Dependent Species Sampling Models for Spatial Density Estimation

Seongil Jo[*], Jaeyong Lee[†‖], Peter Müller[‡], Fernando A. Quintana[§**], and Lorenzo Trippa[¶††],

Abstract. We consider a novel Bayesian nonparametric model for density estimation with an underlying spatial structure. The model is built on a class of species sampling models, which are discrete random probability measures that can be represented as a mixture of random support points and random weights. Specifically, we construct a collection of spatially dependent species sampling models and propose a mixture model based on this collection. The key idea is the introduction of spatial dependence by modeling the weights through a conditional autoregressive model. We present an extensive simulation study to compare the performance of the proposed model with competitors. The proposed model compares favorably to these alternatives. We apply the method to the estimation of summer precipitation density functions using Climate Prediction Center Merged Analysis of Precipitation data over East Asia.

Keywords: climate prediction, conditional autoregressive model, spatial density estimation, species sampling model.

1 Introduction

Inference on a collection of probability distributions that are related but not identical is a fundamental problem in statistics. Some of the simplest versions of this problem are normal linear regression models, where the distributions are assumed to be normal, with means assumed to be linear functions of covariates. Such models are, however, inadequate to describe more complicated scenarios, such as those arising in inference for climate and meteorological data when observations are indexed by location and estimation of the climate variable distributions at each location is of interest (see, e.g. Gelfand et al., 2005). As a motivating application, we consider data on a 33-year period of summer precipitation over East Asia. One traditional use of such data is the prediction of summer precipitation in the form of probabilities for above average, average, and below average precipitation (Barnston et al., 2003; Tippett et al., 2007). A crucial element

[*]Department of Statistics (Institute of Applied Statistics), Chonbuk National University, Republic of Korea, statjs@jbnu.ac.kr

[†]Department of Statistics, Seoul National University, Republic of Korea, leejyc@gmail.com

[‡]Department of Mathematics, University of Texas at Austin, USA, pmueller@math.utexas.edu

[§]Departamento de Estadísica, Pontificia Universidad Católica de Chile, Chile, quintana@mat.uc.cl

[¶]Department of Biostatistics, Dana-Farber Cancer Institute and Havard University, USA, ltrippa@jimmy.harvard.edu

in such predictions is the adequate modeling of the distribution of climate variables. Modeling of climate variables, however, is not straightforward, as their distributions are poorly approximated by commonly used parametric families, and might vary by covariates such as geographical coordinates. In summary, the problem can be described as spatially varying density estimation.

Density estimation is one of the traditional applications of nonparametric Bayesian methods. See, for example, Müller and Mitra (2013) for a review. Related nonparametric Bayesian models include, in particular, prior models for families of distributions that are indexed by a given set of covariates, $\{G_x : x \in \mathcal{X}\}$. Such models are known as *dependent* random probability measures and have been a focal point of recent research efforts. The study of such models started with the seminal work by MacEachern (1999, 2000), who proposed a covariate-dependent version of the Dirichlet process (DP) by Ferguson (1973).

The key idea of the dependent DP construction is based on the stick-breaking representation by Sethuraman (1994). A random probability measure G with DP prior can be represented as

$$G(B) = \sum_{h=1}^{\infty} w_h \delta_{\theta_h}(B), \qquad B \in \mathcal{B}, \tag{1}$$

where $w_h = V_h \prod_{j=1}^{h-1}(1 - V_j)$ with $V_h \overset{\text{iid}}{\sim} \text{Beta}(1, M)$ and $\theta_h \overset{\text{iid}}{\sim} G_0$, independent of the w_h. Here, $\mathcal{B}$ is a Borel σ-field on the space where G_0 is supported. In most applications, this space is Euclidean. In summary, the DP prior model is indexed by two parameters, M and G_0. The parameter $M > 0$ is known as the *total mass parameter*, and G_0 is called the *centering distribution*. We write $G \sim \text{DP}(M, G_0)$.

MacEachern's dependent DP (DDP) considers a family of random probability measures $G_x(B) = \sum_h w_{x,h} \delta_{\theta_{x,h}}(B)$, where the support points $\theta_{x,h}$ and weights $w_{x,h}$ are modeled as stochastic processes indexed by covariates x. DDPs have interesting general properties, such as continuity and large support under suitable conditions (Barrientos et al., 2012). The construction of the DDP is such that marginally each G_x follows a DP distribution for every $x \in \mathcal{X}$. A commonly used variation of the DDP model is obtained when dependence is introduced through support points only, keeping the weights $w_{x,h} = w_h$ common across x. This is usually referred to as *common or single weights* DDP and has been used by many authors, including De Iorio et al. (2004), Gelfand et al. (2005), Rodríguez and ter Horst (2008), and Jara et al. (2010) to name just a few. One of the advantages of models based on common weights DDPs is that posterior simulation algorithms for models based on regular DP priors can usually be adapted with few modifications.

Although popular, common weight DDPs have been argued to have some limitations. See, e.g. the discussion in Griffin and Steel (2006), Duan et al. (2007) and Rodríguez and Dunson (2011). This has led some researchers to consider *common or single atoms* DDPs where dependence is introduced in the weights only. Griffin and Steel (2006) proposed the order-based DP in which one set of weights and support points are shared by all random measures G_x, but with different random order. Chung and Dunson (2011) defined the local DP. The local DP utilizes one set of weights and support points as in

the order-based DP, but each probability measure G_x is defined by a random subset of the weights and support points. Dunson and Park (2008) proposed the kernel stick-breaking process in which $[0,1]$-valued random variables that define a stick-breaking process are multiplied by a kernel factor to reflect location effects.

Reich and Fuentes (2007) and Fuentes and Reich (2013) used the kernel stick-breaking process in modeling wind speed to define spatially varying random distributions, with uniform and elliptical kernels used as location-specific multiplicative factors. Li et al. (2015) used a similar idea in the boundary detection problem where the kernel factor is replaced by logit transformed random variables from a latent autoregressive model. Rodríguez and Dunson (2011) defined the probit stick-breaking process in which the Beta random variables of the stick-breaking process are replaced by probit-transformed Gaussian processes. Alternatively, Ren et al. (2011) and Ding et al. (2012) used Beta random variables defined by a logit transformation of covariates. These approaches allow for more flexible variations of random probability measures than the common weights DDP.

All these approaches include the stick-breaking construction, and the type of dependence structure introduced is counterintuitive and foreign to the spatial dependence literature, with the exception of Rodríguez and Dunson (2011) and Li et al. (2015). Another critical limitation of DP- or DDP-based models is that the weights in the stick-breaking construction are stochastically ordered in an exponentially descending way. This implies a particular clustering structure that favors a small group of large clusters, which is unnatural in many applications. To deal with these limitations, we propose a class of dependent random probability measures whose definition is based on two main ingredients: species sampling models (SSM) and conditional autoregressive (CAR) models. In particular, we do not use the stick-breaking process to define the weights of random probability measures and define the weights directly via normalization.

The first component of the proposed approach is the use of general SSM's (Pitman, 1996). A SSM defines a random discrete probability measure with random weights and support points that are independent of each other. It is the de Finetti measure of a species sampling sequence, that is, an exchangeable sequence of random variables for which the predictive distribution of the next observation is a function of the number of ties and frequencies among the earlier observations. Species sampling sequences are extensions of the Pólya urn sequence (Blackwell and MacQueen, 1973), that is, an exchangeable sequence of random variables whose de Finetti measure is the Dirichlet process. SSMs are random probability measures that involves a probabilistic structure to define a random probability measure. Many well known examples follow this structure.

The second element of the proposed approach is the use of CAR models, which were introduced by Besag (1974) and have been widely used for modeling data with spatial dependencies. For example, Geman and Geman (1984) and Besag et al. (1991) applied CAR models to image data, while Nieto-Barajas (2008) and Lee (2011) used these models to investigate medical records. The main advantages of CAR models are the easy implementation of posterior simulation algorithms and the natural extension to the modeling of spatio-temporal data. We will argue that the combination of these two features provides a rich basis for flexible modeling of data with a spatial structure.

The rest of this paper is organized as follows. In Section 2 we review SSMs and CAR models. In Section 3 we introduce the proposed CAR SSMs. Computational methods to implement posterior simulation under the CAR SSMs are described in Section 4.1. In Section 5 we present a simulation study and the application to the motivating climate variables data for predicting rainfall in the Korean peninsula. Some final comments and conclusions are presented in Section 6.

2   Background

2.1   Species Sampling Models (SSM)

A random probability measure G is called a SSM (Pitman, 1996), if it can be represented as

$$G = \sum_{h=1}^{\infty} P_h \delta_{\theta_h} + RG_0, \tag{2}$$

for a sequence of positive random variables (P_h) and a nonnegative random variable R with $R = 1 - \sum_{h=1}^{\infty} P_h$ with probability 1, and where (θ_h) is a sequence of random samples from a diffuse probability measure G_0 and independent from (P_h). We can immediately see that (1) is a special case of (2). A SSM for which $P(R = 0) = 1$ is termed *proper*. The practical use of SSM as nonparametric priors is restricted to proper SSM's. SSMs admit many popular classes of random probability measures as particular cases. These include the DP, the two-parameter DP or Pitman–Yor process (Pitman and Yor, 1997), the stick-breaking priors (Ishwaran and James, 2001) and the normalized-inverse Gaussian processes (Lijoi et al., 2005). More properties of SSMs can be found in Pitman (1996), Ishwaran and James (2003), Navarrete et al. (2008), James (2008), James et al. (2009), Jang et al. (2010), and Lee et al. (2013).

Recently, Bassetti et al. (2010) and Airoldi et al. (2014) proposed a generalization of species sampling sequences designed for nonexchangeable data which has modeling potential for time series and spatial statistics. Lee et al. (2013) considered a family of SSMs whose weights are normalized positive random variables

$$G = \sum_{h=1}^{\infty} P_h \delta_{\theta_h}, \qquad P_h = \frac{w_h}{\sum_{l=1}^{\infty} w_l}, \qquad h = 1, 2 \ldots, \tag{3}$$

where $(w_h)_{h=1}^{\infty}$ is a sequence of positive random variables and (θ_h) are independent and identically distributed (i.i.d.) random support points distributed according to G_0, and independent of (w_h). A sufficient condition for $\sum_{l=1}^{\infty} w_l < \infty$ a.s. is that $\sum_{l=1}^{\infty} E(w_l) < \infty$ (Lee et al., 2013).

The class of SSMs is a large class of discrete random probability measures subject only to the constraint of having i.i.d. atoms that are independent of the weights. In the following discussion, we consider a specific class of SSMs based on sequences of log-normal random variables: for $h = 1, 2, \ldots,$

$$w_h = e^{u_h} \qquad \text{with} \qquad u_h \sim N\left(\log[1 - \{1 + e^{b-ah}\}^{-1}], \tau^2\right), \tag{4}$$

where a, b, β, τ^2 are positive constants. It is easy to verify that $\sum_{l=1}^{\infty} E(w_l) < \infty$, and thus $\sum_{l=1}^{\infty} w_l < \infty$ a.s. The specification (4) is chosen to define models with relevant differences compared to the Dirichlet process prior. The proposed form for the weights (w_h) allows for a number of moderately large weights of (a priori) comparable size. This is in contrast to the DP prior which assumes a priori stochastically ordered and geometrically decreasing sizes for the weights. This will be important in the context of the spatial dependence across related measures which we will introduce later. With multiple dominant weights it is easier to model the desired heterogeneity across locations.

