Spatio-temporal Bayesian model selection for disease mapping

Peer-reviewed author version

Carroll, Rachel; LAWSON, Andrew; FAES, Christel; Kirby, Russell S.; AREGAY, Mehreteab & WATJOU, Kevin (2016) Spatio-temporal Bayesian model selection for disease mapping. In: ENVIRONMETRICS, 27(8), p. 466-478.

DOI: 10.1002/env.2410
Handle: http://hdl.handle.net/1942/23816

Reference (Published version):
Carroll, Rachel; Lawson, Andrew B.; Faes, Christel; Kirby, Russell S.; Aregay, Mehreteab & Watjou, Kevin(2016) Spatio-temporal Bayesian model selection for disease mapping. In: ENVIRONMETRICS, 27(8), p. 466-478

DOI: 10.1002/env.2410
Handle: http://hdl.handle.net/1942/23816

Spatio-temporal Bayesian model selection for disease mapping

R Carroll [A]*, AB Lawson [A], C Faes [B], RS Kirby [C], M Aregay [A], K Watjou [B]

[A] Department of Public Health, Medical University of South Carolina,

[B] Interuniversity Institute for Statistics and Statistical Bioinformatics, Hasselt University,

[C] Department of Community and Family Health, University of South Florida,

*Corresponding author, Dr. R Carroll

Department of Public Health, Medical University of South Carolina,

135 Cannon St

Charleston, SC 29425

USA

e-mail: mosra@musc.edu

Article Category: Research Article

Short title: Spatio-temporal model selection

Keywords: MCMC; BRugs; Poisson; model selection; melanoma;

Abstract

Spatio-temporal analysis of small area health data often involves choosing a fixed set of predictors prior to the final model fit. In this paper, we propose a spatio-temporal approach of Bayesian model selection to implement model selection for certain areas of the study region as well as certain years in the study time line. Here, we examine the usefulness of this approach by way of a large-scale simulation study accompanied by a case study. Our results suggest that a special case of the model selection methods, a mixture model allowing a weight parameter to indicate if the appropriate linear predictor is spatial, spatio-temporal, or a mixture of the two, offers the best option to fitting these spatio-temporal models. In addition, the case study illustrates the effectiveness of this mixture model within the model selection setting by easily accommodating lifestyle, socio-economic, and physical environmental variables to select a predominantly spatio-temporal linear predictor.

1. Introduction

In environmental health studies it is often important to assess the link between environmental as well as socio-economic predictors and health outcomes. This is usually done by means of linear or generalized linear models where linear predictors are fitted to the health outcome to assess the degree of explanation achieved. These predictors are often chosen based on prior understanding of their role or via some variable selection method. An alternative approach involves the use of model selection as it has been established that model selection methods are useful in determining the most appropriate linear predictor for a set of data. Model selection techniques are often used in lieu of variable selection as they alleviate some of the issues related to variable selection (e.g. co-linearity, an excess of parameters) (Bondell *et al.*, 2010; Garcia *et al.*, 2010; George and Clyde., 2004; Rockova *et al.*, 2014; Scheel *et al.*, 2013; Hoeting *et al.*, 2002). Previous studies have shown that model selection methods can be helpful for spatial data (Carroll *et al.*, 2015.a , Carroll *et al.*, 2016). However, researchers often have access to data across time as well as space. These model selection methods allow for different models to be employed at each temporal unit, spatial unit, or both, depending on how the model is specified (Li *et al.*, 2012; Waldorp *et al.*, 2005).

One challenge indicated by previous studies involving spatio-temporal variable selection includes having more parameters than the possible number of Markov chain Monte Carlo (MCMC) iterations (Lee *et al.*, 2014). Our model selection methods minimize this issue as it is not necessary to apply a selection parameter to each individual predictor to determining its model probability. Rather, the model probability is applied to the linear predictors in order to determine their appropriateness as a whole.

The aim of this paper revolves around fitting our four alternative model selection methods to eight different simulated model scenarios and assessing them for goodness of fit as well as recovery of the simulation ground truth. Additionally, a melanoma cancer of the skin case study example is examined with these methods using melanoma cancer of the skin with a predictor set including race, sunlight exposure, and socio-economic variables. Prior to the model selection execution, we explored fitting the simulated data with the Knorr-Held (Knorr-Held, 2000) model to deduce how well we are able to recover the simulated random effects. Our objective is to determine which model selection method performs best. It is worth noting that these methods are generalizable in the spatio-temporal

disease mapping setting in that the alternative linear predictor space furnished by these four different methods achieve combinations that are comparable and typical for this type of analysis.

This paper is developed as follows. We first describe the proposed model selection methods. Next, we define the simulated data scenarios and display the simulation results. Following that, we provide an application through a case study example. Finally, we discuss and draw conclusions relating to the results.

2. Model Selection Methods

This paper focuses on the context of disease mapping in *m* predefined small areas. For all models defined below, both simulated and fitted, we conventionally assume that the outcome, an aggregated count of disease (y_{ij}), is observed in the i^{th} small area at time j and that these counts are conditionally independent Poisson distributed outcomes. In addition, we assume that the expected rate (e_{ij}) is available in each small area at each time. This is a commonly assumed model for small area counts in disease mapping (Lawson, 2013), and is defined as follows:

$$y_{ij} \mid \mu_{ij} \sim Pois(\mu_{ij})$$
$$\mu_{ij} = e_{ij}\theta_{ij}$$
$$\log\left(\theta_{ij}\right) = X_{ij}^{T}\boldsymbol{\alpha} + R_{ij}$$

where the mean of the Poisson model, μ_{ij} , is defined as the expected rate of disease, e_{ij} , times the risk, θ_{ij} . Further, X_{ij}^{T} is the ij^{th} element of the design matrix, $\boldsymbol{\alpha}$ is the vector of parameter estimates, and R_{ij} is a linear combination of different types of random effects which are described in more detail in the following sections and Table 1 which displays the simulated data scenarios to be discussed in Section 3. Further, this is a basic definition of the model of interest and will be explained in more detail as well as with more options in the following sections. In what follows we describe the initial model selection formulation, our extensions to the methodology, and the supplied alternative linear predictors.

2.1. Initial Model Selection Structure

The initial model selection formula that we consider is a mixture of model linear predictors and is as follows:

$$\log\left(\theta_{ij}\right) = \alpha_0 + \sum_{l=1}^{L} w_l \omega_{lij} \qquad (1)$$

where α_0 is an intercept and w_l provides the weight for the l^{th} of L linear predictor alternatives, notated as ω_{lij}. This weight is in the form of a binary indicator controlling for the inclusion of linear predictors, and can be constant or varying spatially, temporally, or spatio-temporally. We assume that weights w_l follow an independent Bernoulli prior distribution with parameter p_*, i.e. $w_* \sim Bern(p_*)$, where p_* is the model probability. The Bernoulli model probabilities associated with these weights can be defined in a variety of ways and this has been explored for spatial applications previously (Carroll *et al.*, 2016). The weights and thus the model probabilities are considered to be latent variables as they are unmeasured effects that relates the measured effects to the outcome of interest. A higher model probability indicates that the associated alternative linear predictor is more likely to be the true model, and thus, if all models were equally probable, they would all attain the value $p_* = 1/L$. A suitable prior distribution for the model probability can be defined. Initially, we assume a normalized logit score such that $p_* = z_* / \sum_{l=1}^{L} z_l$ and $\text{logit}(z_*) \sim N(0, \tau_*^{-1})$ with a uniform prior for the standard deviation: $\tau_*^{-1/2} \sim Unif(0, C)$. We explore this prior hierarchy choice in preference to independent Beta or Dirichlet distributions as this allows us to extend the specification into dependences that are spatial, temporal, or both. We would also want to allow the set of L linear predictors to have the ability to vary across space and time presenting the option for inclusion of spatial, temporal, or spatio-temporal covariates and associated parameter estimates as well as spatial, temporal, or spatio-temporal random effects.

