Statistics Division of National Health Service Scotland, and relate to the Greater Glasgow and Clyde health board which has a population of around 1 million people. For the purposes of this study it has been split into $I = 271$ intermediate zones (IZ), and data relate to the estimated numbers of children eligible to attend preschool from nonoverlapping 2-year birth cohorts who attended preschool between 1998 and 2014. Here we have the estimated number of children who are susceptible to measles (y_{ij}) and the total number of children (n_{ij}) for area i and time period j, giving a raw proportion $p_{ij} = y_{ij}/n_{ij}$. The raw proportions $\{p_{ij}\}$ are displayed in Fig. 1, which shows the distribution for each 2-year cohort. The figure shows a clear increase in susceptibility between 1998 and 2004 corresponding to the publication and subsequent media attention surrounding the Wakefield et al. (1998) paper. However, since the partial retraction in 2004 susceptibility to measles has declined, reaching an all time low in 2014.

Fig. 1 shows very different levels of spatial variation each year, suggesting that the simple single spatial surface model given by (9) is likely to be inappropriate. Instead, we fit the model proposed by Napier et al. (2016), which was applied to the same data but for all of Scotland rather than just Glasgow. Thus the results here can be contrasted to see if the Glasgow temporal trend differs from the Scottish average. The model is given by

$$\ln(\theta_{ij}) = \beta_0 + \alpha_{ij} + \lambda_j,$$

which is completed by the binomial likelihood model $Y_{ij} \sim$Binomial(n_{ij}, θ_{ij}). Here, the spatio-temporal pattern in the proportion susceptible is modeled

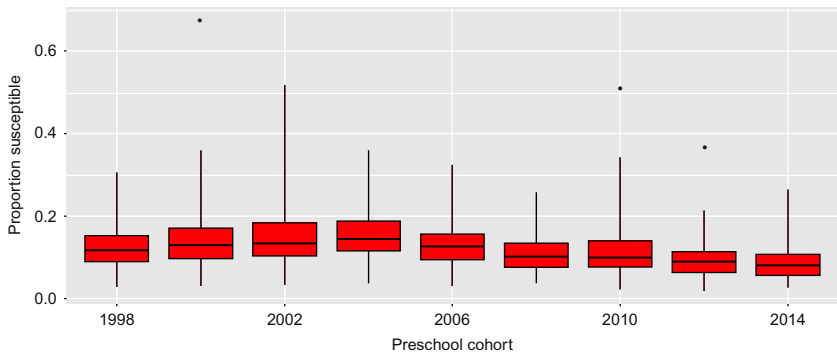

FIG. 1 Boxplots showing the distribution of the proportion of children who are susceptible to measles in each preschool cohort.

FIG. 2 Line graph showing the posterior median and 95% credible intervals for the average region-wide estimated proportion susceptible for each time period.

by an overall region-wide time trend ($\{\lambda_j\}$) and spatial surfaces for each time period ($\{\alpha_{ij}\}$), the latter being modeled by independent CAR priors for each time period with different variance parameters. The model is fitted using the `ST.CARsepspatial()` function in the `R` package `CARBayesST` (Lee et al., 2016). Inference for the model is based on 10,000 MCMC samples, which were obtained following a burn-in period of 20,000 samples and thinning the next 100,000 samples by 10 to reduce their autocorrelation. The region-wide temporal pattern is displayed in Fig. 2, which shows the posterior median and 95% credible interval. The figure shows a clear peak in measles susceptibility in 2004 just before the Wakefield et al. (1998) article was partially retracted and when media attention was at its highest, with an estimated Glasgow-wide proportion of 0.156 (15.6%). This differs from the Scottish average of 0.175 reported by Napier et al. (2016), although the shape of the temporal trend over time is the same. The figure also shows that measles susceptibility changed substantially over the time period, and appears to have reduced once the article was partially retracted in 2004 (0.156) compared to 2014 (0.088).