2.2 Gaussian Conditional Autoregressive (CAR) Models

The CAR model is one of the most popular models for a large number of dependent random variables. Under a CAR model, the joint distribution of a set of random variables is specified via the full conditional distributions. However, not every set of full conditional distributions leads to a legitimate joint distribution. See further details in Besag (1974), Cressie (1993), Kaiser and Cressie (2000), Cressie and Wikle (2011) and references therein.

We focus on Gaussian CAR models. Let $\mathcal{D} = \{s_i, i = 1, \ldots, n\}$ be a set of spatial locations and let $u = (u_1, \ldots, u_n)^T$ be a collection of random variables with u_i corresponding to location s_i for $i = 1, 2, \ldots, n$. We define

$$u_i \mid u_j, j \neq i \sim N\Big(\mu_i - \sum_{j:j\neq i}^{n} \xi_{ij}(u_j - \mu_j), \tau_i^2\Big), \quad i = 1, \ldots, n, \tag{5}$$

where $\mu_i, \xi_{ij} \in \mathbb{R}$ and $\tau_i^2 > 0$ for all $i, j = 1, \ldots n$ and let $\Lambda = (\lambda_{ij})$ with

$$\lambda_{ii} = 1/\tau_i^2, \quad \lambda_{ij} = \xi_{ij}/\tau_i^2, \quad i, j = 1, \ldots, n.$$

A key result for (5) to define a proper model is the following. If Λ is symmetric and positive definite, then the joint distribution of y is $N(\mu, \Lambda^{-1})$. See Rue and Held (2005) for further details, or Besag (1974), Cressie (1993), and Banerjee et al. (2004). Specific choices for ξ_{ij} will be discussed below, when we combine (3) through (5) to construct the desired spatially dependent SSMs.

3 Spatial CAR SSMs

We combine the definitions of SSM's and CAR models to construct a class of spatially $\mp$ dependent SSMs $\{G_s : s \in \mathcal{D}\}$ on a set of locations $\mathcal{D} = \{s_1, s_2, \ldots, s_n\}$. The domain $\mathcal{D}$ can be a set of grid points defined on a spatial region, such as the area including South Korea for our motivating application, or a set of time points, etc. We define a probability measure G_s at each location $s \in \mathcal{D}$ as

$$G_s = \sum_{h=1}^{\infty} P_{s,h} \delta_{\theta_h}, \tag{6}$$

where

$$P_{s,h} = \frac{w_{s,h}}{\sum_{k=1}^{\infty} w_{s,k}}, \quad s \in \mathcal{D}, \qquad h \geq 1,$$

$w_{s,h}$ are nonnegative random variables with $\sum_{h \geq 1} w_{s,h} < \infty$ a.s., and $(\theta_h, \, h \geq 1)$ is a random sample from a nonatomic distribution G_0, independent of $w_{s,h}$. The dependence of the random probability measures G_s across s is introduced through the dependence of $(w_{s,h}, \, s \in \mathcal{D})$. We use the CAR model to define the joint distributions of $(w_{s,h}, \, s \in \mathcal{D})$. We also note that (6) highlights the discrete nature of the model for G_s. This implies that a sample from any G_s will have ties with positive probability, giving thus rise to a particular clustering structure whose distribution is controlled by the distributional assumptions on the set of weights $(P_{s,h})$ for $s \in \mathcal{D}$, in turn determined by the $(w_{s,h})$ quantities.

We combine (4) and (5) to define $w_{s_i,h} = e^{u_{i,h}}$, with the joint distribution of $(u_{i,h})$ determined by the CAR model

$$u_{i,h} \mid u_{\ell,h}, \; \ell \neq i \sim N\Big(m_{i,h} - \sum_{\ell:\ell \neq i}^{n} \xi_{i\ell}(u_{\ell,h} - m_{\ell,h}), \tau^2\Big), \quad i = 1, \ldots, n \qquad (7)$$

where $\tau^2 > 0$, and

$$m_{i,h} = \log\{1 - (1 + e^{b-ah})^{-1}\}, \quad a > 0, \;\; b > 0, \quad i = 1, \ldots, n. \qquad (8)$$

A key element of model (7) is the specification of the autoregressive coefficients $(\xi_{i\ell})$. We consider here two alternative ways of defining $(\xi_{i\ell})$, which differ in the extent and strength of the induced dependence structure. We will later compare the relative merits of both specifications.

Mercer CAR model: Suppose $K(s,t)$ is a Mercer kernel with $K(s,s) = 1$ for all $s \in D$. Define

$$\xi_{i\ell} = K(s_i, s_\ell), \qquad i \neq \ell, \qquad i, \ell = 1, \ldots, n,$$

then $\Lambda = (\lambda_{i\ell})_{i,\ell=1}^{n}$, with $\lambda_{i\ell} = K(s_i, s_\ell)/\tau^2$ is symmetric and positive definite. Thus, $\boldsymbol{u}_h \sim N(\boldsymbol{m}_h, \Lambda^{-1})$, where $\boldsymbol{m}_h = (m_{1,h}, \ldots, m_{n,h})$ and $\boldsymbol{u}_h = (u_{1,h}, \ldots, u_{n,h})$. We specifically use the Gaussian kernel

$$K(s_i, s_\ell) = \exp(-\rho||s_i - s_\ell||^2), \qquad i, \ell = 1, \ldots, n,$$

where ρ is positive. We call the spatial SSM defined by a Mercer kernel a Mercer CAR SSM, and denote it as $(G_s, s \in \mathcal{D}) \sim \text{MCS}(G_0, a, b, \tau^2, \rho)$. The model does not include the independent model with $\boldsymbol{u}_h$ having diagonal covariance matrix. But it can be arbitrarily close to the independent model.

Clayton–Kaldor CAR model (Clayton and Kaldor, 1987; Sun et al., 1999): Let $N(s_i)$ be the set of neighbors of the i-th location s_i. Define

$$\xi_{i\ell} = \begin{cases} -\rho & \ell \in N(s_i) \\ 0 & \ell \notin N(s_i), \end{cases}$$

where $\rho \in (\psi_{(1)}^{-1}, \psi_{(n)}^{-1})$, $\psi_{(1)} \leq \psi_{(2)} \leq \cdots \leq \psi_{(n)}$ are the ordered eigenvalues of an adjacency matrix $C = [c_{ij}]$ with $c_{ii} = 0$, $i = 1, \ldots, n$ and $c_{i\ell} = 1$ if $\ell \in N(s_i)$ (Sun et al., 1999). This guarantees that $\Lambda = I - \rho C$ is positive definite and $\boldsymbol{u}_h \sim N(\boldsymbol{m}_h, \tau^2 \Lambda^{-1})$. We specifically consider the neighborhood of s_i defined by

$$N(s_i) := \{s_\ell : ||s_i - s_\ell||^2 < B, \ \ell \neq i\},$$

for some fixed positive constant B. We refer to this spatial SSM as the CK CAR SSM and denote it as $(G_s, s \in \mathcal{D}) \sim \text{CKCS}(G_0, a, b, \tau^2, \rho, B)$. The parameter ρ covers a spectrum of dependence among $u_{i,h}$'s. For example, the independence model can be specified by setting $\rho = 0$.

The main difference between the two CAR SSM models is the nature of the precision matrix that is implied by the two models, which in turn results in different partial correlation structures. The precision matrix of CK CAR is sparse and thus its partial correlation structure is local, i.e. only neighboring sites have nonzero partial correlation. On the other hand, the precision matrix of Mercer CAR model is dense and its partial correlation structure is global, i.e., any two sites have nonzero partial correlation. Although the two models have little difference in fitted mean models, the difference in partial correlation structure can have actual impacts in other data analysis. As $\rho \longrightarrow \infty$, the Mercer CAR can be arbitrarily close to the independent model with independent random probability measures at different locations. But in actual data analysis choice of hyperparameter ρ can be delicate and sometimes the posterior includes long range dependencies. See Figure 10. Thus, if the densities are expected to possibly change abruptly, the Mercer CAR needs to be used with caution. Also, the sparsity of the precision matrix of CK CAR may be used to speed up the computation for large data set.

In a recent paper, Li et al. (2015) considered the use of CAR models to define a spatial stick-breaking prior. In the CAR SSM the weights of random probability measures are directly defined via normalization of CAR models, while Li et al. (2015) modeled the weights with a stick-breaking process defined by logit transformations of CAR models.

Nonparametric prior models, by definition, are designed to be flexible and to capture departures from simpler parametric constructions. A direct approach to verify this property consists of characterizing the weak support of the prior model. It is common in the Bayesian nonparametric literature to prove that the prior model has full weak support in order to emphasize its flexibility. The prior is said to have full weak support if the support of the prior under the weak topology is the same as the parameter space. This approach has been used for example to study prior models for unknown densities (see, e.g. Wu and Ghosal, 2008) and prior models for families of unknown distributions indexed by covariates (see, e.g. Barrientos et al., 2012). We refer to Ghosh and Ramamoorthi (2003) for a discussion on the use of the weak metric in Bayesian nonparametrics. The following proposition shows that the CAR SSMs have full weak support. The proof is given in Appendix A of the Supplementary Material (Jo et al., 2016).

Proposition 1. If G_0 has full support on $\mathbb{R}$, then the Mercer CAR SSM, $\text{MCS}(G_0, a, b, \tau^2, \rho)$, and the CK CAR SSM, $\text{CKCS}(G_0, a, b, \tau^2, \rho, B)$, have full weak supports.

4   Mixtures of CAR SSM

In this section, we discuss how the CAR SSMs can be used to model spatially located density functions simultaneously. Consider a grid $\mathcal{D} = \{s_i : i = 1, 2, \ldots, n\}$, and a set of corresponding observations at each point, $\mathcal{Y} = \{(y_{s_i,j}, j = 1, \ldots, N_i) : s_i \in \mathcal{D}\}$. In many applications, such as in the motivating example, the responses $\mathcal{Y}$ are continuous variables. The discrete nature of the MCS and the CKCS model would be inappropriate for such data.