To allow for spatial dependency in the inclusion model, the dependency of the weights is extended to w_{*i} such that $p_{*i} = z_{*i} / \sum_{l=1}^{L} z_{li}$, $\lambda_{*i} = \text{logit}(z_{*i})$, $\lambda_{*i} \sim N(\bar{\lambda}_{\delta_i}, \tau_\lambda^{-1} / n_{\delta_i})$ where $\bar{\lambda}_{\delta_i} = \sum_{d \in \delta_i} \lambda_d / n_{\delta_i}$ and δ_i is the set of adjacent neighbors of the i^{th} area. This is an intrinsic conditional autoregressive (ICAR) model (Besag *et al.*, 1991), and we denote this prior distribution subsequently as $CAR(\tau_*^{-1})$. For this extension, model (1) becomes $\log(\theta_{ij}) = \alpha_0 + \sum_{l=1}^{L} w_{li} \omega_{lij}$. To additionally consider temporal variation for the weights by allowing them to vary randomly over time, a time labelled CAR prior distribution, i.e. $CAR(\tau_{*j}^{-1})$, is assumed.

The following subsections and sections expand on the specific details involved in this study. The four explored extensions to the initial model section methods (F1 up to F4) are defined in Section 2.2. The first three are most akin to this initial method while the fourth is an even further extension of it by way of a mixture model. Next, the specifics of the alternative linear predictors as well as assumptions concerning prior distributions for regression parameters and precisions for the fitted models are given in Section 2.3. Also, note that Table 2 displays fitted models F1 up to F4 as well as their associated alternative linear predictors.

2.2. Model Selection Extensions

F1 model: The F1 model is a fairly simple form of model selection, but because of the spatio-temporal structure of the weights, it actually becomes a very computationally intense model with a reasonable amount of flexibility. It also includes a random effect interaction term and is described as follows:

$$\log\left(\theta_{ij}\right) = \alpha_0 + \sum_{l=1}^{L} w_{lij}\omega_{lij} + \psi_{ij}$$

where α_0 is a fixed intercept term and ψ_{ij} is a spatio-temporal interaction that functions as a random noise term. The hierarchical prior distributions are such that $\alpha_0 \sim N\left(0, \tau_\alpha^{-1}\right)$, $\tau_\alpha^{-1/2} \sim Unif\left(0, \mathrm{C}\right)$, $\psi_{ij} \sim N\left(0, \tau_\psi^{-1}\right)$, and $\tau_\psi^{-1/2} \sim Unif\left(0, \mathrm{C}\right)$ where C is chosen to be large so to ensure non-informativeness. These priors will continue to apply for the rest of the fitted models when appropriate. For the models, $i = 1, \ldots, I$ represents the spatial unit and $j = 1, \ldots, J$ represents the temporal unit. Additionally, the indicators, w_{*ij}, vary over space and time, thus different linear predictors could be chosen for each of the counties as well as years. Explicitly, the weights are defined such that $\lambda_{*ij} \sim CAR(\tau_{*j}^{-1})$ and the remainder of the parameters follow what was described in the previous section. Hence, a higher value of p_{*ij} indicates the preference of the associated alternative linear predictor for county i and time j.

F2 model: F2 offers separate linear predictors for the spatial and spatio-temporal components of the model in combination with the spatio-temporal interaction term. Separating the two components offers more flexibility than the model presented in F1 in some respects given that fewer linear predictors can be supplied to each of the components that together offer $L*K$ combinations of alternative linear predictors where L and K indicate the number of linear predictors associated with the spatial and spatio-temporal model sets, ω_{li}^{S} and ω_{kij}^{ST} respectively. This model is described with the following notation:

$$\log\left(\theta_{ij}\right) = \alpha_{0j} + \sum_{l=1}^{L} w_l^S \omega_{li}^S + \sum_{k=1}^{K} w_k^{ST} \omega_{kij}^{ST} + \psi_{ij}$$

As in the initial model selection formulation and F1, the best model is selected based on the model probabilities associated with indicators w_l^S and w_k^{ST}. Since these model probabilities do not vary over space or time in this extension, they each have the same distribution and it is described as follows, with w_* being an example:

$w_* \sim Bern(p_*)$, $p_* = z_* / \sum_{l=1}^{L} z_l$, $\text{logit}(z_*) \sim N\left(0, \tau_*^{-1}\right)$, and $\tau_*^{-1/2} \sim Unif(0, C)$; this will be the model probability distribution associated with the remainder of the fitted models. The additional parameters are defined in the same way as described for model F1 and the initial model selection formulation.

F3 model: F3 is an extension of F1, but the bulk of the extension involves the specification of the alternative linear predictors, and this will be discussed in Section 3.2. Hence, comparable, albeit slightly more complex, alternative linear predictors are supplied here, and because of this increase in complexity among the alternative linear predictors, we have limited the indicators, w_*, such that they no longer have spatio-temporal variation. Thus, the model is now described as follows:

$$\log\left(\theta_{ij}\right) = \alpha_0 + \sum_{l=1}^{L} w_l \omega_{lij} + \psi_{ij}$$

Again, the parameters within this model are as defined previously.

F4 model: For F4 we consider a situation that applies a type of mixture model, which allows the data to indicate the appropriate relationship. This formulation is similar in structure to the BaySTDetect model proposed by Li *et al.* (2012). The BaySTDetect model has the ability to detect departure patterns in small area health data, and here, we apply it such that these detections can suggest the appropriate model for the data: spatial, spatio-temporal, or a mixture of the two. Further, F4 is considered to be a special case of the model selection methods as selection can still occur if the results indicate that one of the mixture components is dominating the model. Additionally, this model avoids the need for the specification of Bernoulli prior weights. It is described such that:

$$\log\left(\theta_{ij}\right) = \alpha_0 + p\omega_i^S + (1-p)\omega_{ij}^{ST} + \psi_{ij}$$

where ω_i^S and ω_{ij}^{ST} are the mixture components, linear predictors supplied to this mixture model. For this, the mixture components are weighted similarly to F1-F3 via a model probability p which is defined such that $\text{logit}(p) \sim Norm\left(0, \tau_p^{-1}\right)$ and $\tau_p^{-2} \sim Unif(0, C)$. Thus, p is large if a spatial model is the preferred fit and small if a spatio-temporal model is favored. The contents of ω_i^S and ω_{ij}^{ST} will be explained in Section 3.2 along with the alternative linear predictors for model F1 through F3, but the linear predictors associated with F4 differ in that there is only a single linear predictor supplied to each component as no sum across the L or K alternative linear predictors is included in the above formula.

