

## **BAYESIAN GAUSSIAN MODELS FOR POINT REFERENCED SPATIAL AND SPATIO-TEMPORAL DATA**

K. SHUVO BAKAR\*

*Centre for Social Research and Methods*  
*The Australian National University, Canberra, 2601, Australia.*  
*Email: shuvo.bakar@anu.edu.au*

PHILIP KOKIC

*Climate Change Institute*  
*The Australian National University, Canberra, 2601, Australia.*  
*Email: philip.kokic@anu.edu.au*

### SUMMARY

When data is correlated both spatially and temporally, spatial and spatio-temporal modelling is useful for meaningful interpretation of the parameters of the covariates and for reliable predictions. In this paper we discuss some modelling strategies for point referenced spatial and spatio-temporal data. We describe Gaussian models in this context and use Bayesian hierarchical approaches for model based inference and predictions through the Markov chain Monte Carlo (MCMC) algorithm. Yearly average precipitation data from Western Australia is used to illustrate the models.

*Keywords and phrases:* Bayesian inference, spatial and spatio-temporal prediction, dynamic model.

## **1 Introduction**

Practitioners and researchers are increasingly analysing data that are geographically referenced, and both spatially and temporally correlated. For example, in environmental science, there is a growing need to model climate data (e.g., temperature, precipitations) that are measured at a network of monitoring stations with regular or irregular time-series of observations at each location. One of the important intentions of modelling spatially referenced data is to obtain spatial predictions at locations where no data has been observed, and also for spatio-temporal data to obtain forecasts to future time-points. With the advancement of Bayesian methods and high powered computing in the last few decades, spatial and spatio-temporal statistical modelling have become feasible and indeed popular tools for analysing such data.

In this paper we present a brief overview of the models used for spatially referenced data. We elaborate the models through real-life practical data applications using R (R Core Team, 2017), a popular programming language tool for statistical applications. The models discussed in this

---

\* Corresponding author

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

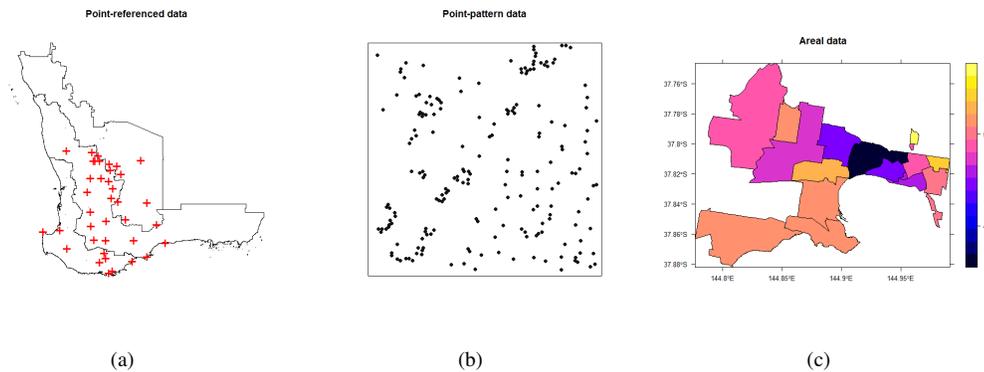


Figure 1: Three types of spatial data: (a) Temperature monitoring locations in Western Australia, (b) California redwood trees in a sampling region, and (c) Female population distribution in western suburbs of Melbourne.

paper are hierarchical in nature and use Bayesian methods to obtain model-based inference and predictions. We avoid technical theorems and proofs, and refer to those in books and articles when necessary. In addition, we present a brief overview of Bayesian hierarchical techniques for readers.

Perhaps the most well-known book on spatial statistical analysis is Cressie (1993). The book includes high level mathematics of the subject, but represents less on modern hierarchical Bayesian methods. However, the recent books of Cressie and Wikle (2011), Banerjee et al. (2014) and Diggle (2013) bring together modern aspects of spatial and spatio-temporal modelling theory. They cover the effective use of Markov chain Monte Carlo (MCMC) computing for analysing spatial and spatio-temporal data with complex and sophisticated hierarchical models, and extend the methodology to non-Gaussian and multivariate data. The books also address the modelling problems associated with spatial and temporal misalignments of data.

Spatially dependent data are often classified into three major types (Banerjee et al., 2014): (i) point-referenced data (ii) point pattern data, and (iii) areal data. Point referenced data is also known as the geostatistical data, where the random observation is measured at a location, which varies continuously over the study region. Theoretically the number of locations in the study region is infinite. For point pattern data the study domain is random and its index set gives the locations of random events that describe the observed spatial point patterns. The third type of spatial data is known as areal data, where the study domain is a fixed subset with regular or irregular shapes, but partitioned into a finite number of areal units with well-defined boundaries. Figure 1(a) is an example of the point-referenced data where precipitation is measured at 38 high-quality weather monitoring locations in Western Australia (WA). Here the monitoring locations are fixed; however, the precipitation measurements are random. An example of point-pattern data is shown in Figure 1(b), where Californian redwood trees are distributed over a sampling region (Strauss, 1975). Here the sampling region is fixed; however, the position of the trees are random. Figure 1(c) is an example of areal data, where

the distribution of the percentage of females in the population are observed at postcode levels in the western part of Melbourne. This paper will focus on methods for analysing point referenced spatial data only. Furthermore, we restrict attention to Gaussian modelling strategies for analysing point referenced data.

In 1951 D. G. Krige, a South African mining engineer, developed a method for spatial prediction using a small amount of Gaussian data. Later, Matheron (1963) formalised the method and named it ‘‘Kriging’’ in honour of D. G. Krige. Kriging is basically a method of ‘optimal’ spatial prediction at unobserved locations when the spatial covariance function is known. Krige developed a range of statistical estimation methods for mining applications termed geo-statistical techniques. One of his aims was to quantify the spatial variability of a variable of interest that varied spatially. He mentioned that the spatial technique allows one to solve some important problems in mining exploration.

The rest of this paper is organised as follows. In Section 2 we describe the basics of point referenced data and process. The technique ‘kriging’ is described with a practical data application. Section 3 explains the basic idea of Bayesian analysis and modelling. This is followed by a description of spatial and spatio-temporal modelling techniques with examples in Sections 4 and 5. In Section 6 we explain the concept of dynamic modelling with example. Finally, in Section 7 we provide a conclusion with some discussions on the big- $n$  problem for spatial and spatio-temporal hierarchical models.

## 2 Spatial Processes and Kriging

Let  $\mathbf{s}$  be any spatial location within a study region  $\mathcal{S} \subset \mathfrak{R}^d$ , where  $\mathfrak{R}^d$  is the  $d$  dimensional space (e.g., Euclidean). The spatial process is written as:

$$Z(\mathbf{s}) : \mathbf{s} \in \mathcal{S} \subset \mathfrak{R}^d,$$

where  $Z(\mathbf{s})$  is a (univariate) measurement of the attribute of interest at location  $\mathbf{s}$ . Notationally, for  $n$  different locations, the measurements can be written as,  $\mathbf{Z}(\mathbf{s}) = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$ .

Before further model discussions, we define the terms stationarity and isotropy. The idea of stationarity comes from the general theory of stochastic processes. Consider two spatial locations,  $\mathbf{s}$  and  $\mathbf{s} + \mathbf{h}$ , where  $\mathbf{h} \in \mathfrak{R}^d$ . A spatial process is called strictly stationary if, for any given  $n \geq 1$ , any set of  $n$  locations  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$  and for any  $\mathbf{h} \in \mathfrak{R}^d$ , the joint distributions of  $Z(\mathbf{s})$  and  $Z(\mathbf{s} + \mathbf{h})$  are same, i.e.:

$$\pi(Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)) = \pi(Z(\mathbf{s}_1 + \mathbf{h}), \dots, Z(\mathbf{s}_n + \mathbf{h})).$$

Assume that the process has a valid covariance function  $\text{Cov}(Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h}))$ . The process  $Z(\mathbf{s})$  is known as second-order stationary (also known as weak stationary), if

$$\text{Cov}(Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h})) = C(\mathbf{h}), \quad \forall \mathbf{s} \in \mathcal{S}, \quad \mathbf{h} \in \mathfrak{R}^d,$$

where  $C(\mathbf{h})$  is a function that depends on the difference in the spatial locations,  $\mathbf{h}$ , only. For non-stationary spatial process either or both the above type of stationarity conditions do not hold. The

function  $C(\mathbf{h})$ , is known as a covariogram. The intrinsic stationary defines only the first and second moments of the differences  $Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})$  but not anything about their distributions. We can write the covariance function  $C(\mathbf{h})$  as:

$$C(\mathbf{h}) = \sigma^2 \kappa(\mathbf{s}_i, \mathbf{s}_j; \Phi),$$

where  $\sigma^2$  is the common variance term and  $\kappa(\mathbf{s}_i, \mathbf{s}_j; \Phi)$  is the spatial correlation between locations  $\mathbf{s}_i$  and  $\mathbf{s}_j$  with smoothness and decay parameters  $\Phi = (\nu, \phi)'$ .