A common extension of discrete random probability measures to accommodate continuous data is the use of an additional convolution with a continuous kernel (see, e.g. Lo, 1984). Specifically, we propose a probability model for data $\mathcal{Y}$ with spatial structure $\mathcal{D}$ which can be described as a mixture with respect to a CAR SSMs. We use either the Mercer or the CK constructions described above. The model is stated as follows.

$$(G_{s_i}, s_i \in \mathcal{D}) \sim \text{MCS}(G_0, a, b, \tau^2, \rho) \text{ or } \text{CKCS}(G_0, a, b, \tau^2, \rho, B)$$

$$(y_{s_i,j}, j = 1, 2, \ldots, N_i) \mid G_{s_i} \overset{\text{iid}}{\sim} \int f(y|\theta) dG_{s_i}(\theta), \text{ for } i = 1, 2, \ldots, n, \tag{9}$$

where $f(y|\theta)$ is a kernel density with parameter θ, and $a, b, \tau^2, \rho, B > 0$. While this kernel may be chosen to be any distribution supported on the real numbers, for convenience we often specify it as the normal density $f(\cdot|\theta) \equiv N(\cdot; \mu, \sigma^2)$ with $\theta = (\mu, \sigma^2)$, in which case the conjugate (Normal Inverse Gamma) base distribution G_0 is chosen as

$$G_0(\mu, \sigma^2; \mu_0, \alpha, \nu, \psi) = N(\mu; \mu_0, \sigma^2/\alpha) \, \text{Ga}(\sigma^{-2}; \nu/2, \nu\psi/2), \tag{10}$$

where $\mu_0 \in \mathbb{R}$ and $\alpha, \nu, \psi > 0$, and $\text{Ga}(a, b)$ refers to a gamma distribution with mean a/b and variance a/b^2. The hyperparameter ψ plays the role of a bandwidth in the kernel density estimation and controls the smoothness of the density estimates. It can affect posterior sensitivity. We fix hyper parameters μ_0, α, and ν, but place a hyperprior on ψ:

$$\psi \sim Ga\Big(\psi; \frac{\nu_0}{2}, \frac{\nu_0 \psi_0^{-1}}{2}\Big), \tag{11}$$

where ν_0 is a positive constant.

To facilitate posterior sampling, we use the following equivalent hierarchical model, which is obtained by introducing latent variables θ_{ij} and replacing the integral in (9) with a hierarchical prior on θ_{ij}:

$$y_{s_i,j} \mid \theta_{ij} \overset{\text{ind}}{\sim} f(\cdot|\theta_{ij}), \quad j = 1, \ldots, N_i,$$

$$\theta_{ij} \mid G_{s_i} \overset{\text{iid}}{\sim} G_{s_i}, \tag{12}$$

$$G_{s_i} = \sum_{h=1}^{\infty} P_{s_i,h} \delta_{\phi_h}, \quad P_{s_i,h} \propto w_{s_i,h} \quad \text{and} \quad \phi_h \mid \psi \overset{\text{iid}}{\sim} G_0(\cdot \mid \psi),$$

where $(w_{s_i,h}, s_i \in \mathcal{D})$ are defined by either Mercer or CK CAR models, and the hyperprior for ψ is given in (11). We refer to (12) as either the MCS mixture or CKCS mixture model.

4.1   Posterior Computation

We introduce a posterior sampling algorithm for the mixtures of CAR SSM (12). For the posterior computation, the SSMs are truncated to the first K terms in the infinite sum. Similar truncations are often used in infinite mixture models similar to the ones considered here. See, e.g. Ishwaran and James (2001) and Rodríguez and Dunson (2011).

The proposed algorithm is based on the blocked Gibbs sampling algorithm (Ishwaran and James, 2001) and a data augmentation method for the multinomial logit model (Scott, 2011) which is an extension of the algorithm of Albert and Chib (1993) for the probit model. To be specific we use again the normal kernel.

We introduce an additional set of latent variables to help identifying cluster allocation. Let $z_{i,j} = h$ if $\theta_{ij} = \phi_h$, for $i = 1, \dots, n$, $j = 1, \dots, N_i$ and $h = 1, \dots, K$, where $\phi_h = (\mu_h, \sigma_h^2)$. Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$, and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_K^2)$. Let $u_{i,h} = \log(w_{i,h})$, $i = 1, \dots, n$, $h = 1, \dots, K$, $\boldsymbol{u}_h = (u_{i,h}, i = 1, \dots, n)$ and $\boldsymbol{u} = \{\boldsymbol{u}_h, h = 1, \dots, K\}$. The mixtures of CAR SSM model (12) can be equivalently written as

$$
\begin{aligned}
y_{i,j} \mid \phi, z_{i,j} &\overset{\text{ind}}{\sim} f(\cdot | \phi_{z_{i,j}}), & j &= 1, \dots, N_i, \\
p(z_{i,j} = h \mid \boldsymbol{u}) &= \frac{e^{u_{i,h}}}{\sum_{k=1}^{K} e^{u_{i,k}}}, & h &= 1, \dots, K, \\
\boldsymbol{u}_h \mid \eta &\sim p(\boldsymbol{u}_h \mid \eta), & h &= 1, \dots, K, \\
\phi_h &\overset{\text{iid}}{\sim} G_0(\cdot | \psi), & \psi &\sim \pi(\psi),
\end{aligned}
\tag{13}
$$

where $p(\boldsymbol{u}_h \mid \eta)$ is the density of $\boldsymbol{u}_h$ under the Mercer or CK CAR models with hyperparameter $\eta = (a, b, \tau^2, \rho)$ and ψ is the parameter for G_0 with hyperprior $\pi(\psi)$. For the normal kernel f, $\pi(\psi)$ is given by (11). The details of the full conditionals and MCMC steps are given in Appendix B of the Supplementary Material (Jo et al., 2016).

Finally, we comment on computation times. Using the described posterior MCMC simulation strategy we find that inference for problems with up to some 100s sites can be implemented with reasonable computing times. See, for example, the computing times reported in Table 2. For the data analysis reported later, in Section 5.2., we could finish the computation of the proposed CAR SSMs within a few hours for inference with 6000 observations over more than 200 sites. In summary, computational effort for the proposed CAR SSMs compares favorably against available methods, but there remains a limiting factor.

4.2   Hyperparameters

Inference appears to be robust with respect to most of the hyperparameters. We therefore recommend to fix hyperparameters except ψ. Below we give guidelines for specific choices that we find to work well. Let K be the number of mixture components to be used in the MCMC sampler. The value of K needs to be large enough for the model to be flexible to fit the data, but not too large for it to slow down the MCMC sampler. Thus, small K which provides enough flexibility of the model is ideal. We choose

$K = C \max_{1 \leq i \leq N} N_i$ with $C = \frac{1}{2}$, where N_i is the number of the observations at site i. From our experience, this appears to work for most data sets. If the computation time is scarce, smaller K with $C = \frac{1}{3}$ or $\frac{1}{4}$ can be used. The values of a, b and τ are chosen to allow for a sufficient number of non-negligible weights, but at the same time to represent a preference for a small number of effective components. If m_{ih} decreases too rapidly with h, the number of effective mixture components gets too small and the mixture model may not be flexible enough. We use $\log m_{i,K/2} = 1/2$ and $\log m_{i,3K/4} = 1/4$, and $Pr(e^{u_{i,K/2}} \in [\frac{1}{4}, \frac{3}{4}]) \geq \frac{1}{2}$ which renders $\tau = 0.6$.

The parameter ψ in the base measure plays a similar role as the bandwidth in the kernel density estimation. We put a prior on ψ,

$$\psi \sim Ga\left(\frac{\nu_0}{2}, \frac{\nu_0 \psi_0^{-1}}{2}\right)$$

where $\psi_0 = 0.7 \hat{\sigma} \bar{N}^{-1/5}$, $\nu_0 = 2$, $\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}_i$ and $\hat{\sigma}_i$ is the standard deviation of observations at site s_i. The proposed value for ψ_0 is motivated by the rule of thumb for a kernel density estimation bandwidth parameter, suggested by Silverman (1986) as $1.06 \hat{\sigma} N^{-1/5}$.

We choose

$$\mu_0 = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i, \qquad \alpha = 0.7^2 n^{-2/5} \qquad \text{and} \qquad \nu = 2,$$

which are chosen so that the mean and variance of $N(\mu_0, \sigma^2/\alpha)$ in the base measure match the mean and variance of observations.

The partial correlation of the $u_{i,h}$ variables is controlled by ρ and B. In the Mercer CAR model the partial correlation between $u_{i,h}$ and $u_{j,h}$ is $-e^{-\rho d^2}$, with $d = ||s_i - s_j||$. For example, if $\rho \approx 0$, the partial correlation between two sites are close to 1 and the model may fail to cope with abrupt changes of the densities. Thus, one needs to be careful in choosing the value of ρ. We suggest the following steps to choose ρ. We first determine the fixed distance d_0 at which the partial correlation is 0.5, i.e., $e^{-\rho d_0^2} = 0.5$. Then, solve the equation to get the value of $\rho = (\log 2)/d_0^2$. In the above steps, 0.5 can be replaced by a different value. For the Clayton–Kaldor model we choose ρ and B similarly.

5 Examples

5.1 Simulation Studies

We carried out two simulation studies to compare the proposed CAR SSMs versus (i) a spatially independent SSM, (ii) the spatially dependent Dirichlet process mixture model (sDP) proposed by Gelfand et al. (2005), (iii) the generalized spatial Dirichlet process model (gsDP) of Duan et al. (2007), and (iv) the spatial dependent probit stick-breaking process model (sPSBP) of Rodríguez and Dunson (2011). In the first simulation study,

we carry out repeat experimentation with 100 independent data sets for 13 locations over a univariate grid, while the second simulation study considers a large scale data set over a bivariate grid with $10 \times 10 = 100$ grid points.

The sDP, gsDP and sPSBP are implemented using JAGS (Plummer, 2003), which can be freely downloaded from `http://mcmc-jags.sourceforge.net`. For speed, we coded posterior sampling algorithms for the independent SSMs and the proposed CAR SSMs in `Fortran` functions that we called from `R`. (We are developing a user-friendly R package to implement our models; a preliminary version is available from `http://sites.google.com/site/joseongil/software/`). For example, for a two-dimensional large scale data with 100 locations, posterior simulations of independent SSM and the proposed CAR SSMs take about 2 hours and 5 hours to run 30,000 iterations (10,000 burn-in, 20,000 iterations after burn-in period), respectively. By contrast, the sPSBP takes approximately 11 hours and the sDP and gsDP take approximately 10 hours and 72 hours in JAGS, respectively. This may be due to the fact that the algorithms of the sDP, gsDP and sPSBP are implemented through JAGS.

We evaluate the performance of the density estimators by integrated absolute error (IAE)

$$\text{IAE} = \int_{-\infty}^{\infty} |f_s(y) - \hat{f}_s(y)| dy, \tag{14}$$

where $f_s(y)$ is the density under the simulation truth at location s and $\hat{f}_s(y)$ is the corresponding density estimate. The integral was approximated numerically using Simpson's 3/8 rule over a grid of 200 equally-spaced points in the interval $[-6, 6]$. We also report the deviance information criterion (DIC), introduced by Spiegelhalter et al. (2002) and the log-pseudo marginal likelihood (LPML) defined as the sum of log conditional predictive ordinates (CPO) (Geisser and Eddy, 1979). Celeux et al. (2006) discussed 8 versions of DICs for models with missing variables. When there are many latent variables in the models, due to poor performance of the estimators, the effective number of parameters p_D can be negative. We compute DIC as

$$\text{DIC} = -4\mathbb{E}[\log f(y|\theta)|y] + 2\log \mathbb{E}[f(y|\theta)|y],$$

which is one of two recommended versions by Celeux et al. (2006). Although Celeux et al. (2006) cautioned against this version because its p_D can be negative in some models, we use it because our focus of estimation is $f(y|\theta)$ and in our examples p_D was never negative.

