Dynamic Multiscale Spatiotemporal Models for Poisson Data

Thaís C. O. Fonseca[1] and Marco A. R. Ferreira[2]

Abstract

We propose a new class of dynamic multiscale models for Poisson spatiotemporal processes. Specifically, we use a multiscale spatial Poisson factorization to decompose the Poisson process at each time point into spatiotemporal multiscale coefficients. We then connect these spatiotemporal multiscale coefficients through time with a novel Dirichlet evolution. Further, we propose a simulation-based full Bayesian posterior analysis. In particular, we develop filtering equations for updating of information forward in time and smoothing equations for integration of information backward in time, and use these equations to develop a forward filter backward sampler for the spatiotemporal multiscale coefficients. Because the multiscale coefficients are conditionally independent a posteriori, our full Bayesian posterior analysis is scalable, computationally efficient, and highly parallelizable. Moreover, the Dirichlet evolution of each spatiotemporal multiscale coefficient is parametrized by a discount factor that encodes the relevance of the temporal evolution of the spatiotemporal multiscale coefficient. Therefore, the analysis of discount factors provides a powerful way to identify regions with distinctive spatiotemporal dynamics. Finally, we illustrate the usefulness of our multiscale spatiotemporal Poisson methodology with two applications. The first application examines mortality ratios in the state of Missouri, and the second application considers tornado reports in the American Midwest.

Keywords: Areal data; Bayesian dynamic models; Massive data sets; MCMC; Multiscale modeling; Time series models for counts.

[1]Department of Statistics, Federal University of Rio de Janeiro, CEP: 21945-970, Brazil, thais@im.ufrj.br

[2]Department of Statistics, Virginia Tech, Blacksburg, VA, USA, marf@vt.edu

1 Introduction

Advances in data acquisition technology have led to an explosive increase in the size of scientific data sets. This in turn has led to the need of statistical methods that scale well with data set size. Several methods have been proposed for the analysis of large Gaussian point-referenced data sets (e.g., Fuentes, 2007; Banerjee et al., 2008; Paciorek and McLachlan, 2009; Lemos and Sansó, 2009). With respect to areal data (e.g., Banerjee et al., 2004), several spatiotemporal models and methods have been developed for disease mapping (e.g., see Bernardinelli et al., 1995; Waller et al., 1997; Knorr-Held, 2000; Schmid and Held, 2004; Tzala and Best, 2008, and references therein). However, these models and methods usually do not scale well with data set size. A possible way to develop methods that scale well for large data sets is through multiscale spatiotemporal models, such as the models for Gaussian data developed by Berliner et al. (1999), Johannesson et al. (2007), and Ferreira et al. (2010, 2011); however, their methodology is not directly applicable to Poisson data. Here, we propose a new class of dynamic multiscale models for Poisson spatiotemporal processes.

To develop new dynamic multiscale spatiotemporal Poisson (MSSTP) models, we couple a multiscale Poisson factorization with a novel Dirichlet temporal evolution. Specifically, at each time point we apply the Kolaczyk-Huang multiscale spatial Poisson factorization (Kolaczyk and Huang, 2001) to decompose the mean or intensity function of the Poisson process into spatiotemporal multiscale coefficients. Further, we connect the spatiotemporal multiscale coefficients through time with a novel Dirichlet temporal evolution. This temporal evolution is related to conditionally conjugate temporal evolution for binomial, negative binomial, Poisson, and multinomial observations (Smith, 1979, 1981) (see also Harvey, 1989; Prado and West, 2010; Gamerman et al., 2013). Our novel Dirichlet temporal evolution depends on discount factors related to how fast the spatiotemporal multiscale coefficients change through time.

For the analysis of these new dynamic MSSTP models, we propose a simulation-based full Bayesian posterior analysis. In particular, we develop filtering equations for updating of

information forward in time and smoothing equations for integration of information backward in time. We then use these filtering and smoothing equations to develop a forward filter backward sampler for the spatiotemporal multiscale coefficients. Further, we provide marginal posterior densities for the discount factor parameters. Hence, to simulate from the joint posterior distribution of all the unknown quantities in the model, we use a composite approach. First, we simulate a sample of the discount factor parameters from their marginal posterior distributions. After that, for each simulated discount factor we simulate a realization of the spatiotemporal multiscale coefficients using our forward filter backward sampler. As a result, we obtain a sample from the joint posterior distribution of discount factors and spatiotemporal multiscale coefficients.

Research on multiscale spatiotemporal models is in its infancy, and the relatively few articles published to date focus on Gaussian data. Berliner et al. (1999) developed a hierarchical model for turbulence using a wavelet decomposition for an underlying latent turbulence process and allowed the wavelet coefficients to evolve through time using state-space equations. Johannesson et al. (2007) have proposed a multiscale decomposition that uses a coarse to fine construction, assumes a fixed number of children subregions at each resolution level, and assumes temporal dynamics only at the aggregated coarse level. More recently, Ferreira et al. (2010, 2011) have developed multiscale spatiotemporal models based on a non-wavelet-based multiscale decomposition that allows a different number of children for each subregion at any resolution level, allows non-constant variance across the region of interest, and assumes temporal dynamics at all scales of resolution. However, these previous multiscale spatiotemporal models are not directly applicable to Poisson data.

Our MSSTP methodology decomposes large Poisson data sets into many smaller components called empirical multiscale coefficients. For inhomogeneous spatiotemporal Poisson processes, given the latent spatiotemporal multiscale coefficients, the empirical multiscale coefficients are conditionally independent. Further, we assume a separate Dirichlet temporal evolution for each latent spatiotemporal multiscale coefficient, that is, we assume that these

latent coefficients are conditionally independent a priori. As a result, the multiscale Poisson factorization of the likelihood function leads the spatiotemporal multiscale coefficients to be conditionally independent a posteriori. As a practical consequence, each spatiotemporal multiscale coefficient and its corresponding discount factor can be analyzed separately, leading our full Bayesian posterior analysis to be scalable, computationally efficient, and highly parallelizable.

Our multiscale spatiotemporal framework performs smoothing simultaneously in space and time. This point is made clear in Section 4, where we present results on the spatial and spatiotemporal dependence structures of our proposed models. To further investigate this point, Section 5.3 presents for two real data applications a comparison of our MSSTP model to two competing models. First, to access the ability of our MSSTP framework to borrow strength spatially, we compare our model to a model that assumes that each finest level subregion has its own temporal evolution. Second, to access our framework's ability to incorporate spatiotemporal dynamics, we compare our model to a widely used spatiotemporal model based on Markov random fields. We perform these model comparisons using two criteria: the conditional Bayes factor (e.g., see Ghosh et al., 2006; Vivar and Ferreira, 2009) that compares the predictive performance of the competing models, and the deviance information criterion (DIC) (Spiegelhalter et al., 2002). As Section 5.3 reports, for both applications our MSSTP model is substantially superior to the two competing models. Therefore, by borrowing strength spatially and by incorporating spatiotemporal dynamics, our MSSTP model for Poisson data provides superior results.

We advocate the analysis of the discount factors as a powerful way to identify regions with important spatiotemporal dynamics. Specifically, as spatiotemporal data sets grow increasingly larger, it is not feasible to visualize the entire data set. Our MSSTP framework offers great opportunity in terms of prioritizing what subregions of the region of interest should be analyzed more thoroughly. In particular, each discount factor is related to how fast the corresponding spatiotemporal multiscale coefficient changes through time. Hence,

each discount factor encodes the relevance of the temporal evolution of the corresponding spatiotemporal multiscale coefficient. Therefore, smaller discount factors indicate regions where the mean or intensity function changes more rapidly through time and thus contain spatiotemporal dynamics that warrant further investigation.

The remainder of this article is organized as follows. Section 2 describes the Poisson multiscale factorization and introduces the proposed dynamic multiscale spatiotemporal model for Poisson data. Section 3 develops a simulation-based full Bayesian posterior analysis methodology. Section 4 presents results on the spatial and spatiotemporal dependence structures of our proposed models. Section 5 illustrates the usefulness of our novel methodology with applications to two real data sets. Finally, Section 6 presents conclusions and future developments. For convenience of exposition, we present all proofs in the Appendix.

2 Multiscale Poisson spatiotemporal modeling

To develop new dynamic multiscale models for Poisson spatiotemporal processes, we couple a multiscale Poisson factorization with a novel Dirichlet temporal evolution. Specifically, at each time point we apply the multiscale spatial Poisson factorization proposed by Kolaczyk and Huang (2001) to decompose the mean or intensity function of the Poisson process into spatiotemporal multiscale coefficients. In addition, we apply the same factorization to the Poisson data set and obtain empirical spatiotemporal multiscale coefficients that are naïve estimators of the latent spatiotemporal multiscale coefficients. For completeness and to establish notation, Section 2.1 presents the multiscale factorization of Kolaczyk and Huang (2001) (see also Ferreira and Lee, 2007, Chapter 9) that we specialize for spatiotemporal Poisson data. After that, Section 2.2 describes the temporal evolution of the mean process at the coarsest resolution level and of the spatiotemporal multiscale coefficients at the various resolution levels. Finally, in Section 2.3 we complete our MSSTP model for Poisson areal data by specifying the priors for the discount factors and for the latent process at time $t = 0$.

2.1 Poisson multiscale factorization

Assume that interest lies in an inhomogeneous spatiotemporal Poisson process with rate $\{\lambda_t(s) : s \in \mathcal{S}, t \in \mathbb{Z}\}$ on a spatial domain $\mathcal{S} \subset \mathbb{R}^k$. Here, k is typically less than or equal to 3. Moreover, assume that because of measurement, resources or confidentiality restrictions, data are available only up to a given scale of resolution L on a partition of the domain $\mathcal{S}$. Denote this partition by $\{B_{L1}, \ldots, B_{L,n_L}\}$, with $B_{Lj} \in \mathcal{S}$, $j = 1, \ldots, n_L$, $B_{Li} \cap B_{Lj} = \emptyset$, $i \neq j$, and $\cup_{j=1}^{n_L} B_{Lj} = \mathcal{S}$. For example, because of confidentiality concerns data on occurrences of a disease may only be available at the county level; in this case, $B_{L1}, \ldots, B_{L,n_L}$ would be the counties within the domain $\mathcal{S}$. On another example, when considering point process data, because of limitations of computational resources it may be more advantageous to bin the data. In this case, $B_{L1}, \ldots, B_{L,n_L}$ would be the bins at the finest resolution level. We illustrate this latter case with the tornado report application presented in Section 5.2.

For each subregion B_{Lj} there is a count y_{tLj} of the number of occurrences of the event of interest at time t, $t = 1, \ldots, T$. Moreover, the expected number of counts on B_{Lj} at time t is $\mu_{tLj} = E(y_{tLj}) = \int_{B_{Lj}} \lambda_t(s)ds$, $j = 1, \ldots, n_L$. Then, the model for the number of counts at the finest resolution level L at time t is

$$y_{tLj}|\mu_{tLj} \sim \text{Poisson}(\mu_{tLj}). \tag{1}$$

Further, similarly to Kolaczyk and Huang (2001) we assume that $y_{tL1}, \ldots, y_{tL,n_L}$ are conditionally independent given $\mu_{tL1}, \ldots, \mu_{tL,n_L}$, $t = 1, \ldots, T$. In what follows, the latent spatiotemporal process $\lambda_t(s)$ is constructed in a way that this latent process will be spatiotemporally correlated. Therefore, this spatiotemporal dependence will transfer to the counts $y_{tL1}, \ldots, y_{tL,n_L}$ and lead their marginal distribution to contain spatiotemporal dependence.

In addition to the mean process at the Lth resolution level, we are also interested in the process at aggregated coarser scales. At the lth scale of resolution, the domain $\mathcal{S}$ is partitioned in n_l subregions $B_{l1}, \ldots, B_{ln_l}$, $l = 1, \ldots, L-1$. Moreover, the partition at level l

is assumed to be a refinement of the partition at level $l+1$; that is, $B_{lj} = \cup_{(l+1,j')\in D_{lj}} B_{l+1,j'}$, where D_{lj} is the set of descendants of subregion j at level l, and $D_{lj} \cap D_{li} = \emptyset, i \neq j$. Additionally, let $A_l(L,j)$ be the ancestor at resolution level l of subregion (L,j). Finally, denote by d_{lj} the number of descendants of subregion (l,j).

We assume that we observe data at the finest resolution level L and aggregate the data to obtain the data at resolution levels 1 to $L-1$. Let G be a set of subregions and denote by $\mathbf{y}_{tG}$ and $\boldsymbol{\mu}_{tG}$ the corresponding vectors of counts and expected values at time t, respectively. For example, $\mathbf{y}_{tD_{lj}}$ denotes the vector of counts observed in the descendants of subregion (l,j) at time t. Further, let $\mathbf{1}_m$ denote the m-dimensional vector of ones. Then, the aggregated counts at the lth level of resolution at time t are recursively defined as $y_{tlj} = \mathbf{1}'_{d_{lj}} \mathbf{y}_{tD_{lj}}$ with corresponding aggregated mean process $\mu_{tlj} = \mathbf{1}'_{d_{lj}} \boldsymbol{\mu}_{tD_{lj}}$. Usually, the mean μ_{tlj} may be written as $\mu_{tlj} = \lambda_{tlj} e_{tlj}$, where λ_{tlj} is the relative risk on subregion (l,j) at time t and e_{tlj} is either known or unknown up to a low-dimensional parameter vector. Examples of known e_{tlj} include the case when $e_{tlj} = 1$ and the case when e_{tlj} is the known population size of subregion (l,j) at time t. We illustrate these two cases of known e_{tLj} with the application presented in Section 5.1. In contrast, an example of e_{tLj} unknown up to a low-dimensional parameter vector is $e_{tLj} = \exp(\mathbf{x}'_l \boldsymbol{\beta}_{Lj})$, where $\mathbf{x}_t$ is a known vector of regressors common to all regions at time t and $\boldsymbol{\beta}_{Lj} = \boldsymbol{\beta}_{A_1(L,j)}$, and we recall that $A_l(L,j)$ is the ancestor at resolution level l of subregion (L,j). We demonstrate this latter case of unknown e_{tLj} with the application presented in Section 5.2. Similarly to the observed counts, e_{tlj} may be aggregated as $e_{tlj} = \mathbf{1}'_{d_{lj}} \mathbf{e}_{tD_{lj}}$. In that case, the aggregation for the mean process implies that the relative risk process is aggregated as $\lambda_{tlj} = e_{tlj}^{-1} \boldsymbol{\lambda}'_{tD_{lj}} \mathbf{e}_{tD_{lj}}$.

We now apply the Poisson multiscale factorization of Kolaczyk and Huang (2001) to the intensity function and to the data at time t. Because the observations at the finest resolution level L given the latent process λ are conditionally independent and have Poisson

distributions, the likelihood function admits the multiscale factorization

$$\prod_{j=1}^{n_L} p(y_{tLj}|\mu_{tLj}) = \prod_{j=1}^{n_1} p(y_{t1j}|\mu_{t1j}) \prod_{l=1}^{L-1} \prod_{j=1}^{n_l} p(\mathbf{y}_{tD_{lj}}|y_{tlj}, \boldsymbol{\omega}_{tlj}), \tag{2}$$

where $y_{t1j}|\mu_{t1j} \sim \text{Poisson}(\mu_{t1j})$. Further, $\mathbf{y}_{tD_{lj}}|y_{tlj}, \boldsymbol{\omega}_{tlj} \sim \text{Multinomial}(y_{tlj}, \boldsymbol{\omega}_{tlj})$, where y_{tlj} plays the role of the sample size parameter of the multinomial distribution, and $\boldsymbol{\omega}_{tlj} = \boldsymbol{\mu}_{tD_{lj}}/\mu_{tlj}$ is the vector of probabilities. Hence, $\boldsymbol{\omega}_{tlj}$ describes how the counts y_{tlj} at subregion (l, j) at time t are expected to be distributed among the descendants in D_{lj}. Because $\boldsymbol{\omega}_{tlj}$ connects coarser to finer resolution levels, in analogy to wavelet analysis (Vidakovic, 1999) we refer to $\boldsymbol{\omega}_{tlj}$ as the spatiotemporal multiscale coefficient. Finally, note that the factorization (2) reparameterizes the Poisson model initially parameterized by the mean process at the finest level $\boldsymbol{\mu}_{tL,1:n_L}$ in terms of the mean process at the coarsest level and the multiscale coefficients that connect the several resolution levels $(\boldsymbol{\mu}_{t1,1:n_1}, \boldsymbol{\omega}_{t1,1:n_1}, \ldots, \boldsymbol{\omega}_{t,L-1,1:n_{L-1}})$.

Let $\mathcal{D}_0$ denote the information available at time $t = 0$ and recursively define $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{\mathbf{y}_t\}$ to be the information available up to time t. We follow the terminology used by West and Harrison (1997) for describing the propagation of information through time. Specifically, the distribution of μ_{tlj} given the data up to time $t-1$, $p(\mu_{tlj}|\mathcal{D}_{t-1})$, is referred to as the prior distribution. After incorporating the data observed at time t, $p(\mu_{tlj}|\mathcal{D}_t)$ is referred to as the posterior or filtered distribution. Finally, the distribution of μ_{tlj} conditional on the data observed at all time points, $p(\mu_{tlj}|\mathcal{D}_T)$, is referred to as the smoothing distribution.

2.2 Temporal evolution

This section describes the temporal evolution of the mean process at the coarsest resolution level and of the spatiotemporal multiscale coefficients at the various resolution levels.