2.3. Alternative Linear Predictors

The alternative linear predictors described here are associated with the four different model selection methodologies described in Section 2.2. The alternative linear predictors differ per fitted model selection extension and are chosen such that they represent common spatial and spatio-temporal models as well as offer some situations of misspecification in addition to appropriate specification in terms of the simulated data scenarios which are described in the subsequent section and Table 1. Further, some of these alternative linear predictors are similar to each other to assess the different fitted models' abilities to decipher between the two. The hierarchical prior distributions used for the parameters included in this section are as follows: $\alpha_h \sim N\left(0, \tau_{\alpha_h}^{-1}\right)$, $\tau_{\alpha_h}^{-1/2} \sim Unif(0, \text{C})$, $\alpha_{gj} \sim N\left(0, \tau_g^{-1}\right)$, $\tau_{\alpha_g}^{-1/2} \sim Unif(0, \text{C})$, $\alpha_{fi} \sim N\left(0, \tau_f^{-1}\right)$, $\tau_{\alpha_f}^{-1/2} \sim Unif(0, \text{C})$, $u_i \sim N\left(0, \tau_u^{-1}\right)$, $\tau_u^{-1/2} \sim Unif(0, \text{C})$, $v_i \sim N(\bar{u}_{\delta_i}, \tau_v^{-1} / n_{\delta_i})$, $\tau_v^{-1/2} \sim Unif(0, C)$, $\gamma_j \sim N\left(\gamma_{j-1}, \tau_\gamma^{-1}\right)$, and $\tau_\gamma^{-1/2} \sim Unif(0, \text{C})$. These parameters will be explained in more detail in what follows. Additionally, the equations given in this section include notes indicating which simulated data scenarios correspond to the alternative linear predictors. The simulated data scenario names in parentheses indicate that the alternative linear predictor reflects that specific scenario meaning that the fixed and random effects are specified in the exact same way. Further, if the alternative linear predictor leads to a slight overparameterization of the simulated data scenario, it is indicated by italicizing the scenario name. For F2 and F4, the spatial and spatio-temporal effects are considered separately. Table 3 offers another view of which alternative linear predictors reflect which simulated data scenario. Simulated data scenarios are distinguished from fitted models with an 'S' preceding the model number (as seen in Table 1) rather than an 'F.'

For F1, we take the stance that there are only spatially varying predictors with outcomes that are measured over space and time. The set of alternative linear predictors is as follows:

$$\omega_{ij} = \begin{cases} \alpha_1 x_{1i} + \alpha_2 x_{2i} & = \omega_{1ij} \quad \text{(S3)} \\ \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} & = \omega_{2ij} \\ \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} + \gamma_j & = \omega_{3ij} \\ \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} + u_i + v_i + \gamma_j & = \omega_{4ij} \quad \text{(S8)} \\ \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i + \gamma_j & = \omega_{5ij} \quad \text{(S1,S2)} \end{cases}$$

These linear predictor alternatives are specified such that they are true for a selection of the simulated data scenarios defined in Table 1. Although spatio-temporal covariates are excluded from these alternative linear predictors, the j subscript on the fixed parameter estimates indicates that the spatially varying covariate's relationship with the outcome has the ability to change over time.

The sets of linear predictors for F2 are as follows:

$$\omega_i^S = \begin{cases} \alpha_1 x_{1i} + \alpha_2 x_{2i} & = \omega_{1i}^S \quad \text{(S3,S6)} \\ \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i & = \omega_{2i}^S \quad \text{(S1)} \\ \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i & = \omega_{3i}^S \quad \text{(S2,S7)} \end{cases} \qquad \omega_{ij}^{ST} = \begin{cases} \alpha_1 x_{3ij} + \alpha_2 x_{4ij} & = \omega_{1i}^{ST} \quad \text{(S4)} \\ \alpha_{1j} x_{3ij} + \alpha_{2j} x_{4ij} & = \omega_{2i}^{ST} \quad \text{(S5,S6,S7)} \\ \alpha_{1j} x_{3ij} + \alpha_{2j} x_{4ij} + \gamma_j & = \omega_{3i}^{ST} \quad \text{(S1,S2,S3)} \end{cases}$$

These predictors nearly cover the same alternative linear predictor space as those in F1, and they also cover a considerable amount of the simulated data scenarios, particularly the spatial ones.

As mentioned in Section 2.1., F3 is an extension of F1. This extension is such that, for example, α_* in F1 is replaced by $\alpha_* + \alpha_{*i} + \alpha_{*j}$ in F3. In this expression, there are fixed, spatial, and temporal components to the fixed effect parameter estimates within the alternative linear predictors via α_1, α_{li}, and α_{lj} respectively. This allows the parameter estimates to vary over space and time rather than being only fixed or temporally varying as they were in F1. Explicitly, the linear predictors are defined such that:

$$\omega_{ij} = \begin{cases} x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) & = \omega_{1ij} \quad \text{(S3)} \\ x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) + \gamma_j & = \omega_{2ij} \\ x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) + u_i + v_i + \gamma_j & = \omega_{3ij} \quad \text{(S1,S2,S8)} \end{cases}$$

This augmented alternative linear predictor specification ultimately leads to overparameterization in all cases because of the fixed parameter estimates but also covers the same linear predictor space with a smaller number of alternative linear predictors.

Since F4 is a mixture model, there is simply a single alternative linear predictor supplied to the spatial and spatio-temporal components each, as mentioned in Section 2.2. The two linear predictors are defined as follows:

$$\omega_i^S = \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i \qquad \omega_{ij}^{ST} = \alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij} + \gamma_j$$

This model is complex enough to cover much of the same predictor space as the other fitted models. While this complexity could lead to overparameterization for many cases, we believe that the nature of the mixture model has the ability to compensate for that.

3. Simulation Study

The specific methodologies described above are investigated via a simulation study. This section first describes the simulated data generated under 8 different scenarios to reflect a range of possible spatio-temporal data settings (Section 3.1). Different combinations of these simulated data scenarios and the fitted models described in the previous section are used for the simulation study based on their appropriateness, these are noted in Table 2. A fitted model is deemed appropriate for a specific simulated data scenario if it can accommodate spatio-temporal covariates when necessary. Thereafter, some computational considerations are discussed in Section 3.3., and lastly, the simulation study results are presented in Section 3.4.

3.1. Simulated Data

To simulate the data, we fixed the expected rate (e_{ij}) for the areas across the simulated data sets as a single realization of a $Gam(1,1)$ distribution. This allows for no necessary assumption to be made about the populations at each ij^{th} unit as the expected rate already takes into consideration the expected count adjusted by the total population. Further, it sets the primary focus on the estimation of risk θ_{ij}. To complete the parameterization, a relative risk which is parameterized with a range of different risk models is assumed.

The predictors used to simulate the outcome data come from the Area Health Resource Files (AHRF) (http://ahrf.hrsa.gov/) data set for the state of Georgia, USA which consists of 159 counties. Hence, $i = (1,...,159)$ for this

county set. This analysis covers the ten year time frame 1998-2007. The chosen ecological predictors are as follows: median household income (in thousands of dollars), percent persons below poverty level (PPBPL), unemployment rate of those aged 16 or greater (UER), and percent African American (AA) population. The geographical distributions of these predictors for year 2003 are displayed in Figure 1, and qualitatively, these data appear to have some spatial structure. These predictors are denoted as x_1, x_2, x_3, and x_4 respectively; additionally, they are standardized for simulation purposes. Depending on the model of interest, some of these predictors (x_1 and x_2) only vary spatially while others (x_3 and x_4) vary temporally as well as spatially. For the strictly spatially varying covariates, we use a central year of the data, year 2003 or $j = 6$ and assume that this covariate remains the same across the study time frame. The purpose of restricting the covariates in this way is simply to examine spatially-only varying covariates in this simulation study.