A spatial process  $Z(\mathbf{s})$  is called isotropic if its covariance function  $C(\mathbf{h})$  depends only on the distance  $|\mathbf{h}|$  between the two locations  $\mathbf{s}$  and  $\mathbf{s} + \mathbf{h}$ . Covariance functions of anisotropic processes exhibit different behaviour in different directions. However, isotropic processes are popular because of their simplicity, and easy interpretability. There are some common parametric isotropic covariance models available in spatial analysis (Banerjee et al., 2004, p.27). These models are in simple parametric form and are available as candidates for the variogram  $2\gamma(\mathbf{h})$ , where the variogram is defined as:

$$E[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})]^2 = \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = 2\gamma(\mathbf{h})$$

The term  $\gamma(\mathbf{h})$  is also known as the semivariogram (Banerjee et al., 2004, p.22).

A special case of the function  $\kappa(\mathbf{s}_i, \mathbf{s}_j; \Phi)$  is the parametric Matérn class correlation function (Matérn, 1986) defined as:

$$\kappa(\mathbf{s}_i, \mathbf{s}_j; \phi, \nu) = \frac{1}{2^{\nu-1} \Gamma(\nu)} (2\sqrt{\nu} \|\mathbf{s}_i - \mathbf{s}_j\| \phi)^\nu K_\nu(2\sqrt{\nu} \|\mathbf{s}_i - \mathbf{s}_j\| \phi), \quad \phi > 0, \nu > 0,$$

where  $K_\nu$  is the modified Bessel function of the second kind with order  $\nu$ , and  $\|\mathbf{s}_i - \mathbf{s}_j\|$  is the distance between locations  $\mathbf{s}_i$  and  $\mathbf{s}_j$ . Special cases of this correlation function include the exponential and Gaussian form, obtained by setting  $\nu = 1/2$  or  $\nu \rightarrow \infty$ , respectively. The exponential form is:

$$\kappa(\mathbf{s}_i, \mathbf{s}_j; \phi) = \exp(-\phi \|\mathbf{s}_i - \mathbf{s}_j\|), \quad \phi > 0,$$

and the Gaussian form is:

$$\kappa(\mathbf{s}_i, \mathbf{s}_j; \phi) = \exp(-\phi \|\mathbf{s}_i - \mathbf{s}_j\|^2), \quad \phi > 0.$$

We can also write the term  $\kappa(\mathbf{s}_i, \mathbf{s}_j; \phi)$  as a parametric spherical correlation function:

$$\kappa(\mathbf{s}_i, \mathbf{s}_j; \phi) = 1 - \frac{3}{2} \phi \|\mathbf{s}_i - \mathbf{s}_j\| + \frac{1}{2} (\phi \|\mathbf{s}_i - \mathbf{s}_j\|)^3, \quad 0 < \|\mathbf{s}_i - \mathbf{s}_j\| < 1/\phi.$$

A summary of the covariance functions are given in Banerjee et al. (2004)[p.24]. For more details see Banerjee et al. (2014); Cressie (1993); Cressie and Wikle (2011).

## 2.1 Kriging

In spatial statistics the goal is to obtain an optimal spatial prediction (or kriging) for the variable  $Z$  at a location  $\mathbf{s}_0$ , where no data has been observed. For squared error loss the best linear prediction

minimizes  $E[Z(\mathbf{s}_0) - (\sum l_i Z(\mathbf{s}_i) + \delta_0)]^2$  over  $\delta_0$  and  $l_i$ . We let  $\sum l_i = 1$  for a constant mean process and the term  $\delta_0$  becomes zero under unbiased and intrinsic stationarity assumptions. Using Lagrange multipliers for constrained optimisation it can be shown that this becomes a function of the semivariogram  $\gamma(\mathbf{h})$  (Cressie, 1993). Thus, we obtain an ordinary kriging estimate by estimating  $\gamma$ . The beauty of ordinary kriging is that it does not require any distributional assumption for the  $Z(\mathbf{s}_i)$ . However, it comes with limitations of a constant mean, and it fails to estimate the uncertainty of the prediction.

Kriging can also be structured in a Gaussian context. Let

$$\mathbf{Z} = \mu \mathbf{1} + \boldsymbol{\eta}, \quad \text{where } \boldsymbol{\eta} \sim N(\mathbf{0}, \Sigma),$$

where with a nugget effect,  $\Sigma = \sigma^2 \kappa(\cdot) + \tau^2 \mathbf{I}$ , and without a nugget effect,  $\Sigma = \sigma^2 \kappa(\cdot)$ . Note that the term  $\sigma^2$  is known as the sill or spatial variance and the nugget effect ( $\tau^2$ ) measures the variation of the Gaussian white-noise. If we include covariates  $\mathbf{x} = (x(\mathbf{s}_1), \dots, x(\mathbf{s}_n))'$  in the model, then:

$$\mathbf{Z} = X\boldsymbol{\beta} + \boldsymbol{\eta}, \quad \text{where } \boldsymbol{\eta} \sim N(\mathbf{0}, \Sigma),$$

$X$  is an  $n \times p$  design matrix and  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of parameters. The spatial prediction at  $\mathbf{s}_0$  is:

$$E[Z(\mathbf{s}_0) | \mathbf{z}] = \mathbf{x}'_0 \hat{\boldsymbol{\beta}} + C' \Sigma^{-1} (\mathbf{Z} - X\hat{\boldsymbol{\beta}}),$$

where  $C' = \text{Cov}(Z(\mathbf{s}), Z(\mathbf{s}_0))$  and  $\mathbf{x}_0$  is the observed covariates at the prediction location  $\mathbf{s}_0$ . In this case the prediction error can be estimated.

## 2.2 Example

Consider the average precipitation data in Figure 1(a) collected from September–December 2009 at 38 locations in Western Australia (WA). To check the presence of the spatial correlation in these data, we first plot the empirical variogram. Note that the data consists of 8 missing values. These are removed, but can be treated as a part of the prediction problem. However, in Sections 4 and 5 a Bayesian approach can also be used to address the missing value problem for both spatial and spatio-temporal models. Instead of analysing the actual precipitation observations, which are highly skewed, the data are first transformed using a logarithm function. This ensures that the Gaussian assumption is approximately satisfied (see details in Section 5.2).

Figure 2(a) shows the semivariogram plot. Observe that, as expected, the semi-variance increases with distance. It is not straightforward to interpret a variogram; however, we can have an idea on the spatial range and sill ( $\sigma^2$ ) from the plot. For example, from Figure 2(a) one can deduce that the effective range is 250 kms with a sill of about 0.5 and nugget variance of 0.02. Note that there is a cyclicity in the semivariogram, which may occurred due to underlying geological periodicity or a limited amount of data (here we only have 30 data points). Clearly the data exhibits spatial correlation so we can apply Kriging to predict to locations where there are no measurements. For example, predictions to a regular grid covering the region can be made. Figure 2(b) shows a map of the spatial prediction based on ordinary kriging (i.e., no covariates are used in the kriging method). We see that the predictions approximately match the actual data in different locations. However, in

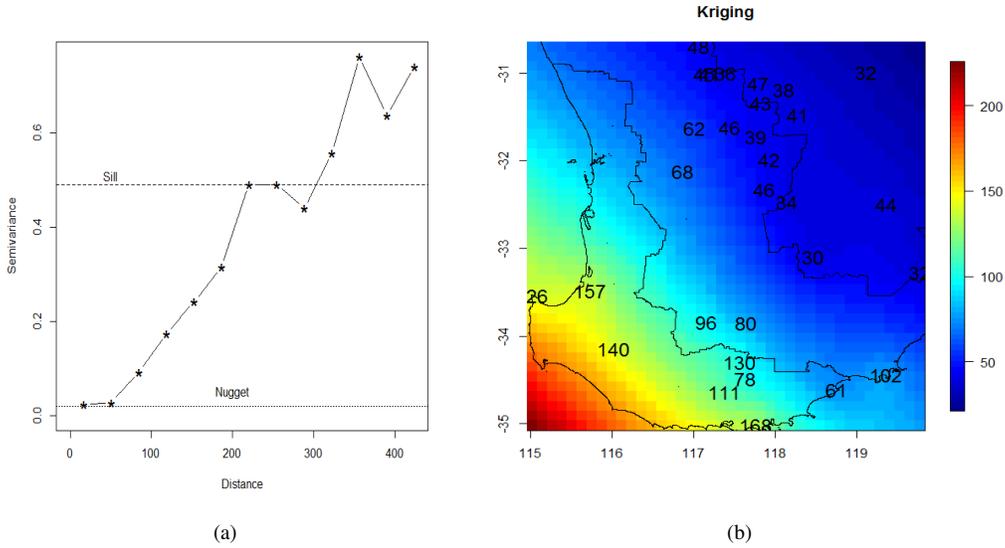


Figure 2: (a) Semivariogram plot of the temperature data measured in Western Australia. (b) Spatial prediction at grid points. The actual measurements are superimposed in the plot.

some places the predictions are inaccurate. This can possibly be improved by further modelling, e.g., the use of covariates.