Both, the conditional density of the CPO statistic and DIC, are evaluated for

$$f(y_{s_i,j}|\theta) = \sum_{h=1}^{K} P_{s_i,h} N(y_{s_i,j}|\mu_h, \sigma_h^2), \qquad i = 1, \ldots, n, \quad j = 1, 2, \ldots, N_i. \tag{15}$$

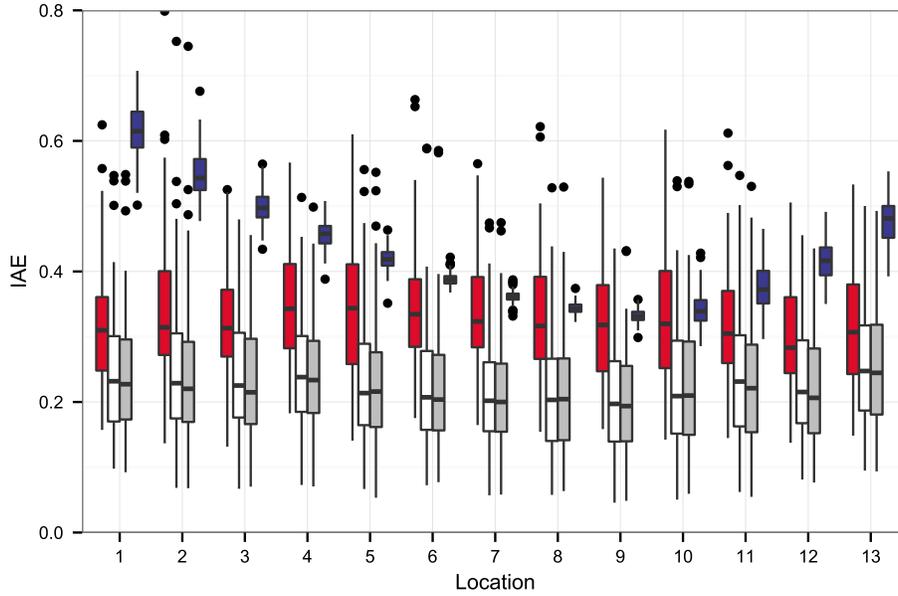Higher LPML values and lower DIC values are indicative of better models.

Figure 1: Simulation 1. Boxplots of IAE values under the CAR SSMs, an independent SSM and a sDP for 100 data sets generated from simulation model defined in (16c). The boxplots show results under the following models: $\text{CKCS}(G_0, a, b, \tau^2, \rho, B)$ (white), $\text{MCS}(G_0, a, b, \tau^2, \rho)$ (gray), Independent SSM (red), and sDP (blue).

One-Dimensional Data

In the first simulation study, we compare the proposed CAR SSMs with inference under a spatially independent SSM and inference under the sDP. The hypothetical data are generated from three simulation truths consisting of mixtures of two normal distributions, a modified version of the simulation model used by Rodríguez and ter Horst (2008). The first two models only have spatially varying weights or atoms, while the third model has both, spatially varying locations and weights. Also, in all models we fix the scale parameters.

$$\texttt{Model 1:} \quad f_s(y) = w_s N(y; -1.5, 1) + (1 - w_s)N(y; 1.5, 1), \tag{16a}$$

$$\texttt{Model 2:} \quad f_s(y) = 0.7N(y; \mu_{s,1}, 1) + 0.3N(y; \mu_{s,2}, 1), \tag{16b}$$

$$\texttt{Model 3:} \quad f_s(y) = w_s N(y; \mu_{s,1}, 1) + (1 - w_s)N(y; \mu_{s,2}, 1), \tag{16c}$$

where $w_s = 0.15 + 0.05s$, $\mu_{s,1} = -2 + 0.05s$ and $\mu_{s,2} = 2 - 0.05s$, for $s = 1, \ldots, 13$. We generated 100 independent data sets, each of which consists of 260 observations for all 13 locations, i.e., 20 observations for each location.

Figure 1 shows box plots of IAE values across all locations for `Model 3` (16c) and Table 1 reports the LPML and DIC values, obtained for the generated datasets of the first simulation study under the proposed CAR SSMs, the independent SSM, and the sDP.

The density estimates under both mixture of CAR SSMs perform favorably compared to estimates under the competing models. The results from these examples suggest that the proposed methods utilize information across space and estimate spatially varying densities effectively. The proposed models also provide a better fit to the data. The two proposed CAR SSMs perform comparably with each other. However, the sDP performs less favorably for these data. The model is not flexible enough to identify the two modes of both densities. This may be due to the fact that the dependency of the sDP is modeled through atoms only and not through weights.

	Model 1		Model 2		Model 3	
Models	LPML	DIC	LPML	DIC	LPML	DIC
Independent SSM	− 553.91	1097.89	− 547.55	1086.06	− 556.08	1102.82
sDP	− 547.70	1095.26	− 537.12	1074.07	− 555.13	1110.09
Mercer CAR SSM	**− 529.29**	**1056.64**	**− 522.83**	**1043.85**	**− 531.21**	**1060.44**
CK CAR SSM	− 529.43	1056.92	-523.42	1044.84	− 532.25	1062.41

Table 1: Simulation 1. Goodness of fit for simulated data under (16).

To have a closer look at the density estimates, we present in Figure 2 density estimates for one simulated data set. For comparison the densities under the respective simulation truth are overlaid in the same plot. Generally speaking, the proposed estimates are closer to the simulation truth than under the competing models. The independent SSM overfits the data by following the empirical distribution (histograms) too closely. For some locations the density estimate has three modes, while each of the proposed CAR SSM density estimates correctly identify two modes.

Two-Dimensional Data

We consider data indexed by bivariate spatial coordinates. We generate data with spatial correlation on a 10×10 two-dimensional lattice. In the data set, each location has 20 observations and each observation follows a mixture of two normals with spatially varying weights and locations, similar to `Model 3` of the first simulation study,

$$f_{s_1,s_2}(y) = w_1(s_1, s_2)N\{y; \mu_1(s_1, s_2), 1\} + w_2(s_1, s_2)N\{y; \mu_2(s_1, s_2), 1\}, \quad (17)$$

where $w_1(s_1, s_2) = 0.14 + 0.08(s_1 - 1)$, $w_2(s_1, s_2) = 1 - w_1(s_1, s_2)$, $s_1, s_2 = 1, \ldots, 10$ and

$$\mu_1(s_1, s_2) = -2 + 0.06s_2, \quad \mu_2(s_1, s_2) = 2 - 0.06s_2, \quad s_1, s_2 = 1, \ldots, 10.$$

The data set consists of 2,000 observations for all 100 locations, i.e., 20 observations for each location. In this study, we compare inference under the CAR SSMs with inference under the sDP, the gsDP, and the sPSBP models.

Table 2 shows LPML, DIC and run times. Figure 3 displays box plots of IAE values, obtained from data generated under the simulation truth (17). In Table 2, the CK CAR SSM appears to be the best by both criteria, while the Mercer CAR SSM is slightly worse than gsDP. However, the proposed CAR SSMs have the smaller IAE values (Figure 3)
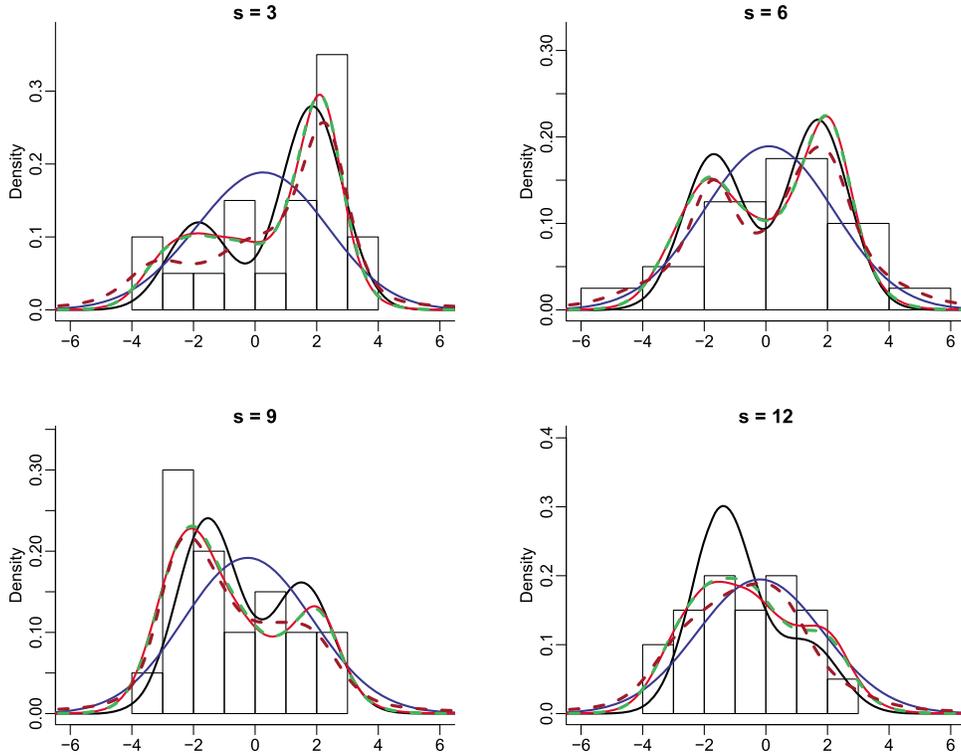
Figure 2: Simulation 1. Density estimates and simulation truth under (16c): true densities (black solid), $\text{MCS}(G_0, a, b, \tau^2, \rho)$ (green dashed), $\text{CKCS}(G_0, a, b, \tau^2, \rho, B)$ (red solid), independent SSM (brown dashed) and sDP (blue solid).