[Fig 1]

The eight basic models for risk (S1 up to S8) are presented in Table 1. The parameters that make up these models have the same prior distributions for each scenario, but over the 200 simulated data sets simulated for each of the risk models, we allow the temporal trend (γ_j), the spatio-temporal interaction term (ψ_{ij}), and the uncorrelated heterogeneous (UH) term (u_i) to vary from one simulated data set to the next. The respective hierarchical prior distributions of these random effects are as follows: $N(\gamma_{j-1}, \tau_\gamma^{-1})$, $N(0, \tau_\psi^{-1})$, and $N(0, \tau_u^{-1})$. The correlated heterogeneous (CH) term, $v_i \sim CAR(\tau_v)$, is unlike the other random effects in that it remains constant across the risk model scenarios as one realization of this distribution so that the spatial correlation is known and the same for all models and simulated data sets. Additionally, the CH term is defined by the same type of distribution which described the spatially varying model probabilities, and while the distribution is the same, the role of the parameters is much different. The UH and CH terms together form a convolution term to provide segregated random noise to the simulated data sets and allow some of this random noise to be unstructured (UH) while the rest is spatially correlated (CH). The standard deviations associated with these prior distributions are all uniformly distributed over the interval $[0, C]$.

As far as the regression parameter estimates are concerned, the α's are fixed across the 200 simulated data sets. For model scenarios with the non-temporally varying parameter estimates (S1 to S4), they are set to be

$(0.2,\ -0.3,\ 0.2,\ -0.4)$ corresponding to x_1 up to x_4, respectively, so that they are forced to be fairly large and thus have a detectable relationship with the outcome, yet not so large resulting in simulated outcome counts that tend towards infinity. For the scenarios where some or all of the α's vary over time (S5-S8), a single realization of a random walk trend, $\alpha_{k,j} \sim N\left(\alpha_{k,j-1}, 0.5\right)$, is employed for the temporally varying parameters such that the rows following matrix represents the parameter estimates that are fixed across the simulated data sets and associated with x_1 up to x_4, respectively:

$$\begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ -0.3 & -0.3 & -0.3 & -0.3 & -0.3 & -0.3 & -0.3 & -0.3 & -0.3 & -0.3 \\ 0.2 & 0.598 & 1.082 & 1.352 & 1.996 & 1.357 & 1.542 & 1.606 & 1.800 & 1.514 \\ -0.4 & -0.616 & 0.421 & -0.383 & -0.738 & -1.376 & -1.413 & -0.889 & -0.763 & -0.010 \end{bmatrix}.$$

This matrix is the fixed set of values for the scenarios that allow for some temporally varying parameter estimates (S5 to S8). Fixing the α's is for simplicity, consistency, and because they are not the main focus of the model selection application to begin with. Previous studies suggest that model selection should be performed ultimately to select the best linear predictors, and re-fits of the chosen models should be executed to gain appropriate parameter estimates (Carroll *et al.*, 2016). Additionally, S8 differs from the other models that include only spatially varying predictors in that the parameter estimates are allowed to vary over time; in this instance, we simply use the bottom two rows of fixed parameter estimates from the matrix above for simulation purposes. For all simulated model scenarios, the intercept, α_0, is set to be 5. These eight different scenarios furnish a wide range of models, and thus, a good basis to judge which fitted models perform best under the different conditions. Some models contain all of the predictors while others contain only a subset or none at all; we expect certain fitted models to perform better for specific scenarios based on this fact alone.

3.2. Summary of Simulated Data and Fitted Model Combinations

Table 2 summarizes the simulation study in terms of the simulated models associated with the different fitted models, the model description of the fitted models, and the linear predictor alternatives. Through these different combinations of simulated and fitted models, we believe that a wide range of scenarios are being tested and compared for inference. Additionally, Table 3 displays the appropriate linear predictor from each fitted model with respect to the simulated data scenarios. This table clearly identifies which linear predictor should be selected for each simulated data scenario and fitted model combination. Further, for all fitted models, we assumed C, as indicated in the prior distributions for the standard deviations of the fixed and random effects, to be four. This leads

to a reasonably non-informative prior distribution for models such as these on the log scale, and it was chosen so that the estimates are contained and unable to become too large. We also examined this prior for sensitivity. Further, other studies have shown that relative risks are not overly influenced by prior distributions placed on the variance parameters (Bernardinelli *et al.*, 1995).

To compare these modeling options among the different scenarios, we use a combination of qualitative and quantitative criteria. Maps and figures of model probabilities will be presented for comparison in a qualitative way. For a quantitative approach, we calculate *DIC* and its components. These provide goodness of fit summaries for a range of models. First, the DIC estimates are defined as follows:

$$D(\theta) = -2\sum_i \log\left(p(y_i|\theta)\right)$$
$$pD = Var(D(\theta))/2$$
$$DIC = \overline{D(\theta)} + pD$$

where $\sum_i \log\left(p(y_i|\theta)\right)$ is the log likelihood. The effective number of parameters, pD, in this definition involves the more conservative calculation using the variance of the deviance (Gelman *et al.*, 2004) rather than the mean deviance minus the deviation of the means (Spiegelhalter *et al.*, 2002). We chose this definition of *pD* because it is strictly positive, although, it does tend to produce a conservative estimate of the number of predictors. This is considered a penalty term for the models as models tend to fit better when more parameters are effectively used. The parameter estimates produced for all of these comparison measures are based on the posterior mean. This measure of location is defined such that $\bar{\eta} = \int \eta\, p(\eta|\mathbf{y})\, d\eta$, and it minimizes the squared loss (Lesaffre and Lawson, 2013). The use of DIC as a goodness of fit criterion can be criticized when mixtures with unknown numbers of components arise (Spiegelhalter *et al.*, 2002). In our case, we have fixed component numbers, thus DIC is a valid measure.

3.3. Computational Considerations

A common problem with large scale MCMC simulation is computation time. Computationally fast spatio-temporal analyses can be performed using approximate Bayesian inference via the R package INLA (Blangiardo *et al.*, 2013; Martins *et al.*, 2013; Schrödle and Held, 2010; Schrödle and Held 2011; Ugarte *et al.*, 2014; Rue *et al.*, 2009; R Core Team, 2015), however this package does not offer the degree of flexibility required for our model selection methods (Carroll *et al.*, 2015.b). Alternatively, to improve the computation time in this instance, the R package

snowfall which allows for parallelizing code can be implemented (Knaus, 2013). Thus, implementation of Bayesian model selection via the BRugs software package which calls OpenBUGS from the statistical software program R (Lunn *et al.*, 2013; Thomas *et al.*, 2014) is an appropriate computing choice for these methodologies. Specifically, running the same single dataset with F1 up to F4 for the same number of iterations and requesting the same types of parameters takes the following respective amounts of time: 1.3 hours, 59 minutes, 1.3 hours, and 23.8 minutes. However, these times are dependent on the computer processing power, the other processes being performed simultaneously, the number of iterations requested, and the number of parameters requested.

In the OpenBUGS sampler, the default sampling algorithms were utilized such that the discrete slice updater was used for the indicator weight parameters, a slice updater for the standard deviations parameters, a chain updater for the CH terms as well as the model probability parameter for F1, and current point metropolis type updaters for all other parameters. Additionally, it appears that the more parameter-heavy methods, F1 and F3, switch to a Gibbs type updater for a portion of the univariate normal conditional distributions, e.g. the fixed effect parameter estimates. For the two chains of these simulations, we ran 30000 iterations and sampled 2000 of them, thus a burn-in of 28000 was used. Furthermore, when these models are fitted, the initial values supplied to each chain of the MCMC are such that they are the expected values of the associated prior distributions. Finally, for convergence, we referred to $\hat{R}$ values and trace plots for a subset of the simulated data sets. Examples of these diagnostics using the melanoma data set are included as Figure 5 as well as Table 7 (all found in the online supporting information).