The temporally varying spatial surfaces of the proportion susceptible is displayed in Fig. 3 for 1998, 2004, 2008, and 2014, from which the overall temporal trend from Fig. 2 is evident. Additionally, Fig. 3 shows dramatically different levels of spatial variation across the four years, with interquartile ranges in the estimated proportions susceptible (over Glasgow) of: 1998, 0.0234; 2004, 0.0375; 2008, 0.0231; and 2014, 0.0218. These results suggest that the spatial inequality in the proportion susceptible was at its highest in 2004, which is because some areas appear to have been affected more by the MMR scare than others. However, by 2014 the scare is long since over, and susceptibility rates are almost uniformly low in all areas. The final point is that a sizeable proportion of the areas that exhibit elevated susceptibility rates in 2004 also show slightly elevated rates in the other years compared to other areas, suggesting that the scare has not changed the areas at greatest susceptibility in Glasgow.

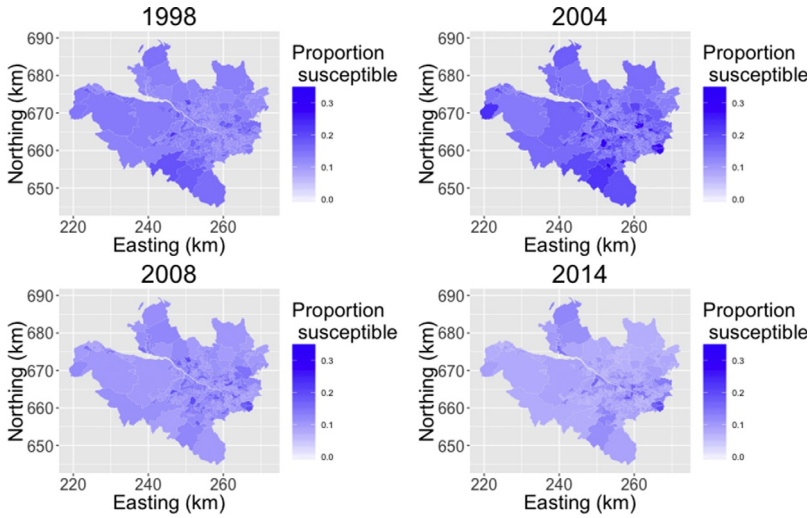

FIG. 3 Maps showing the estimated proportion susceptible in 1998, 2004, 2008, and 2014.

12 DISCUSSION AND FUTURE DIRECTIONS

In this chapter we have provided a summary and basic introduction to a range of topics within the sphere of Bayesian disease mapping for Public Health. Our focus has been on the main topic areas which are currently of importance and have active research activity. Our focus on modeling has precluded discussion of testing-based methods of which there are many, especially in cluster detection and surveillance. We believe, however, that Bayesian modeling provides a significantly more flexible and integrated paradigm applied to public health problems.

ACKNOWLEDGMENTS

The work of the first author is supported by the NIH grant NCI R01CA172805. The work of the second author was funded by the UK Medical Research Council (MRC) grant number MR/L022184/1.

REFERENCES

Al-Hadhrami, A., Lawson, A.B., 2011. Bayesian hierarchical modeling of latent period switching in small-area putative health hazard studies. Stat. Methods Med. Res. 20, 5–28.

Anderson, C., Lee, D., Dean, N., 2014. Identifying clusters in Bayesian disease mapping. Biostatistics 15, 457–469.

Anderson, C., Lee, D., Dean, N., 2016. Spatial clustering of average risks and risk trends in Bayesian disease mapping. Biometrical J. 59, 4156. https://doi.org/10.1002/bimj.201600018.

Azarian, T., Winn, S., Zaheer, S., Buehler, J., Hopkins, R., 2009. Utilization of syndromic surveillance with multiple data sources to enhance public health response. Adv. Dis. Surv. 7, 1–6.

Banerjee, S., 2016. Spatial survival models. In: Lawson, A.B., Banerjee, S., Haining, R., Ugarte, L. (Eds.), Handbook of Spatial Epidemiology. CRC Press, New York.

Banerjee, S., Carlin, B.P., 2003. Semiparametric spatio-temporal frailty modeling. Environmetrics 14, 523–535.

Banerjee, S., Wall, M.M., Carlin, B.P., 2003. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. Biostatistics 4, 123–142.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC Press, London.

Bastos, L., Gamerman, D., 2006. Dynamical survival models with spatial frailty. Lifetime Data Anal. 12, 441–460.