### 3 Bayesian Modelling

A Bayesian approach is often preferable to the traditional frequentist approaches since it lets us deal with uncertainty in the model parameters. In the Bayesian framework prior information of the parameters (represented by their distribution) is updated using Bayes theorem to obtain their posterior distribution. Let  $f(\mathbf{z} | \boldsymbol{\theta})$  be the likelihood function of parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$  based on the observed data  $\mathbf{z} = (z_1, \dots, z_n)'$ . If we have a prior distribution  $\pi(\boldsymbol{\theta})$  for the parameters, then from Bayes theorem the posterior distribution is:

$$\pi(\boldsymbol{\theta} | \mathbf{z}) = \frac{f(\mathbf{z} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\Theta} f(\mathbf{z} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}},$$

where the denominator of the above equation is the integral over the parameter space  $\Theta$ . This integral is also known as the marginal likelihood of the data  $\mathbf{z}$  and it is free of the parameters  $\boldsymbol{\theta}$ , hence it can be treated as a constant. That is why the posterior distribution is often written as proportional to the product of the likelihood and the prior distribution, i.e.,

$$\pi(\boldsymbol{\theta} | \mathbf{z}) \propto f(\mathbf{z} | \boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

Finally, inference is based on the updated information in the posterior distribution.

The choice of the prior distributions is one of the important steps in any Bayesian analysis. The most attractive choice of prior distributions should be the one that best takes into account any previous knowledge. These types of priors are known as informative priors. However, there is often no clear choice of prior distributions for unknown parameters. For various likelihood functions there exists prior distributions that lead to a posterior distribution from the same distribution family as the prior. These types of prior distributions are known as the conjugate priors. For example, the prior distribution conjugate to a Bernoulli likelihood is a Beta distribution, and for a Poisson likelihood the conjugate prior distribution is a Gamma distribution. Choice of this type of prior distribution is attractive because of its straightforward computation. However, for many posterior and prior distribution families this type of prior might not be available. To overcome this conjugacy problem, priors with large variance are often used. These are referred to as noninformative priors. For example, the inverse-gamma distribution with large variance is often used for the non-negative variance parameters in Bayesian models. For the noninformative prior, the uniform distribution is also commonly used. In many situations Jeffrey's rule is applied to obtain the noninformative prior distribution (Gelman et al., 2014)[p. 36], that is, it is proportional to the square root of the determinant of Fisher's information matrix. In this paper, we illustrate the examples in Sections 4.2, 5.2 and 6.1 considering conjugate proper prior with large variance.

For making posterior inference from a Bayesian model, Markov chain Monte Carlo (MCMC) methods are popular. The MCMC generates a sequence of samples from the joint probability distribution of random variables. For example, the MCMC algorithm generates a Markov chain,  $\{\boldsymbol{\theta}^{(j)}\}_{j=1}^m$ , from the posterior distribution  $\pi(\boldsymbol{\theta} | \mathbf{z})$ , where  $m$  is the number of MCMC samples. Then, the samples are used to estimate integrals using Monte Carlo methods as:

$$\hat{E}(h(\boldsymbol{\theta})) = \int_{\Theta} h(\boldsymbol{\theta})\pi(\boldsymbol{\theta} | \mathbf{z})d\boldsymbol{\theta},$$

where  $h(\boldsymbol{\theta})$  is a function of  $\boldsymbol{\theta}$ . Different MCMC simulation techniques are available, for example, Metropolis and Metropolis-Hastings algorithm (Hastings, 1970), random walk Metropolis, and Gibbs sampling (Gelfand and Smith, 1990). Details of these techniques are found in Gelman et al. (2014, Ch. 11). To check the convergence of the MCMC algorithm several diagnostic methods exist (Gelman and Rubin, 1992; Roberts et al., 1997); however, a time series plot of the MCMC samples is a very popular way to check the status of the convergence visually.

## 4 Bayesian Spatial Modelling

In this section we present Bayesian spatial models and explain how spatial prediction (kriging) at unobserved locations can be derived from such models. A vast literature on spatial prediction from a Bayesian perspective can be found in Banerjee et al. (2014)[ch. 5] and Cressie and Wikle (2011)[ch. 4.1]. In this paper we will restrict attention to univariate linear models. See Banerjee et al. (2004)[ch. 7] for further details on bivariate and multivariate spatial models, although the spatio-temporal model described in Section 5 can also be treated as a multivariate spatial model.

A linear univariate spatial model can be written as:

$$\mathbf{Z} = X\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\epsilon}. \quad (4.1)$$

Here, the residual is partitioned into spatial  $\eta(\mathbf{s})$  and white-noise  $\epsilon(\mathbf{s})$  components. As defined earlier  $\eta(\mathbf{s})$  is a stationary Gaussian process that contains the partial sill ( $\sigma^2$ ), smoothness parameter ( $\nu$ ), and spatial decay parameter ( $\phi$ ). White-noise adds the nugget ( $\tau^2$ ) effect. Under the Bayesian paradigm, setting  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \tau^2, \phi, \nu)'$ , and defining the prior distribution  $\pi(\boldsymbol{\theta})$ , the posterior is  $\pi(\boldsymbol{\theta} | \mathbf{z}) \propto f(\mathbf{z} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})$ . Typically one chooses independent priors, i.e.,  $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(\phi)\pi(\nu)$ . Commonly used candidates are the multivariate normal for  $\boldsymbol{\beta}$ , and inverse gamma for  $\sigma^2$  and  $\tau^2$ . Prior distributions for  $\phi$  and  $\nu$  depend on the choice of the  $\kappa(\cdot)$  function and usually a uniform or discrete prior is assigned. Bayesian models are hierarchical in nature. This means we can split the univariate spatial model defined in eq. (4.1) into three levels (Cressie and Wikle, 2011)[p. 21]:

$$\text{Data Model: } \mathbf{Z} | \boldsymbol{\theta}, \boldsymbol{\eta} \sim N(X\boldsymbol{\beta} + \boldsymbol{\eta}, \tau^2\mathbf{I})$$

$$\text{Process Model: } \boldsymbol{\eta} | \sigma^2, \phi, \nu \sim GP(\mathbf{0}, \sigma^2\kappa(\cdot))$$

Parameter Model: Priors on  $\boldsymbol{\theta}$

The marginal posterior  $\pi(\boldsymbol{\theta} | \mathbf{z})$  under the hierarchical setting is  $f(\mathbf{z} | \boldsymbol{\theta}, \boldsymbol{\eta})\pi(\boldsymbol{\eta} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})$ . However, Metropolis updates are needed for  $\sigma^2$ ,  $\tau^2$ ,  $\phi$  and  $\nu$ , because closed distributional forms can not be obtained for these parameters. The Metropolis MCMC updates often converge faster than full Gibbs (Gelfand and Smith, 1990) updates for a conditional model; however, precise tuning is required to obtain rapid convergence.

The model defined in eq. (4.1) can be restructured in a different way. A conjugate prior can be used for the variance parameters that allows for a full conditional model. Thus,

$$\mathbf{Z} = \mathbf{O} + \boldsymbol{\epsilon} \quad (4.2)$$

$$\mathbf{O} = X\boldsymbol{\beta} + \boldsymbol{\eta}, \quad (4.3)$$

where  $\mathbf{O} = (O(\mathbf{s}_1), \dots, O(\mathbf{s}_n))'$  is the true process corresponding to  $\mathbf{Z}$ , and in the second stage we model the true process as a function of the covariates  $X$ . The hierarchical structure for this restructured model is:

$$\text{Data Model: } \mathbf{Z} | \mathbf{O}, \boldsymbol{\theta} \sim N(\mathbf{O}, \tau^2\mathbf{I})$$

$$\text{Process Model: } \mathbf{O} | \boldsymbol{\beta}, \sigma^2, \phi, \nu \sim GP(X\boldsymbol{\beta}, \sigma^2\kappa(\cdot))$$

Parameter Model: Priors on  $\boldsymbol{\theta}$

In this case full conditional distributions of the parameters  $\sigma^2$  and  $\tau^2$  can be obtained in closed form; however, it still requires a Metropolis update for the smoothness  $\nu$  parameter. Estimation of  $\nu$  is often complex and a popular choice is to replace  $\nu$  with  $1/2$ , which turns the Matérn into an exponential covariance model. The spatial decay parameter  $\phi$  also requires a Metropolis update. Alternatively, add-hoc estimated values can be calculated based on the distances between the locations. The effective distance (of spatial dependency) is computed as  $[-\log(0.05)/d_{\max}]$  (Finley et al., 2007), where  $d_{\max}$  is the maximum distance calculated from the model fitting locations.