At the coarsest level $l = 1$, we have the aggregated observations $\mathbf{y}_{t1,1:n_1} = (y_{t11}, \ldots, y_{t1n_1})$. Because of the additive property of the Poisson distribution, it follows from Equation (1)

that the aggregated observations also have a Poisson distribution, that is

$$y_{t1j}|\mu_{t1j} \sim \text{Poisson}(\mu_{t1j}),\ j = 1, \ldots, n_1. \tag{3}$$

We are interested in making inference for the relative risk at region $(1, j)$ at time t which is defined as $\lambda_{t1j} = \mu_{t1j}/e_{t1j}$, where e_{t1j} is the expected number of counts at region $(1, j)$ and time t. The expected count e_{t1j} might be assumed known or estimated using covariates by $e_{t1j} = \exp\{\mathbf{x}'_{tj}\boldsymbol{\beta}_j\}$, where $\mathbf{x}_{tj}$ is a vector of covariates and $\boldsymbol{\beta}_j$ is a vector of unknown regression coefficients. For the stochastic temporal evolution of the coarse level relative risk λ_{t1j}, $j = 1, \ldots, n_1$, we assume the beta temporal evolution first proposed by Smith and Miller (1986) in the context of state-space models for univariate Poisson observations. Other state-space models for Poisson observations could be used as well, such as for example dynamic generalized linear models (West et al., 1985; Ferreira and Gamerman, 2000). However, here we are concerned with speed and scalability of computations and the model of Smith and Miller (1986) leads to fast conjugate-analysis filters and samplers. The following definition adapts the model of Smith and Miller (1986) to our context.

Definition 2.1 (Beta evolution for coarse level risk) *Let* $\eta_{tj}|\mathcal{D}_{t-1}, \gamma_j \sim Beta(\gamma_j a_{t-1,j}, (1-\gamma_j)a_{t-1,j})$, *where* $0 < \gamma_j \leq 1$ *is a discount factor parameter, and* $a_{t-1,j} > 0$. *Then, the beta temporal evolution for* λ_{t1j} *is defined as*

$$\lambda_{t1j} = \lambda_{t-1,1j}\gamma_j^{-1}\eta_{tj}. \tag{4}$$

From the first and second moments of the beta distribution, we obtain $E(\gamma_j^{-1}\eta_{tj}|\mathcal{D}_{t-1}, \gamma_j) = 1$ and $Var(\gamma_j^{-1}\eta_{tj}|\mathcal{D}_{t-1}, \gamma_j) = (\gamma_j^{-1} - 1)/(a_{t-1,j} + 1)$. Thus, the closer γ_j is to one and the larger the value of $a_{t-1,j}$, the closer λ_{t1j} will be to $\lambda_{t-1,1j}$. Further, as we discuss in Section 3.1, a_{tj} summarizes the information up to time t about λ_{t1j}. In addition, the discount factor γ_j provides the rate of information loss from time $t-1$ to time t about the risk for the coarse level subregion $(1, j)$. Finally, the discount factor γ_j will be estimated thus providing data-adaptive smoothing for the underlying relative risk process λ.

Now let us consider the temporal evolution of the spatiotemporal multiscale coefficients at levels of resolution $l = 1, \ldots, L-1$. Recall that $\boldsymbol{\omega}_{tlj} = \boldsymbol{\mu}_{tD_{lj}}/\mu_{tlj}$ is the spatiotemporal multiscale coefficient that connects the mean process at subregion (l, j) to the mean process at its descendants at time t. Further, let $\boldsymbol{\omega}^e_{tlj} = \mathbf{y}_{tD_{lj}}/y_{tlj}$ be an estimator of $\boldsymbol{\omega}_{tlj}$. We refer to $\boldsymbol{\omega}^e_{tlj}$ as the empirical spatiotemporal multiscale coefficient for subregion (l, j) at time t. The multiscale decomposition given in Equation (2) implies for $\mathbf{y}_{tD_{lj}} = y_{tlj}\boldsymbol{\omega}^e_{tlj}$ the model

$$y_{tlj}\boldsymbol{\omega}^e_{tlj}|y_{tlj}, \boldsymbol{\omega}_{tlj} \sim \text{Multinomial}(y_{tlj}, \boldsymbol{\omega}_{tlj}). \tag{5}$$

Definition 2.2 below describes our novel Dirichlet temporal evolution for the spatiotemporal multiscale coefficient $\boldsymbol{\omega}_{tlj}$. Let $\odot$ denotes the Hadamard product (p. 45, Magnus and Neudecker, 1999), that is, the operation that returns the vector of element-wise products. As we establish in Theorem 3.3, Definition 2.2 implies an integration of information forward in time based on the Dirichlet distribution.

Definition 2.2 (Dirichlet evolution for spatiotemporal multiscale coefficients) *Let* $\boldsymbol{\phi}_{tlj} = (\phi_{tlj1}, \ldots, \phi_{tlj,d_{lj}})'$, *where* $\phi_{tlj1}, \ldots, \phi_{tlj,d_{lj}}$ *are independent with* $\phi_{tlji} \sim Beta(\delta_{lj}c_{t-1,lji}, (1-\delta_{lj})c_{t-1,lji})$, $S_{tlj} = \boldsymbol{\phi}'_{tlj}\boldsymbol{\omega}_{t-1,lj}$, $0 < \delta_{lj} \leq 1$ *is a discount factor parameter, and* $c_{t-1,lji} > 0$, $i = 1, \ldots, d_{lj}$. *Then, the Dirichlet temporal evolution for* $\boldsymbol{\omega}_{tlj}$ *is defined as*

$$\boldsymbol{\omega}_{tlj} = \frac{1}{S_{tlj}}\boldsymbol{\phi}_{tlj} \odot \boldsymbol{\omega}_{t-1,lj}. \tag{6}$$

As it will become clear in Section 3.1, $\mathbf{c}_{tlj} = (c_{tlj1}, \ldots, c_{tlj,d_{lj}})$ summarizes the information up to time t about $\boldsymbol{\omega}_{tlj}$. Moreover, the discount factor δ_{lj}, which will be estimated from the data, provides the rate of information loss from time $t-1$ to time t about the spatiotemporal multiscale coefficient for subregion (l, j). Other state-space models for multinomial observations could be used for the evolution of $\boldsymbol{\omega}_{tlj}$, such as for example the conditionally Gaussian dynamic models of Cargnoni et al. (1997). This would allow the inclusion of covariates at multiple scales of resolution, and variable selection in that context may be performed using spike and slab priors. Research on inclusion of covariates at multiple scales of resolution is

proceeding and will be reported elsewhere. However, here we prefer our Dirichlet temporal evolution that leads to a conjugate-analysis-based forward filter and backward sampler that is computationally fast and scalable.

2.3 Initial conditions and priors for the discount factors

We complete our proposed MSSTP model for Poisson areal data by specifying the priors for the discount factors and for the latent process at time $t = 0$.

The priors for the latent process at time $t = 0$ represent the initial information about the process before obtaining any observation. For the latent mean process at the coarsest level of resolution λ_{01j}, $j = 1, \ldots, n_1$, we assume conditionally independent conjugate gamma prior distributions $\lambda_{01j}|\mathcal{D}_0 \sim \text{Gamma}(a_{0j}, b_{0j})$, $j = 1, \ldots, n_1$, where $a_{0j} > 0$ and $b_{0j} > 0$ are known and represent the prior information at time $t = 0$ about λ_{01j}. The prior mean and variance of λ_{01j} are a_{0j}/b_{0j} and a_{0j}/b_{0j}^2 respectively. Typically, a_{0j} and b_{0j} are chosen to be small values to indicate vague prior information.

For the latent spatiotemporal multiscale coefficient for subregion (l, j) at time $t = 0$, $\boldsymbol{\omega}_{0lj}$, $j = 1, \ldots, n_j$, we assume conditionally independent conjugate Dirichlet prior distributions $\boldsymbol{\omega}_{0lj}|\mathcal{D}_0 \sim \text{Dirichlet}(\mathbf{c}_{0lj})$, where $\mathbf{c}_{0lj}$ is a known vector with elements $c_{0lji} > 0, i = 1, \ldots, d_{lj}$. The vector $\mathbf{c}_{0lj}$ represents the prior information about $\boldsymbol{\omega}_{0lj}$. Typically, the elements of $\mathbf{c}_{0lj}$ are chosen to be small values to indicate vague prior information.

Similarly to discount factors in Gaussian state-space models (West and Harrison, 1997), the discount factors in our proposed model describe the loss of information through time before the next observation is obtained. The discount factors are parameters that may assume values in the interval $(0, 1)$. As we discuss in Section 3, a discount factor close to 1 indicates a smaller rate of loss of information through time and therefore a more stable latent process. Because discount factors belong to the interval $(0, 1)$, we assume for them beta prior distributions. Specifically, we assume *a priori* $\gamma_j \sim Beta(a_\gamma, b_\gamma), j = 1, \ldots, n_1$, and $\delta_{lj} \sim Beta(a_\delta, b_\delta), l = 1, \ldots, L-1, j = 1, \ldots, n_l$. Typical choices for the hyperparameters

of the discount factor priors are $a_\gamma = b_\gamma = 1$ and $a_\delta = b_\delta = 1$ which imply noninformative uniform priors on the interval $(0, 1)$.

3 Posterior Analysis

In this section we develop simulation-based Bayesian posterior analysis for our proposed MSSTP model for Poisson data. The first two sections, 3.1 and 3.2, consider the case when the discount factors and regression coefficients are fixed. Specifically, Section 3.1 presents results for temporal filtering, that is, the updating of information through time for λ_{t1j} and $\boldsymbol{\omega}_{tlj}$. In addition, Section 3.2 presents results for spatiotemporal smoothing, that is, the integration of spatiotemporal information backward in time. For the case when the discount factors and the regression coefficients are unknown, Section 3.3 presents their marginal posterior distributions. Finally, Section 3.4 introduces algorithms for the implementation of simulation-based full Bayesian posterior analysis.

The following theorem states that the posterior analysis of our MSSTP model may be broken down into many independent parts. As a result, the analysis may be implemented with a divide-and-conquer strategy that leads to a highly parallelizable algorithm.

Theorem 3.1 *Consider the MSSTP model defined by Equations (1), (2), (3), (4), (5), and (6). Given the discount factor parameters* $\gamma_j, j = 1, \ldots, n_1$, *and* $\delta_{lj}, l = 1, \ldots, L-1, j = 1, \ldots, n_l$, *the vectors* $\boldsymbol{\lambda}_{1:T,11}, \ldots, \boldsymbol{\lambda}_{1:T,1n_1}, \boldsymbol{\omega}_{1:T,11}, \ldots, \boldsymbol{\omega}_{1:T,1n_1}, \ldots, \boldsymbol{\omega}_{1:T,L-1,1}, \ldots, \boldsymbol{\omega}_{1:T,L-1,n_{L-1}}$ *are conditionally independent a posteriori.*

In what follows, we present results and algorithms for the simulation of the parameters $\gamma_1, \ldots, \gamma_{n_1}$, $\delta_{11}, \ldots, \delta_{1n_1}$, $\ldots$, $\delta_{L-1,1}, \ldots, \delta_{L-1,n_{L-1}}$, $\boldsymbol{\lambda}_{1:T,11}, \ldots, \boldsymbol{\lambda}_{1:T,1n_1}$, $\boldsymbol{\omega}_{1:T,11}, \ldots,$ $\boldsymbol{\omega}_{1:T,1n_1}$, $\ldots$, $\boldsymbol{\omega}_{1:T,L-1,1}, \ldots,$ $\boldsymbol{\omega}_{1:T,L-1,n_{L-1}}$ from their posterior distribution.

3.1 Temporal filtering

This section presents results on temporal filtering for $\boldsymbol{\lambda}_{1:T,1j}$ and $\boldsymbol{\omega}_{1:T,lj}$. First, let us consider filtering for $\boldsymbol{\lambda}_{1:T,1j}$. As explained in Section 2.3, at time $t = 0$ we assume the distribution $\lambda_{01j}|\mathcal{D}_0 \sim \text{Gamma}(a_{0j}, b_{0j})$. Then, Theorem 3.2 provides the filtering equations for $\boldsymbol{\lambda}_{1:T,1j}$.

Theorem 3.2 *Assume the initial distribution $\lambda_{01j}|\mathcal{D}_0 \sim Gamma(a_{0j}, b_{0j})$ and consider the Observation Equation (3) and the beta temporal evolution for λ_{t1j} given by Equation (4) in Definition 2.1. Then, for $t = 1, \ldots, T$:*

(i) Posterior for $\lambda_{t-1,1j}$: $\lambda_{t-1,1j}|\mathcal{D}_{t-1}, \gamma_j, \boldsymbol{\beta}_j \sim Gamma(a_{t-1,j}, b_{t-1,j})$.

(ii) Prior for λ_{t1j}: $\lambda_{t1j}|\mathcal{D}_{t-1}, \gamma_j, \boldsymbol{\beta}_j \sim Gamma(a_{t|t-1,j}, b_{t|t-1,j})$,
where $a_{t|t-1,j} = \gamma_j a_{t-1,j}$ and $b_{t|t-1,j} = \gamma_j b_{t-1,j}$.

(iii) Posterior for λ_{t1j}: $\lambda_{t1j}|\mathcal{D}_t, \gamma_j, \boldsymbol{\beta}_j \sim Gamma(a_{tj}, b_{tj})$,
where $a_{tj} = \gamma_j a_{t-1,j} + y_{t1j}$ and $b_{tj} = \gamma_j b_{t-1,j} + e_{t1j}$.

Hence, a_{tj} and b_{tj}, $t = 1, \ldots, T$, are recursively computed based on Theorem 3.2 and summarize the information about λ_{t1j} up to time t. Note that the prior mean $E(\lambda_{t1j}|\mathcal{D}_{t-1}, \gamma_j)$ is equal to $E(\lambda_{t-1,1j}|\mathcal{D}_{t-1}, \gamma_j)$ while the prior variance $V(\lambda_{t1j}|\mathcal{D}_{t-1}, \gamma_j)$ is equal to $\gamma_j^{-1} V(\lambda_{t-1,1j}|\mathcal{D}_{t-1}, \gamma_j)$. Thus, the discount factor parameter γ_j inflates the variance from time $t-1$ to time t by a factor equal to γ_j^{-1}. Therefore, values of γ_j closer to one imply a smaller rate of information loss through time and therefore a more stable latent process. Finally, in the limit when $\gamma_j \to 1$ the latent process $\{\lambda_{t1j}\}$ would be constant through time.

From the posterior distribution $\lambda_{t1j}|\mathcal{D}_t, \gamma_j, \beta_j$, it follows that an estimator of λ_{t1j} at time t is given by the filtered posterior mean $E(\lambda_{t1j}|\mathcal{D}_t, \gamma_j, \boldsymbol{\beta}_j) = (\gamma_j a_{t-1,j} + y_{t1j})/(\gamma_j b_{t-1,j} + e_{t1j})$ and a measure of uncertainty is given by the filtered posterior variance $Var(\lambda_{t1j}|\mathcal{D}_t, \gamma_j, \boldsymbol{\beta}_j) = (\gamma_j a_{t-1,j} + y_{t1j})/(\gamma_j b_{t-1,j} + e_{t1j})^2$.

Now let us consider filtering for $\boldsymbol{\omega}_{1:T,lj}$. First, recall from Section 2.3 that the initial distribution for $\boldsymbol{\omega}_{0lj}$ at time $t = 0$ is $\boldsymbol{\omega}_{0lj}|\mathcal{D}_0 \sim \text{Dirichlet}(\mathbf{c}_{0lj})$. Then, Theorem 3.3 provides the filtering equations for $\boldsymbol{\omega}_{1:T,lj}$.

Theorem 3.3 *Assume the initial distribution* $\boldsymbol{\omega}_{0lj}|\mathcal{D}_0 \sim$ *Dirichlet*$(\mathbf{c}_{0lj})$*, and consider the Observation Equation (5) and the Dirichlet evolution for the spatiotemporal multiscale coefficient* $\boldsymbol{\omega}_{tlj}$ *given by Equation (6) in Definition 2.2. Then, for* $t = 1, \ldots, T$*:*

(i) Posterior for $\boldsymbol{\omega}_{t-1,lj}$*:* $\quad \boldsymbol{\omega}_{t-1,lj}|\mathcal{D}_{t-1}, \delta_{lj} \sim$ *Dirichlet*$(\mathbf{c}_{t-1,lj})$*.*

(ii) Prior for $\boldsymbol{\omega}_{tlj}$*:* $\quad \boldsymbol{\omega}_{tlj}|\mathcal{D}_{t-1}, \delta_{lj} \sim$ *Dirichlet*$(\mathbf{c}_{t|t-1,lj})$*, where* $\mathbf{c}_{t|t-1,lj} = \delta_{lj}\mathbf{c}_{t-1,lj}$*.*

(iii) Posterior for $\boldsymbol{\omega}_{tlj}$*:* $\quad \boldsymbol{\omega}_{tlj}|\mathcal{D}_{t}, \delta_{lj} \sim$ *Dirichlet*$(\mathbf{c}_{tlj})$*, where* $\mathbf{c}_{tlj} = \delta_{lj}\mathbf{c}_{t-1,lj} + y_{tlj}\boldsymbol{\omega}^e_{tlj}$*.*

Hence, $\mathbf{c}_{tlj}$, $t = 1, \ldots, T$, are recursively computed based on Theorem 3.3 and summarize the information up to time t about the spatiotemporal multiscale coefficient $\boldsymbol{\omega}_{tlj}$. From Theorem 3.3, it follows that $E(\boldsymbol{\omega}_{tlj}|\mathcal{D}_{t-1}, \delta_{lj}) = E(\boldsymbol{\omega}_{t-1,lj}|\mathcal{D}_{t-1}, \delta_{lj}) = \mathbf{c}_{t-1,lj}/\ (\mathbf{1}'_{d_{lj}}\mathbf{c}_{t-1,lj})$, where d_{lj} is the number of descendants of subregion (l, j). Thus, conditional on the data up to time $t-1$ the expected value of $\boldsymbol{\omega}_{tlj}$ is equal to that of $\boldsymbol{\omega}_{t-1,lj}$. Moreover, the prior covariance matrix of $\boldsymbol{\omega}_{tlj}$ is

$$V(\boldsymbol{\omega}_{tlj}|\mathcal{D}_{t-1}, \delta_{lj}) = \delta_{lj}^{-1}\frac{\mathbf{1}'_{d_{lj}}\mathbf{c}_{t-1,lj} + 1}{\mathbf{1}'_{d_{lj}}\mathbf{c}_{t-1,lj} + \delta_{lj}^{-1}}V(\boldsymbol{\omega}_{t-,lj}|\mathcal{D}_{t-1}, \delta_{lj}).$$

Hence, similarly to γ_j, the discount factor δ_{lj} impacts the information flow through time by inflating the covariance matrix of $\boldsymbol{\omega}_{tlj}$ by a factor that, for large values of $\mathbf{1}'_{d_{lj}}\mathbf{c}_{t-1,lj}$, is approximately equal to δ_{lj}^{-1}. Therefore, values of δ_{lj} closer to one imply a smaller rate of information loss through time and therefore a more stable latent spatiotemporal multiscale coefficient. Finally, from Theorem 3.3 it follows that an estimator of $\boldsymbol{\omega}_{tlj}$ at time t is given by the filtered or posterior mean $E(\boldsymbol{\omega}_{tlj}|\mathcal{D}_t, \delta_{lj}) = (\delta_{lj}\mathbf{c}_{t-1,lj} + y_{tlj}\boldsymbol{\omega}^e_{tlj})/(\delta_{lj}\mathbf{1}'_{d_{lj}}\mathbf{c}_{t-1,lj} + y_{tlj})$.