Models	LPML	DIC	run time	software
Independent SSM	-3997.19	7890.92	2 hours	R and Fortran
sDP	-4118.85	8237.02	10 hours	rjags
gsDP	-3871.71	7727.65	72 hours	rjags
sPSBP	-3881.06	7761.94	11 hours	rjags
Mercer CAR SSM	-3883.51	7746.96	5 hours	R and Fortran
CK CAR SSM	$\mathbf{-3837.46}$	**7672.76**	5 hours	R and Fortran

Table 2: Simulation 2. Goodness of fit measures for simulated 10×10 grid data defined in (17) and computation times. The computation times reported here are not for comparison of computation times of the models, because Fortran is considerably faster than rjags.

for most locations. Figure 4 shows density estimates for one of the simulated data sets. Similar to the first simulation study, the proposed estimates are closer to the simulation truth than under the competing models.
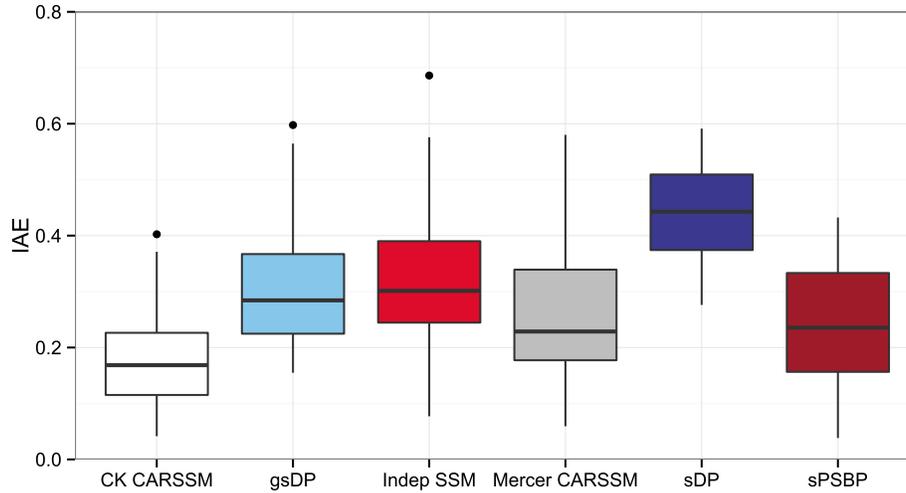
Figure 3: Simulation 2. Boxplots of IAE values under the CAR SSMs, an independent SSM, a sDP, and a gsDP under simulation truth (17). The boxplots show results under the following models: $CKCS(G_0, a, b, \tau^2, \rho, B)$ (white), $MCS(G_0, a, b, \tau^2, \rho)$ (gray), independent SSM (red), sDP (blue), gsDP (skyblue), and sPSBP (brown).

Figure 5 shows the number of dominant weights, which is the smallest number of mixture components with the sum of weights larger than 0.9,for the Mercer CAR SSM (black), the CK CAR SSM (red), and gsDP (green) at each location. The proposed CAR SSMs use more dominant weights than gsDP at most locations. In particular, the CK CAR SSM uses more dominant weights than gsDP at all locations. The gsDP captures the spatial structure through both the weights and atoms, while CAR SSMs fits the spatial dependencies with spatially varying weights only. A larger number of non-negligible weights provides more flexibility to model the desired heterogeneity across locations.

We also examine the prediction performance of the proposed CAR SSMs for 13 unobserved new sites. Figure 6 displays the posterior predictive densities for five selected locations and box plot of IAE values for 13 new sites. Predictive inference under the proposed models is quite accurate, even with the small sample sizes. The proposed methods borrow information across space effectively.

5.2   Precipitation over East Asia

We use the proposed CAR SSM models to analyze summer precipitation data collected for 216 sites over East Asia over 33 years, from 1979 to 2011. Figure 7 displays the 216 observation sites over East Asia. For each grid point, we have only 33 observations to construct probability density estimates, which is not enough for precise estimates. However, we expect the nearby densities are similar to each other and therefore expect better estimates if neighborhood information is utilized, as it is implemented in the
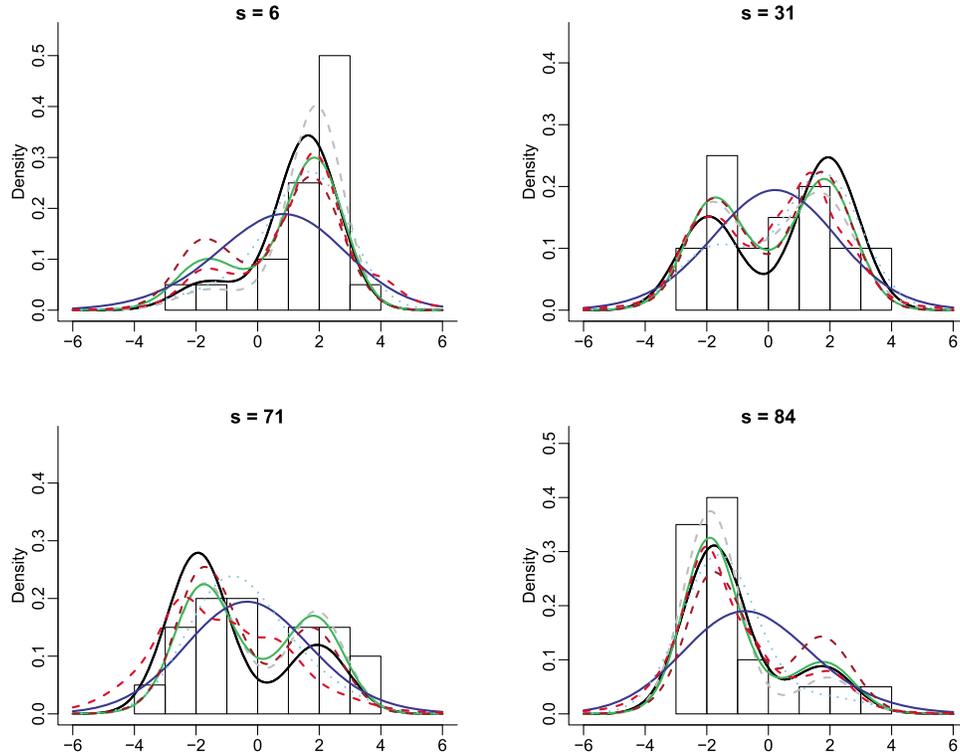
Figure 4: Simulation 2. Density estimates under simulation truth (17): true densities (black solid), $\text{MCS}(G_0, a, b, \tau^2, \rho)$ (gray dashed), $\text{CKCS}(G_0, a, b, \tau^2, \rho, B)$ (green solid), independent SSM (red dashed), sDP (blue solid), gsDP (skyblue dotted), and sPSBP (brown dashed).

CAR SSM models. Good inference for summer precipitation is of critical importance in South Korea. The country frequently suffers floods in summer that can cause large monetary loss as well as human casualties. For this reason, at the beginning of each season the Korean meteorology administration (KMA) publishes precipitation estimates as probabilities for three categories (below, normal and above) for subregions of Korea. The category "normal" means that in the coming season the precipitation will be in the middle 1/3 probability range of the past precipitations. The validity of these forecast probabilities hinges on good density estimates that make efficient use of the available data while honestly reporting relevant uncertainties.

In the following analysis we show how the proposed CAR SSM models can provide such density estimates (still without a temporal component for prediction). We carry out inference under the proposed models to obtain the desired density estimates using recorded summer (i.e. June, July and August) precipitation for grid points over East Asia ($21.25°\text{N} - 48.75°$, $101.25°\text{E} - 143.75°\text{E}$). We use the Climate Prediction Center
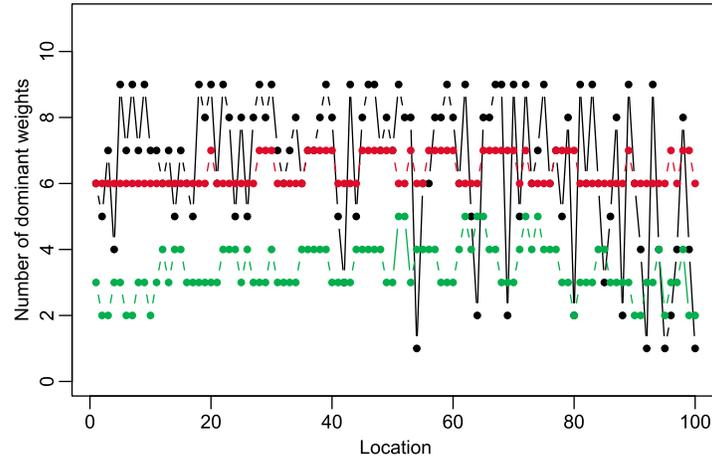
Figure 5: Simulation 2. The number of dominant weights under the Mercer CAR SSM (black), the CK CAR SSM (red) and under the gsDP (green) for data under simulation truth (17).

Merged Analysis of Precipitation (CMAP) data set. See Figure 7 for the grid points.

As before we fix the hyperparameters of the CAR SSMs. We use $B = 13$ so that the neighborhood of each grid point consists of 8 points around it; and $K = 16$ ($K = n/2$) for the discrete probability measures. Inference is based on 20,000 posterior samples from Markov chain Monte Carlo (MCMC) posterior simulation after a burn-in period of 10,000 samples.

Table 3 shows LPML and DIC values for fits under both CAR SSMs, $\mathrm{MCS}(G_0, a, b, \tau^2, \rho)$ and $\mathrm{CKCS}(G_0, a, b, \tau^2, \rho, B)$, and under sDP, and sPSBP. The Mercer CAR SSM is favored by both criteria. Both CAR SSMs provide better fits than the competitors. The results suggest that the proposed methods utilize information across space and estimate spatially varying densities effectively. Figure 8 shows the posterior mean estimates under the proposed CAR SSMs, sDP and sPBSP. Figure 9 shows the estimated 33.33 % and 66.66 % percentiles of the precipitation density functions, based on the density estimates, on each grid.

Figure 10 shows a predictive density estimate for a single site whose observations were excluded from the analyses. That is, no observations at this site were used for the fit. The predictive density estimates show the difference of Mercer CARSSM and CK CARSSM. The smaller mode in the density estimate of the Mercer CARSSM appears to be an effect from distant observations; while the density estimate of the CK CARSSM seems to have limited effect from distant observations.