## 4.1 Prediction Details

Suppose we want to predict  $Z$  at location  $\mathbf{s}_0$ . The covariates at this location are  $\mathbf{x}_0$ . The posterior predictive distribution for  $Z(\mathbf{s}_0)$  is obtained by integrating over the parameters with respect to the joint posterior distribution:

$$\pi(Z(\mathbf{s}_0) | \mathbf{z}, X, \mathbf{x}_0) = \int \pi(Z(\mathbf{s}_0) | \mathbf{z}, \boldsymbol{\theta}, X, \mathbf{x}_0) \pi(\boldsymbol{\theta} | \mathbf{z}, X) d\boldsymbol{\theta}. \quad (4.4)$$

This integral can be estimated through Monte Carlo simulation using composition with Gibbs samples. The prediction details for the restructured model in eqs. (4.2) and (4.3) can be written as:

$$\begin{aligned} \pi(Z(\mathbf{s}_0) | \mathbf{z}, X, \mathbf{x}_0) &= \int \pi(Z(\mathbf{s}_0) | O(\mathbf{s}_0), \mathbf{O}, \mathbf{z}, \boldsymbol{\theta}, X, \mathbf{x}_0) \pi(O(\mathbf{s}_0) | \mathbf{O}, \mathbf{z}, \boldsymbol{\theta}, X, \mathbf{x}_0) \\ &\quad \times \pi(\mathbf{O}, \boldsymbol{\theta} | \mathbf{z}, X) dO(\mathbf{s}_0) d\mathbf{O} d\boldsymbol{\theta} \end{aligned} \quad (4.5)$$

According to (4.2),

$$Z(\mathbf{s}_0) \sim N(O(\mathbf{s}_0), \sigma_\epsilon^2),$$

where  $O(\mathbf{s}_0)$  is the true prediction values at location  $\mathbf{s}_0$ , and the samples for  $O(\mathbf{s}_0)$  are obtained from:

$$\begin{pmatrix} O(\mathbf{s}_0) \\ \mathbf{O} \end{pmatrix} \sim N \left[ \begin{pmatrix} X(\mathbf{s}_0)\boldsymbol{\beta} \\ X\boldsymbol{\beta} \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & S_{12} \\ S_{21} & S \end{pmatrix} \right],$$

where  $S_{12}$  is  $1 \times n$  with  $i^{\text{th}}$  entry given by,  $\kappa(\mathbf{s}_i, \mathbf{s}'; \phi, \cdot)$ ,  $i = 1, \dots, n$  and  $S_{12} = S'_{21}$ . In summary, we draw samples  $\boldsymbol{\theta}^{(j)}$ ,  $j \geq 1$ , from the full conditional posterior distributions and then draw  $O^{(j)}(\mathbf{s}_0)$  and  $Z^{(j)}(\mathbf{s}_0)$ .

## 4.2 Example

The R package `spBayes` (Finley et al., 2015) can be used for the spatial model defined in eq. (4.1), where Metropolis MCMC updates are applied for the variance parameters in the model. For Gibbs sampling the `spTimer` (Bakar and Sahu, 2015) package is useful. In this section `spTimer` is used to fit the spatial model described in eqs. (4.2) and (4.3). The precipitation dataset described in Section 2.2, used for the variogram and kriging problems, are compared here with the Bayesian model-based fitting and predictions. The dataset is randomly divided into training (33 locations) and test (5 locations) datasets, and then the Bayesian model is fitted to the training dataset and validated with the test dataset. The following R output was obtained for when fitting the Gaussian process spatial model in the `spTimer` package.

---

```
Model: GP
Call: precp ~ 1
Iterations: 5000
nBurn: 1000
```

Acceptance rate for phi (%): 39.6

```
-----
                Goodness.of.fit Penalty  PMCC
values:                3.41    18.12 21.53
-----
```

Computation time: 3.28 - Sec.

Parameters:

```
-----
                Mean Median      SD Low2.5p Up97.5p
(Intercept)  4.1590 4.1142 0.7522  2.6764  5.7581
sig2eps      0.2293 0.1877 0.1730  0.0781  0.6326
sig2eta      1.0571 0.5397 1.5769  0.2018  4.8904
phi          0.0527 0.0052 0.2622  0.0011  0.5836
-----
```

Here, we ran 5,000 MCMC iterations and discarded the first 1,000 samples as burn-in. The summary statistics presented are based on the remaining 4,000 MCMC samples. Observe that the spatial variance, i.e. the sill, is estimated as 0.54 (median) and the variance for the white noise is estimated as 0.19 (median). The Bayesian model estimates of the decay parameter is 0.0052, which represents an effective range of  $-\log(0.05)/0.0052 = 576.1$  KM (Finley et al., 2013). The estimate for the sill approximately matches the initial estimates that we obtained from the semivariogram plot in Figure 2(a). However, the range and the nugget variance are estimated higher by the Bayesian model. The predictive model choice criteria (PMCC) (Gelfand and Ghosh, 1998) is automatically reported by `spTimer`. It can be used as a model selection criteria: lower values of PMCC indicate better model fit (see details in Sahu et al., 2015; Crimp et al., 2015).

For validation we predict to the test locations and then calculate statistical validation criteria, e.g., mean squared error (MAE), root mean squared error (RMSE), etc. The `spTimer` package gives the following results for the validation statistics. Note that the validation statistics were calculated using the logarithmic value, and in this case the RMSE (of prediction) is 0.33. We can also obtain spatial predictions to a set of grid points. Figures 3(a) and (b) show the median and standard deviation (sd) of the predicted surface of the spatial process. Code for this analysis is provided in the Appendix.

## 5 Bayesian Spatio-Temporal Modelling

The spatial process  $Z(\mathbf{s})$  can be extended to a spatio-temporal process with the addition of an index,  $t$ , for time. That is:

$$Z(\mathbf{s}, t) : \mathbf{s} \in \mathcal{S} \subset \mathbb{R}^d, t \in \mathcal{T} \subset \mathbb{R}$$

is used to denote the spatio-temporal process of interest. If the  $Z(\mathbf{s}, t)$  process is observed at  $n$  spatial locations  $\mathbf{s}_1, \dots, \mathbf{s}_n$  at  $T$  different time points, then the spatio-temporal process can be written as  $\mathbf{Z}(\mathbf{s}, t) = (Z(\mathbf{s}_1, t), \dots, Z(\mathbf{s}_n, t))'$ ,  $1 \leq t \leq T$ . From a mathematical perspective, we can also

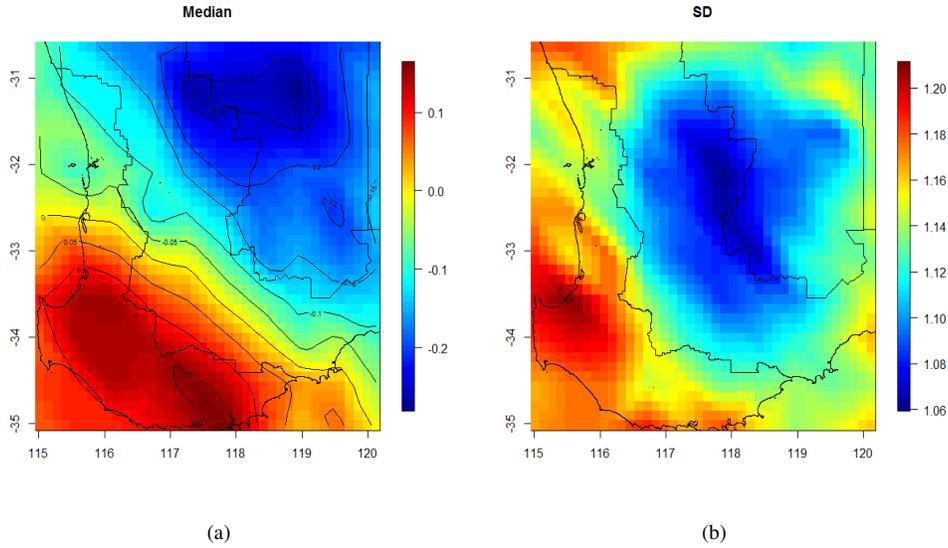


Figure 3: Predicted surface of the spatial process at grid level: (a) median and (b) standard deviation (sd).

represent the spatio-temporal process as a multivariate spatial process with dimension  $T$ , (see e.g., Finley et al., 2015; Dou et al., 2010). They argue that every time point of the spatio-temporal process can be regarded as a separate spatial random field.