3.2 Temporal smoothing

This section presents results on temporal smoothing for $\boldsymbol{\lambda}_{1:T,1j}$ and $\boldsymbol{\omega}_{1:T,lj}$. First, let us consider smoothing for $\boldsymbol{\lambda}_{1:T,1j}$.

Proposition 3.1 *Assume the Observation Equation (3) and the beta temporal evolution for* λ_{t1j} *given by Equation (4) in Definition 2.1. Then, the conditional smoothing distribution of* $\lambda_{t-1,1j}$ *given* λ_{t1j} *is equal to*

$$\begin{aligned} p(\lambda_{t-1,1j}|\mathcal{D}_T, \lambda_{t1j}, \gamma_j) &= \frac{b_{t-1,j}^{(1-\gamma_j)a_{t-1,j}}}{\Gamma((1-\gamma_j)a_{t-1,j})}(\lambda_{t-1,1j} - \gamma_j\lambda_{t1j})^{(1-\gamma_j)a_{t-1,j}-1} \\ &\quad \times \exp\{-b_{t-1,j}(\lambda_{t-1,1j} - \gamma_j\lambda_{t1j})\}, \qquad \lambda_{t-1,1j} - \gamma_j\lambda_{t1j} > 0. \end{aligned} \tag{7}$$

We use Proposition 3.1 to develop a backward sampling scheme for λ_{t1j} that is described in Algorithm 3.1.

Let us now consider smoothing for the spatiotemporal multiscale coefficient $\boldsymbol{\omega}_{1:T,lj}$. Recall from Equation (6) that the auxiliary variable S_{tlj} is defined as $S_{tlj} = \sum_{i=1}^{d_{lj}} \phi_{tlji}\omega_{t-1,lji}$. Then, the following proposition provides a way to simulate from the joint distribution of S_{tlj} and $\boldsymbol{\omega}_{t-1,lj}$.

Proposition 3.2 *Consider the Observation Equation (5) and the Dirichlet temporal evolution for the spatiotemporal multiscale coefficient* $\boldsymbol{\omega}_{tlj}$ *given by Equation (6) in Definition 2.2. Then,*

(i) $\boldsymbol{\omega}_{t-1,lj}|\mathcal{D}_T, S_{tlj}, \boldsymbol{\omega}_{tlj}, \delta_{lj} \sim$ *Mod-Dirichlet*$((1-\delta_{lj})\mathbf{c}_{t-1,lj}, S_{tlj}\boldsymbol{\omega}_{tlj})$

(ii) $S_{tlj}|\mathcal{D}_T, \boldsymbol{\omega}_{tlj}, \delta_{lj} \sim \text{Beta}(\delta_{lj}\tilde{c}_{t-1,lj}, (1-\delta_{lj})\tilde{c}_{t-1,lj})$

where $\tilde{c}_{t-1,lj} = \sum_{i=1}^{d_{lj}} c_{t-1,lji}$ *and Mod-Dirichlet denotes the modified Dirichlet distribution defined in Definition A.1.*

We use Proposition 3.2 to develop a backward sampling scheme for $\boldsymbol{\omega}_{1:T,lj}$ that is described in Algorithm 3.2.

3.3 Marginal posterior distributions for hyperparameters

This section presents the marginal posterior distributions for the discount factors $\boldsymbol{\gamma}_{1:n_1}$, $\boldsymbol{\delta}_{1,1:n_1}, \ldots, \boldsymbol{\delta}_{L-1,1:n_{L-1}}$, and for the regression coefficients $\boldsymbol{\beta}_{1:n_1}$.

At the coarsest level $l = 1$, the Poisson-gamma model implies a negative binomial predictive distribution for the coarse level observations. The likelihood function for the discount factor γ_j and the regression coefficients $\boldsymbol{\beta}_j$, $j = 1, \ldots, n_1$ is

$$\begin{aligned} p(\mathbf{y}_{1:T,1j}|\gamma_j, \boldsymbol{\beta}_j, \mathcal{D}_0) &= \prod_{t=1}^{T} p\left(y_{t1j}|\mathcal{D}_{t-1}, \gamma_j, \boldsymbol{\beta}_j\right) \\ &= \prod_{t=1}^{T} \int_0^\infty p\left(y_{t1j}|\gamma_j, \boldsymbol{\beta}_j, \lambda_{t1j}\right) p\left(\lambda_{t1j}|\mathcal{D}_{t-1}, \gamma_j, \boldsymbol{\beta}_j\right) d\lambda_{t1j} \\ &= \prod_{t=1}^{T} \left\{ \frac{\Gamma(\gamma_j a_{t-1,j} + y_{t1j})}{\Gamma(\gamma_j a_{t-1,j})\Gamma(y_{t1j}+1)} \left(\frac{\gamma_j b_{t-1,j}}{e^{x'_{tj}\boldsymbol{\beta}_j}}\right)^{\gamma_j a_{t-1,j}} \right. \\ &\qquad \left. \left(1 + \frac{\gamma_j b_{t-1,j}}{e^{x'_{tj}\boldsymbol{\beta}_j}}\right)^{-(\gamma_j a_{t-1,j} + y_{t1j})} \right\}. \end{aligned} \tag{8}$$

Therefore, by Bayes Theorem the marginal posterior distribution for $(\gamma_j, \boldsymbol{\beta}_j)$ is given by $p(\gamma_j, \boldsymbol{\beta}_j|\mathcal{D}_T) \propto p(\mathbf{y}_{1:T,1j}|\mathcal{D}_0, \gamma_j, \boldsymbol{\beta}_j)p(\gamma_j, \boldsymbol{\beta}_j)$.

The multinomial-Dirichlet model implies that the likelihood function for the finer resolution discount factors δ_{lj}, $l = 1, \ldots, L-1$, $j = 1, \ldots, n_l$, is given by

$$\begin{aligned} p(\mathbf{y}_{1:T,D_{lj}}|\mathcal{D}_0, \mathbf{y}_{1:T,lj}, \delta_{lj}) &= \prod_{t=1}^{T} p(\mathbf{y}_{t,D_{lj}}|\mathcal{D}_{t-1}, y_{tlj}, \delta_{lj}) \\ &= \prod_{t=1}^{T} \int p(\mathbf{y}_{t,D_{lj}}|y_{tlj}, \boldsymbol{\omega}_{tlj}) p(\boldsymbol{\omega}_{tlj}|\mathcal{D}_{t-1}, \delta_{lj}) d\boldsymbol{\omega}_{tlj} \\ &= \prod_{t=1}^{T} \left\{ \frac{\Gamma(\sum_i \delta_{lj} c_{t-1,lji})\Gamma(y_{tlj}+1)}{\Gamma(\sum_i \delta_{lj} c_{t-1,lji} + y_{tlj})} \right. \\ &\qquad \left. \prod_{i=1}^{d_{lj}} \frac{\Gamma(\delta_{lj} c_{t-1,lji} + y_{tlj}\omega^e_{tlji})}{\Gamma(y_{tlj}\omega^e_{tlji}+1)\Gamma(\delta_{lj} c_{t-1,lji})} \right\}. \end{aligned} \tag{9}$$

Therefore, by Bayes Theorem the marginal posterior distribution for δ_{lj} is $p(\delta_{lj}|\mathcal{D}_T) \propto p(\mathbf{y}_{1:T,D_{lj}}|\mathcal{D}_0, \mathbf{y}_{1:T,lj}, \delta_{lj})p(\delta_{lj})$.

Because we assume $a_0 > 0$, $b_0 > 0$, and $c_0 > 0$ (see Section 2.3), the densities that appear in Equations (8) and (9) are always proper. In that case, the products in those equations should start from $t = 1$. However, if a practitioner decides to use improper priors with $a_0 = 0$

and $b_0 = 0$, then the products in Equation (8) have to start from the first time τ for which $p\left(\lambda_{\tau 1 j}|\mathcal{D}_{\tau-1}, \gamma_j, \boldsymbol{\beta}_j\right)$ is a proper density. Similarly, if the practitioner uses improper priors with $c_0 = 0$ then the products in Equation (9) have to start from the first time τ for which $p(\boldsymbol{\omega}_{\tau l j}|\mathcal{D}_{\tau-1}, \delta_{lj})$ is a proper density.

3.4 Posterior simulation

This section presents algorithms for the simulation from the joint posterior distribution of the parameters of our MSSTP model for Poisson data defined by Equations (1), (2), (3), (4), (5), and (6). Note that as a result of Theorem 3.1, simulation from the joint posterior distribution may be broken down into many independent parts. Specifically, simulation of $(\gamma_1, \boldsymbol{\beta}_1, \boldsymbol{\lambda}_{1:T,1})$, ..., $(\gamma_{n_1}, \boldsymbol{\beta}_{n_1}, \boldsymbol{\lambda}_{1:T,n_1})$, $(\delta_{1,1}, \boldsymbol{\omega}_{1:T,1,1})$, ..., $(\delta_{1,n_1}, \boldsymbol{\omega}_{1:T,1,n_1})$, ..., $(\delta_{L-1,1}, \boldsymbol{\omega}_{1:T,L-1,1})$, ..., $(\delta_{L-1,n_{L-1}}, \boldsymbol{\omega}_{1:T,L-1,n_{L-1}})$, may be performed in parallel. As a consequence, each draw from the posterior distribution has computations of the order $\mathcal{O}(T\{n_1 + \sum_{l=1}^{L-1}\sum_{j=1}^{n_l} d_{lj}\})$, which is typically substantially smaller than the computations of the order $\mathcal{O}(Tn_L^3)$ associated with a spatiotemporal model based on state-space equations with a full system equation covariance matrix. Finally, the use of a parallel machine with number of nodes larger than the number of spatiotemporal multiscale coefficients will have computational time of the order $\mathcal{O}(Tmax_{l,j}d_{lj})$.

At the coarsest level of resolution, sampling from $p(\gamma_j, \boldsymbol{\beta}_j, \boldsymbol{\lambda}_{1:T,1j}|\mathcal{D}_T)$, $j = 1, \ldots, n_1$, may be performed through a composite sampling algorithm. First, we sample $(\gamma_j, \boldsymbol{\beta}_j)$ from the marginal posterior distribution $p(\gamma_j, \boldsymbol{\beta}_j|\mathcal{D}_T)$ using a Metropolis-Hastings within Gibbs algorithm. After that, for the draws of $(\gamma_j, \boldsymbol{\beta}_j)$ obtained after convergence, we sample $\boldsymbol{\lambda}_{1:T,1}$ from $p(\boldsymbol{\lambda}_{1:T,1}|\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathcal{D}_T)$. This latter part is performed with a forward filter backward sampler scheme that we describe below.

Algorithm 3.1 *(Forward Filter Backward Sampler for $\boldsymbol{\lambda}_{1:T,j}$). To sample from $p(\boldsymbol{\lambda}_{1:T,j}|\mathcal{D}_T, \gamma_j, \boldsymbol{\beta}_j)$, $j = 1, \ldots, n_1$, we proceed as follows.*

1. *Recursively compute the filtering distributions* $\lambda_{t1j}|\mathcal{D}_t, \gamma_j, \boldsymbol{\beta}_j \sim Gamma(a_{tj}, b_{tj})$, $t = 1, \ldots, T$, *as given in Theorem 3.2.*

2. *Sample* λ_{T1j} *from* $\lambda_{T1j}|\mathcal{D}_T, \gamma_j, \boldsymbol{\beta}_j \sim Gamma(a_{Tj}, b_{Tj})$.

3. *For* t *from* T *to* 2, *recursively sample* $\lambda_{t-1,1j}$ *from* $p(\lambda_{t-1,1j}|\mathcal{D}_T, \lambda_{t1j}, \gamma_j)$ *which is presented in Proposition 3.1:*

   - *Sample* u *from* $Gamma((1-\gamma_j)a_{t-1,j}, b_{t-1,j})$;
   - *Set* $\lambda_{t-1,1j} = u + \gamma_j \lambda_{t1j}$.

At the l-th level of resolution, sampling from $p(\delta_{lj}, \boldsymbol{\omega}_{1:T,lj}|\mathcal{D}_T)$, $l = 1, \ldots, L-1$, $j = 1, \ldots, n_l$, may be accomplished with a composite sampling algorithm. Because δ_{lj} is one-dimensional and its prior distribution has compact support, we find that a straightforward sampling-importance-resampling algorithm works well in simulating from the marginal posterior distribution of δ_{lj}. After that, for the draws of δ_{lj} obtained after convergence, we sample $\boldsymbol{\omega}_{1:T,lj}$ from $p(\boldsymbol{\omega}_{1:T,lj}|\boldsymbol{\delta}_{lj}, \mathcal{D}_T)$. This latter part is performed with the novel forward filter backward sampler scheme that we describe below.

Algorithm 3.2 *(Forward Filter Backward Sampler for* $\boldsymbol{\omega}_{1:T,lj}$*).* *To sample from* $p(\boldsymbol{\omega}_{1:T,lj}|\mathcal{D}_T, \delta_{lj})$, $l = 1, \ldots, L-1$, $j = 1, \ldots, n_l$, *we proceed as follows.*

1. *Recursively compute the filtering distributions* $\boldsymbol{\omega}_{tlj}|\mathcal{D}_t, \delta_{lj} \sim Dirichlet(\mathbf{c}_{tlj})$, $t = 1, \ldots, T$, *as given in Theorem 3.3.*

2. *Sample* $\boldsymbol{\omega}_{Tlj}$ *from* $\boldsymbol{\omega}_{Tlj}|\mathcal{D}_T, \delta_{lj} \sim Dirichlet(\mathbf{c}_{Tlj})$.

3. *For* t *from* T *to* 2, *recursively sample* $\boldsymbol{\omega}_{t-1,lj}$ *from* $p(\boldsymbol{\omega}_{t-1,lj}|\mathcal{D}_T, S_{t-1,lj}, \boldsymbol{\omega}_{tlj}, \boldsymbol{\omega}^e_{1:T,lj}, \delta_{lj})$ *which is a Mod-Dirichlet*$((1-\delta_l)\mathbf{c}_{t-1,lj}, s\boldsymbol{\omega}_{tlj})$*:*

   - *Sample* S_{tlj} *from* $p(S_{tlj}|\mathcal{D}_T, \boldsymbol{\omega}_{tlj}, \boldsymbol{\omega}^e_{1:T,lj}, \delta_{lj})$ *which is a* $Beta(\delta_{lj}\tilde{c}_{t-1,lj}, (1-\delta_{lj})\tilde{c}_{t-1,lj})$ *as presented in Proposition 3.2;*

- *Sample* $\mathbf{u}$ *from Dirichlet*$((1-\delta_{lj})\mathbf{c}_{t-1,lj})$*;*
- *Set* $\boldsymbol{\omega}_{t-1,lj} = S_{tlj}\boldsymbol{\omega}_{tlj} + (1-S_{tlj})\mathbf{u}$.

Frequently, we will be interested in making inference for the expected number of counts or for the relative risk at subregion (l, j) at time t for resolution levels finer than the coarsest resolution. Note that the algorithms in this section will provide a sample from the posterior distribution of $(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\lambda}_{1:T,1}, \ldots, \boldsymbol{\lambda}_{1:T,n_1}, \boldsymbol{\omega}_{1:T,11}, \ldots, \boldsymbol{\omega}_{1:T,1n_1}, \ldots, \boldsymbol{\omega}_{1:T,L-1,1}, \ldots, \boldsymbol{\omega}_{1:T,L-1,n_{L-1}})$. Then, a sample from the posterior distribution of the expected number of counts at the coarsest resolution level can be obtained using the equality $\mu_{t1j} = \lambda_{t1j}e_{t1j}$, $t = 1, \ldots, T$, $j = 1, \ldots, n_1$. After that, a sample from the posterior distribution of the expected number of counts at the various resolution levels can be obtained recursively with the equality $\boldsymbol{\mu}_{tD_{lj}} = \boldsymbol{\omega}_{tlj}\mu_{tlj}$, $t = 1, \ldots, T$, $l = 1, \ldots, L-1$, $j = 1, \ldots, n_l$. In addition, draws from the posterior distribution of $\mathbf{e}_{tlj}$ can be easily obtained as functions of the simulated values of $\boldsymbol{\beta}$. Let $\oslash$ denote the elementwise division operator. Finally, a sample from the posterior distribution of the relative risk at the descendants of subregion (l, j) at time t can be obtained with the equality $\boldsymbol{\lambda}_{tD_{lj}} = \boldsymbol{\mu}_{tD_{lj}} \oslash \mathbf{e}_{tD_{lj}}$.

4 Spatiotemporal dependence structure

In this section we present the spatiotemporal dependence structure implied by the MSSTP model we propose in Section 2. For simplicity of exposition, in this section we assume $e_{t1j} = 1$ implying that $\mu_{t1j} = \lambda_{t1j}$. In addition, we slightly abuse notation and denote by $\log \mathbf{v}$ the vector that results from applying the logarithm to each element of vector $\mathbf{v}$. First we show that, in the spirit of Bayesian dynamic models (West and Harrison, 1997), our MSSTP model implies a spatiotemporal mean process at all resolution levels that is nonstationary. In addition, to understand the spatial dependence structure, we provide results on the mean vector and covariance matrix of $\log \boldsymbol{\mu}_{t1,1:n_1}$, $\log \boldsymbol{\omega}_{tlj}$, and $\log \boldsymbol{\mu}_{tl,1:n_l}$ conditional on $\mathcal{D}_t$, the information up to time t. Further, in terms of spatiotemporal structure we show that the log

mean process at the finest resolution level may be partitioned in two components: a common temporal mean to all subregions that share the same coarsest level ancestor; and dynamic spatiotemporal random effects that may be affected by distinct dynamics at specific regions at multiple resolution levels. Finally, to better understand the spatiotemporal dependence structure of our MSSTP model, we compare the implied dependence structure with that of a widely used spatiotemporal model.