Belitz, C., Brezger, A., Kneib, T., Lang, S., 2012. BayesX Software for Bayesian Inference in Structured Additive Regression Models. Version 2.1. http://www.BayesX.org/.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., Songini, M., 1995. Bayesian analysis of space-time variation in disease risk. Stat. Med. 14, 2433–2443.

Besag, J., Higdon, D., 1999. Bayesian analysis of agricultural field experiments. J. R. Stat. Soc. Ser. B 61, 691–746.

Besag, J., York, J., Mollie, A., 1991. Bayesian image restoration with two applications in spatial statistics. Ann. Inst. Stat. Math. 43, 1–59.

Best, N., Richardson, S., Thomson, A., 2005. A comparison of Bayesian spatial models for disease mapping. Stat. Methods Med. Res. 14, 35–59.

Biggeri, A., Dreassi, E., Catelan, D., Rinaldi, L., Laglazio, C., Cringoli, G., 2006. Disease mapping in veterinary epidemiology: a Bayesian geostatistical approach. Stat. Methods Med. Res. 15, 337–352.

Bivand, R., Piras, G., 2015. Comparing implementations of estimation methods for spatial econometrics. J. Stat. Softw. 63, 18.

Blangiardo, M., Finazzi, F., Cameletti, M., 2016. Two-stage Bayesian model to evaluate the effect of air pollution on chronic respiratory diseases using drug prescriptions. Spat. Spatio-temporal Epidemiol. 18, 1–12.

Breslow, N., Clayton, D., 1993. Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. 88, 9–25.

Brewer, M., Nolan, A., 2007. Variable smoothing in Bayesian intrinsic autoregressions. Environmetrics 18, 841–857.

Brookmeyer, R., 2006. Modeling an outbreak of anthrax. In: Peck, R., Casella, G., Cobb, G., Hoerl, R., Nolan, D., Starbuck, R., Stern, H. (Eds.), Statistics: A Guide to the Unknown. Thomson, New York, pp. 197–210.

Buckeridge, D., Owens, D., Switzer, P., Frank, J., Musen, M., 2006. Evaluating detection of an inhalational anthrax outbreak. Emerg. Infect. Dis. 12, 1942–1949.

Carlin, B.P., Hodges, J., 1999. Hierarchical proportional hazards regression models for highly stratified data. Biometrics 55, 1162–1170.

Charras-Garrido, M., Abrial, D., de Goer, J., 2012. Classification method for disease risk mapping based on discrete hidden Markov random fields. Biostatistics 13, 241–255.

Charras-Garrido, M., Azizi, L., Forbes, F., Doyle, S., Peyrard, N., Abrial, D., 2013. On the difficulty to delimit disease risk hot sports. J. Appl. Earth Obs. Geoinf. 22, 99–105.

Chen, M., Shao, Q., Ibrahim, J., 2000. Monte Carlo Methods in Bayesian Computation. Springer Verlag, New York.

Clayton, D.G., Bernardinelli, L., Montomoli, C., 1993. Spatial correlation in ecological analysis. Int. J. Epidemiol. 22, 1193–1202. https://doi.org/10.1093/ije/22.6.1193.

Cooner, F., Banerjee, S., McBean, M., 2006. Modelling geographically referenced survival data with a cure fraction. Stat. Methods Med. Res. 15 (4), 307–324.

Corberán-Vallet, A., Lawson, A.B., 2011. Conditional predictive inference for online surveillance of spatial disease incidence. Stat. Med. 30, 3095–3116.

Corberán-Vallet, A., Lawson, A.B., 2014. Prospective analysis of infectious disease surveillance data using syndromic information. Stat. Methods Med. Res. 23 (6), 572–590.

Croissant, Y., Millo, G., 2008. Panel data econometrics in R: the plm package. J. Stat. Softw. 27 (2), 1–43. http://www.jstatsoft.org/v27/i02/.

Diggle, P.J., 2005. Spatio-temporal point processes, partial likelihood, foot and mouth disease. Stat. Methods Med. Res. 15, 325–336.

Diggle, P., 2007. Spatio-temporal point processes: methods and applications. In: Finkenstadt, B., Held, L., Isham, V. (Eds.), Statistical Methods for Spatio-Temporal Systems. CRC Press, London, pp. 1–45 (Chapter 1).