Similar to the spatial processes, the spatio-temporal process  $Z(\mathbf{s}, t)$  is considered to be mean stationary within its spatio-temporal domain  $\mathcal{S} \times \mathcal{T}$ , if its mean process is constant within the spatio-temporal domain (Bruno et al., 2009). For weak stationary spatio-temporal processes, the mean function is assumed to be constant and the covariance function is assumed to depend on spatial and temporal covariances. Note that mean stationary only implies weak stationary if the first two moments i.e., mean and variance exist, whereas weak stationary only implies strict stationary if the spatio-temporal random process  $Z(\mathbf{s}, t)$  is a Gaussian process.

One of the key conceptual aspects of spatio-temporal models is an understanding of the separability of the space-time covariance function. The separability of models refers to formation of the spatio-temporal covariance function as a product of the spatial and temporal covariance functions (Rouhani and Myers, 1990; Diggle et al., 1998). A separable spatio-temporal covariance function is defined as:

$$C(Z(\mathbf{s}, t); Z(\mathbf{s}', t')) = C_S(\mathbf{s}, \mathbf{s}')C_T(t, t'),$$

where  $\mathbf{s}$  and  $\mathbf{s}'$  are the spatial locations and  $t$  and  $t'$  are the temporal points, and the terms  $C_S(\mathbf{s}, \mathbf{s}')$  and  $C_T(t, t')$  represent the spatial and temporal covariance functions, respectively. A space-time covariance function is called nonseparable if it cannot be represented as the product of spatial and

temporal functions. For a separable process, the space and time covariance function can be modelled separately.

The main advantage of assuming separability is computational convenience, since the spatio-temporal covariance matrix can be written as the Kronecker product of two smaller dimensional matrices. However, there are many nonseparable models available. For example, a set of nonseparable stationary covariance functions for spatio-temporal data are discussed in Cressie and Huang (1999). They used Fourier transforms in their approach to guarantee positive definiteness for the covariance function. There are also other approaches for constructing non-stationary covariance function (see for example, Stein, 2005; Schmidt and O'Hagan, 2003; Fuglstad et al., 2015; Cunha et al., 2017).

We write the spatio-temporal model as:

$$\mathbf{Z}_t = X_t\boldsymbol{\beta} + \boldsymbol{\eta}_t + \boldsymbol{\epsilon}_t; \quad \boldsymbol{\eta}_t \sim GP(\mathbf{0}, \Sigma_t), \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \tau^2\mathbf{I}). \quad (5.1)$$

Note that eq. (4.1) and eq. (5.1) are almost the same except for the index  $t$ . Here, the term  $\boldsymbol{\eta}_t$  is a spatio-temporal process that can have either separable or nonseparable spatio-temporal covariance.

Another way of including the time-series correlation is by adding dynamic autoregressive parameters or processes in the modelling hierarchy. For example, a model can include the autoregressive dynamics at the spatial error process level  $\boldsymbol{\eta}_t$  (Sahu and Bakar, 2012b):

$$\mathbf{Z}_t = X_t\boldsymbol{\beta} + \boldsymbol{\eta}_t + \boldsymbol{\epsilon}_t \quad (5.2)$$

$$\boldsymbol{\eta}_t = \sum_{k=1}^r \rho_k \boldsymbol{\eta}_{(t-k)} + \boldsymbol{\delta}_t, \quad \boldsymbol{\delta}_t \sim N(\mathbf{0}, \sigma_\delta^2 \mathbf{I}), \quad (5.3)$$

where the term  $\rho_k$  is the  $k^{\text{th}}$  autoregressive parameter. It is also possible to developed a model where the true process is used as a lagged autoregressive parameter. For example, Sahu et al. (2007) developed the model:

$$\mathbf{Z}_t = \mathbf{O}_t + \boldsymbol{\epsilon}_t \quad (5.4)$$

$$\mathbf{O}_t = \rho \mathbf{O}_{t-1} + X_t\boldsymbol{\beta} + \boldsymbol{\eta}_t, \quad (5.5)$$

where  $\mathbf{O}_0$  (at time  $t = 0$ ) is an initial value that follows a Gaussian process with zero mean and a known or unknown variance that may be introduced through prior information in the model. They used a lag one model; however, the model can be extended for  $r$  lags by replacing  $\rho \mathbf{O}_{t-1}$  with  $\sum_{k=1}^r \rho_k \mathbf{O}_{t-k}$ . A simpler version of models in eq. (5.4) and (5.5) can be obtained when  $\rho = 0$ :

$$\mathbf{Z}_t = \mathbf{O}_t + \boldsymbol{\epsilon}_t \quad (5.6)$$

$$\mathbf{O}_t = X_t\boldsymbol{\beta} + \boldsymbol{\eta}_t, \quad (5.7)$$

where no autoregression has been used.

## 5.1 Prediction Details

Prediction for spatio-temporal models are two types: (1) prediction to an unobserved location  $s_0$  and (2) prediction to a future time point  $T + 1$ . In the rest of this paper time prediction will be referred to

as a forecast. Spatial prediction for models defined in eq. (5.6) and (5.7) are similar to that defined in eq. (4.5); however, it will include an index for time-series  $t$ :

$$\pi(Z(\mathbf{s}_0, t) | \mathbf{z}, X, \mathbf{x}_0) = \int \pi(Z(\mathbf{s}_0, t) | O(\mathbf{s}_0, t), \mathbf{O}, \mathbf{z}, \boldsymbol{\theta}, X, \mathbf{x}_0) \pi(O(\mathbf{s}_0, t) | \mathbf{O}, \mathbf{z}, \boldsymbol{\theta}, X, \mathbf{x}_0) \pi(\mathbf{O}, \boldsymbol{\theta} | \mathbf{z}, X, \mathbf{x}_0) dO(\mathbf{s}_0, t) d\mathbf{O} d\boldsymbol{\theta} \quad (5.8)$$

Note that the input in the above integral changes for the different models described by eqs. (5.1) to (5.5).

Now, the one-step ahead forecast distribution of  $Z(\mathbf{s}_0, T + 1)$  at any unobserved location at time  $T + 1$  for the model defined in eq. (5.6) and (5.7) is:

$$\pi(Z(\mathbf{s}_0, T + 1) | \mathbf{z}, X, \mathbf{x}_{0, T+1}) = \int \pi(Z(\mathbf{s}_0, T + 1) | O(\mathbf{s}_0, T + 1), \mathbf{O}, \mathbf{z}, \boldsymbol{\theta}, X, \mathbf{x}_{0, T+1}) \pi(O(\mathbf{s}_0, T + 1) | \mathbf{O}, \mathbf{z}, \boldsymbol{\theta}, X, \mathbf{x}_{0, T+1}) \pi(\mathbf{O}, \boldsymbol{\theta} | \mathbf{z}, X, \mathbf{x}_{0, T+1}) dO(\mathbf{s}_0, T + 1) d\mathbf{O} d\boldsymbol{\theta} \quad (5.9)$$

Similar to spatial prediction, the one-step ahead forecast distribution is constructed by combining samples from the joint posterior MCMC samples. Note that the marginal distribution for the forecast distribution is used instead of the conditional distribution because we have already obtained the conditional distribution given the observed information up to time  $T$  at the locations  $\mathbf{s}_1, \dots, \mathbf{s}_n$ . In this case, at the future time point  $T + 1$  no further information is needed except the covariates  $\mathbf{x}(\mathbf{s}_0, T + 1)$ . It is also possible to obtain forecast distributions at any observed location  $\mathbf{s}_i$  at time  $T + 1$  by iterative sampling from the conditional distribution  $O(\mathbf{s}_i, T + 1) | \cdot \sim N(\mathbf{x}'(\mathbf{s}_i, T + 1)\boldsymbol{\beta}, \sigma^2)$  and then  $Z(\mathbf{s}_i, T + 1) | \cdot \sim N(O(\mathbf{s}_i, T + 1), \tau^2)$ .