Let us consider the logarithm of the mean process. Note that Equation (4) implies the temporal evolution $\log \mu_{t1j} = \log \mu_{t-1,1j} + \log(\eta_{tj}/\gamma_j)$, where $E(\eta_{tj}/\gamma_j) = 1$ and $Var(\eta_{tj}/\gamma_j) = (1-\gamma_j)/\{\gamma_j(a_{t-1,j}+1)\}$. Hence, the temporal evolution at the coarsest level of $\log \mu_{t1j}$ is a random walk. In addition, Equation (6) implies the temporal evolution $\log \boldsymbol{\omega}_{tlj} = \log \boldsymbol{\omega}_{t-1,lj} + \log(\boldsymbol{\phi}_{tlj} \ / S_{tlj})$, where $S_{tlj} = \boldsymbol{\omega}'_{t-1,lj}\boldsymbol{\phi}_{tlj}$ guarantees that $\mathbf{1}'_{d_{lj}}\boldsymbol{\omega}_{tlj} = 1$. Thus, $\log \boldsymbol{\omega}_{tlj}$ follows a nonGaussian singular random walk (compare to Gaussian singular random walks used by Ferreira et al., 2011). Note that the log-mean process may be disaggregated from level l to level $l+1$ with the disaggregation equation $\log \boldsymbol{\mu}_{tD_{lj}} = \mathbf{1}_{d_{lj}} \log \mu_{tlj} + \log \boldsymbol{\omega}_{tlj}$, $l = 1, \ldots, L-1$. Because the temporal evolution of both the log-mean at the coarsest level and the log spatiotemporal multiscale coefficients are random walks, the implied spatiotemporal mean process at all resolution levels is nonstationary.

In addition, the disaggregation equation may be expanded to relate all the subregions at levels $l-1$ and l as

$$\log \boldsymbol{\mu}_{tl,1:n_l} = \mathbf{B}_{l-1} \log \boldsymbol{\mu}_{t,l-1,1:n_{l-1}} + \log \boldsymbol{\omega}_{t,l-1,1:n_{l-1}}, \tag{10}$$

where $\mathbf{B}_{l-1}$ is an $n_{l-1} \times n_l$ block-diagonal matrix with jth diagonal block equal to a vector of ones with length $d_{D_{l-1,j}}$ and off-block-diagonal elements equal to zero. Even though our model is nonstationary, the vector of expectations and the covariance matrix of the log-mean process at time t conditional on $\mathcal{D}_t$, the information up to time t, are well defined. In particular, Equation (10) may be used to recursively compute the vector of expectations and the covariance matrix of the log-mean process at time t at all resolution levels. Let $\psi(a) = \frac{d}{da}\log\Gamma(a)$ and $\psi'(a) = \frac{d}{da}\psi(a)$ be the digamma and trigamma functions, respectively.

We provide the recursive formulas for expectations and covariance matrices in the following theorem.

Theorem 4.1 *Consider the MSSTP model for Poisson data defined by Equations (1), (2), (3), (4), (5), and (6). Then:*

(i) $E(\log \boldsymbol{\mu}_{t1,1:n_1}|\mathcal{D}_t, \boldsymbol{\gamma}) = (\psi(a_{t1}), \ldots, \psi(a_{tn_1}))' - \log(b_{t1}, \ldots, b_{tn_1})'$.

(ii) $Cov(\log \boldsymbol{\mu}_{t1,1:n_1}|\mathcal{D}_t, \boldsymbol{\gamma}) = \text{diag}(\psi'(a_{t1}), \ldots, \psi'(a_{tn_1}))$.

(iii) $E(\log \boldsymbol{\omega}_{tlj}|\mathcal{D}_t, \boldsymbol{\delta}) = (\psi(c_{tlj,1}), \ldots, \psi(c_{tlj,d_{lj}}))' - \psi(\mathbf{1}'_{d_{lj}}\mathbf{c}_{tlj})\mathbf{1}_{d_{lj}}$.

(iv) $Cov(\log \boldsymbol{\omega}_{tlj}|\mathcal{D}_t, \boldsymbol{\delta}) = \text{diag}(\psi'(c_{tlj,1}), \ldots, \psi'(c_{tlj,d_{lj}})) - \psi'(\mathbf{1}'_{d_{lj}}\mathbf{c}_{tlj})\mathbf{1}_{d_{lj}}\mathbf{1}'_{d_{lj}}$.

(v) $E(\log \boldsymbol{\mu}_{tl,1:n_l}|\mathcal{D}_t, \boldsymbol{\gamma}, \boldsymbol{\delta}) = \mathbf{B}_{l-1}E(\log \boldsymbol{\mu}_{t,l-1,1:n_{l-1}}|\mathcal{D}_t, \boldsymbol{\gamma}, \boldsymbol{\delta}) + E(\log \boldsymbol{\omega}_{t,l-1,1:n_{l-1}}|\mathcal{D}_t, \boldsymbol{\delta})$.

(vi) $Cov(\log \boldsymbol{\mu}_{tl,1:n_l}|\mathcal{D}_t, \boldsymbol{\gamma}, \boldsymbol{\delta}) = \mathbf{B}_{l-1}E(\log \boldsymbol{\mu}_{t,l-1,1:n_{l-1}}|\mathcal{D}_t, \boldsymbol{\gamma}, \boldsymbol{\delta})\mathbf{B}'_{l-1} + Cov(\log \boldsymbol{\omega}_{t,l-1,1:n_{l-1}}|\mathcal{D}_t, \boldsymbol{\delta})$,
with $Cov(\log \boldsymbol{\omega}_{t,l-1,1:n_{l-1}}|\mathcal{D}_t, \boldsymbol{\delta}) = \text{diag}(Cov(\log \boldsymbol{\omega}_{t,l-1,1}|\mathcal{D}_t, \boldsymbol{\delta}), \ldots, Cov(\log \boldsymbol{\omega}_{t,l-1,n_{l-1}}|\mathcal{D}_t, \boldsymbol{\delta}))$.

Theorem 4.1 may be used to obtain the covariance matrix of the log mean process at the finest resolution level L. For that purpose, note that Equation (10) may be recursively applied to obtain the mean process at the finest level in terms of the mean process at the coarsest level and the spatiotemporal multiscale coefficients. For $l_2 \geq l_1$, let $\mathbf{A}_{l_1:l_2} = \prod_{k=l_1}^{l_2} \mathbf{B}_k$, that is, $\{\mathbf{A}_{l_1:l_2}\}_{ij} = 1$ if the ith subregion at level $l_2 + 1$ is a descendant of the jth subregion at level l_1, and $\{\mathbf{A}_{l_1:l_2}\}_{ij} = 0$ otherwise. Moreover, for $l_2 = l_1$ let $\mathbf{A}_{l_1:l_2} = \mathbf{I}_{n_{l_2+1}}$. Then, we may write the log mean process at the finest resolution level L as

$$\log \boldsymbol{\mu}_{tL,1:n_L} = \mathbf{A}_{1:(L-1)} \log \boldsymbol{\mu}_{t1,1:n_1} + \sum_{l=1}^{L-1} \mathbf{A}_{(l+1):(L-1)} \log \boldsymbol{\omega}_{tl,1:n_l}. \tag{11}$$

Hence, spatiotemporal random effects of finest level subregions that share the same ancestor at resolution level l, $l = 1, \ldots, L-1$, will be impacted in the same manner by the corresponding lth level spatiotemporal multiscale coefficient. Therefore, Theorem 4.1 implies that

the covariance matrix of the log mean process at the finest resolution level L conditional on $(\mathcal{D}_t, \boldsymbol{\gamma}, \boldsymbol{\delta})$ is

$$\begin{aligned} Cov(\log \boldsymbol{\mu}_{tL,1:n_L} | \mathcal{D}_t, \boldsymbol{\gamma}, \boldsymbol{\delta}) &= \mathbf{A}_{1:(L-1)} \text{diag}(\psi'(a_{t1}), \ldots, \psi'(a_{tn_1})) \mathbf{A}'_{1:(L-1)} \\ &\quad + \sum_{l=1}^{L-1} \mathbf{A}_{(l+1):(L-1)} \text{diag}_{j=1,\ldots,n_l} \Big(\text{diag}(\psi'(c_{tlj,1}), \ldots, \psi'(c_{tlj,d_{lj}})) \\ &\qquad - \psi'(\mathbf{1}'_{d_{lj}} \mathbf{c}_{tlj}) \mathbf{1}_{d_{lj}} \mathbf{1}'_{d_{lj}} \Big) \mathbf{A}'_{(l+1):(L-1)}. \end{aligned} \tag{12}$$

To better understand the covariance structure implied by Equation (12), Figure 1 presents the posterior spatial correlation function for the log mean process at the finest resolution level for different choices of the discount factor parameters. This correlation function was computed at time $t = T = 200$ for data simulated from a model with 8 levels of resolution in a one-dimensional spatial domain where each parent region has two children subregions. The discount factor parameters were parameterized as $\delta_{lj} = 1 - ae^{-bl}$. Hence, in this specification the discount factor parameters decrease for higher resolution levels. Figure 1 presents correlation functions for the four combinations of $a = 1, 0.5$ and $b = 1, 2$. Note that ae^{-b} is the magnitude of the change in the discount factors from one resolution level to the next. A smaller value of b implies smaller values for the discount factors at higher resolutions. That in turn leads to less smooth processes that are consistent, as shown in Figure 1, with the spatial correlation function decreasing faster for $b = 1$ than for $b = 2$. Further, the correlation function is a step function with step sizes that depend not only on the discount factors but also on the number of resolution levels. More resolution levels would lead to an ability to model more details and the spatial correlation function would have smaller steps. Finally, even for a distance of 250 there is still substantial spatial correlation. Therefore, our MSSTP model is able to accommodate spatial long range dependence.

Another important consequence of Equation (11) is that the spatiotemporal evolution of

the log mean process at the finest resolution level L may be written as

$$\begin{aligned}\log \boldsymbol{\mu}_{t+1,L,1:n_L} &= \mathbf{A}_{1:(L-1)} \left\{\log \boldsymbol{\mu}_{t1,1:n_1} + \log(\boldsymbol{\eta}_{t+1} \oslash \boldsymbol{\gamma}_{1:n_1})\right\} \\ &\quad + \sum_{l=1}^{L-1} \mathbf{A}_{(l+1):(L-1)} \left\{\log \boldsymbol{\omega}_{tl,1:n_l} + \log(\boldsymbol{\phi}_{t+1,lj}/S_{t+1,lj})\right\}. \qquad (13)\end{aligned}$$

In the right hand side of Equation (13), the first term shows the temporal evolution of the temporal mean that is common to all subregions at the finest level that share the same ancestor at the coarsest level. The second term shows the spatiotemporal evolution of

$$\sum_{l=1}^{L-1} \mathbf{A}_{(l+1):(L-1)} \log \boldsymbol{\omega}_{tl,1:n_l}. \qquad (14)$$

Hence, $\sum_{l=1}^{L-1} \mathbf{A}_{(l+1):(L-1)} \log \boldsymbol{\omega}_{tl,1:n_l}$ is the vector of dynamic spatiotemporal random effects at the finest resolution level. Note that these spatiotemporal random effects at the finest resolution level are impacted by the multiscale coefficients of ancestral nodes at all resolution levels. Therefore, our MSSTP model implies at the finest resolution level a spatiotemporal evolution composed of two components: a common temporal mean to all subregions that share the same coarsest level ancestor; and dynamic spatiotemporal random effects that may be affected by distinct dynamics at multiple resolution levels.

To better understand the spatiotemporal dependence structure, we compare our MSSTP model to a widely used spatiotemporal model based on conditional autoregressions (STCAR) for Poisson data. Specifically, in the STCAR model let y_{tj} be the observed number of counts at the finest resolution level at time t and subregion (L, j) and assume $y_{tj}|\mu_{tj} \sim Poisson(\mu_{tj})$, with log link $\log \mu_{tj} = \alpha_0 + a_j + b_t$, $j = 1, \ldots, n_L$, $t = 1, \ldots, T$. For this model, the common mean follows a random walk $b_t = b_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, W)$. In addition, the spatial random effects $a_1, \ldots, a_{n_L}$ follow a proper conditional autoregressive (CAR) structure (Besag, 1974; Ferreira and de Oliveira, 2007) that assumes for the vector of spatial random effects $\mathbf{a} = (a_1, \ldots, a_{n_L})'$ conditional distributions $a_j|a_{i\neq j} \sim N(\bar{a}_j, \{(m_j + d)\tau_c\}^{-1})$, where $\bar{a}_j = (m_j + d)^{-1} \sum_{i\in\partial_j} a_i$, ∂_j is the set of neighbors of region j, d is a propriety parameter as implemented in the R package INLA (www.r-inla.org), τ_c is a precision parameter, and m_j

is the number of neighbors of region j. Note that in this STCAR model the spatial random effects do not vary through time.

A particular case of our MSSTP model that has a structure similar to that of the STCAR model is the degenerate case when all the discount factors for the multiscale coefficients are equal to 1, that is, $\delta_{lj} = 1$ for all $l = 1, \ldots, L-1, j = 1, \ldots, n_l$. In that case, all the multiscale coefficients are fixed through time, that is, $\boldsymbol{\omega}_{tlj} = \boldsymbol{\omega}_{0lj}$, $t = 1, \ldots, T$, for all $l = 1, \ldots, L-1, j = 1, \ldots, n_l$. As a result, the spatiotemporal random effects given in Equation (14) are fixed through time and play the role of the spatial random effects $\mathbf{a}$ from the STCAR model. However, an important difference is that instead of a CAR prior that assumes reasonably smooth spatial random effects, the spatial random effects in our MSSTP model follow a multiscale spatial prior that may accommodate sharp transitions among neighboring subregions. This multiscale spatial prior is similar to wavelet basis priors that have been successfully used to account for spatial dependence in functional magnetic resonance imaging (Flandin and Penny, 2007; Sanyal and Ferreira, 2012). Further, let us assume that there is only one subregion at the coarsest level. Then, in this particular case of our MSSTP model the log-mean at the coarsest level $\log \mu_{t1j}$ corresponds to $\alpha_0 + b_t$ in the STCAR model. Therefore, in the degenerate case when all the discount factors for the spatiotemporal multiscale coefficients are equal to 1, our MSSTP model has a log-mean spatiotemporal process that is the sum of a common temporal log-mean that follows a random walk and spatial random effects that follow a multiscale spatial prior. Finally, because the $\boldsymbol{\omega}_{tlj}$'s allocate intensity in a multiscale manner to the subregions at the several resolution levels, in the general case $\delta_{lj} \in (0, 1)$ controls how much this intensity allocation varies through time. Therefore, the dynamic spatiotemporal random effects in our MSSTP model may incorporate distinct dynamics not only at the several resolution levels but also at specific regions at each resolution level.

To the best of our knowledge, our MSSTP model does not nest other more traditional spatiotemporal models. However, as shown in the two applications in Section 5, our model

has the capacity to perform much better than CAR-based spatiotemporal models. We think that CAR-based models may be more adequate in cases when the underlying spatial random effects are fairly smooth, and the smoothness is about the same in the entire region under study. In contrast, our MSSTP will be more adequate when the spatial random effects are not very smooth or in the case when the smoothness varies across the region of interest.

5 Applications

In this section we illustrate the usefulness of our MSSTP methodology with two applications. Section 5.1 examines mortality ratios in the state of Missouri and illustrates the case when there is a known offset in the model, which in this application is the population size. Section 5.2 considers tornado reports in the American Midwest and illustrates the case when there is a regressor that at each time point is common to all regions. Finally, Section 5.3 presents a comparison of our MSSTP model for Poisson data to two competing models. The first competing model assumes Poisson temporal evolution for each subregion at the finest resolution level. The second competing model is the CAR-based model described in Section 4.

In all applications presented in this section, we have used the same default prior distributions. For the initial distributions $\lambda_{01j}|\mathcal{D}_0$ and $\boldsymbol{\omega}_{0lj}|\mathcal{D}_0$, we assume $a_{0j} = b_{0j} = 0.01$ and $\mathbf{c}_{0lj} = 0.01\mathbf{1}_{d_{lj}}$, that imply vague prior information. In addition, for the priors for the discount factors, we assume $a_\gamma = b_\gamma = 1$ and $a_\delta = b_\delta = 1$ which imply noninformative uniform priors on the interval $(0, 1)$. Application of these default priors to a simulated data set (not shown) leads our estimation methodology to recover the true values of the parameters. In addition, we have performed a sensitivity study that has shown that results are not very sensitivity to the choice of prior hyperparameters. Finally, we find that the default priors proposed here work well in our two real data applications.

5.1 Mortality in Missouri

This section presents an application of our MSSTP methodology to a data set of mortality per county in the State of Missouri. The data set was downloaded from the website www.dhss.mo.gov/DeathMICA/indexcounty.html of the Missouri Department of Health and Senior Services. Specifically, here we consider the annual total number of deaths in the age group from 45 to 64 years old from 1990 to 2009. The finest resolution level in this application is the county level. Moreover, we assume the multiscale structure estimated by Ferreira et al. (2011) for the state of Missouri. Specifically, we assume 3 levels of resolution ($l = 1, 2, 3$) with a total of 115 counties at the finest level of resolution as presented in Figure 2. The total number of deaths per county per year range from 0 to 1537, and several counties have annual number of deaths less than 10. Because of these low counts, the use of a Gaussian approximation would not be adequate. Thus, we apply here our MSSTP methodology for Poisson data. In this application, y_{tLj} is the number of deaths occurred in county j in year t. Moreover, we assume $y_{tLj}|\lambda_{tLj} \sim \text{Poisson}(\lambda_{tLj}e_{tLj})$, where e_{tLj} is the population size divided by 100,000 in county j in year t, and λ_{tLj} is the risk of death per 100,000 inhabitants in county j in year t.