Diggle, P., Knorr-Held, L., Rowlingson, B., Su, T., Hawtin, P., Bryant, T., 2004. On-line monitoring of public health surveillance data. In: Brookmeyer, R., Stroup, D.F. (Eds.), Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance. Oxford University Press, Oxford, pp. 233–266 (Chapter 9).

Elliman, D., Bedford, H., 2007. MMR: where are we now? Arch. Dis. Child. 92 (12), 1055–1057.

Forbes, F., Charras-Garrido, M., Azizi, L., Doyle, S., Abrial, D., 2013. Spatial risk mapping for rare disease with hidden Markov fields and variational EM. Ann. Appl. Stat. 7, 1192–1216.

Fricker Jr., R., 2011. Some methodological issues in biosurveillance. Stat. Med. 30, 403–415.

Frisen, M., 1992. Evaluations of methods for statistical surveillance. Stat. Med. 11, 1489–1502.

Frisen, M., Mare, J.D., 1991. Optimal surveillance. Biometrika 78, 271–280.

Frisén, M., Sonesson, C., 2005. Optimal surveillance. In: Lawson, A.B., Kleinman, K. (Eds.), Spatial and Syndromic Surveillance for Public Health. Wiley, New York (Chapter 3).

Gangnon, R., Clayton, M., 2000. Bayesian detection and modeling of spatial disease clustering. Biometrics 56, 922–935.

Gelfand, A., Ghosh, S., 1998. Model choice: A minimum posterior predictive loss approach. Biometrika 85, 1–11.

Gelfand, A., Vounatsou, P., 2003. Proper multivariate conditional autoregressive models for spatial data. Biostatistics 4, 11–25.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D., 2004. Bayesian Data Analysis. Chapman and HAll/CRC Press, London.

Gelman, A., Huang, J., Vehtari, A., 2014. Understanding predictive information criteria in Bayesian models. Stat. Comput. 24, 997–1016.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 4. Oxford University Press, New York, pp. 169–193.

Green, P., Richardson, S., 2002. Hidden Markov models and disease mapping. J. Am. Stat. Assoc. 97, 1055–1070.

Greven, S., Dominici, F., Zeger, S., 2011. An approach to the estimation of chronic air pollution effects using spatio-temporal information. J. Am. Stat. Assoc. 106, 396–406.

Grigg, O., spiegelhalter, D., 2007. A simple risk-adjusted exponentially weighted moving average. J. Am. Stat. Assoc. 102, 140–152.

Gustafson, P., 1998. Flexible Bayesian modelling for survival data. Lifetime Data Anal. 4, 281–299.

Held, L., Natario, I., Fenton, S., Rue, H., Becker, N., 2005. Towards joint disease mapping. Stat. Methods Med. Res. 14, 61–82.

Held, L., Hofmann, L., Hohle, M., Schmid, V., 2006. A two-component model for counts of infectious diseases. Biostatistics 7, 422–437.

Henderson, R., Shimakura, S., Gorst, D., 2002. Modeling spatial variation in leukaemia survival data. J. Am. Stat. Assoc. 97, 965–972.

Hughes, J., Haran, M., 2013. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. J. R. Stat. Soc. Ser. B 75, 139–160.

Ibrahaim, J., Chen, M., Sinha, D., 2000. Bayesian Survival Analysis. Springer, New York.

Jin, X., Carlin, B., Banerjee, S., 2005. Generalized hierarchical multivariate CAR models for areal data. Biometrics 61, 950–961.

Katz, R., May, L., Baker, J., Test, E., 2011. Redefining syndromic surveillance. J. Epidemiol. Glob. Health 1, 21–31.

Kim, H., Sun, D., Tsutakawa, R., 2001. A bivariate Bayes method for improving the estimates of mortality rates with a twofold autoregressive model. J. Am. Stat. Assoc. 96, 1506–1521.

Kissling, W., Carl, G., 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. Glob. Ecol. Biogeogr. 17, 59–71.

Knorr-Held, L., 2000. Bayesian modelling of inseparable space-time variation in disease risk. Stat. Med. 19, 2555–2567.

Knorr-Held, L., Raßer, G., 2000. Bayesian detection of clusters and discontinuities in disease maps. Biometrics 56, 13–21.