## 5.2 Example

To explore the applications of the spatio-temporal models we use the same precipitation data from WA. However, unlike having only one time point (i.e., 2009) the dataset now consists of average yearly precipitation measurements from 1960 to 2009. In this example we also use two climatic covariates in the model. These are: (i) Indian Ocean Dipole (IOD) and (ii) El Nino-Southern Oscillation (ENSO) anomaly (Anom3.4). It is well established that these indices have influence on the onshore climatic measurements in Australia (Crimp et al., 2015).

Figure 4(a) shows a scatter plot of the precipitation measurements by years for all climate monitoring locations. We also superimposed the trends in precipitations, which show a decreasing pattern over time. The histogram of the precipitation measurements in transformed scale (logarithm) is presented in Figure 4(b), which show a symmetric bell shape distribution compared to the distribution of precipitation without any scale transformation.

The full precipitation dataset is divided into three parts, one for model fitting, and the other two for validation using spatial and temporal predictions. The `spTimeR` package is used in this context and the models described in eq. (5.6) and (5.7) are fitted. We randomly selected 5 locations and the last 5 years data are set aside as a validation dataset. The model was fitted to 33 locations and 45

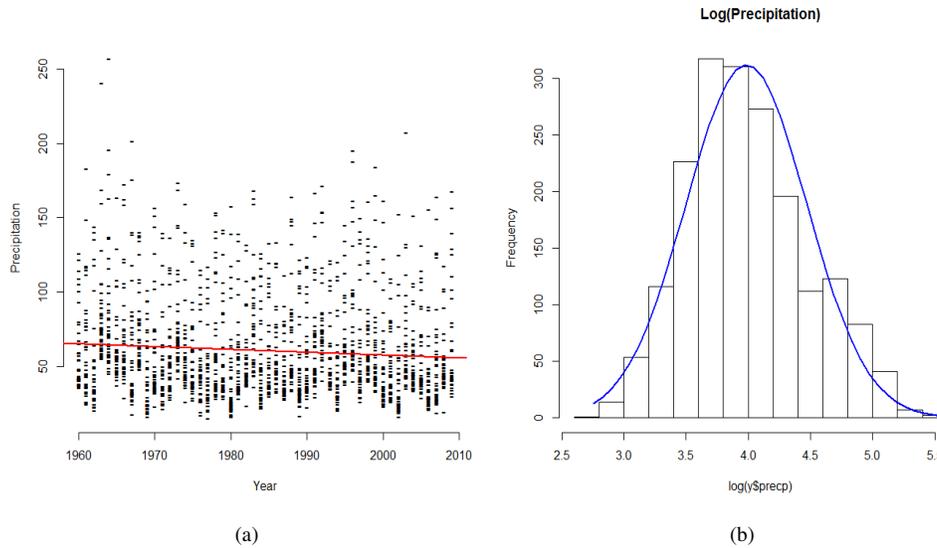


Figure 4: (a) Trends in precipitation measurements. (b) Distribution of precipitation in logarithmic scale.

years of data. Hence, the data we use for model fitting contains  $33 \times 45 = 1485$  observations. From R the following summary statistics of the model parameters were obtained.

```
-----
Model: GP
Call: precp ~ IOD + Anom3.4
Iterations: 5000
nBurn: 1000
Acceptance rate for phi (%): 27.78
-----
```

```
Goodness.of.fit Penalty PMCC
values:          185.36 269.94 455.3
-----
```

```
Computation time: 10.43 - Sec.
-----
```

```
Parameters:
          Mean  Median    SD Low2.5p Up97.5p
(Intercept) 4.2695 4.2589 0.0811 4.1393 4.4552
IOD         -0.0251 -0.0249 0.0675 -0.1602 0.1125
Anom3.4     -0.0488 -0.0472 0.1181 -0.2954 0.1931
```

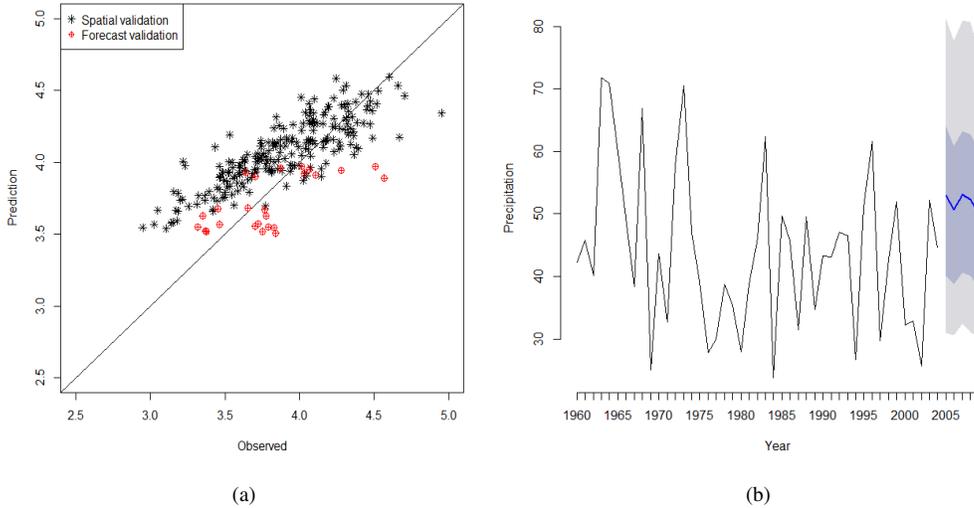


Figure 5: (a) Scatterplot of the predicted versus observed precipitation data in transformed scale. (b) 5-year ahead forecast in original scale.

sig2eps	0.0075	0.0072	0.0013	0.0063	0.0103
sig2eta	0.4165	0.2566	0.5098	0.1362	1.5758
phi	0.0032	0.0030	0.0013	0.0011	0.0060

Note that the covariates IOD and Anom3.4 are not statistically significant as the 95% credible interval contains zero. A possible explanation for this is that the covariates are observed globally, for example, the IOD observations are measured from Indian Ocean sea surface temperature and ENSO anomaly (Anom3.4) is measured from the mid-Pacific sea surface temperature. Hence, these values are same for each monitoring location. This situation actually introduces a spatial misalignment between the observed precipitation and the covariates, which will be discussed further in Section 6. To validate the model we predict and compare with the test validation datasets that were created for both spatial prediction and forecast (temporal prediction) at the unobserved spatial locations. Figure 5(a) shows a plot of the actual and predicted values in transformed scale. It can be seen that the predictions match with the actual observations. The forecast values with credible 95% and 50% intervals are given in Figure 5(b). As expected, the forecast intervals are wider compared to the spatial prediction intervals. Also, the RMSE validation statistic was 0.29 for spatial prediction and 0.26 for temporal prediction at the unobserved locations. It is also possible to obtain temporal predictions at the spatial locations where data has already been observed. Detailed code to reproduce the results are given in the Appendix.

## 6 Dynamic Models

The models so far discussed in this paper are almost linear in nature. On the other hand the dynamic models that will be described in this section are non-linear. Spatial models can be dynamic in form by introducing spatially varying coefficients. Similarly a spatio-temporal dynamic model can have both space and time varying coefficients.

Often it is necessary to understand the spatial effects of the parameters of the covariates. For example, the influence of the covariates may vary across locations. In particular, these models are useful when the covariate measurements are fixed over space, which is a common situation in environmental applications. For example, environmental indices such as IOD and ENSO anomaly (Anom3.4) used in Section 5.2 for the precipitation example vary over time only, and their impact may vary across space. These models can also be used to downscale (Berrocal et al., 2010; Kang et al., 2012), where output from gridded numerical deterministic models is calibrated with point level observational data. This type of calibration is also known as a change of support problem (Cressie and Wikle, 2011). Among the popular approaches for solving the point-to-area problem are block kriging (Gotway and Young, 2007), Bayesian melding (Fuentes and Raftery, 2005), fusion modelling (Gelfand and Sahu, 2009) and the use of spatially varying coefficients (Berrocal et al., 2010; Gelfand et al., 2003). Sometimes covariates are spatially misaligned with the predictand (Lopiano et al., 2011; Szpiro et al., 2011) and dynamic models are also beneficial for this situation (Bakar et al., 2015; Cai et al., 2013).