We advocate the analysis of the discount factors as a powerful way to identify regions with spatiotemporal dynamics that warrant further investigation. For example, Figure 3 presents the posterior densities for the discount factors γ_j, $\delta_{1j}, j = 1, \ldots, n_1$, and $\delta_{2j}, j = 1, \ldots, n_2$. In particular, Figure 3(a) presents the posterior densities for the discount factor parameters for the risk at the coarsest level. Subregion (1,1) has the smallest discount factor and therefore the least smooth temporal evolution with faster changes in the risk level through time. In addition, Figure 3(b) presents the posterior densities for the discount factor parameters that relate the coarsest level subregions with their descendants at the intermediate level. In Figure 3(b), subregion (1,1) has again the smallest discount factor which indicates that its descendants might have large changes in mortality risk through time. Furthermore, Figure 3(c) presents the posterior density of discount factors for each descendant of subregion (1,1)

at the intermediate resolution level. In particular, in Panel (c) the left-most posterior density is for $\delta_{2,6}$ indicating that the descendants of subregion $(2, 6)$, that is, counties within the Saint Louis metropolitan area, have the most temporal change in relative importance. To further investigate this issue, Figure 4 presents the estimated risk per 100,000 inhabitants for the 45 to 64 years old age group for counties within the Saint Louis metropolitan area. Several counties have decreasing risk and, in particular, panel (j) shows a substantial and steady rate of decrease in risk through time for the City of Saint Louis.

Finally, Figure 5 presents maps of estimated risk per county of the State of MIssouri per 100,000 inhabitants for the 45 to 64 years old age group. Specifically, Figure 5 presents the maps of observed standardized mortality ratio (left panels), posterior medians (central panels), and standard errors (right panels) for years 1993, 1997, 2001, 2005, and 2009. As intuitively expected, the standard errors decrease with county population size. Moreover, because the observations have Poisson distribution, the standard errors increase with county mean risk level. In addition, the standard errors are mostly below 100, and typically around 50. Hence, the use of color-coded intervals of length 94 for the observed and fitted maps allows for meaningful interpretation.

Analysis of the fitted maps shows that our MSSTP methodology is able to provide fitted maps much smoother than the observed maps. At the same time, our methodology respects sharp transitions between neighboring regions. For example, the fitted maps preserve the much higher risk of the City of Saint Louis when compared with Saint Louis County. From an epidemiological perspective, the fitted maps show that the northern counties have lower death rates than the southern counties. In particular, several counties in the southeast part of Missouri have persistently high death rates.

5.2 Tornadoes in the American Midwest

This section presents a MSSTP analysis of annual tornado report data from 1953 to 2010 in the United States. Tornadoes may be classified in the Fujita scale rating system from

F1 (weaker) to F5 (stronger). Here we consider the more intense tornadoes from F3 to F5. For each tornado we consider only the year and the location of the first touchdown. The region of interest is the tornado alley which is the area with the highest incidence of tornadoes in the United States. Specifically, here we consider the rectangle region defined by the north-most boundary of South Dakota, the west-most boundary of South Dakota, the south-most boundary of Mississippi and the east-most boundary of Michigan, as presented in Figure 6. Furthermore, we consider three spatial resolution levels: a coarse resolution with 4 subregions, an intermediate resolution with 16 subregions, and a fine resolution with 64 subregions. Thus, y_{tLj} is the number of tornado touchdowns in year t in the j-th subregion at the finest resolution level.

To account for the relationship between tornado occurrence and the El Niño and La Niña phenomena (Monfredo, 1999; Wikle and Anderson, 2003), we assume $y_{tLj}|\lambda_{tLj}, \beta_{Lj} \sim$ Poisson$(\lambda_{tLj} e_{tLj})$, where $e_{tLj} = \exp(x_t \beta_{Lj})$, x_t is the average Southern Oscillation Index (SOI) for the months of March, April and May in year t, and $\beta_{Lj} = \beta_{A_1(L,j)}$, with $A_1(L, j)$ being the ancestor at resolution level $l = 1$ of subregion (L, j). That is, we assume that the El Niño and La Niña phenomena may have a different impact on each of the four coarse level regions. For simplicity of notation, we denote $\beta_j = \beta_{1j}$, $j = 1, \ldots, 4$. Figure 7(a) presents the posterior densities for β_j, $j = 1, \ldots, 4$, which indicate that increases in the SOI are related to increases in the risk of tornado in subregions (1,2), (1,3), and (1,4) at the coarsest resolution level. In addition, risk of tornado in subregion (1,1) does not seem to be related to the SOI. Thus, its northern position and proximity to the Rocky Mountains seem to isolate subregion (1,1) from the influence of the El Niño and La Niña phenomena. Finally, Figure 7(b) presents the posterior densities of γ_j, $j = 1, \ldots, 4$. These posterior densities are located below 0.6 which indicate that, after accounting for the SOI, there are fast temporal changes in the risk of tornado for all four coarse level regions.

Figure 7(c) presents the posterior densities of δ_{1j}, $j = 1, \ldots, 4$, which indicate that the relative risk of tornado among the descendants of subregions (1,3) and (1,4) may be changing

through time. This suggests that we should investigate in more detail the descendants of subregions (1,3) and (1,4). Figure 8(a) presents the posterior densities for δ_{2j} for $j \in D_{1,3}$, whereas Figure 8(b) presents the posterior densities for δ_{2j} for $j \in D_{1,4}$. On Panel (a), the notably smaller discount factor corresponds to subregion (2,10); specifically, the posterior mean of $\delta_{2,10}$ is equal to 0.81. This indicates that the spatiotemporal multiscale coefficient $\boldsymbol{\omega}_{t,2,10}$ that relates subregion $(2, 10)$ to its descendants changes substantially through time and should be investigated in more detail.

Figure 9 shows the time series plot of posterior median (solid line) and 95% credible interval (dashed line) for each element of $\boldsymbol{\omega}_{t2,10}$. Each element of $\boldsymbol{\omega}_{t2,10}$ corresponds to one descendant of subregion $(2, 10)$. These descendents are $(3, 37)$ (panel a), $(3, 38)$ (panel b), $(3, 39)$ (panel c) and $(3, 40)$ (panel d). For reference, note that if the relative risk was the same for all four descendants then the spatiotemporal multiscale coefficient would have all elements equal to 0.25. Thus, the posterior probability that an element of the spatiotemporal multiscale coefficient is larger than 0.25 provides a measure of whether the corresponding region has relative risk higher or lower than its siblings. Let us call this probability the exceedance probability. As a consequence, two features are worth mention: the first feature is that the exceedance probability for subregion $(3, 39)$ is above 0.9 from 1950 to 1980. That is, during that period subregion $(3, 39)$ had substantially higher risk than its siblings. The second feature is that the relative risks between subregions (3,37), (3,38), (3,39), and (3,40) have been changing substantially since 1980. This may indicate that subregion $(2, 10)$ may be particularly susceptible to the effects of climate change.

Additional evidence of the importance of subregion $(2, 10)$ for studies of climate change is given by considering the number of F5 tornadoes. F5 are the strongest tornadoes and are much less likely to be underreported. Here we consider a comparison of the period from 1950 to 2010 (used in our multiscale analysis) to the period from 2011 to 2014 (not included in our multiscale analysis). During the period from 1950 to 2010 there were 52 F5 tornadoes in the United States and 8 of those, or about 15.4%, occurred in subregion $(2, 10)$. In contrast,

during the period from 2011 to 2014 there were 7 F5 tornadoes in the United States and 3 of those, or 42.8%, occurred in subregion $(2, 10)$. Hence, Earth atmospheric variables conducive to the formation of tornadoes seem to be changing substantially within subregion $(2, 10)$. Therefore, a detailed future study of subregion $(2, 10)$ may shed light on what are these Earth atmospheric variables and how they are being impacted by global warming.

While one of the three tornadoes that occurred in subregion $(2, 10)$ from 2011 to 2014 hit a less populated area, the other two tornadoes hit more densely populated areas. Specifically, these two tornadoes hit Joplin, Missouri, in 2011, and Moore, Oklahoma, in 2013, causing estimated losses of about 4.8 billion dollars and 182 deaths. For reference, the lower right panel of Figure 6 shows the locations of Joplin (red dot) and Moore (blue dot). Therefore, a detailed future study of subregion $(2, 10)$ may not only help identify important Earth atmospheric variables impacted by global warming, but may also help improve tornado alert systems to reduce property and human life losses.

5.3 Model comparison

This section presents a comparison of our MSSTP model to two competing models. First, to access the ability of our multiscale spatiotemporal framework to borrow strength spatially, we compare our model to a model that assumes that each finest level subregion has its own independent temporal evolution (ITE). Second, to access our framework's ability to incorporate spatiotemporal dynamics, we compare our model to the STCAR model described in Section 4. We perform these model comparisons using conditional Bayes factors (e.g., see Ghosh et al., 2006) and the deviance information criterion (DIC) (Spiegelhalter et al., 2002) for both Missouri and tornado data sets.

We have implemented the conditional Bayes factor for both models following the simulation-based approach proposed by Vivar and Ferreira (2009). This approach uses the data observed at the first t^* times points as a training sample. Specifically, let $p_q(\boldsymbol{\mu}_{1:t,L,1:n_L}, \boldsymbol{\gamma}, \boldsymbol{\delta}|\mathcal{D}_t)$ denote the joint posterior density of $\boldsymbol{\mu}_{1:t,L,1:n_L}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$ under model

q given the data up to time t. Then, the one-step-ahead predictive density is

$$p_q(\mathbf{y}_{t+1}|\mathcal{D}_t) = \int p_q(\mathbf{y}_{t+1}|\boldsymbol{\mu}_{t+1,L,1:n_L}, \boldsymbol{\gamma}, \boldsymbol{\delta}) p_q(\boldsymbol{\mu}_{t+1,L,1:n_L}|\boldsymbol{\mu}_{t,L,1:n_L}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \mathcal{D}_t) \times p_q(\boldsymbol{\mu}_{1:t,L,1:n_L}, \boldsymbol{\gamma}, \boldsymbol{\delta}|\mathcal{D}_t) d\boldsymbol{\mu}_{1:(t+1),L,1:n_L} d\boldsymbol{\gamma} d\boldsymbol{\delta}.$$

The posterior simulation algorithms presented in Section 3 can be easily used to estimate the above one-step-ahead predictive density. Then, the joint predictive density of $\mathbf{y}_{t^*+1}, \ldots, \mathbf{y}_T$ can be computed as $p_q(\mathbf{y}_{t^*+1}, \ldots, \mathbf{y}_T|\mathcal{D}_{t^*}) = \prod_{t=t^*+1}^{T} p_q(\mathbf{y}_t|\mathcal{D}_{t-1})$ (e.g., see p. 135 Prado and West, 2010) Finally, the conditional Bayes factor of model q_1 against model q_2 is equal to

$$B_{q_1,q_2} = \frac{p_{q_1}(\mathbf{y}_{t^*+1}, \ldots, \mathbf{y}_T|\mathcal{D}_{t^*})}{p_{q_2}(\mathbf{y}_{t^*+1}, \ldots, \mathbf{y}_T|\mathcal{D}_{t^*})}.$$

Therefore, the conditional Bayes factor uses the one-step-ahead predictive densities and, thus, effectively compares the predictive performance of the competing models.

First, we compare our MSSTP model to an ITE model that assumes that each finest level subregion has its own temporal evolution according to the Poisson state-space model given by the observational equation (1) and the beta evolution equation (4). Here we have used $t^* = 10$ for the Missouri data set and $t^* = 18$ for the tornado report data set. These values for t^* provide stable analyses and allow for each data set many time points to be used in the model comparison. For the data set on mortality ratio in Missouri, the logarithm of the conditional Bayes factor of our MSSTP model against the ITE model is equal to 39.24 which provides strong evidence in favor of our MSSTP model. Further, the DIC of our MSSTP model is 15,120.23 whereas the DIC of the ITE model is 15,145.10. Hence, the DIC agrees with the conditional Bayes factor and chooses our MSSTP model for the Missouri data set. For the tornado report data, the logarithm of the conditional Bayes factor is equal to 326.21, again providing strong evidence in favor of our MSSTP model. In addition, the DIC of our MSSTP model is 7,385.67 and the DIC of the ITE model is 7,690.95. Once again, the DIC agrees with the conditional Bayes factor and chooses our MSSTP model for the tornado data set. Hence, in both applications the conditional Bayes factor and the DIC provide strong evidence of the superiority of our MSSTP model. Because this is a comparison with a model

that does not use spatial information, this result implies that our MSSTP model superiority arises from its ability to borrow strength spatially.

Second, we compare our MSSTP model to the STCAR model considered in Section 4 that assumes Poisson distribution for the observations, and a logarithm link function that connects the mean of each observation to a sum of a temporal trend and spatial random effects. We assume the priors $\phi \sim Beta(1,1)$, $p(\alpha_0) \propto 1$, $\tau_c \sim Ga(0.5, 0.0005)$, $b_0 \sim N(0, 2000)$, $\log d \sim Ga(1,1)$, $W \sim Ga(0.5, 0.0005)$. As in the previous model comparison, in the computation of the conditional Bayes factor we have used $t^* = 10$ for the Missouri data set and $t^* = 18$ for the tornado report data set. The logarithm of the conditional Bayes factor of our MSSTP model against the STCAR model is equal to 10.28 for the data set on mortality ratio in Missouri, and is equal to 84.27 for the tornado report data set. Meanwhile, the DICs for the STCAR model are 15,307.37 for the Missouri data set and 7,874.39 for the tornado data set. Therefore, in both applications the conditional Bayes factor and the DIC provide evidence of the superiority of our MSSTP model to incorporate spatiotemporal dynamics.

A fair question is whether or not our MSSTP model is identifiable. Or in other words, are the conditional Bayes factor and the DIC able to correctly distinguish amongst the three models considered above. To answer this question, we have performed a simulation study where we have simulated 10 data sets from each of the three models. In this simulation study, we have used the multiscale structure of the state of Missouri as depicted in Figure 2 in which $n_0 = 5$, $n_1 = 22$ and $n_2 = 115$. To simulate the data sets from our MSSTP model we have used $\gamma = (0.98, 0.98, 0.95, 0.95, 0.98)$, $\delta_1 = (0.98, 0.98, 0.98, 0.95, 0.95)$ and $\delta_2 = 0.981_{n_1}$. To simulate the data sets from the ITE model, the discount factors for the 115 counties were simulated independently from a uniform distribution on the interval $(0.4, 1)$. To simulate data from the STCAR model, we have assumed $\tau_c = 1$, $d = 1$, and $W = 0.1$. We assess the performance of the conditional Bayes factor and the DIC in terms of the success rate in choosing the correct model. For the three models, the DIC selected the correct model for 100% of the simulated data sets. Meanwhile, the conditional Bayes factor selected the

correct model for 100% of the data sets simulated from the ITE and STCAR models. For data sets simulated from our MSSTP model, the conditional Bayes factor chose the correct model 80% of the time, and in the remaining 20% of the time the conditional Bayes factor selected the STCAR model. Hence, the conditional Bayes factor and the DIC are able to correctly distinguish amongst the three models. Therefore, our MSSTP model is identifiable.

The excellent ability of the DIC and, to a lesser extent of the conditional Bayes factor, to select amongst the three models increases our assurance on the results for the two real data sets favorable to the MSSTP model. Therefore when compared against the ITE and STCAR models, our proposed MSSTP model has superior ability to explain the spatiotemporal behavior of the Missouri and tornado data sets.

6 Conclusions

Our multiscale spatiotemporal framework for Poisson processes has five distinctive features. First, the evolution of the spatiotemporal multiscale coefficients describes changes in relative risk among sibling spatial subregions. Second, our methodology produces estimated risk maps that accomplish a nice balance between the conflicting objectives of smoothing out the noise and at the same time respecting existing sharp transitions between neighbor regions. Third, our framework decomposes large Poisson spatiotemporal processes into many small components that may be analyzed separately. This divide-and-conquer modeling and analysis strategy leads to algorithms that are fast, scalable, and highly parallelizable. Fourth, the analysis of the discount factors may help identify subregions that possess distinctive spatiotemporal dynamics. Finally, as shown in the model comparison in Section 5.3, by borrowing strength spatially and by incorporating spatiotemporal dynamics, our Poisson multiscale spatiotemporal model provides practically useful results.

Our MSSTP framework uses conditional independence assumptions that allow fast and scalable algorithms. However, a fair question is whether these conditional independence assumptions are reasonable. To verify model adequacy, we may perform analysis of residuals

to check if they still exhibit excess dependence. Such excess dependence may indicate some sort of model misspecification, such as for example inadequacy of the assumption that the latent spatiotemporal multiscale coefficients are conditionally independent a priori. Another possible model misspecification is an inadequate number of levels of resolution. Specifically, excess residuals dependence may occur when the model does not include enough fine scale resolutions. We recommend that practitioners use residual analysis to check model adequacy and use conditional Bayes factors to compare competing models.

Spatial smoothing at time t is controlled by the vectors $\mathbf{c}_{tlj}$. To see that, consider the case when $\mathbf{c}_{tlj} = c\mathbf{1}_{d_{lj}}$, where d_{lj} is the number of descendants of subregion (l, j) and c is a scalar. In the limit when $c \to \infty$, the multiscale coefficient $\boldsymbol{\omega}_{tlj} \to d_{lj}^{-1}\mathbf{1}_{d_{lj}}$. That would imply that all descendants of subregion (l, j) would have the same mean, the ultimate spatial smoothness for those subregions. In addition, note that each parent region has its own $\mathbf{c}_{tlj}$, hence this spatial smoothness is adaptive both by scale and location. Further, the $\mathbf{c}_{tlj}$s are computed as given in Theorem 3.3 and thus are data dependent. Therefore, our framework allows adaptive data-dependent spatial smoothness. In the two applications considered here, there are three levels of resolution. In all those three levels, $\mathbf{c}_{tlj}$ is informed by many occurrences of events and the choice of prior for $\boldsymbol{\omega}_{tlj}$ at time 0 is, as shown by a sensitivity study, relatively unimportant. However, in other applications one may have many more levels of resolution with possibly sparse data at the finest resolution levels. For those applications, the spatial smoothness will crucially depend on the prior for $\boldsymbol{\omega}_{tlj}$ at time 0. In those applications, instead of vague priors for $\boldsymbol{\omega}_{tlj}$ at time 0, we may assign level-dependent informative priors. For example, we may assume $\boldsymbol{\omega}_{0lj}|\mathcal{D}_0, \delta_{lj} \sim Dirichlet(\mathbf{c}_{0lj})$ with $\mathbf{c}_{0lj} = c\mathbf{1}_{d_{lj}}$ and $c = \xi_1 \exp(\xi_2 l)$, where ξ_1 and ξ_2 are parameters to be estimated. This and other ways to use informative priors to induce more spatial smoothness will be the subject of future research.