We investigate sensitivity with respect to some prior specifications. We first repeated the analysis using $(a, b) \in \{(1, 10), (0.5, 20)\}$ and $\tau \in \{0.1, 0.5, 1, 4\}$ for both CAR SSMs. The values were chosen to correspond to prior mean weights shown in Figure 11. In all cases, the remaining parameters were fixed as in the simulation study. Figure 12
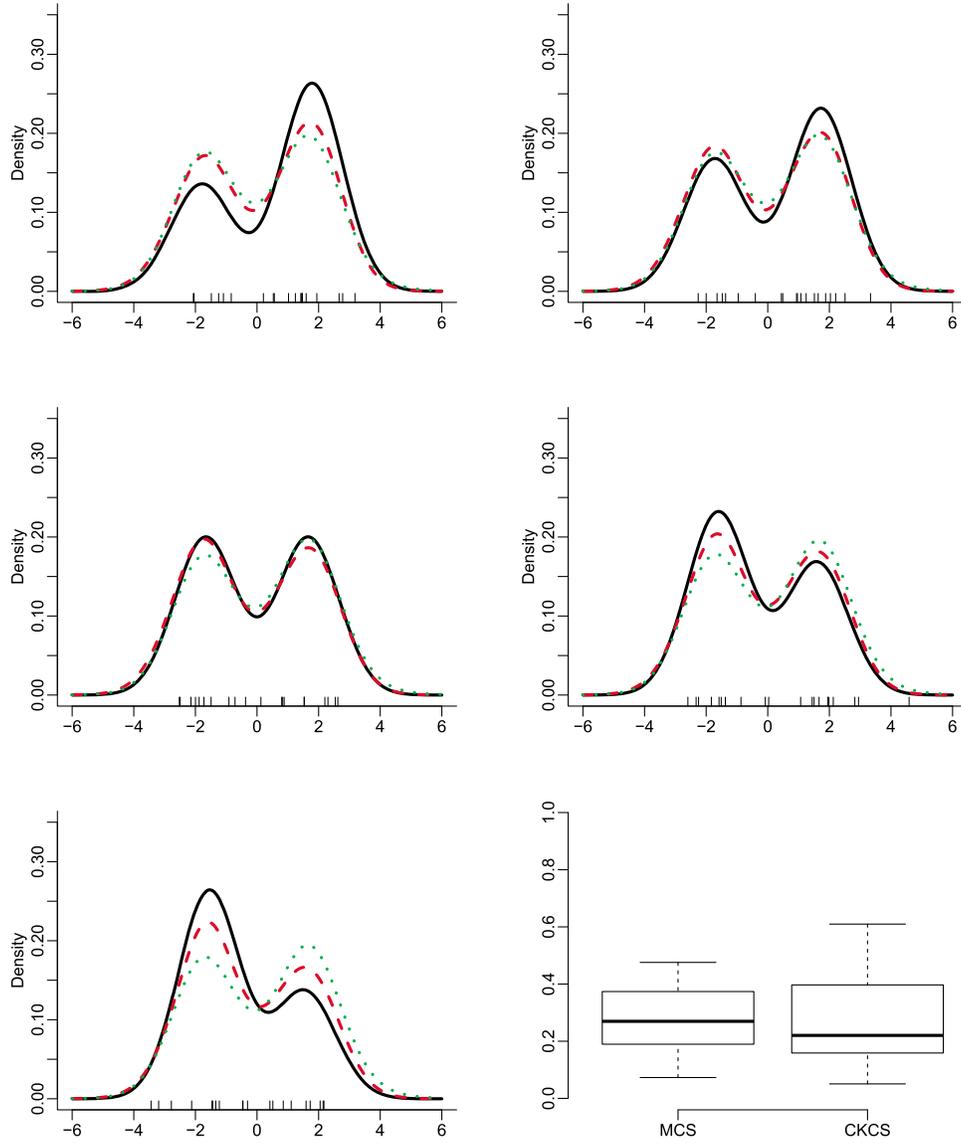
Figure 6: Simulation 2. Predictive density estimates and true densities for selected 5 new sites and box plot of IAE values for 13 new sites: true densities (black solid), $\text{MCS}(G_0, a, b, \tau^2, \rho)$ (green dotted) and $\text{CKCS}(G_0, a, b, \tau^2, \rho, B)$ (red dashed).

compares density estimates for a location over the Korean peninsula in Figure 7 and the posterior distribution for the number of clusters. Both are sensitive with respect to the choices of a, b and τ.
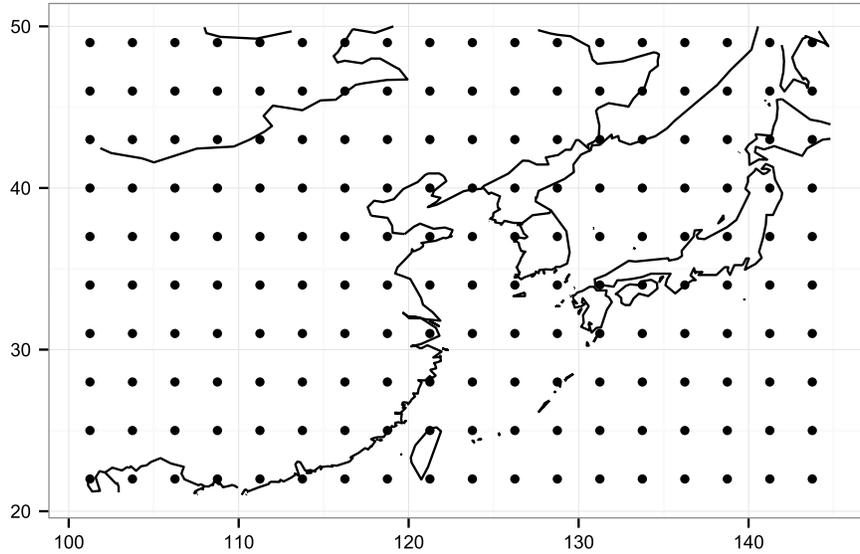
Figure 7: East Asia; resolution of grid point is $2.5° \times 2.5°$.

Models	LPML	DIC
sDP	−13847.85	27800.25
sPSBP	−12863.77	25721.10
Mercer CAR SSM	**−11264.91**	**22513.65**
CK CAR SSM	−11640.27	23269.93

Table 3: East Asia precipitation. Goodness of fit measures.

We also assessed sensitivity of posterior inference with respect to changing the type of kernel in the Mercer CAR SSM from Gaussian, to exponential, Cauchy, or Matérn (Whittle, 1954) kernels and varying the value of the B parameter that specifies the neighborhood size, from $B = 6.25$ to $B = 31.25$. The grid for B corresponds to neighborhood sizes from 4 to 12. We find little change in the reported density estimates. Summaries are shown in Figure 13. We conclude that posterior inference is robust with respect to the choice of the kernel and B.

6   Conclusion

In this paper we proposed dependent species sampling models as an alternative class of nonparametric priors for the estimation of spatially varying densities. The proposed CAR SSMs introduce the desired spatial dependence in the random weights of discrete probability measures. We use CAR models to construct such spatially dependent weights. Since the dependence is modeled through weights, the CAR SSMs are flexible to accommodate spatial variations of densities. This is confirmed by the simulation stud-
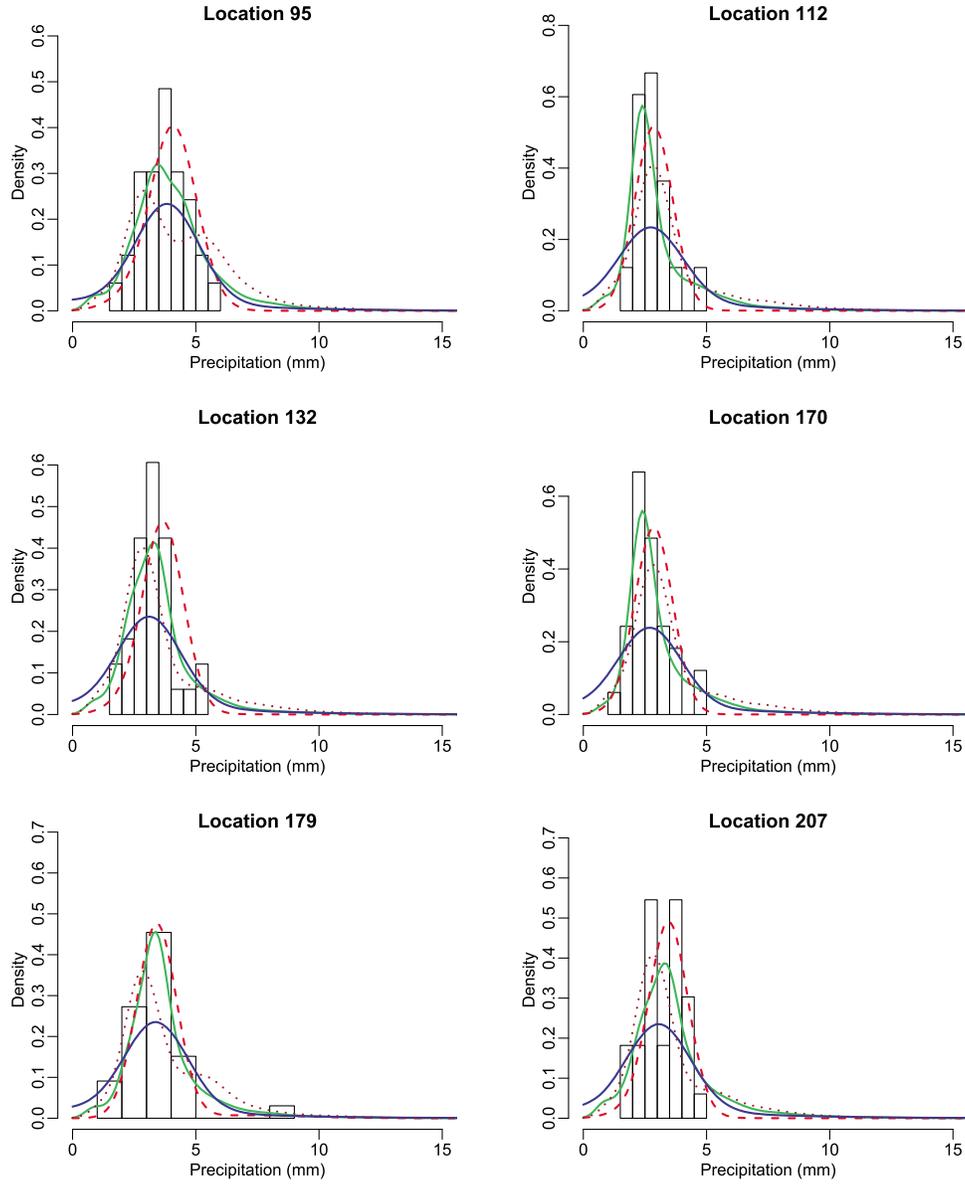
Figure 8: East Asia precipitation. Fits under the MCS$(G_0, a, b, \tau^2, \rho)$ (red dashed), CKCS$(G_0, a, b, \tau^2, \rho, B)$ (green solid), sDP (blue solid), and sPSBP (brown dotted).