3.4. Simulation Results

Table 4 displays the goodness of fit measures associated with all model combinations of the fitted model selection methods averaged over the 200 simulated data sets. We can only truly compare these DIC measures across the fitted models as the data varies across the simulated models. Fitted model F1 does not appear to perform the best for any of the simulated data scenarios indicating that the spatio-temporal dependency on the model probabilities is unnecessary. F2 appears to be a good option for some of the simulated data scenarios; it has the flexibility to fit models with spatio-temporal parameters and produces the lowest DIC values for scenarios S5 and S8. F2 also provides the lowest DIC estimate for S6, but it is not significantly lower than the estimate produced by F4. F3 and F4 attain much smaller *pD* values for the predominantly spatial models, with F3 performing the best overall for S1 and S2 while F4 performs the best for S3. Additionally, F4 produces the lowest DIC values for S4 and S7. From

these estimates, we note that the deviances are quite close in value, and the large differences lie in the effective number of parameters, the penalty component to the DIC calculation. Altogether, these measurements accompanied by ability to fit a wide range of models indicates that the mixture model (F4) produces better fitting results overall since the DICs produced for this fitted model are consistently either the best or second best for all models except S8.

The results for F1 are difficult to assess because the model probabilities vary over both space and time meaning that there are 159*10 model probability estimates. We have included county maps displaying the model probabilities across time in the online supporting information (Figure 6). The maps show that there is not much variability across the counties, or across space, as the model probabilities are all very close in value.

Figure 2 plots the model probabilities associated with all simulated data set fits with fitted model F2. Most obviously, these show that there is more variance associated with the spatio-temporal linear predictors than with the spatial linear predictors. Additionally, the misspecified models (S1F2, S2F2, and S8F2) appear to exhibit more variance overall than the appropriately specified models (S3F2 up to S7F2). Of the models that are truly specified, they seem to be somewhat well identified. Models S1F2 and S2F2 are only slightly misspecified in that they match spatial linear predictors, ω_2^S and ω_3^S respectively, but additionally have a temporal component, γ_j. This temporal component is included in ω_3^{ST}, and that linear predictor does appear to be selected for these models. We also note that a large amount of models stay around the 0.333 range in probability estimates meaning that no linear predictor is favored. S7F2 is behaving differently from expected in that, based on the simulated model definition in Table 1, it seems as though ω_3^S and ω_2^{ST} should be selected, and then the spatio-temporal interaction random effect would pick up the rest of the noise present in the data due to misspecification. Rather, it appears that the interaction term is attempting to explain all of the temporal noise since the temporal linear predictor that includes the appropriate spatio-temporal components, ω_3^{ST}, is not being selected. Alternatively, S8F2 does appear to be selecting ω_3^{ST} to account for the temporal variation in γ_j.

[Fig 2]

Figure 3 displays the model probabilities associated with fitted model F3. The distribution of the model probabilities are similar for all model scenarios. For S2F3 and S8F3, the median for p_3 is higher, though only slightly, than p_2;

this is appropriate since the linear predictor associated with p_3 is closer to the simulation scenarios for S2 and S8. Additionally, the estimates produced for scenario S3F3 appear to be much more stable than the others. Both of these issues are likely due to the fact that the first linear predictor alternative is less complex and does not involve random effects which are present in the second and third linear predictor alternatives. Thus, the second two linear predictors dominate the selection process and there is more variation present when random effects are a part of the simulated data scenario.

[Fig 3]

Figure 4 displays violin plots of the model probabilities associated with the F4 model fits. These are produced using the vioplot package in R (Adler, 2005) and allow for viewing the distributions associated with the model probabilities for each scenario. Each violin plot image is a boxplot with a kernel density plot on each side such that the white circle is the median and the black rectangle represents the quartiles. This plot also displays a horizontal line at $probability = 0.5$ to aid in identifying the mixing effect of this fitted model. S1F4 and S2F4 are appropriately represented as spatial models with a bit of temporal mixture; their distributions are quite similar and this is to be expected because they only differ by one random effect, v_i; its inclusion in S2F4 causes that model to tend slightly more towards a predominantly spatial model. Alternatively, S3F4 is a simple, strictly spatial model that seems to be having a difficult time identifying that. Here, S4F4 and S5F4 are appropriately identified as spatio-temporal models based on their associated model probabilities; of these two models, S5F4 has a much smaller spread to its distribution because it is more truly specified within this model than S4F4 since the parameter estimates also vary temporally. Interestingly, S6F4 appears to have a nearly identical distribution to S5F4, and while they do have common parameters, S6F4 additionally has spatial only covariates. Thus, it is expected to be represented as more of a mixed model. S7F4 and S8F4 also behave as expected by displaying mixture effects between the spatial and spatio-temporal components.

[Fig 4]

4. Melanoma Cancer of the Skin Data Example

For our melanoma example, we use incidence of melanoma cancer of the skin obtained from the Georgia Center for Cancer Statistics (http://web1.sph.emory.edu/GCCS/cms/index.html) as an outcome. Predictors for melanoma are chosen from environmental and lifestyle/socio-economic factors. For the environmental predictor, we include average daily sunlight hours collected from the North America Land Data Assimilation System (http://wonder.cdc.

gov/) as a predictor, and this is likely to be an important environmental predictor with respect to melanoma. However, there could be a delay effect present between sun exposure and melanoma incidence but we believe that an ecological relationship should be present. The three socio-economic predictors are a subset of those included in the simulation study. These data overlap for a study time line of 1999 to 2007; as before, the spatial only varying covariates come from a central year of the study time line, 2003. Explicitly, the spatial only varying predictors are proportion AA population (x_1) and PPBPL (x_2) while the spatio-temporal predictors are average daily sunlight (x_3) and UER (x_4). Past studies suggest that melanoma has a spatial structure and is related to these predictors. However, they are not all the main risk factors associated with this cancer (National Institutes of Health, http://seer.cancer.gov/statfacts/html/melan.html; Ananthaswamy, 2001; American Cancer Society, http://www.cancer.org/cancer/skincancer-melanoma/detailedguide/melanoma-skin-cancer-what-causes; National Institutes of Health, http://www.cancer.gov/research/areas/disparities; Giovannucci, 2005; American Cancer Society, http://seer.cancer.gov/statfacts/html/lungb.html), and we are looking to explore the temporal structure as well. We apply these data using fitted models F2-F4 with the alternative linear predictors specified in Section 3.2. as these models appear to perform the best in the simulation study.

The goodness of fit results from these different model selection alternatives are displayed in Table 5. In terms of DIC and its components, F3 appears to be the best fitting model for these data. Table 6 displays the model probabilities produced as well as the resulting selected linear predictors. F4 is the only model that produces very strong results in that the model probability is large and well estiamted. These results indicate that this model can be described exclusively by a spatio-temporally structured linear predictor, and this could explain the large DIC value produced as well. The F2 model produces model probabilities of nearly identical magnitude for its spatial and spatio-temporal components and this suggests that both a spatial and spatio-temporal component are needed. While the F2 results do appear to be indicating selection for one of the spatial and one of the spatio-temporal linear predictor alternatives, the selection is not nearly as strong as the selection seen with F4. Additionally, F3 results are not clearly indicating one linear predictor over the others.