There are several other promising directions for future research. In this paper, we have considered the case when regressor values are common for all regions at each time point and

the regression coefficient is the same for all descendants of each coarsest level subregion. Thus, a promising direction is the extension of our framework to cases when regressors values vary in space and time at all resolution levels. A particularly important extension would be to consider the modifiable areal unit problem (Mugglin and Carlin, 1998; Banerjee et al., 2004; Cressie and Wikle, 2011), that is, the case when regressors values are known at a resolution level finer than the finest resolution level at which the dependent variable is observed. Another promising direction is the extension of our framework to accommodate overdispersion. This can be accomplished by expanding the MSSTP model with gamma distributed overdispersion parameters. The simulation of the overdispersion parameters may be easily incorporated within our MCMC scheme, but that will likely reduce the scalability of our computational approach.

Another promising direction is the development of alternative priors for the discount factors. For example, we may include an additional layer of modeling for the hyperparmeters a_γ, b_γ, a_δ, and b_δ to take scale and location of the respective discount factors into account. These hyperparameters would then be estimated and that would allow borrowing of strength across subregions for the estimation of the discount factors. That development may be of interest when the number of time points is small. Another alternative is the development of mixture priors for discount factors and the associated model selection procedures for pattern identification. As Poisson spatiotemporal data sets become increasingly larger, automatic methods to identify regions of interest will become of paramount importance. In particular, automatic model selection procedures based on mixture priors for the discount factors may identify regions for which there is significant spatiotemporal dynamics. The researcher will then be able to focus on the scientifically relevant most promising regions. These mixture-priors-based model selection procedures may also be used for spatiotemporal data compression. Specifically, these procedures may identify sibling subregions that are homogeneous, and therefore their data may be combined to reduce data set size.

Finally, while we have assumed that the multiscale structure was known a priori, the de-

velopment of methods for the estimation of the multiscale structure is an important future research direction. Note that this includes both estimation of how many resolution levels to include in the model and estimation of the multiscale partition. To perform those tasks, we may borrow ideas from the multiscale spatiotemporal clustering method for Gaussian data of Ferreira et al. (2011), the stochastic search approach for Poisson spatial data of Hui and Bradlow (2012), and the adaptive Pólya tree approach of Ma (2015). We may use these ideas in conjunction with the mixture-priors-based model selection procedures discussed in the above paragraph to estimate the number of resolution levels and the multiscale partition. Note that if the model selection procedure indicates that all spatiotemporal multiscale coefficients at resolutions finer than a certain resolution level are equal to zero, then we do not need to consider those finer resolution levels. Moreover, it may be the case that spatiotemporal multiscale coefficients at a certain resolution level are different than zero only in a certain spatial region. In that case, the model selection procedures will indicate that resolution refinement is needed only for a part of the region of interest. Thus, these model selection procedures will provide a data-driven way to decide how many levels of resolution are necessary to describe the process of interest in different regions. Therefore, this will lead to procedures that will provide data-driven spatially varying adaptive resolution refinement. Work in this area is proceeding and will be reported in the future.

Appendix A: Auxiliary Facts and Theorems

Let $\psi(a) = \frac{d}{da}\log\Gamma(a)$ and $\psi'(a) = \frac{d}{da}\psi(a)$ be the digamma and trigamma functions, respectively. Then:

Auxiliary Fact A.1 *(Johnson et al., 1994) Let* $Y \sim Gamma(a, b)$ *and* $Z = \log Y$. *Then* $E(Z) = \psi(a) - \log b$ *and* $Var(Z) = \psi'(a)$.

Auxiliary Fact A.2 *(Johnson et al., 1995) Let* $Y \sim Beta(a, b)$ *and* $Z = \log Y$. *Then* $E(Z) = \psi(a) - \psi(a + b)$ *and* $Var(Z) = \psi'(a) - \psi'(a + b)$.

Auxiliary Fact A.3 *Let* $\mathbf{Y} = (Y_1, \ldots, Y_d)' \sim$ *Dirichlet*$(c_1, \ldots, c_d)$, $\mathbf{Z} = (Z_1, \ldots, Z_d)' = (\log Y_1, \ldots, \log Y_d)'$, *and* $\tilde{c} = \sum_{j=1}^{d} c_j$. *Then* $E(Z_j) = \psi(c_j) - \psi(\tilde{c})$, $Var(Z_j) = \psi'(c_j) - \psi'(\tilde{c})$, *and* $Cov(Z_i, Z_j) = -\psi'(\tilde{c})$.

Proof: Auxiliary Fact A.3 follows from representing a Dirichlet random vector in terms of independent gamma random variables (Kotz et al., 2000), known relationships between Dirichlet and beta distributions, and direct application of Auxiliary Facts A.1 and A.2. □

Definition A.1 *(Modified Dirichlet distribution). Assume that* $\mathbf{X}$ *is a* k*-dimensional random vector with Dirichlet distribution parametrized by a vector* $\mathbf{c}$ *of positive reals and satisfying* $\sum_{i=1}^{k} X_i = 1$. *Let* $\mathbf{Y} = (\mathbf{b} - \mathbf{a}) \odot \mathbf{X} + \mathbf{a}$, *where* $\mathbf{a}$ *and* $\mathbf{b}$ *are* k*-dimensional vectors of positive reals such that* $\sum_{i=1}^{k} a_i \leq 1$, $a_i \leq b_i$. *Then,* $\mathbf{Y}$ *is said to have a modified Dirichlet distribution with parameters* $\mathbf{c}$, $\mathbf{a}$ *and* $\mathbf{b}$. *Notation:* $\mathbf{Y} \sim$ *Mod-Dirichlet*$(\mathbf{c}, \mathbf{a}, \mathbf{b})$.

Theorem A.1 (Modified Dirichlet density) *If* $\mathbf{Y} \sim$ *Mod-Dirichlet*$(\mathbf{c}, \mathbf{a}, \mathbf{b})$, $\mathbf{a} \in \mathbb{R}^k$, $\mathbf{b} \in \mathbb{R}^k$, $\mathbf{c} \in \mathbb{R}^k$, *then its density is given by*

$$g(\mathbf{y}) = \frac{\Gamma(\sum_{i=1}^{k} c_i)}{\prod_{i=1}^{k} \Gamma(c_i)} \prod_{i=1}^{k} \left(\frac{y_i - a_i}{b_i - a_i} \right)^{c_i - 1} \frac{1}{\prod_{i=1}^{k-1} (b_i - a_i)}, \tag{A.1}$$

with support in the set of k*-dimensional vectors* $\mathbf{y}$ *whose entries* y_i *are real numbers in the interval* (a_i, b_i), $i = 1, \ldots, k$.

Proof: Let $\mathbf{X}$ be a variable with Dirichlet distribution. Define $\mathbf{Y} = (\mathbf{b} - \mathbf{a}) \odot \mathbf{X} + \mathbf{a}$. The Jacobian of the transformation is $|J| = \prod_{i=1}^{k} (b_i - a_i)^{-1}$. Then, by the standard Jacobian transformation method the density of $\mathbf{Y}$ follows. □

Corollary A.1 (Modified Dirichlet moments) *If* $\mathbf{Y} \sim$ *Mod-Dirichlet*$(\mathbf{c}, \mathbf{a}, \mathbf{b})$ *then* $E(\mathbf{Y}) = (\mathbf{b} - \mathbf{a}) \odot \mathbf{c}\tilde{c}^{-1} + \mathbf{a}$, $V(Y_i) = (b_i - a_i)^2 c_i (\tilde{c} - c_i) \{\tilde{c}^2 (\tilde{c} + 1)\}^{-1}$, $i = 1, \ldots, k$, *and* $Cov(Y_i, Y_j) = -(b_i - a_i)(b_j - a_j) c_i c_j \{\tilde{c}^2 (\tilde{c} + 1)\}^{-1}$, $i, j = 1, \ldots, k$, $i \neq j$, *where* $\tilde{c} = \sum_{i=1}^{k} c_i$.

Proof: The results follow directly from the properties of the Dirichlet distribution. In particular, if $\mathbf{X} \sim Dirichlet(\mathbf{c})$ then $E(\mathbf{X}) = \mathbf{c}/\tilde{c}$, $V(X_i) = c_i(\tilde{c} - c_i)/\{\tilde{c}^2(\tilde{c}+1)\}$, and $Cov(X_i, X_j) = -c_i c_j/\{\tilde{c}^2(\tilde{c}+1)\}$, $i \neq j$. □

Definition A.2 *(Beta4 distribution): If $X \sim Beta(\alpha, \beta)$ with shape parameters α and β and a and b are constantes such that $0 \leq a < b \leq 1$ then $Y = (b-a)X + a$ has distribution Beta4 with parameters a, b, α, β. Notation: $Y \sim Beta4(\alpha, \beta, a, b)$.*

Corollary A.2 *If $Y \sim Beta4(\alpha, \beta, a, b)$ then its density is given by*

$$p(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{(y-a)^{\alpha-1}(b-y)^{\beta-1}}{(b-a)^{\alpha+\beta-1}},\ y \in (a,b). \tag{A.2}$$

Note that this is a particular case of the Modified Dirichlet distribution with dimension $k = 2$, and parameters $c_1 = \alpha$, $c_2 = \beta$, $a_1 = a$, $a_2 = 1-b$, $b_1 = b$, and $b_2 = 1-a$.

Appendix B: Proofs of Main Results

Proof of Theorem 3.1. By Bayes' Theorem, the conditional posterior distribution of $\boldsymbol{\lambda}_{1:T,11}, \ldots, \boldsymbol{\lambda}_{1:T,1n_1}$, $\boldsymbol{\omega}_{1:T,11}, \ldots, \boldsymbol{\omega}_{1:T,1n_1}$, $\ldots$, $\boldsymbol{\omega}_{1:T,L-1,1}, \ldots, \boldsymbol{\omega}_{1:T,L-1,n_{L-1}}$, given the discount factor parameters $\gamma_j, j = 1, \ldots, n_1$, and $\delta_{lj}, l = 1, \ldots, L-1, j = 1, \ldots, n_l$, is proportional to the product of the prior densities for $\lambda_{01j}, j = 1, \ldots, n_1$, and $\boldsymbol{\omega}_{0lj}, l = 1, \ldots, L-1, j = 1, \ldots, n_l$, multiplied by the product from time $t = 1$ to time $t = T$ of the multiscale factorized likelihood given by Equation (2), times the transition densities implied by the stochastic temporal evolution equations (4) and (6), that is

$$\left\{\prod_{j=1}^{n_1} p(\lambda_{01j}) \prod_{l=1}^{L-1}\prod_{j=1}^{n_l} p(\boldsymbol{\omega}_{0lj})\right\} \prod_{t=1}^{T}\left\{\prod_{j=1}^{n_1} p(y_{t1j}|\lambda_{t1j}) \prod_{l=1}^{L-1}\prod_{j=1}^{n_l} p(\mathbf{y}_{t,D_{lj}}|y_{tlj}, \boldsymbol{\omega}_{tlj})\right\}$$
$$\times \prod_{t=1}^{T}\left\{p(\lambda_{t1j}|\lambda_{t-1,1j}, \gamma_j) \prod_{l=1}^{L-1}\prod_{j=1}^{n_l} p(\boldsymbol{\omega}_{tlj}|\boldsymbol{\omega}_{t-1,lj}, \delta_{lj})\right\}.$$

After reorganizing the terms, the above expression becomes

$$\prod_{j=1}^{n_1}\left[p(\lambda_{01j})\prod_{t=1}^{T}\{p(y_{t1j}|\lambda_{t1j})p(\lambda_{t1j}|\lambda_{t-1,1j},\gamma_j)\}\right]$$
$$\times\prod_{l=1}^{L-1}\prod_{j=1}^{n_l}\left[p(\boldsymbol{\omega}_{0lj})\prod_{t=1}^{T}\left\{p(\mathbf{y}_{t,D_{lj}}|y_{tlj},\boldsymbol{\omega}_{tlj})p(\boldsymbol{\omega}_{tlj}|\boldsymbol{\omega}_{t-1,lj},\delta_{lj})\right\}\right].$$

Hence, conditional on the discount factors $\gamma_j, j = 1,\ldots,n_1$, and $\delta_{lj}, l = 1,\ldots,L-1, j = 1,\ldots,n_l$, the posterior density can be factorized in a product of terms that involve only $\boldsymbol{\lambda}_{1:T,1j}$ for each $j = 1,\ldots,n_1$, and terms that involve only $\boldsymbol{\omega}_{1:T,lj}$ for each combination of $l = 1,\ldots,L-1$ and $j = 1,\ldots,n_l$, implying conditional independence *a posteriori*. □

Proof of Theorem 3.2 Theorem 3.2 follows directly from the work by Smith and Miller (1986) on state-space models for Poisson data. □

Proof of Theorem 3.3. Proof of Theorem 3.3 is by induction. First, assume the truth of the distribution in (i). Then, the proof of (ii) is as follows. For simplicity of notation and ease of exposition, let $\delta = \delta_{lj}$, $c_i = c_{t-1,lji}$, $n = d_{lj}$, $\phi_i = \phi_{tlji}$, $\boldsymbol{\omega}_t = \boldsymbol{\omega}_{tlj}$, $S = S_{tlj}$. Let $Z_i = \phi_i\omega_{t-1,i}$. Because $\phi_1,\ldots,\phi_n$ are i.i.d. $Beta(\delta c_i,(1-\delta)c_i)$, we have that $Z_1,\ldots,Z_n$ are conditionally independent given $\omega_{t-1,i}$ and $Z_i|\boldsymbol{\omega}_{t-1} \sim Beta4(\delta c_i,(1-\delta)c_i,0,\omega_{t-1,i})$, where $Beta4(\alpha,\beta,a,b)$ is the Beta distribution with 4 parameters from Definition A.2.

Let $\mathbf{Z} = (Z_1,\ldots,Z_n)'$ and $S = \sum_{i=1}^{n} Z_i$. It is straightforward to show that the Jacobian of the transformation from $\mathbf{Z}$ to $(S,\boldsymbol{\omega}_t)$ is s^{n-1}. Thus, it follows from the conditional independence of $Z_1,\ldots,Z_n$ and from Equation A.2 that the joint conditional density of S and $\boldsymbol{\omega}_t$ given $\boldsymbol{\omega}_{t-1}$ is

$$\begin{aligned}
p(s,\boldsymbol{\omega}_t|\boldsymbol{\omega}_{t-1}) &= s^{n-1}\prod_{i=1}^{n}p(z_i|\boldsymbol{\omega}_{t-1}) \\
&= s^{n-1}\prod_{i=1}^{n-1}\frac{\Gamma(c_i)}{\Gamma(\delta c_i)\Gamma((1-\delta)c_i)}\,\frac{(s\omega_{ti})^{\delta c_i-1}(\omega_{t-1,i}-s\omega_{ti})^{(1-\delta)c_i-1}}{\omega_{t-1,i}^{c_i-1}}\times \\
&\quad \frac{\Gamma(c_n)}{\Gamma(\delta c_n)\Gamma((1-\delta)c_n)}\,\frac{[s(1-\sum_{i=1}^{n-1}\omega_{ti})]^{\delta c_n-1}[\omega_{t-1,n}-s(1-\sum_{i=1}^{n-1}\omega_{ti})]^{(1-\delta)c_n-1}}{\omega_{t-1,n}^{c_n-1}},
\end{aligned}$$

with $0 < s < 1$ and $0 < z_i = s\omega_{ti} < \omega_{t-1,i} < 1 - s\sum_{k\neq i}\omega_{tk},\ i = 1, \ldots, n$.