ies and the real data example. An appealing feature of the proposed models is that the modeling of dependence remains intuitive, as the CAR models are a well-understood class of statistical models. The posterior sampling algorithm we described is greatly
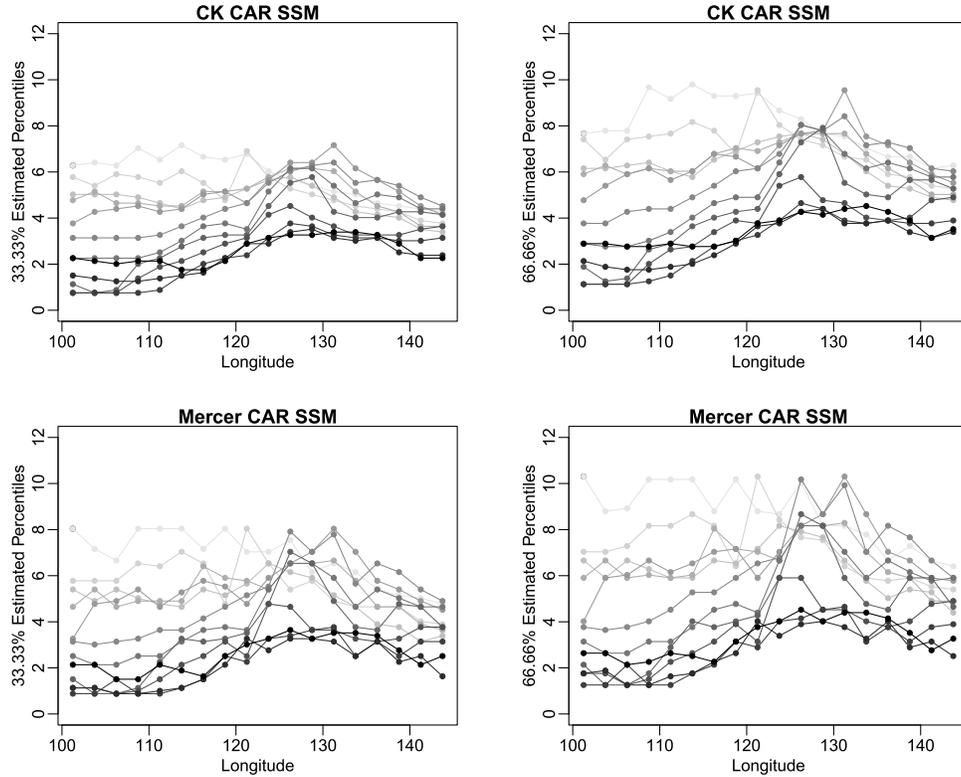
Figure 9: East Asia precipitation. A connected line represents the estimated 33.33 % (left) and 66.66 % (right) percentiles of the precipitation density functions at a latitude. The darkest and brightest lines correspond to the latitudes of 48.75° and 21.25° degrees, respectively.

simplified after introducing the data augmentation via auxiliary latent variables. CAR SSMs can be applied in any problem where a group of distributions indexed by locations or time need to be estimated simultaneously. One advantage of the posterior sampling algorithm for CAR SSMs is that the two components of the MCMC iterations, (i) cluster membership imputation of each observation and (ii) conditional sampling of the mixture weights $P_{s,h}$, present strict similarities with extensively studied computational procedures. The first component is present in a large portion of algorithms dedicated to Dirichlet mixture models, and the second component is nearly identical to widely used MCMC strategies for posterior sampling under logit binary regression models. In different words, procedures designed to solve these two important sampling problems can be plugged in and structure the MCMC for CAR SSMs. Additionally, the computational advantages associated to the conditional independence structure of CK CAR models can be inherited in the posterior sampling of the unnormalized weights $w_{s,h}$. In particular, the conditional independence structure in the models can be capitalized in
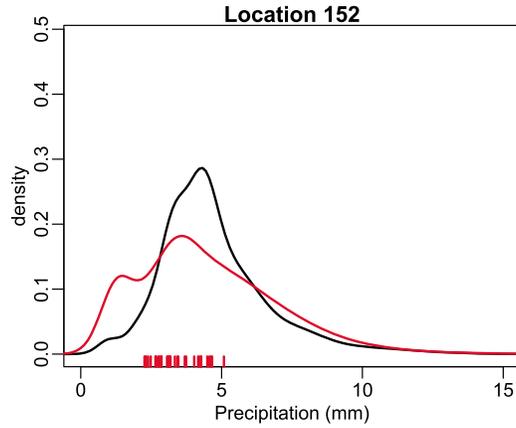
Figure 10: East Asia precipitation. Predictive density estimates under the proposed CAR SSMs; $\text{MCS}(G_0, a, b, \tau^2, \rho)$ (red solid) and $\text{CKCS}(G_0, a, b, \tau^2, \rho, B)$ (black solid). Vertical lines (red) on x-axis correspond to actual observations.
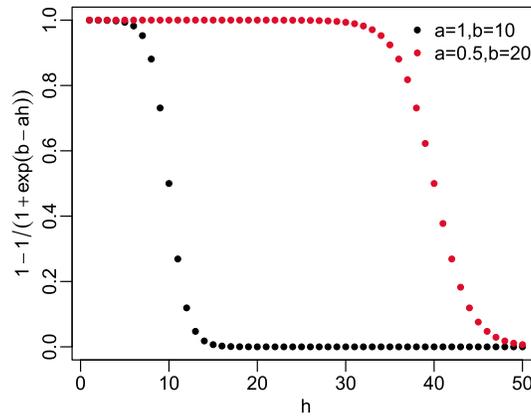

Figure 11: Mean weights (8) *a priori* versus h, under different choices of (a, b).

the implementation of multiprocessing or parallel posterior simulations.

In this paper we included an application to inference for precipitation density regression. But CAR SSMs can be applied in many other problems, including other climate variables such as temperature, ozone or CO_2, density estimation for housing prices, stock market prices etc.

In future work we plan to generalize the precipitation density regression with the inclusion of temporal components. An important remaining hurdle is computation time. New computational methods such as parallel MCMC or variational methods are promising directions. Other important generalizations are the inclusion of discrete and contin-
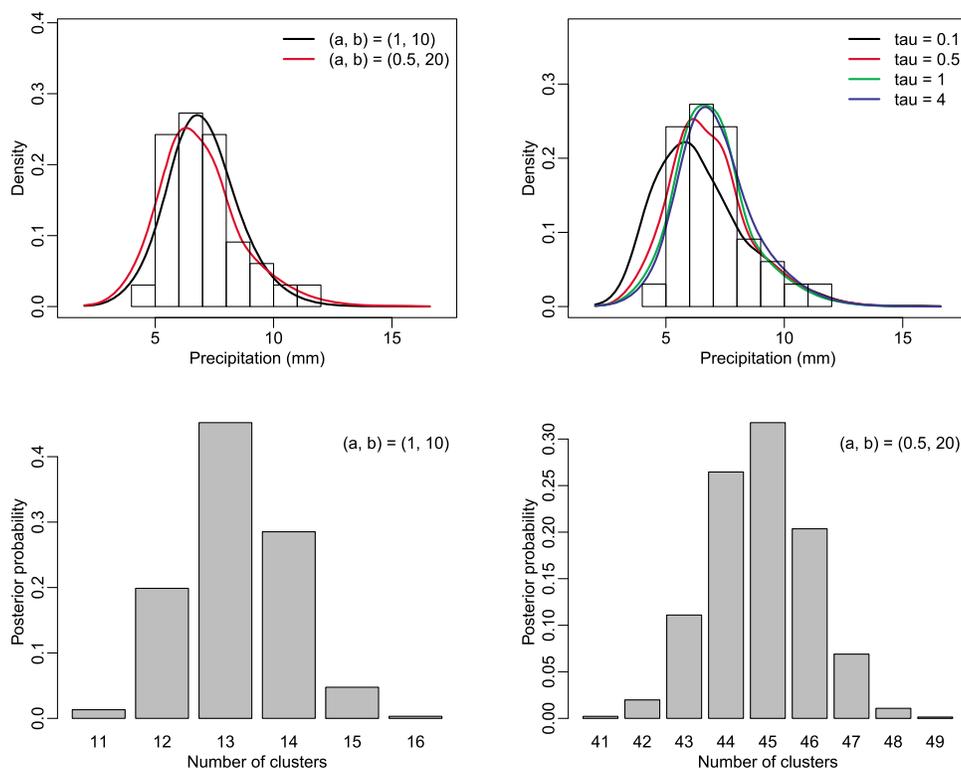
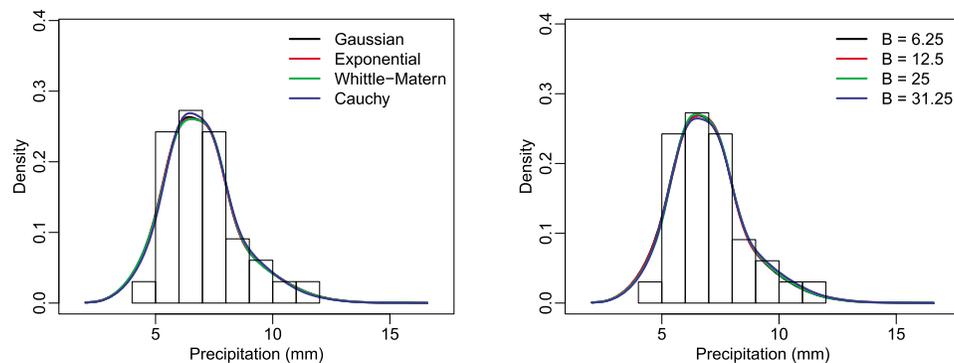Figure 12: Sensitivity. Inference under alternative hyperparameter choices (a, b, τ).

Figure 13: Sensitivity. Inference under alternative kernels for the Mercer CAR SSM and alternative B for the CK CAR SSM.

uous components in the distributions as well as a regression on covariates. The latter can easily be achieved by introducing a regression in the sampling model. Suppose response $y_{s_i,j}$ has associated covariate $x_{s_i,j}$. The response $y_{s_i,j}$ could then be modeled as $y_{s_i,j} = x'_{s_i,j}\beta + \epsilon_{s_i,j}$, where β is a regression coefficient and $\epsilon_{s_i,j}$ follows the mixtures of CAR SSMs. When the covariate is discrete, assuming only a few values, the mixtures of CAR SSMs could be applied with covariates being accommodated in the model in the same way as spatial locations. For example, see, e.g., Gelfand et al. (2005) and Duan et al. (2007).

Supplementary Material

Supplementary Material for Dependent Species Sampling Models for Spatial Density Estimation (DOI: 10.1214/16-BA1006SUPP; .pdf).