For the selected model re-fit, we have only chosen to use F2 and F4. F3 was eliminated because it did not make a clear selection of one of the alternative linear predictors. The re-fits include selected alternative linear predictors

associated with fitted model F2 and the spatio-temporal term of F4 since it appears as though that term is dominating the model. The goodness of fit measures produced in the re-fits show that F2 performs slightly better in comparison to its model selection fit DIC. On the other hand, F4 performs much better in terms of goodness of fit when only the spatio-temporal component is included; the re-fit of F4 offers the best overall fit in terms of DIC for this example. These goodness of fit measures can be found in Table 5 below the model selection fit estimates. The $\log(\theta_{ij})$ estimates produced for these model re-fits produce a nearly identical display to those for the model selection fits of these models. Lastly, the parameter estimates produced for these models are contained in Table 8 (found in the online supporting information). These parameter estimates show that several predictors are important since they are well estimated in selected linear predictor re-fits. From these estimates, it is evident that the models suggest that melanoma has a negative relationship with sun, UER and PPBPL while a positive relationship is present with proportion of AA population. However, re-fitting these models could be considered a re-use of the data, thus adjustments to these estimates may be necessary. We have not explored adjustment options as this paper is focused on model selection methods rather than inference.

Additional tables and figures included in the appendix display additional parameter estimates produced with selected linear predictors. Table 9 (found in the online supporting information) displays the temporal random walk term, γ_j. The estimates associated with F3 are much smaller than those associated with the other models and this could be due to the fact that there are two terms of this type in separate alternative linear predictors. Further, neither of these alternative linear predictors are the selected one. The F4 estimates are very large because they are being scaled back by the model probability weight (see Table 6). The maps of the log risk are displayed in Figures 7, 8, and 9 (found in the online supporting information). These maps show some, though not much variation from one year or model to the next, and there appears to be an elevated risk associated with the counties surrounding Atlanta, GA in the northwestern region of the state as well as some scattered elevation in the more rural counties.

In conclusion, this case study demonstrates that the implementation of these Bayesian model selection methods prove to be very useful. The methods effectively selected appropriate linear predictors for an optimal re-fit so that better inference could be made. Selected linear predictors were not the same; however, they only differed by the inclusion of the spatially varying predictors, and according to DIC, the final model fits best when these predictors

are not included. This overall best final model suggests that the incidence of melanoma is increasing over time based on the γ_j random effect and that when UER increases, the incidence of melanoma decreases. Both of these results agree with previous research involving melanoma incidence (Singh *et al.*, 2011; Erdei and Torres, 2010).

5. Discussion

Based on the results in the previous sections, the mixture model, F4, is the best option for the majority of model scenarios, but it is not without fault. Fitted models F2 and F3 also perform well for certain scenarios, but each have their own weaknesses. F2 offers the ability to fit more combinations of linear predictor alternatives by allowing for separate selection between the spatial and spatio-temporal components. F3, on the other hand, allows for more complex linear predictor alternatives within the spatial setting. The more complex linear predictor alternatives offered by both of these fitted model scenarios lead to slow fitting models and more difficulty with convergence. Fitted model F1 does not perform as well as these other models, and this is ultimately because of the complexity within the model probability estimates.

Concerning measures of goodness of fit for the simulation study, most models fit best with F4 according to DIC. However, a few of the spatial and spatio-temporal models still perform better with F3 and F2 respectively. When taking into consideration the models' ability to appropriately identify the true alternative linear predictors, it appears that while F2 performs well when the models are appropriately specified, F4 seems to do best at identifying spatial only, spatio-temporal only, or both. The case study also demonstrated F4's ability to select the best fitting linear predictor.

The conventional model selection models are all quite sensitive to model misspecification, and this is to be expected since the models are very intricate. Fitted model F4 also superceeds the others in terms of misspecification because almost all models are truly specified when supplied to F4; there may be additional, unnecessary terms included with certain model fits, but this mixture model proves able to handle that. Through the mixture component, F4 is able to accommodate most of the linear predictor alternatives the other fitted models offer. Fitted model F2 gives the worst, most variable results for misspecified models, particularly when temporal misspecification is present. This indicates that F2 is not as flexible as or able to support overparameterization in the same way as F4.

Two additional issues among this study involve the random effects. The first of these involves the CH effect. Other studies mention that the inclusion of a correlated random effect can alter the fixed parameter estimates (Reich *et al.*, 2006; Hodges and Reich, 2010), but our focus here is model selection, not parameter estimation. Therefore, we did not explore this limitation further with regard to its impact on our alternative linear predictors. Second, with the amount of random effects included in these models, particularly for F1, F2, and F3, there can be issues with identifiability. Futhermore, all of the fitted models include a spatio-temporal interaction term as an uncorrelated random effect, and this may also be producing some indentification issues with the model selection process. An example of this was distinguished involving model S7F2 selecting the inappropriate alternative linear predictor and letting the spatio-temporal interaction term explain the variation instead.

One negative feature of fitted model F4 involves its lack of flexibility through the restriction of supplying only two linear predictor alternatives that must be determined a priori. Our melanoma example illustrates that this may not be a problem for every application of this method, but the results associated with model scenarios S3F4 and S6F4 illustrate why this can be an issue. S3 is fully spatial, but the median for that model probability over the 200 simulated data sets for S3F4 is below 0.5, however not significantly so. Correspondingly, S6 is spatio-temporal with S3 as a nested model indicating a strong spatial component, but S6F4 does not appear to detect those spatial trends as the distribution of its model probabilities are quite stable and well below 0.5. This may be a fault in the data as the spatial only varying true parameter estimates do have smaller magnitudes than the temporally varying ones due to the random walk that was placed on these parameters.

6. Conclusion

Following these study results, we believe that fitted model F4, the mixture model, is the best option of the model selection alternatives. This model is fast fitting, the least vulnerable to misspecification, the least vulnerable to identification issues, accommodates a wide range of linear predictors, and the one that appears to fit the best overall. Following F4, F2 is another worthwhile option as this model also accommodates an even wider range of linear predictor alternatives and offers the best fit in certain situations. With respect to implementation in public health and environmental studies, our results suggest that both F4 and F2 are adequate starting points for spatio-temporal disease mapping situations in which there is a fixed set of appropriate linear predictors that can be determined a

priori. These alternative linear predictors 1) may be supposed optimal combinations of important predictors, 2) could be important in terms of the study design, and 3) can offer a comparison of alternating collinear predictors.

Additional information and supporting material for this article is available online at the journal's website.

Acknowledgements

This research was supported in part by funding under grant NIH R01CA172805.

References

Adler D. 2005. vioplot: Violin plot. http://wsopuppenkiste.wiso.uni-goettingen.de/~dadler.