Moreover, $\boldsymbol{\omega}_{t-1}|\mathcal{D}_{t-1}, \delta \sim Dirichlet(\mathbf{c}_{t-1})$. Therefore, the prior density of $\boldsymbol{\omega}_t$ at time $t-1$ is given by

$$\begin{aligned} p(\boldsymbol{\omega}_t|\mathcal{D}_{t-1}, \delta) &= \int_0^1 \left\{ \int_{s\omega_{t1}}^{1-s\sum_{j\neq 1}\omega_{tj}} \cdots \int_{s\omega_{t,n-1}}^{1-s\sum_{j\neq n-1}\omega_{tj}} p(s, \boldsymbol{\omega}_t|\boldsymbol{\omega}_{t-1}) p(\boldsymbol{\omega}_{t-1}|\mathcal{D}_{t-1}, \delta) d\boldsymbol{\omega}_{t-1} \right\} ds \\ &= \kappa_1 \prod_{i=1}^{n} \omega_{ti}^{\delta c_i - 1} \int_0^1 s^{\delta\sum_{i=1}^n c_i - 1} \int_{R_{\boldsymbol{\omega}_{t-1}}} \prod_{i=1}^{n} (\omega_{t-1,i} - s\omega_{ti})^{(1-\delta)c_i - 1} d\boldsymbol{\omega}_{t-1} ds, \end{aligned}$$

where $\omega_{tn} = 1 - \sum_{i=1}^{n-1}\omega_{ti}$ and $\kappa_1 = \Gamma(\sum_{i=1}^n c_i)\{\prod_{i=1}^n \Gamma(\delta c_i)\Gamma((1-\delta)c_i)\}^{-1}$. The integrand in the inner integral is proportional to a modified Dirichlet density with parameters $(1-\delta)\mathbf{c}$, $s\boldsymbol{\omega}_t$, and $1 - s\sum_{j\neq i}\omega_{jt}$. Thus, after solving the inner integral we obtain

$$p(\omega_t|\mathcal{D}_{t-1}, \delta) = \kappa_2 \prod_{i=1}^{n} \omega_{ti}^{\delta c_i - 1} \int_0^1 s^{\delta\sum_{i=1}^n c_i - 1} (1 - s\sum_{i=1}^{n}\omega_{ti})^{(1-\delta)\sum_{i=1}^n c_i - 1} ds,$$

where $\kappa_2 = \Gamma(\sum_{i=1}^n c_i)\{\Gamma((1-\delta)\sum_{i=1}^n c_j)\prod_{i=1}^n \Gamma(\delta c_i)\}^{-1}$. Finally, solving the above integral we obtain

$$p(\boldsymbol{\omega}_t|\mathcal{D}_{t-1}, \delta) = \frac{\Gamma(\delta\sum_{i=1}^n c_i)}{\prod_{i=1}^n \Gamma(\delta c_i)} \prod_{i=1}^{n} \omega_{ti}^{\delta c_i - 1},$$

that implies Item (ii), that is, $\omega_t|\mathcal{D}_{t-1}, \delta \sim Dirichlet(\delta\mathbf{c})$. Finally, (iii) follows from Bayes Theorem and the conjugacy of the Dirichlet and multinomial distributions. □

Proof of Proposition 3.1. Because $\lambda_{t,1j} = \gamma_j^{-1}\lambda_{t-1,1j}\eta_{tj}$, the Jacobian of the transformation from η_{tj} to λ_{t1j} is $\gamma_j\lambda_{t-1,1j}^{-1}$. Hence,

$$p(\lambda_{t1j}|\lambda_{t-1,1j}, \mathcal{D}_{t-1}, \gamma_j) = \frac{\Gamma(a_{t-1,j})\gamma_j\lambda_{t-1,1j}^{-1}}{\Gamma(\gamma_j a_{t-1,j})\Gamma((1-\gamma_j)a_{t-1,j})} \left(\frac{\gamma_j\lambda_{t,1j}}{\lambda_{t-1,1j}}\right)^{\gamma_j a_{t-1,j}-1} \left(1 - \frac{\gamma_j\lambda_{t,1j}}{\lambda_{t-1,1j}}\right)^{(1-\gamma_j)a_{t-1,j}-1}.$$

From the Markovian property assumed for the temporal evolution of λ_{t1j}, it follows that $\lambda_{t-1,1j}$ and $y_{t:T,1j}$ are conditionally independent given λ_{t1j}. Thus, $p(\lambda_{t-1,1j}|\lambda_{t,1j}, \mathcal{D}_T, \gamma_j) = p(\lambda_{t-1,1j}|\lambda_{t,1j}, \mathcal{D}_{t-1}, \gamma_j)$. Moreover, $\lambda_{t-1,1j}|\mathcal{D}_{t-1}, \gamma_j \sim \text{Gamma}(a_{t-1,j}, b_{t-1,j})$ and $\lambda_{t,1j}|\mathcal{D}_{t-1}, \gamma_j \sim \text{Gamma}(\gamma_j a_{t-1,j}, \gamma_j b_{t-1,j})$. Therefore, the result follows by direct application of Bayes theorem:

$$p(\lambda_{t-1,1j}|\lambda_{t,1j}, \mathcal{D}_{t-1}, \gamma_j) = \frac{p(\lambda_{t,1j}|\lambda_{t-1,1j}, \mathcal{D}_{t-1}, \gamma_j)p(\lambda_{t-1,1j}|\mathcal{D}_{t-1}, \gamma_j)}{p(\lambda_{t,1j}|\mathcal{D}_{t-1}, \gamma_j)}.$$ □

Proof of Proposition 3.2. Let $\delta = \delta_{lj}$, $c_i = c_{t-1,lji}$, $n = d_{lj}$, and $\boldsymbol{\omega}_t = \boldsymbol{\omega}_{tlj}$. From proof of Theorem 3.3 we have

$$\begin{aligned} p(s, \boldsymbol{\omega}_t | \boldsymbol{\omega}_{t-1}, \mathcal{D}_{t-1}, \delta) &= s^{n-1} \prod_{i=1}^{n-1} \frac{\Gamma(c_i)}{\Gamma(\delta c_i)\Gamma((1-\delta)c_i)} \frac{(s\omega_{ti})^{\delta c_i - 1}(\omega_{t-1,i} - s\omega_{ti})^{(1-\delta)c_i - 1}}{\omega_{t-1,i}^{c_i - 1}} \times \\ & \frac{[s(1 - \sum_{i=1}^{n-1} \omega_{ti})]^{\delta c_n - 1}[\omega_{t-1,n} - s(1 - \sum_{i=1}^{n-1} \omega_{ti})]^{(1-\delta)c_n - 1}}{\omega_{t-1,n}^{c_n - 1}} \times \\ & \frac{\Gamma(c_n)}{\Gamma(\delta c_n)\Gamma((1-\delta)c_n)} \end{aligned}$$

with $0 < z_i = s\omega_{ti} < \omega_{t-1,i},\ i = 1, \ldots, n$.

Thus,

$$\begin{aligned} p(s, \boldsymbol{\omega}_{t-1} | \boldsymbol{\omega}_t, \mathcal{D}_{t-1}, \delta) &= \frac{p(s, \boldsymbol{\omega}_t, \boldsymbol{\omega}_{t-1} | \mathcal{D}_{t-1}, \delta)}{p(\boldsymbol{\omega}_t | \mathcal{D}_{t-1}, \delta)} = \frac{p(s, \boldsymbol{\omega}_t | \boldsymbol{\omega}_{t-1}, \mathcal{D}_{t-1}, \delta) p(\boldsymbol{\omega}_{t-1,lj} | \mathcal{D}_{t-1}, \delta)}{p(\boldsymbol{\omega}_{tlj} | \mathcal{D}_{t-1}, \delta)} \\ &= \frac{\Gamma(\tilde{c}_{t-1,lj})}{\Gamma(\delta_{lj}\tilde{c}_{t-1,lj})\Gamma((1-\delta_{lj})\tilde{c}_{t-1,lj})} s^{\delta_{lj}\tilde{c}_{t-1,lj} - 1}(1-s)^{(1-\delta_{lj})\tilde{c}_{t-1,lj} - 1} \\ &\times \frac{\Gamma((1-\delta_{lj})\tilde{c}_{t-1,lj})}{\prod_{i=1}^{d_{lj}} \Gamma((1-\delta_l)c_{t-1,lji})} (1-s)^{1 - (1-\delta_{lj})\tilde{c}_{t-1,lj}} \\ &\times \prod_{i=1}^{d_{lj}} (\omega_{t-1,lji} - s\omega_{tlji})^{(1-\delta_{lj})c_{t-1,lji} - 1}, \end{aligned}$$

This is the product of a beta density and a modified Dirichlet density. The result follows. □

Proof of Theorem 4.1. Items (i) and (ii) follow from Theorem 3.2 and Auxiliary Fact A.1. Items (iii) and (iv) follow from Theorem 3.3 and Auxiliary Fact A.3. Item (v) follows from Equation (10) and from the linearity property of expected values. Item (vi) follows directly from Equation (10) and Theorem 3.1. □

Acknowledgments The work of Fonseca was supported in part by a grant from CNPq. The work of Ferreira was supported in part by National Science Foundation Grant DMS-0907064. Part of this research was performed while Ferreira was visiting the Statistical and Applied Mathematical Sciences Institute (SAMSI). We gratefully acknowledge the constructive comments and suggestions made by two anonymous referees and the Associate Editor that led to a substantially improved article.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). "Gaussian predictive process models for large spatial datasets." *Journal of the Royal Statistical Society, Series B*, 70, 825–848.

Berliner, L. M., Wikle, C. K., and Milliff, R. F. (1999). "Multiresolution wavelet analyses in hierarchical Bayesian turbulence models." In *Bayesian Inference in Wavelet Based Models*, eds. P. Müller and B. Vidakovic, 341–359. New York: Springer-Verlag.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. (1995). "Bayesian analysis of space-time variation in disease risk." *Statistics in Medicine*, 14, 2433–2443.

Besag, J. (1974). "Spatial interaction and the statistical analysis of lattice systems (with discussion)." *Journal of the Royal Statistical Society, Series B*, 36, 192–236.

Cargnoni, C., Mller, P., and West, M. (1997). "Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models." *Journal of the American Statistical Association*, 92, 640–647.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Hoboken, New Jersey: Wiley.

Ferreira, M. A. R., Bertolde, A. I., and Holan, S. (2010). "Analysis of economic data with multi-scale spatio-temporal models." In *Handbook of Applied Bayesian Analysis*, eds. A. O'Hagan and M. West. Oxford: Oxford University Press.

Ferreira, M. A. R. and de Oliveira, V. (2007). "Bayesian reference analysis for Gaussian Markov random fields." *Journal of Multivariate Analysis*, 98, 789–812.

Ferreira, M. A. R. and Gamerman, D. (2000). "Dynamic Generalized Linear Models." In *Generalized Linear Models: a Bayesian Perspective*, eds. G. Dey and Mallick, 57–72. New York: Marcel Dekker.

Ferreira, M. A. R., Holan, S. H., and Bertolde, A. I. (2011). "Dynamic Multiscale Spatio-Temporal Models for Gaussian Areal Data." *Journal of the Royal Statistical Society - Series B*, 73, 663–688.

Ferreira, M. A. R. and Lee, H. K. H. (2007). *Multiscale Modeling: A Bayesian Perspective*. Springer Series in Statistics. New York: Springer.

Flandin, G. and Penny, W. D. (2007). "Bayesian fMRI data analysis with sparse spatial basis function priors." *NeuroImage*, 34, 3, 1108–1125.

Fuentes, M. (2007). "Approximate likelihood for large irregularly spaced spatial data." *Journal of the American Statistical Association*, 102, 321–331.

Gamerman, D., Santos, T. R., and Franco, G. C. (2013). "A Non-Gaussian Family of State-Space Models with Exact Marginal Likelihood." *Journal of Time Series Analysis*, 34, 6, 625–645.

Ghosh, J. K., Delampady, M., and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.

Hui, S. K. and Bradlow, E. T. (2012). "Bayesian multi-resolution spatial analysis with applications to marketing." *Quantitative Marketing and Economics*, 10, 4, 419–452.

Johannesson, G., Cressie, N., and Huang, H. (2007). "Dynamic Multi-Resolution Spatial Models." *Environmental and Ecological Statistics*, 14, 5–25.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions, vol. 1*. 2nd ed. New York: Wiley.

— (1995). *Continuous Univariate Distributions, vol. 2*. 2nd ed. New York: Wiley.

Knorr-Held, L. (2000). "Bayesian modelling of inseparable space-time variation in disease risk." *Statistics in Medicine*, 19, 2555–2567.

Kolaczyk, E. D. and Huang, H. (2001). "Multiscale statistical models for hierarchical spatial aggregation." *Geographical Analysis*, 33, 2, 95–118.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions, vol. 1*. 2nd ed. New York: Wiley.

Lemos, R. T. and Sansó, B. (2009). "A spatio-temporal model for mean, anomaly, and trend fields of North Atlantic Sea surface temperature (with discussion)." *Journal of the American Statistical Association*, 104, 5–25.

Ma, L. (2015). "Markov adaptive Polya trees and multi-resolution adaptive shrinkage in nonparametric modeling." *arXiv preprint arXiv:1401.7241*.

Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Revised ed. Chichester: Wiley.

Monfredo, W. (1999). "Relationships between phases of the El Nino-Southern Oscillation and character of the tornado season in the south-central United States." *Physical Geography*, 20, 413–421.

Mugglin, A. S. and Carlin, B. P. (1998). "Hierarchical modeling in Geographic Information Systems: Population interpolation over incompatible zones." *Journal of Agricultural, Biological, and Environmental Statistics*, 3, 111–130.

Paciorek, C. J. and McLachlan, J. S. (2009). "Mapping Ancient Forests: Bayesian Inference for Spatio-Temporal Trends in Forest Composition Using the Fossil Pollen Proxy Record." *Journal of the American Statistical Association*, 104, 608–622.

Prado, R. and West, M. (2010). *Time Series: Modeling, Computation, and Inference*. Boca Raton: CRC Press.

Sanyal, N. and Ferreira, M. A. R. (2012). "Bayesian hierarchical multi-subject multiscale analysis of functional MRI data." *NeuroImage*, 63, 1519–1531.

Schmid, V. and Held, L. (2004). "Bayesian extrapolation of space-time trends in cancer registry data." *Biometrics*, 60, 1034–1042.

Smith, J. Q. (1979). "A generalization of the Bayesian steady forecasting model." *Journal of the Royal Statistical Society, Series B*, 41, 375–387.

— (1981). "The multiparameter steady model." *Journal of the Royal Statistical Society, Series B*, 43, 256–260.

Smith, R. L. and Miller, J. E. (1986). "A Non-Gaussian state space model and application to prediction of records." *Journal of the Royal Statistical Society, Series B*, 48, 79–88.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 4, 583–639.

Tzala, E. and Best, N. (2008). "Bayesian latent variable modelling of multivariate spatio-temporal variation in cancer mortality." *Statistical Methods in Medical Research*, 17, 97–118.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. Wiley.

Vivar, J. C. and Ferreira, M. A. R. (2009). "Spatio-temporal models for Gaussian areal data." *Journal of Computational and Graphical Statistics*, 18, 658–674.

Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997). "Hierarchical spatio-temporal mapping of disease rates." *Journal of the American Statistical Association*, 92, 607–617.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. 2nd ed. Springer-Verlag: New York.

West, M., Harrison, P. J., and Migon, H. S. (1985). "Dynamic generalized linear models and Bayesian forecasting (with discussion)." *Journal of the American Statistical Association*, 80, 73–96.

Wikle, C. K. and Anderson, C. J. (2003). "Climatological analysis of tornado report counts using a hierarchical Bayesian spatio-temporal model." *Journal of Geophysical Research - Atmospheres*, 108 (D24), 9005, doi:10.1029/2002JD002806.

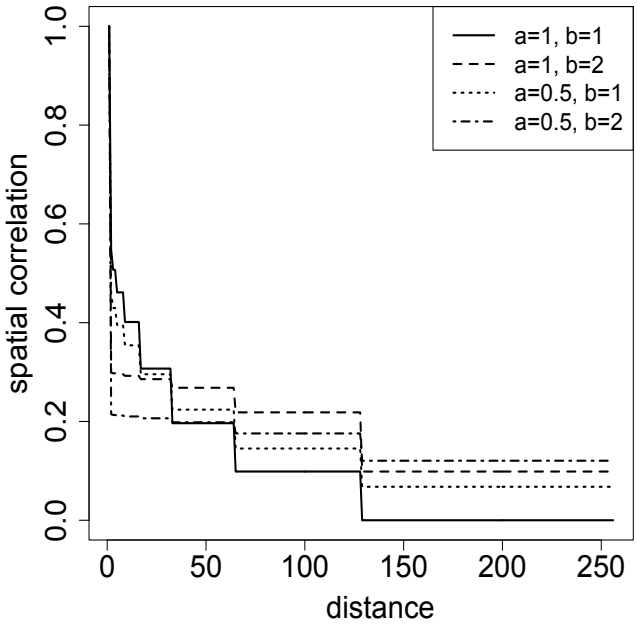
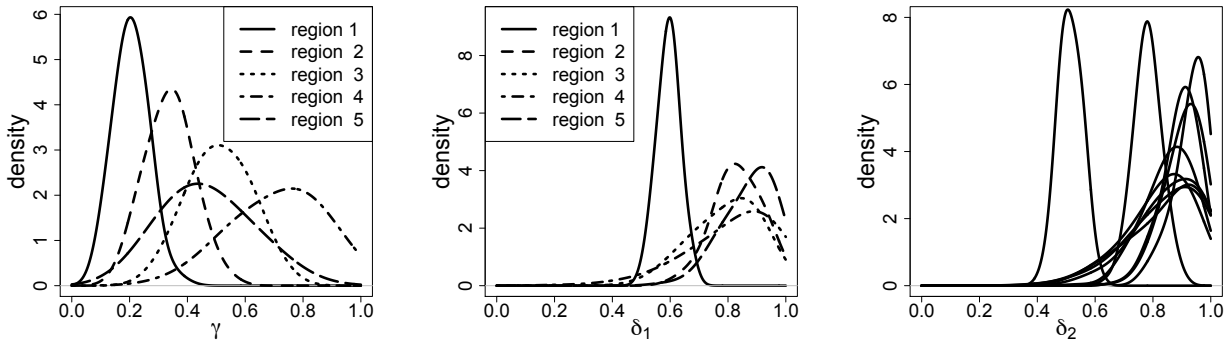


Figure 1: Posterior spatial correlation function for the log mean process at the finest resolution level for different choices of the discount factor parameters. Model with 8 levels of resolution in a one-dimensional spatial domain where each parent region has two children subregions. The discount factor parameters are parameterized as $\delta_{lj} = 1 - ae^{-bl}$ with $a = 1, 0.5$ and $b = 1, 2$.

(a) (b) (c)

Figure 2: Multiscale structure for the state of Missouri (Ferreira et al., 2011) by resolution level: (a) fine (counties), (b) intermediate, and (c) coarse. Panel (b) shows in light grey the Saint Louis metropolitan area (Subregion $(2, 6)$).



(a) (b) (c)

Figure 3: State of Missouri application. Posterior densities for γ_j, $\delta_{1j}, j = 1, \ldots, n_1$, and $\delta_{2j}, j = 1, \ldots, n_2$. In Panel (c) the left-most posterior density is for $\delta_{2,6}$ indicating that the descendants of Subregion $(2, 6)$, that is, counties within the Saint Louis metropolitan area, have the most temporal change in relative importance.