Kulldorff, M., Heffernan, R., Hartman, J., Assuncao, R., Mostashari, F., 2005. A space-time permutation scan statistic for disease outbreak detection. PLoS Med. 2, 216–224.

Lawson, A.B., 2004. Some issues in the spatio-temporal analysis of public health surveillance data. In: Brookmeyer, R., Stroup, D. (Eds.), Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance. Oxford University Press, Oxford, pp. 289–314 (Chapter 11).

Lawson, A.B., 2005. Spatial and spatio-temporal disease analysis. In: Lawson, A.B., Kleinman, K. (Eds.), Spatial and Syndromic Surveillance for Public Health. Wiley, New York (Chapter 4).

Lawson, A.B., 2013. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology, second ed. CRC Press, New York.

Lawson, A., Clark, A., 2002. Spatial mixture relative risk models applied to disease mapping. Stat. Med. 21, 359–370.

Lawson, A.B., Kleinman, K. (Eds.) 2005. Spatial and Syndromic Surveillance for Public Health. Wiley, New York.

Lawson, A.B., Leimich, P., 2000. Approaches to space-time modelling of infectious disease behaviour. IMA J. Math. Appl. Med. Biol. 17, 1–13.

Lawson, A.B., Clark, A.B., Vidal-Rodeiro, C.L., 2004. Developments in general and syndromic surveillance for small area health data. J. Appl. Stat. 31, 951–966.

Lawson, A., Choi, J., Cai, B., Hossain, M., Kirby, R., Liu, J., 2012. Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data. J. Agric. Biol. Environ. Stat. 17, 417–441.

Lawson, A.B., Choi, J., Zhang, J., 2014. Prior choice in discrete latent modeling of spatially referenced cancer survival. Stat. Methods Med. Res. 23, 183–200.

Lawson, A.B., Banerjee, S., Haining, R., Ugarte, L. (Eds.) 2016. Handbook of Spatial Epidemiology. CRC Press, New York.

Leckie, G., Goldstein, H., 2014. A multilevel modelling approach to measuring changing patterns of ethnic composition and segregation among London secondary schools, 2001–2010. J. R. Stat. Soc. Ser. A 178, 405–424.

Lee, D., 2013. CARBayes: an R package for Bayesian spatial modeling with conditional autoregressive priors. J. Stat. Softw. 55, 13.

Lee, D., Lawson, A., 2016. Quantifying the spatial inequality and temporal trends in maternal smoking rates in Glasgow. Ann. Appl. Stat. 10, 1427–1446.

Lee, D., Mitchell, R., 2012. Boundary detection in disease mapping studies. Biostatistics 13, 415–426.

Lee, D., Mitchell, R., 2013. Locally adaptive spatial smoothing using conditional autoregressive models. J. R. Stat. Soc. Ser. C 62, 593–608.

Lee, D., Sarran, C., 2015. Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. Environmetrics 26, 477–487.

Lee, D., Rushworth, A., Sahu, S., 2014. A Bayesian localized conditional autoregressive model for estimating the health effects of air pollution. Biometrics 70, 419–429.

Lee, D., Rushworth, A., Napier, G., 2016. CARBayesST: Spatio-Temporal Generalised Linear Mixed Models for Areal Unit Data. R package version 2.4. https://cran.r-project.org/web/packages/CARBayesST/CARBayesST.pdf.

Lee, D., Mukhopadhyay, S., Rushworth, A., Sahu, S., 2017. A rigorous statistical framework for spatio-temporal pollution prediction and estimation of its long-term impact on health. Biostatistics 18, 370–385. https://doi.org/10.1093/biostatistics/kxw048.

Leroux, B., Lei, X., Breslow, N., 2000. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Halloran, M., Berry, D. (Eds.), Statistical Models in Epidemiology, the Environment and Clinical Trials. Springer-Verlag, New York, pp. 135–178.

Li, P., Banerjee, S., McBean, A., 2011. Mining boundary effects in areally referenced spatial data using the Bayesian information criterion. Geoinformatica 15, 435–454.

Lu, H., Carlin, B., 2005. Bayesian areal wombling for geographical boundary analysis. Geogr. Anal. 37, 265–285.