American Cancer Society. 2015. Cancer Facts & Figures 2015. Atlanta, GA. http://seer.cancer.gov/statfacts/html/lungb.html. [accessed 14 January 2016]

American Cancer Society. 2015. Do we know what causes melanoma skin cancer? Atlanta, GA. http://www.cancer.org/cancer/skincancer-melanoma/detailedguide/melanoma-skin-cancer-what-causes. [accessed 28 January 2016]

Ananthaswamy HN. 2001. Sunlight and Skin Cancer. *Journal of Biomedicine & Biotechnology* **1**(2):49. DOI: 10.1155/S1110724301000122

Area Health Resource Files (AHRF). 2003. Rockville, MD: US Department of Health and Human Services, Health Resources and Services Administration, Bureau of Health Workforce. http://ahrf.hrsa.gov/. [accessed 13 June 2015]

Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. 1995. Bayesian analysis of space—time variation in disease risk. *Statistics in Medicine* **14**(21-22):2433-43. DOI: 10.1002/sim.4780142112

Besag J, York J, Mollié A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**(1):1-20. DOI: 10.1007/bf00116466

Blangiardo M, Cameletti M, Baio G, Rue H. 2013. Spatial and spatio-temporal models with R-INLA. *Spatial and Spatiotemporal Epidemiology* **4**:33-49. DOI: 10.1016/j.sste.2012.12.001

Bondell HD, Krishna A, Ghosh SK. 2010. Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics* **66**(4):1069-77. DOI: 10.1111/j.1541-0420.2010.01391.x

Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. 2015.a. Bayesian model selection methods in modeling small area colon cancer incidence. *Annals of Epidemiology* **26**(1):43-9. DOI: 10.1016/j.annepidem.2015.10.011

Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. 2015.b. Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spatial and Spatiotemporal Epidemiology* **14-15**:45-54. DOI: 10.1016/j.sste.2015.08.001

Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. 2016. Spatially-dependent model selection for disease mapping. *Statistical Methods in Medical Research* In print. DOI: 10.1177/0962280215627298.

Erdei E, Torres SM. 2010. A new understanding in the epidemiology of melanoma. *Expert Review of Anticancer Therapy*. **10**(11):1811-23. DOI: 10.1586/era.10.170

Garcia RI, Ibrahim JG, Zhu H. 2010. Variable selection for regression models with missing data. *Statistica Sinica*. **20**(1):149-65.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2004. *Bayesian Data Analysis* (3rd ed.). CRC Press: Boca Raton.

George EI, Clyde M. 2004. Model uncertainty. *Statistical Science* **19**(1):81-94. DOI: 10.1214/088342304000000035

Georgia Center for Cancer Statistics (GCCS). 1976. Georgia Cancer Registry. Department of Epidemiology, Rollins School of Public Health, Emory University. http://web1.sph.emory.edu/GCCS/cms/index.html. [accessed 14 January 2016]

Giovannucci E. 2005. The epidemiology of vitamin D and cancer incidence and mortality: A review (United States). *Cancer Causes and Control* **16**(2):83-95.

Hodges JS, Reich BJ. 2010. Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician* **64**(4):325-34. DOI: 10.1198/tast.2010.10052

Hoeting JA, Raftery AE, Madigan D. 2002. Bayesian variable and transformation selection in linear regression. *Journal of Computational and Graphical Statistics* **11**(3):485:507. DOI: 10.1.1.35.1365

Knaus J. 2013. snowfall: Easier cluster computing (based on snow). R package version 1.84-6. http://CRAN.R-project.org/package=snowfall.

Knorr-Held L. 2000. Bayesian modeling of inseparable space-time variation in disease risk. *Statistics in Medicine* **9**(17-18):2555-67. DOI: 10.1.1.477.4180

Lawson AB. 2013. *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology* (2nd ed.). CRC Press: Boca Raton.

Lee KJ, Jones GL, Caffo BS, Bassett SS. 2014. Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. *Bayesian Analysis* **9**(3):699-732. DOI: 10.1214/14-BA873

Lesaffre E, Lawson AB. 2013. *Bayesian Biostatistics* (1st ed.). Wiley: West Sussex.

Li G, Best N, Hansell AL, Ahmed I, Richardson S. 2012. BaySTDetect: detecting unusual temporal patterns in small area data via Bayesian model choice. *Biostatistics* **13**(4):695-710. DOI: 10.1093/biostatistics/kxs005.

Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. 2013. *The BUGS book: A practical introduction to Bayesian analysis* (1st ed.). CRC Press: Boca Raton.

Martins TG, Simpson D, Lindgren F, Rue H. 2013. Bayesian computing with INLA: New features. *Computational Statistics and Data Analysis* **67**:68-83. DOI: 10.1016/j.csda.2013.04.014

National Cancer Institute. 2008. Cancer Health Disparities. Rockville, MD. http://www.cancer.gov/research/areas/disparities [accessed 14 Jan 2016].

National Insitutes of Health. SEER Stat Fact Sheets: Melanoma of the Skin. Rockville, MD. http://seer.cancer.gov/statfacts/html/melan.html [accessed 14 Jan 2016].

North America Land Data Assimilation System (NLDAS). 2013. Daily Sunlight (insolation) for years 1979-2011 on CDC WONDER Online Database. Centers for Disease Control and Prevention. http://wonder.cdc.gov/ [accessed 20 January 2016]

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org/

Reich BJ, Hodges JS, Zadnik V. 2006. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* **62**(4):1197-206. DOI: 10.1111/j.1541-0420.2006.00617.x

Rockova V, George EI. 2014. Negotiating multicollinearity with spike-and-slab priors. *Metron* **72**(2):217-29. DOI: 10.1007/s40300-014-0047-y

Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society: Series B* **71**:319-92. DOI: 10.1111/j.1467-9868.2008.00700.x

Scheel I, Ferkingstad E, Frigessi A, Haug O, Hinnerichsen M, Meze-Hausken E. 2013. A Bayesian hierarchical model with spatial variable selection: The effect of weather on insurance claims. *Journal of the Royal Statistical Society: Series C (Applied statistics)*. **62**(1):85-100. DOI: 10.1111/j.1467-9876.2012.01039.x

Schrödle B, Held L. 2010. A primer on disease mapping and ecological regression using INLA. *Computational Statistics* **26**(2):241-58. DOI: 10.1007/s00180-010-0208-2

Schrödle B, Held L. 2011. Spatio-temporal disease mapping using INLA. *Environmetrics* **22**(6):725-34. DOI: 10.1002/env.1065

Singh SD, Ajani UA, Johnson CJ, Roland KB, Eide M, Jemal A, Negoita S, Bayakly RA, Ekwueme DU. 2011. Association of cutaneous melanoma incidence with area-based socioeconomic indicators-United States, 2004-2006. *Journal of the American Academy of Dermatology* **65**(5 Suppl 1):S58-68. DOI: 10.1016/j.jaad.2011.05.035

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B* **64**(4):583-639. DOI: 10.1111/1467-9868.00353.

Thomas A, Best N, Lunn D, Arnold R, Spiegelhalter D. 2014. GeoBUGS user manual. http://www.openbugs.net/Manuals/GeoBUGS/Manual.html.

Ugarte MD, Adin A, Goicoa T, Militino AF. 2014. On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statistical Methods in Medical Research* **23**(6):507-30. DOI: 10.1177/0962280214527528

Waldorp LJ, Huizenga HM, Nehorai A, Grasman RP, Molenaar PC. 2005. Model selection in spatio-temporal electromagnetic source analysis. *IEEE Transactions on Bio-medical Engineering* **52**(3):414-20. DOI: 10.1109/TBME.2004.842982

Table 1: Simulated model scenarios

Model	Contents
S1	$\log\left(\theta_{ij}\right) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + \gamma_j$
S2	$\log\left(\theta_{ij}\right) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i + \gamma_j$
S3	$\log\left(\theta_{ij}\right) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i}$
S4	$\log\left(\theta_{ij}\right) = \alpha_0 + \alpha_3 x_{3ij} + \alpha_4 x_{4ij}$
S5	$\log\left(\theta_{ij}\right) = \alpha_0 + \alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij}$
S6	$\log\left(\theta_{ij}\right) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij}$
S7	$\log\left(\theta_{ij}\right) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij} + u_i + v_i + \psi_{ij}$
S8	$\log\left(\theta_{ij}\right) = \alpha_0 + \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} + u_i + v_i + \gamma_j$