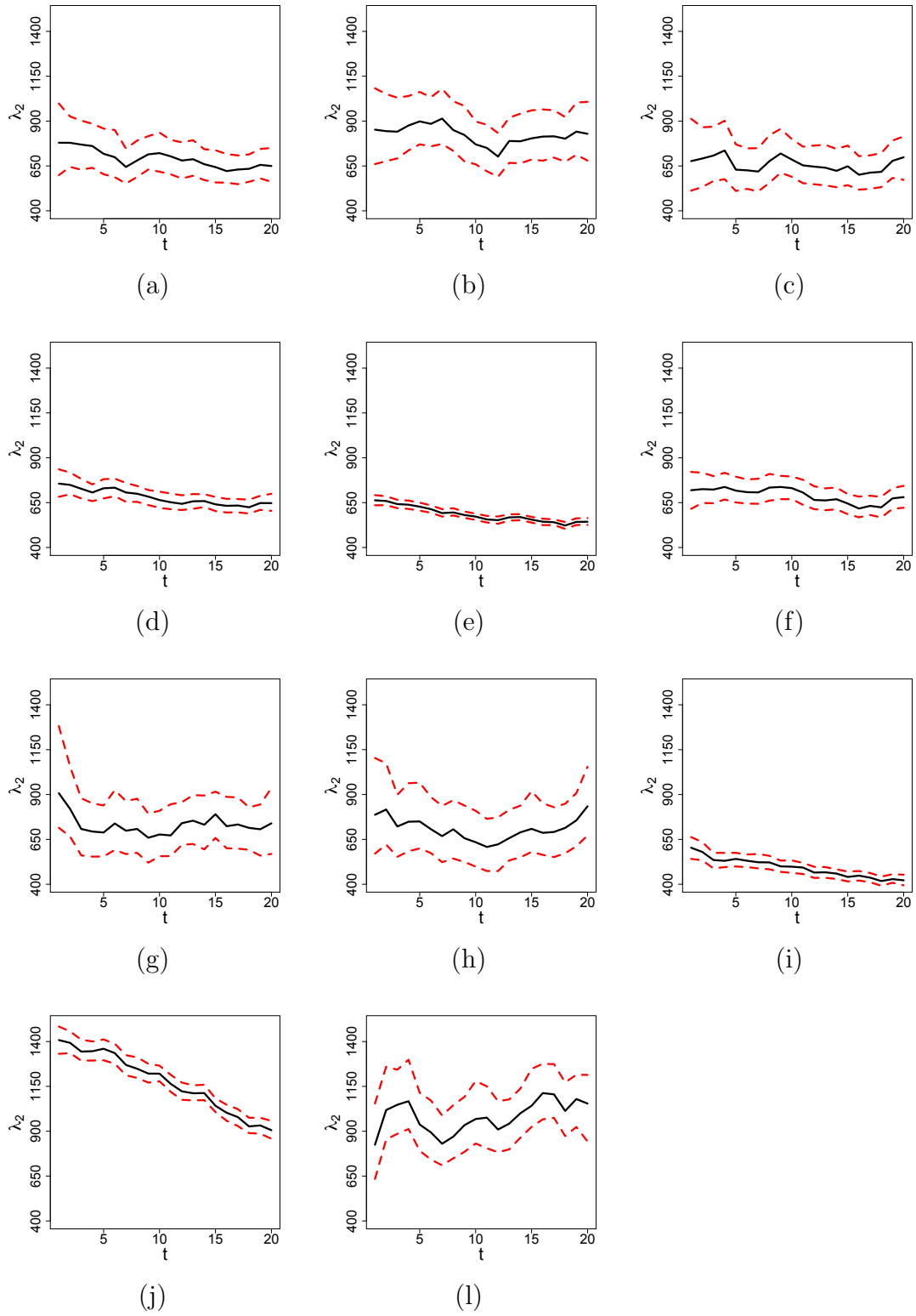


Figure 4: State of Missouri application. Risk per 100,000 inhabitants for 45 to 64 years old age group. Estimated risk level for counties within the Saint Louis metropolitan area. Posterior median (solid line) and 95% credible interval (dashed line) for $\lambda_{t3,D_{2,6}}$. Panel (j) shows a steady rate of decrease in risk through time for the City of Saint Louis.

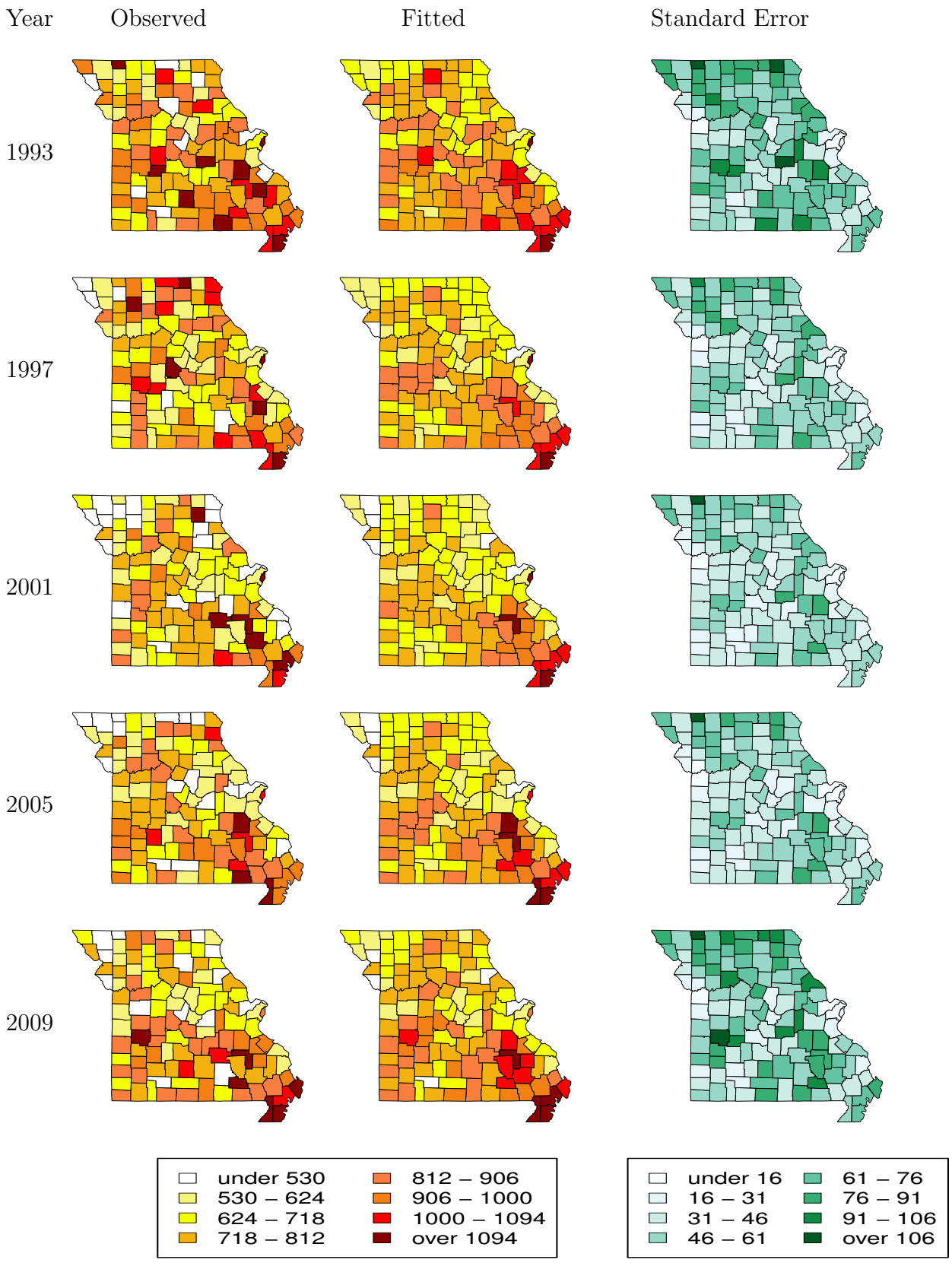


Figure 5: State of Missouri. Standardized mortality ratio per 100,000 inhabitants for 45 to 64 years old age group. Observed (left panels), fitted (central panels), and standard error (right panels).

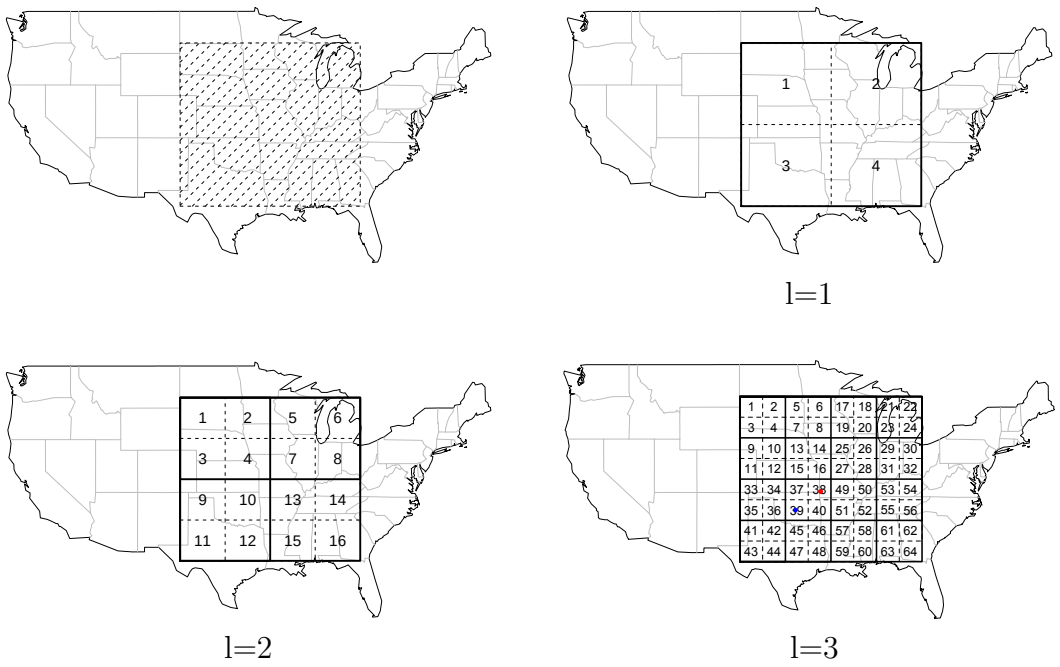
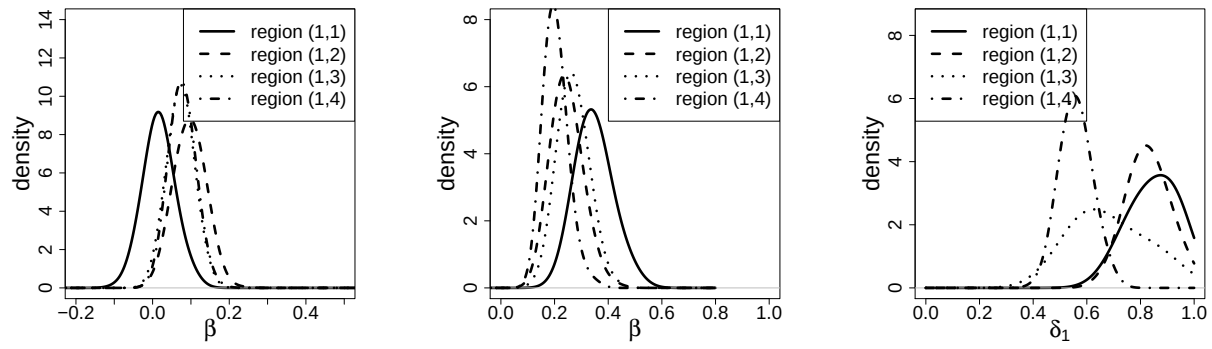
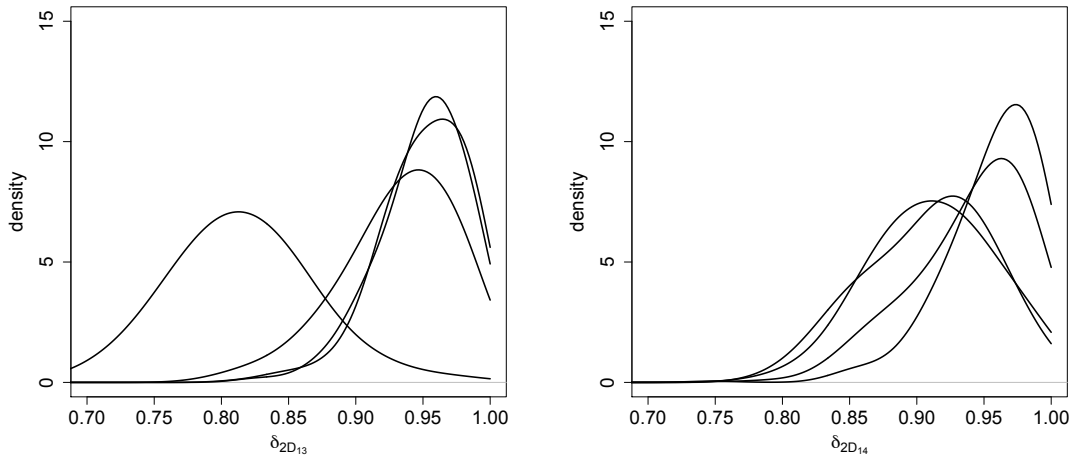


Figure 6: USA Tornado alley and multiscale structure for each of the three levels of resolution. On the lower right panel, the red dot in 38 (region (3,38)) indicates the city of Joplin, Missouri, whereas the blue dot in 39 (region (3,39)) indicates the city of Moore, Oklahoma. In our notation, (l, j) denotes region j at level l.



(a) (b) (c)

Figure 7: Tornado reports application. Posterior densities for β_j, γ_j, and δ_{1j}, $j = 1, \ldots, 4$.



(a) (b)

Figure 8: Tornado reports application. Posterior densities for the discount factors δ_{2j} corresponding to the descendants of region (1,3) (Panel a) and region (1,4) (Panel b). On Panel (a), the notably smaller discount factor corresponds to region (2,10).

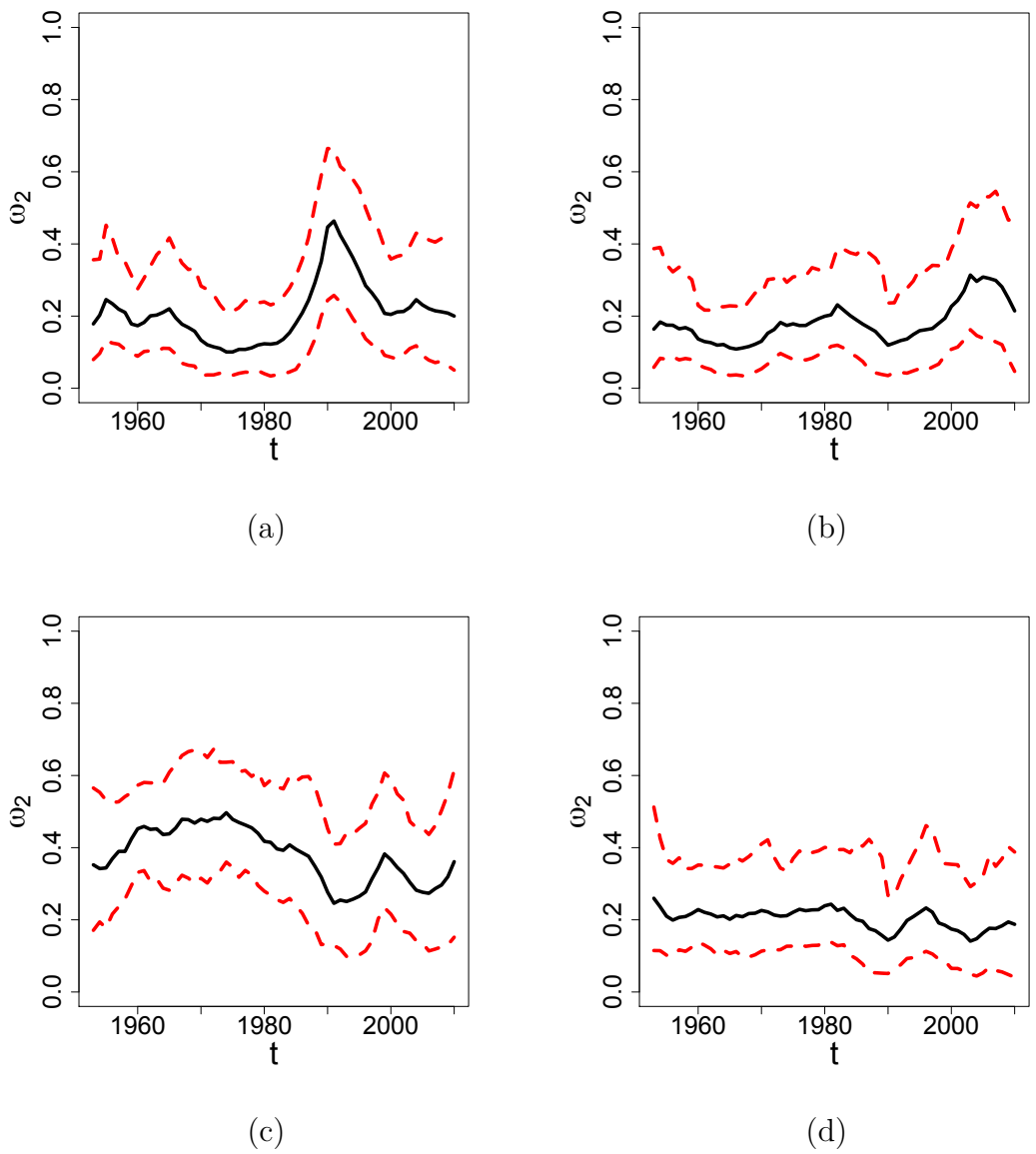


Figure 9: Tornado reports application. Posterior median (solid line) and 95% credible interval (dashed line) for each element of $\boldsymbol{\omega}_{t2,10}$. Each element of $\boldsymbol{\omega}_{t2,10}$ corresponds to one descendant of subregion $(2, 10)$. These descendents are $(3, 37)$ (panel a), $(3, 38)$ (panel b), $(3, 39)$ (panel c) and $(3, 40)$ (panel d). Region $(3, 38)$ indicates an increase in risk from 1990 to 2010. Region $(3, 38)$ includes the city of Joplin, Missouri.