Lu, H., Reilly, C., Banerjee, S., Carlin, B., 2007. Bayesian areal wombling via adjacency modelling. Environ. Ecol. Stat. 14, 433–452.

Lunn, D., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Stat. Comput. 10, 325–337.

Ma, H., Carlin, B., 2007. Bayesian multivariate areal wombling for multiple disease boundary analysis. Bayesian Anal. 2, 281–302.

Ma, H., Carlin, B., Banerjee, S., 2010. Hierarchical and joint site-edge methods for medicare hospice service region boundary analysis. Biometrics 66, 355–364.

Mackenbach, J., Kunst, A., Cavelaars, A., Groenhof, F., Geurts, J., 1997. Socioeconomic inequalities in morbidity and mortality in western Europe. Lancet 349, 1655–1659.

MacNab, Y., Dean, C., 2001. Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. Biometrics 57, 949–956.

Martinez-Beneito, M., Conessa, D., Lopez-Quilez, A., Lopez-Maside, A., 2008. Bayesian Markov switching models for the early detection of influenza epidemics. Stat. Med. 27, 4455–4468.

Michael Höhle, S.M., Paul, M., 2015. Surveillance: Temporal and Spatio-Temporal Modeling and Monitoring of Epidemic Phenomena. R package version 1.9.1. http://CRAN.R-project.org/package=surveillance.

Millo, G., Piras, G., 2012. splm: Spatial panel data models in R. J. Stat. Softw. 47 (1), 1–38. http://www.jstatsoft.org/v47/i01/.

Molina, R., Katsaggelos, A., Mateos, J., 1999. Bayesian regularization methods for hyperparameter estimation in image restoration. IEEE Trans. Image Process. 8, 231–246.

Mugglin, A., Cressie, N., Gemmell, I., 2002. Hierarchical statistical modelling of influenza epidemic dynamics in space and time. Stat. Med. 21, 2703–2721.

Napier, G., Lee, D., Robertson, C., Lawson, A., Pollock, K., 2016. A model to estimate the impact of changes in MMR vaccine uptake on inequalities in measles susceptibility in Scotland. Stat. Methods Med. Res. 25, 1185–1200.

Onicescu, G., Lawson, A.B., Zhang, J., Gebregziabher, M., Wallace, K., Eberth, J., 2016. Bayesian accelerated failure time model for space-time dependency in a geographically augmented survival model. Stat. Methods Med. Res. https://doi.org/10.1177/0962280215596186.

Paciorek, C., 2010. The importance of scale for spatial confounding bias and precision of spatial regression estimators. Stat. Sci. 25, 107–125.

Paul, M., Held, L., 2011. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. Stat. Med. 30, 1118–1136.

Paul, M., Held, L., Toschke, A., 2008. Multivariate modelling of infectious disease surveillance data. Stat. Med. 27, 6250–6267.

Reich, B., Hodges, J., Zadnik, V., 2006. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics 62, 1197–1206.

Richardson, S., Thomson, A., Best, N., Elliott, P., 2004. Interpreting posterior relative risk estimates in disease mappling studies. Environ. Health Perspect. 112, 1016–1025.

Riebler, A., Sørbye, S., Simpson, D., Rue, H., 2016. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. Stat. Methods Med. Res. 25, 1145–1165.

Robert, C., Casella, G., 2010. Introducing Monte Carlo Methods With R, first ed. Springer, New York.

Ronnegard, L., Shen, X., Alam, M., 2010. hglm: A package for fitting hierarchical generalized linear models. The R J. 2, 20–28.

Rotejanaprasert, C., Lawson, A.B., Bolick, S., Hurley, D., 2016. Spatial Bayesian surveillance for small area case event data. Stat. Methods Med. Res. 25, 1101–1117.

Rotejanaprsasert, C., Lawson, A.B., 2017. Bayesian prospective detection of small area health anomalies using Kullback-Leibler divergence. Stat. Methods Med. Res. https://doi.org/10.1177/0962280216652156.

Rowlingson, B., Diggle, P., Su, T.-L., 2005. Point process methodology for on-line spatio-temporal disease surveillance. Environmetrics 16, 423–434.

Royle, A., Berliner, M., 1999. A hierarchical approach to multivariate spatial modeling and prediction. J. Agric. Biol. Environ. Stat. 4, 29–56.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). J. R. Stat. Soc. Ser. B 71, 319–392.