Table 2: Summary of the fitted models

Model	Associated Simulated Models	Description	Linear Predictor Alternatives
F1	S1-S3, S8	$\log\left(\theta_{ij}\right)=\alpha_0+\sum_{l=1}^{L}w_{lij}\omega_{lij}+\psi_{ij}$	$\omega_{ij}=\begin{cases}\alpha_1 x_{1i}+\alpha_2 x_{2i}\\ \alpha_{1j} x_{1i}+\alpha_{2j} x_{2i}\\ \alpha_{1j} x_{1i}+\alpha_{2j} x_{2i}+\gamma_j\\ \alpha_{1j} x_{1i}+\alpha_{2j} x_{2i}+u_i+v_i+\gamma_j\\ \alpha_1 x_{1i}+\alpha_2 x_{2i}+u_i+v_i+\gamma_j\end{cases}$
F2	S1-S8	$\log\left(\theta_{ij}\right)=\alpha_{0j}+\sum_{l=1}^{L}w_l^S\omega_{li}^S+\sum_{k=1}^{K}w_k^{ST}\omega_{kij}^{ST}+\psi_{ij}$	$\omega_i^S=\begin{cases}\alpha_1 x_{1i}+\alpha_2 x_{2i}\\ \alpha_1 x_{1i}+\alpha_2 x_{2i}+u_i\\ \alpha_1 x_{1i}+\alpha_2 x_{2i}+u_i+v_i\end{cases}\quad \omega_{ij}^{ST}=\begin{cases}\alpha_1 x_{3ij}+\alpha_2 x_{4ij}\\ \alpha_{1j} x_{3ij}+\alpha_{2j} x_{4ij}\\ \alpha_{1j} x_{3ij}+\alpha_{2j} x_{4ij}+\gamma_j\end{cases}$
F3	S1-S3, S8	$\log\left(\theta_{ij}\right)=\alpha_0+\sum_{l=1}^{L}w_l\,\omega_{lij}+\psi_{ij}$	$\omega_{ij}=\begin{cases}x_{1i}'(\alpha_1+\alpha_{1i}+\alpha_{1j})+x_{2i}'(\alpha_2+\alpha_{2i}+\alpha_{2j})\\ x_{1i}'(\alpha_1+\alpha_{1i}+\alpha_{1j})+x_{2i}'(\alpha_2+\alpha_{2i}+\alpha_{2j})+\gamma_j\\ x_{1i}'(\alpha_1+\alpha_{1i}+\alpha_{1j})+x_{2i}'(\alpha_2+\alpha_{2i}+\alpha_{2j})+u_i+v_i+\gamma_j\end{cases}$
F4	S1-S8	$\log\left(\theta_{ij}\right)=\alpha_0+p\omega_i^S+(1-p)\omega_{ij}^{ST}+\psi_{ij}$	$\omega_i^S=\alpha_1 x_{1i}+\alpha_2 x_{2i}+u_i+v_i$ $\omega_{ij}^{ST}=\alpha_{3j} x_{3ij}+\alpha_{4j} x_{4ij}+\gamma_j$

Table 3: Description of appropriate linear predictors from the fitted models (F1 up to F4) with respect to the simulated data scenarios (S1 up to S8).

		F1	F2	F3*	F4
S1	S	(ω_{5ij})	ω_{2i}^{S}	(ω_{3ij})	(ω_{i}^{S})
	ST		(ω_{3ij}^{ST})		(ω_{ij}^{ST})
S2	S	ω_{5ij}	ω_{3i}^{S}	ω_{3ij}	ω_{i}^{S}
	ST		(ω_{3ij}^{ST})		(ω_{ij}^{ST})
S3	S	ω_{1i}	ω_{1i}^{S}	ω_{1ij}	(ω_{i}^{S})
S4	ST	---	ω_{1ij}^{ST}	---	(ω_{ij}^{ST})
S5	ST	---	ω_{2ij}^{ST}	---	(ω_{ij}^{ST})
S6	S	---	ω_{1i}^{S}	---	(ω_{i}^{S})
	ST		ω_{2ij}^{ST}		(ω_{ij}^{ST})
S7*	S	---	ω_{3i}^{S}	---	ω_{i}^{S}
	ST		ω_{2ij}^{ST}		(ω_{ij}^{ST})
S8	S	ω_{4ij}	X	ω_{3ij}	X
	ST		(ω_{3ij}^{ST})		ω_{ij}^{ST}

'X' indicates that there is no appropriate linear predictor supplied and ( ) around the linear predictor indicate that this linear predictor is over parameterized with respect to the simulated data scenario. The * by S7 and F3 indicates that these differ from the other simulated data and fitted model scenarios respectively and the ( ) are used apart from those differences. They differ in that S7 has an extra parameter, ψ_{ij}, that the fitted models do not include in the alternative linear predictors and F3 offers more complex fixed parameter estimates compared to the ground truth from the simulated data scenarios.

Table 4: DIC, effective number of parameters, and mean deviance measures

Model	F1			F2			F3			F4		
	DIC	*pD*	$\overline{D(\theta)}$	*DIC*	*pD*	$\overline{D(\theta)}$	*DIC*	*pD*	$\overline{D(\theta)}$	*DIC*	*pD*	$\overline{D(\theta)}$
S1	12777.94	1581.51	12777.94	12123.06	855.73	11267.32	11652.70	319.72	11332.97	11765.04	449.72	11315.32
S2	13067.43	1583.05	11484.38	12167.12	859.34	11307.78	10732.28	306.37	10425.91	11916.57	491.10	11425.47
S3	12404.88	978.86	11426.02	12629.71	1017.19	11612.52	12026.54	438.54	11587.99	11752.24	175.59	11576.65
S4	---	---	---	12674.43	1040.80	11633.64	---	---	---	11784.63	207.30	11577.33
S5	---	---	---	13262.81	1638.36	11624.45	---	---	---	13281.36	1659.66	11621.70
S6	---	---	---	13227.34	1607.54	11619.79	---	---	---	13230.00	1611.58	11618.42
S7	---	---	---	15035.67	3395.29	11640.38	---	---	---	14648.07	3012.52	11635.54
S8	13727.27	1877.75	11849.52	10095.13	1323.07	8772.06	12858.08	810.23	12047.85	13322.44	1522.15	11800.29

Table 5: DIC measures for the melanoma example and the re-fits of selected linear predictors.

Model	*DIC*	*pD*	$\bar{D}$
F2	7024.7	1112.61	5912.1
F3	6954.02	1050.50	5903.5
F4	11278.53	5731.43	5547.1
F2 re-fit	6996.26	1065.8	5930.44
F4 re-fit	6951.37	1046.98	5904.38

Table 6: Model probabilities and selected linear predictors for the melanoma example

Model		Model Probabilities			Selected Linear Predictor
		p_1	p_2	p_3	
F2	Spatial	0.41 (0.04,0.88)	0.29 (0.01,0.64)	0.29 (0.01,0.64)	$\alpha_1 x_{1i} + \alpha_2 x_{2i}$
	Spatio-Temporal	0.29 (0.01,0.66)	0.30 (0.01,0.65)	0.40 (0.03,0.86)	$\alpha_{1j} x_{3ij} + \alpha_{2j} x_{4ij} + \gamma_j$
F3		0.33 (0.02,0.69)	0.34 (0.02,0.68)	0.34 (0.01,0.68)	NA
F4		0.17 (0.16,0.18)			$\alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij} + \gamma_j$