A spatially varying dynamic model can be written as:

$$\mathbf{Z}(\mathbf{s}) = \sum_{j=1}^p \tilde{\mathbf{x}}_j(\mathbf{s})\boldsymbol{\beta}_j(\mathbf{s}) + \boldsymbol{\eta}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}), \quad (6.1)$$

where  $\tilde{\mathbf{x}}_j(\mathbf{s})$  is an  $n \times n$  diagonal matrix of the  $j^{\text{th}}$  covariate and  $\boldsymbol{\beta}_j(\mathbf{s})$  is the  $n \times 1$  vector of corresponding parameters. Note that the difference between eq. (4.1) and (6.1) lies only in the formation of the dynamic coefficient  $\boldsymbol{\beta}_j(\mathbf{s})$ . We assume that  $\boldsymbol{\beta}_j(\mathbf{s}) \sim GP(\mathbf{0}, \Sigma_{\boldsymbol{\beta}_j})$ , where  $\Sigma_{\boldsymbol{\beta}_j} = \sigma_{\boldsymbol{\beta}_j}^2 \kappa(\cdot, \cdot)$ . Similar to eq. (5.6) and (5.7), we can write the spatio-temporal version of the model as:

$$\mathbf{Z}(\mathbf{s}, t) = \mathbf{O}(\mathbf{s}, t) + \boldsymbol{\epsilon}(\mathbf{s}, t) \quad (6.2)$$

$$\mathbf{O}(\mathbf{s}, t) = \sum_{j=1}^p \tilde{\mathbf{x}}_j(\mathbf{s}, t)\boldsymbol{\beta}_j(\mathbf{s}) + \boldsymbol{\eta}(\mathbf{s}, t) \quad (6.3)$$

It is also possible to develop a model using temporal dynamics, where instead of the spatially varying parameter we include temporal dynamics with one or more lags for the covariate coefficients. These models are also known as the dynamic linear models (West and Harrison, 1997; Dou et al., 2010). Such models are popular for modelling data with seasonal variations and are mainly used for time-series analysis. A spatio-temporal version of the DLMS incorporating seasonal effects was introduced by Stroud et al. (2001) and Huerta et al. (2004). A number of practical applications have been undertaken using different versions of the spatio-temporal DLMS, e.g., Gamerman et al. (2007); Ghosh et al. (2010); Sahu and Bakar (2012a); Sigrist et al. (2013). This type of model can be written

as:

$$Z(\mathbf{s}_i, t) = O(\mathbf{s}_i, t) + \epsilon(\mathbf{s}_i, t) \quad (6.4)$$

$$O(\mathbf{s}_i, t) = \sum_{j=1}^p x_j(\mathbf{s}_i, t) \zeta_j(t) + \eta(\mathbf{s}_i, t) \quad (6.5)$$

$$\zeta_j(t) = \sum_{k=1}^r \rho_{jk} \zeta_j(t-k) + \delta_j(t) \quad \delta_j(t) \sim N(0, \sigma_{j\delta}^2) \quad (6.6)$$

The models can also include the dynamics in the spatio-temporal process  $\eta(\mathbf{s}_i, t)$  (Gelfand et al., 2005), similar to the way they are defined in eq. (5.3). In this case an additional equation can be added after eq. (6.6) as:

$$\eta(\mathbf{s}, t) = \sum_{k=1}^r \rho_k \eta(\mathbf{s}, t-k) + \omega(\mathbf{s}, t); \quad \omega(\mathbf{s}, t) \sim GP(\mathbf{0}, \sigma_\omega^2 \kappa(\cdot, \cdot)) \quad (6.7)$$

For further reading see Banerjee et al. (2014)[ch. 10] and Cressie and Wikle (2011)[ch. 7]. Predictive distributions from the dynamic models discussed above can also be obtained by integrating the joint distributions and by composite sampling from the joint posterior distribution (Bakar et al., 2015, 2016).

## 6.1 Example

For dynamic model fitting, and spatial and temporal predictions the R package `spTDyn` (Bakar et al., 2016) can be used. The package uses the spatio-temporal dynamics as an additive term for the covariates in the model. Hence the model is of the form:

$$\mathbf{Z}(\mathbf{s}, t) = \mathbf{O}(\mathbf{s}, t) + \boldsymbol{\epsilon}(\mathbf{s}, t) \quad (6.8)$$

$$\mathbf{O}(\mathbf{s}, t) = \sum_{j=1}^p \tilde{\mathbf{x}}_j(\mathbf{s}, t) \boldsymbol{\beta}_j(\mathbf{s}) + \mathbf{x}(\mathbf{s}, t) \boldsymbol{\zeta}(t) + \boldsymbol{\eta}(\mathbf{s}, t) \quad (6.9)$$

$$\boldsymbol{\zeta}(t) = \mathbf{G} \boldsymbol{\zeta}(t-k) + \boldsymbol{\delta}(t), \quad (6.10)$$

where  $\mathbf{x}(\mathbf{s}, t)$  is an  $n \times p$  design matrix of covariates and  $\boldsymbol{\gamma}(t)$  is the  $p \times 1$  parameter vector at time  $t$ . The term  $\mathbf{G}$  is a diagonal matrix of order  $p \times p$  with elements  $\rho_j$ ,  $j = 1, \dots, p$ . Considering  $\rho_j = 1$  leads to a spatio-temporal random-walk model (West and Harrison, 1997). Note that it is also possible to construct a dynamic model that only consists of either spatial or temporal dynamic parameters.

The training precipitation data is fitted and estimates of the dynamic parameters are obtained. Figures 6(a) and (b) represent the estimates for the IOD and Anom3.4 for each spatial locations. As expected, the estimates vary over the space. Similar results can be obtained for the dynamic temporal parameters ( $\boldsymbol{\zeta}$ ). Figure 7 shows the yearly dynamic estimates for 1960 to 2004. In this paper we fitted a random-walk model by assuming  $\rho = 1$ ; however, one can also estimate  $\rho$  using the package `spTDyn`. One can also obtain spatial and temporal predictions and validate it with the validation datasets. See details of the code used in the Appendix.

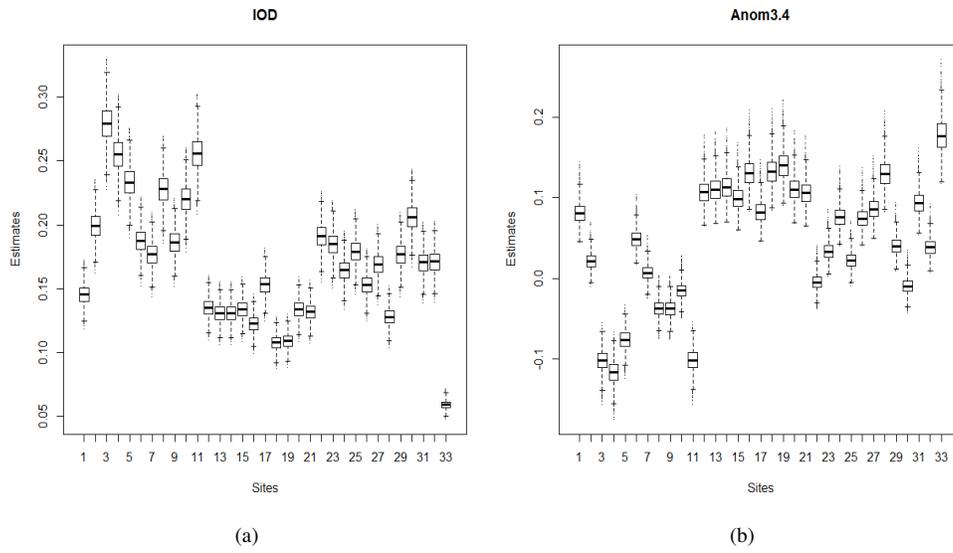


Figure 6: Boxplots of the spatially varying parameter estimates for (a) Indian Ocean Dipole (IOD) and (b) ENSO anomaly.

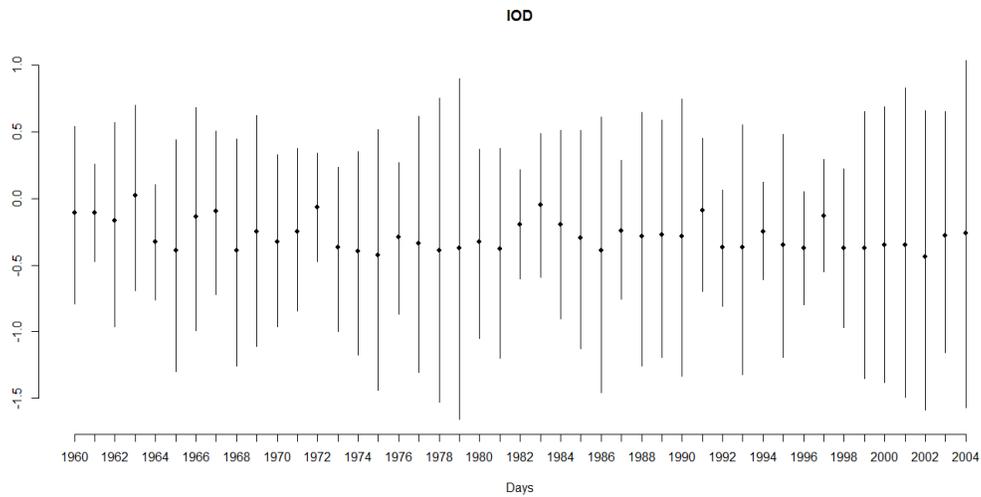


Figure 7: . Estimates for the temporal dynamic parameters from years 1960 to 2004 for the Indian Ocean Dipole (IOD) climatic index.