Rushworth, A., Lee, D., Mitchell, R., 2014. A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. Spat. Spatio-temporal Epidemiol. 10, 29–38.

Rushworth, A., Lee, D., Sarran, C., 2017. An adaptive spatio-temporal smoothing model for estimating trends and step changes in disease risk. J. R. Stat. Soc. Ser. C 66, 141–157. https://doi.org/10.1111/rssc.12155.

Schabenberger, H., 2009. Spatcounts: Spatial Count Regression. R package version 1.1. http://CRAN.R-project.org/package=spatcounts.

Shmeuli, G., Burkom, H., 2010. Statistical challenges facing early outbreak detection in biosurveillance. Technometrics 52, 39–51.

Sosin, D., 2003. Draft framework for evaluating syndromic surveillance systems. J. Urban Health 80 (Suppl. 1), i8–i13.

Spiegelhalter, D., Best, N., Carlin, B., Van der, Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B 64, 583–639.

Stern, H., Cressie, N., 1999. Inference for extremes in disease mapping. In: Lawson, A., Biggeri, D., Boehning, E., Lesaffre, E., Viel, J., Bertollini, R. (Eds.), Disease Mapping and Risk Assessment for Public Health. Wiley, Hoboken, NJ.

Strat, Y.L., Carrat, F., 1999. Monitoring epidemiologic surveillance data using hidden Markov models. Stat. Med. 18, 3463–3478.

Ugarte, D., Goicoa, T., Militino, A., 2010. Spatio-temporal modeling of mortality risks using penalized splines. Environmetrics 21, 270–289.

Ugarte, D., Etxeberria, J., Goicoa, T., Ardanaz, E., 2012. Gender-specific spatio-temporal patterns of colorectal cancer incidence in Navarre, Spain (1990–2005). Cancer Epidemiol. 36, 254–262.

Unkel, S., Farrington, C.P., Garthwaite, P., Robertson, C., Andrews, N., 2012. Statistical methods for the prospective detection of infectious disease outbreaks: a review. J. R. Stat. Soc. 175, 49–82.

Vidal-Rodeiro, C., Lawson, A.B., 2006a. Monitoring changes in spatio-temporal maps of disease. Biometrical J. 48, 463–480.

Vidal-Rodeiro, C., Lawson, A.B., 2006b. Online updating of space-time disease surveillance models via particle filters. Stat. Methods Med. Res. 15, 423–444.

Vieilledent, G., Merow, C., Guélat, J., Latimer, A.M., Kéry, M., Gelfand, A.E., Wilson, A.M., Mortier, F., Silander Jr., J.A., 2014. hSDM: Hierarchical Bayesian Species Distribution Models. R package version 1.4. https://CRAN.R-project.org/package=hSDM.

Wakefield, J., Kim, A., 2013. A Bayesian model for cluster detection. Biostatistics 14, 752–765.

Wakefield, J., Salway, R., 2001. A statistical framework for ecological and aggregate studies. J. R. Stat. Soc. Ser. A 164, 119–137.

Wakefield, A., Murch, S., Anthony, A., Linnell, J., Casson, D., Malik, M., Berelowitz, M., Dhillon, A., Thomson, M., Harvey, P., Valentine, A., Davies, S., Walker-Smith, J., 1998. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet 351 (9103), 637–641.

Wall, M., 2004. A close look at the spatial structure implied by the CAR and SAR models. J. Stat. Plan. Inference 121, 311–324.

Wheeler, B., Lovell, R., Higgins, S., White, M., Alcock, A., Osborne, N., Husk, K., Sabel, C., Depledge, M., 2015. Beyond greenspace: an ecological study of population general health and indicators of natural environment type and quality. Int. J. Health Geogr. 14, 17.

Womble, W., 1951. Differential systematics. Science 114, 315–322.

Zhang, J., Lawson, A., 2011. Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. J. Appl. Stat. 38, 591–603.

Zhou, H., Lawson, A.B., Hebert, J., Slate, E., Hill, E., 2008. Joint spatial survival modelling for the date of diagnosis and the vital outcome for prostate cancer. Stat. Med. 27 (18), 3612–3628.