References

Airoldi, E. M., Costa, T., Bassetti, F., Leisen, F., and Guindani, M. (2014). "Generalized species sampling priors with latent beta reinforcements." *Journal of the American Statistical Association*, 109(508): 1466–1480. MR3293604. doi: http://dx.doi.org/10.1080/01621459.2014.950735. 382

Albert, J. H. and Chib, S. (1993). "Bayesian analysis of binary and polychotomous response data." *Journal of the American Statistical Association*, 88(422): 699–679. MR1224394. 387

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. London: Chapman & Hall. 383

Barnston, A. G., Mason, S. J., Goddard, L., Dewitt, D. G., and Zebiak, S. E. (2003). "Multimodel Ensembling in Seasonal Climate Forecasting at IRI." *Bulletin of the American Meteorological Society*, 84(12): 1783–1796. 379

Barrientos, A. F., Jara, A., and Quintana, F. A. (2012). "On the support of MacEachern's dependent Dirichlet processes and extensions." *Bayesian Analysis*, 7(2): 277–309. MR2934952. doi: http://dx.doi.org/10.1214/12-BA709. 380, 385

Bassetti, F., Crimaldi, I., and Leisen, F. (2010). "Conditionally identically distributed species sampling sequences." *Advances in Applied Probability*, 42(2): 433–459. MR2675111. doi: http://dx.doi.org/10.1239/aap/1275055237. 382

Besag, J. (1974). "Spatial interaction and the statistical analysis of lattice systems (with discussion)." *Journal of the Royal Statistical Society, Series B*, 36(2): 192–236. MR0373208. 381, 383

Besag, J., York, J., and Mollié, A. (1991). "Bayesian image restoration, with two applications in spatial statistics." *Annals of the Institute of Statistical Mathematics*, 43(1): 1–20. MR1105822. doi: http://dx.doi.org/10.1007/BF00116466. 381

Blackwell, D. and MacQueen, J. B. (1973). "Ferguson distributions via Pólya urn schemes." *The Annals of Statistics*, 1: 353–355. MR0362614. 381

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). "Deviance Information Criteria for Missing Data Models." *Bayesian Analysis*, 1(4): 651–674. MR2282197. doi: http://dx.doi.org/10.1214/06-BA122. 389

Chung, Y. and Dunson, D. B. (2011). "The local Dirichlet process." *Annals of the Institute of Statistical Mathematics*, 63(1): 59–80. MR2748934. doi: http://dx.doi.org/10.1007/s10463-008-0218-9. 380

Clayton, D. and Kaldor, J. (1987). "Empirical Bayes estimates of age-standardized relative risks for use in disease mapping." *Biometrics*, 43(3): 671–681. 384

Cressie, N. A. C. (1993). *Statistics for spatial data*. New York: John Wiley & Sons Inc. MR1239641. doi: http://dx.doi.org/10.1002/9781119115151. 383

Cressie, N. A. C. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Hoboken, New Jersey: John Wiley & Sons Inc. MR2848400. 383

De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). "An ANOVA model for dependent random measures." *Journal of the American Statistical Association*, 99(465): 205–215. MR2054299. doi: http://dx.doi.org/10.1198/016214504000000205. 380

Ding, M., He, L., Dunson, D. B., and Lawrence, C. (2012). "Nonparametric Bayesian Segmentation of a Multivariate Inhomogeneous Space-Time Poisson Process." *Bayesian Analysis*, 7(4): 813–840. MR3000015. doi: http://dx.doi.org/10.1214/12-BA727. 381

Duan, J. A., Guindani, M., and Gelfand, A. E. (2007). "Generalized Spatial Dirichlet Process Models." *Biometrika*, 94(4): 809–825. MR2416794. doi: http://dx.doi.org/10.1093/biomet/asm071. 380, 388, 402

Dunson, D. B. and Park, J.-H. (2008). "Kernel stick-breaking processes." *Biometrika*, 95(2): 307–323. MR2521586. doi: http://dx.doi.org/10.1093/biomet/asn012. 381

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *Annals of Statistics*, 1(2): 209–230. MR0350949. 380

Fuentes, M. and Reich, B. (2013). "Multivariate spatial nonparametric modelling via kernel processes mixing." *Statistica Sinica*, 23: 75–97. MR3076159. 381

Geisser, S. and Eddy, W. F. (1979). "A predictive approach to model selection." *Journal of the American Statistical Association*, 74(365): 153–160. MR0529531. 389

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). "Bayesian nonparametric spatial modeling with Dirichlet process mixing." *Journal of the American Statistical Association*, 100(471): 1021–1035. MR2201028. doi: http://dx.doi.org/10.1198/016214504000002078. 379, 380, 388, 402

Geman, S. and Geman, D. (1984). "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6: 721–741.   381

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian nonparametrics*. Springer Series in Statistics. New York: Springer-Verlag. MR1992245.   385

Griffin, J. E. and Steel, M. F. J. (2006). "Order-based dependent Dirichlet processes." *Journal of the American Statistical Association*, 110(473): 179–194.   MR2268037. doi: http://dx.doi.org/10.1198/016214505000000727.   380

Ishwaran, H. and James, L. F. (2001). "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96(453): 161–173.   MR1952729. doi: http://dx.doi.org/10.1198/016214501750332758.   382, 387

Ishwaran, H. and James, L. F. (2003). "Generalized weighted Chinese restaurant processes for species sampling mixture models." *Statistica Sinica*, 13(4): 1211–1235. MR2026070.   382

James, L., Lijoi, A., and Prünster, I. (2009). "Posterior Analysis for Normalized Random Measures with Independent Increments." *Scandinavian Journal of Statistics*, 36(1): 76–97.   MR2508332. doi: http://dx.doi.org/10.1111/j.1467-9469.2008. 00609.x.   382

James, L. F. (2008). "Large sample asymptotics for the two-parameter Poisson–Dirichlet process." In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, volume 3 of *IMS Lecture Notes-Monograph Series*, 187–199. Hayward, CA: Inst. Math. Statist.   MR2459225. doi: http://dx.doi.org/10.1214/074921708000000147.   382

Jang, G., Lee, J., and Lee, S. (2010). "Posterior consistency of species sampling priors." *Statistica Sinica*, 20(2): 581–593. MR2682631.   382

Jara, A., Lesaffre, E., De Iorio, M., and Quintana, F. (2010). "Bayesian semiparametric inference for multivariate doubly-interval-censored data." *Annals of Applied Statistics*, 4(4): 2126–2149.   MR2829950. doi: http://dx.doi.org/10.1214/10-AOAS368. 380

Jo, S., Lee, J., Müller, P., Quintana, F. A., and Trippa, L. (2016). "Supplementary material for dependent species sampling models for spatial density estimation." *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/16-BA1006SUPP.   385, 387

Kaiser, M. S. and Cressie, N. (2000). "The construction of multivariate distribution from Markov random fields." *Journal of Multivariate Analysis*, 73(2): 199–220. MR1763324. doi: http://dx.doi.org/10.1006/jmva.1999.1878.   383

Lee, D. (2011). "A comparison of conditional autoregressive models used in Bayesian disease mapping." *Spatial and Spatio-temporal Epidemiology*, 2(2): 79–89.   381

Lee, J., Quintana, F., Müller, P., and Trippa, L. (2013). "Defining Predictive Probability Functions for Species Sampling Models." *Statistical Science*, 28(2): 209–222. MR3112406.   382

Li, P., Banerjee, S., Hanson, T. A., and McBean, A. M. (2015). "Bayesian models for detecting difference boundaries in areal data." *Statistica Sinica*, 25: 385–402. MR3328821. 381, 385

Lijoi, A., Mena, R. H., and Prünster, I. (2005). "Hierarchical mixture modeling with normalized inverse-Gaussian priors." *Journal of the American Statistical Association*, 100(472): 1278–1291. MR2236441. doi: http://dx.doi.org/10.1198/016214505000000132. 382

Lo, A. Y. (1984). "On a class of Bayesian nonparametric estimates: I. Density estimates." *The Annals of Statistics*, 12: 351–357. MR0733519. doi: http://dx.doi.org/10.1214/aos/1176346412. 386

MacEachern, S. N. (1999). "Dependent Nonparametric Processes." In *ASA Proceedings of the Section on Bayesian Statistical Science*, 50–55. American Statistical Association. 380

MacEachern, S. N. (2000). "Dependent Dirichlet processes." Technical report, Department of Statistics, Ohio State University. 380

Müller, P. and Mitra, R. (2013). "Bayesian nonparametric inference—why and how." *Bayesian Analysis*, 8(2): 269–302. MR3066939. doi: http://dx.doi.org/10.1214/13-BA811. 380

Navarrete, C., Quintana, F. A., and Müller, P. (2008). "Some issues on nonparametric Bayesian modeling using species sampling models." *Statistical Modelling. An International Journal*, 8(1): 3–21. MR2750628. doi: http://dx.doi.org/10.1177/1471082X0700800102. 382

Nieto-Barajas, L. (2008). "A Markov gamma random field for modelling disease mapping data." *Statistical Modelling*, 8(1): 97–114. MR2750633. doi: http://dx.doi.org/10.1177/1471082X0700800107. 381

Pitman, J. (1996). "Some developments of the Blackwell–MacQueen urn scheme." In *Statistics, probability and game theory*, volume 30 of *IMS Lecture Notes-Monograph Series*, 245–267. Hayward, CA: Inst. Math. Statist. MR1481784. doi: http://dx.doi.org/10.1214/lnms/1215453576. 381, 382

Pitman, J. and Yor, M. (1997). "The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator." *Annals of Probability*, 25(2): 855–900. MR1434129. doi: http://dx.doi.org/10.1214/aop/1024404422. 382

Plummer, M. (2003). "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling." In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, 1–10. Austrian Association for Statistical Computing (AASC) and the R Foundation for Statistical Computing. 389

Reich, B. J. and Fuentes, M. (2007). "A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields." *Annals of Applied Statistics*, 1(1): 249–264. MR2393850. doi: http://dx.doi.org/10.1214/07-AOAS108. 381

Ren, L., Du, L., Carin, L., and Dunson, D. B. (2011). "Logistic Stick-Breaking Process." *Journal of Machine Learning Research*, 12: 203–239. MR2773552. 381

Rodríguez, A. and Dunson, D. B. (2011). "Nonparametric Bayesian models through probit stick-breaking processes." *Bayesian Analysis*, 6(1): 145–177. MR2781811. doi: http://dx.doi.org/10.1214/11-BA605.   380, 381, 387, 388

Rodríguez, A. and ter Horst, E. (2008). "Bayesian dynamic density estimation." *Bayesian Analysis*, 3(2): 339–365. MR2407430. doi: http://dx.doi.org/10.1214/08-BA313.   380, 390

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: Theory and applications*. Monographs on Statistics and Applied Probability, 104. Boca Raton: Chapman & Hall. MR2130347. doi: http://dx.doi.org/10.1201/9780203492024.   383

Scott, S. L. (2011). "Data augmentation, frequentist estimation, and the Bayesian analysis of multinomial logit models." *Statistical Papers*, 52(1): 639–650. MR2777069. doi: http://dx.doi.org/10.1007/s00362-009-0205-0.   387

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 4(2): 639–650. MR1309433.   380

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. London: Chapman & Hall. MR0848134. doi: http://dx.doi.org/10.1007/978-1-4899-3324-9.   388

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). "Bayesian measures of model complexity and fit (with discussion)." *Journal of the Royal Statistical Society, Series B*, 64(4): 583–639. MR1979380. doi: http://dx.doi.org/10.1111/1467-9868.00353.   389

Sun, D., Tsutakawa, R. K., and Speckman, P. L. (1999). "Posterior distribution of hierarchical models using CAR(1) distributions." *Biometrika*, 86(2): 341–350. MR1705418. doi: http://dx.doi.org/10.1093/biomet/86.2.341.   384, 385

Tippett, M. K., Barnston, A. G., and Robertson, A. W. (2007). "Estimation of Seasonal Precipitation Tercile-Based Categorical Probabilities from Ensembles." *Journal of Climate*, 20(10): 2210–2228.   379

Whittle, P. (1954). "On stationary processes in the plane." *Biometrika*, 41: 434–449. MR0067450.   397

Wu, Y. and Ghosal, S. (2008). "Kullback Leibler property of kernel mixture priors in Bayesian density estimation." *Electronic Journal of Statistics*, 2: 298–331. MR2399197. doi: http://dx.doi.org/10.1214/07-EJS130.   385