## 7 Conclusion

In this paper, we described some popular Bayesian spatial and spatio-temporal models to analyse geographically point-referenced data. The assumptions of the Bayesian models were discussed and predictions to locations where no data has been observed and forecasts to future time points were obtained. The inclusion of spatial and temporal correlations in the model improves the accuracy of such predictions. We discussed different ways of defining models and explained their relative advantages and disadvantages. The average aggregated yearly precipitation data measured in WA over 50 years from 1960 to 2009 was used as a practical application of the models. The global climatic indices IOD and ENSO anomaly were also used in the analysis as predictor variables. Note that models discussed in this paper are for Gaussian data, however in real-life the data may not always follow a normal distribution. For example, precipitation data is highly skewed even at the monthly level. One quick approach to resolve the issue is known as the scale transformation. Hence as discussed in the example (see Section 5.2) we use a logarithmic transformation of the precipitation observations.

It can be worthwhile, although sometimes more challenging, to implement efficient algorithms to achieve rapid MCMC chain convergence. The dynamic models discussed in Section 6 are non-linear and have autoregressive dependence in both spatial and time-series structure. Thus, besides the Metropolis-within-Gibbs algorithm, it can be worthwhile implementing algorithms such as sequential Monte Carlo (Liu and Chen, 1998; Del Moral et al., 2006), its adaptive version (Zhou et al., 2016), particle MCMC (Andrieu et al., 2010) or forward filtering-backward smoothing (Frühwirth-Schnatter, 1994; Briers et al., 2010) to estimate the dynamic model parameters. These algorithms are not only more efficient, but they can perform better with highly correlated data leading to improved predictive performance of the model. Another approach known as the integrated nested Laplace approximation (INLA) is often popular to obtain inference quickly (Rue et al., 2009). This method is sometimes attractive to solve spatial problems; however, uses a computational approximation of the marginal posterior rather than the MCMC techniques.

In addition to the models discussed in this paper, one important problem that often occurs in spatial and spatio-temporal modelling is due to the presence of a large number of spatial data points. This creates computational complexity when inverting the large  $n$  dimensional variance-covariance matrix and is also known as the big- $n$  problem (Banerjee et al., 2008; Cressie and Johannesson, 2008; Finley et al., 2009; Rue et al., 2009; Cressie et al., 2010; Guhaniyogi et al., 2011). In this situation, the exact likelihood based inference becomes unstable and infeasible. A detail discussion is given in Lasinio et al. (2013); Sahu and Bakar (2012b) and Sahu et al. (2015).

The models and applications used in this paper were for point-referenced data; however, spatial data also come in the form of point-pattern (e.g., distribution of trees in a sampling region) and areal (mortality rate in districts) structure, see Figures 1(b) and (c). The book by Diggle (2013) provides methods for analysing point processes and also spatio-temporal point processes. It also provides applications to real-life data examples including spatial epidemiology (Hossain and Lawson, 2009). For areal data modelling, a discrete space conditional auto-regressive (CAR) specification is often used for defining the area level spatial correlation (Lee, 2011; Lawson et al., 2012; Banerjee et al., 2014, Ch. 3). The CAR approach considers a neighbourhood weight matrix determined with binary

specification that forces correlation between geographically adjacent areas.

## Acknowledgements

The authors wish to acknowledge the support of the CSR&M and the CCI, the Australian National University.

## References

- Andrieu, C., Doucet, A., and Holenstein, R. (2010), "Particle markov chain monte carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 269–342.
- Bakar, K. S., Kokic, P., and Jin, H. (2015), "A spatiodynamic model for assessing frost risk in south-eastern Australia," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64, 755–778.
- (2016), "Hierarchical spatially varying coefficient and temporal dynamic process models using spTDyn," *Journal of Statistical Computation and Simulation*, 86, 820–840.
- Bakar, K. S. and Sahu, S. K. (2015), "spTimer: Spatio-Temporal Bayesian Modelling Using R," *Journal of Statistical Software*, 63, 1–32.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman & Hall/CRC.
- (2014), *Hierarchical modeling and analysis for spatial data*, Crc Press.
- Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008), "Gaussian Predictive Process Models for Large Spatial Data Sets," *Journal of Royal Statistical Society B*, 70, 825–848.
- Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010), "A spatio-temporal downscaler for output from numerical models," *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 176–197.
- Briers, M., Doucet, A., and Maskell, S. (2010), "Smoothing algorithms for state–space models," *Annals of the Institute of Statistical Mathematics*, 62, 61–89.
- Bruno, F., Guttorp, P., Sampson, P. D., and Cocchi, D. (2009), "A simple non-separable, non-stationary spatiotemporal model for ozone," *Environmental and Ecological Statistics*, 16, 515–529.
- Cai, B., Lawson, A. B., Hossain, M., Choi, J., Kirby, R. S., Liu, J., et al. (2013), "Bayesian semi-parametric model with spatially–temporally varying coefficients selection," *Statistics in medicine*, 32, 3670–3685.

- Cressie, N. and Huang, H.-C. (1999), “Classes of nonseparable, spatio-temporal stationary covariance functions,” *Journal of the American Statistical Association*, 94, 1330–1339.
- Cressie, N., Shi, T., and Kang, E. L. (2010), “Fixed rank filtering for spatio-temporal data,” *Journal of Computational and Graphical Statistics*, 19, 724–745.
- Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: John Wiley & Sons.
- Cressie, N. A. C. and Johannesson, G. (2008), “Fixed Rank Kriging for Very Large Spatial Data Sets,” *Journal of the Royal Statistical Society B*, 70, 209–226.
- Cressie, N. A. C. and Wikle, C. K. (2011), *Statistics for Spatio-Temporal Data*, New York: John Wiley & Sons.
- Crimp, S., Bakar, K. S., Kokic, P., Jin, H., Nicholls, N., and Howden, M. (2015), “Bayesian space-time model to analyse frost risk for agriculture in Southeast Australia,” *International Journal of Climatology*, 35, 2092–2108.
- Cunha, M., Gamerman, D., Fuentes, M., and Paez, M. (2017), “A non-stationary spatial model for temperature interpolation applied to the state of Rio de Janeiro,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.
- Del Moral, P., Doucet, A., and Jasra, A. (2006), “Sequential monte carlo samplers,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 411–436.
- Diggle, P. J. (2013), *Statistical analysis of spatial and spatio-temporal point patterns*, CRC Press.
- Diggle, P. J., Tawn, J., and Moyeed, R. (1998), “Model-based geostatistics,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47, 299–350.
- Dou, Y., Le, N. D., and Zidek, J. V. (2010), “Modeling hourly ozone concentration fields,” *Annals of Applied Statistics*, 4, 1183–1213.
- Finley, A. O., Banerjee, S., and Carlin, B. P. (2007), “spBayes: An R Package for Univariate and Multivariate Hierarchical Point-referenced Spatial Models,” *Journal of Statistical Software*, 12, 1–24.
- (2013), *spBayes: Univariate and Multivariate Spatial Modeling*, r package version 0.3-1.
- (2015), “spBayes for Large Univariate and Multivariate Point-Referenced Spatio-Temporal Data Models,” *Journal of Statistical Software*, 63, 1–28.
- Finley, A. O., Sang, H. Y., Banerjee, S., and Gelfand, A. E. (2009), “Improving the performance of predictive process modeling for large datasets,” *Computational Statistics and Data Analysis*, 53, 2873–2884.
- Frühwirth-Schnatter, S. (1994), “Data augmentation and dynamic linear models,” *Journal of time series analysis*, 15, 183–202.

- Fuentes, M. and Raftery, A. E. (2005), "Model Evaluation and Spatial Interpolation by Bayesian Combination of Observations With Outputs From Numerical Models," *Biometrics*, 61, 36–45.
- Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2015), "Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy," *Statistica Sinica*, 115–133.
- Gamerman, D., Salazar, E., and Reis, E. (2007), *Dynamic Gaussian process priors, with applications to the analysis of space-time data*, Oxford University Press, Oxford: Eds. Bernardo, J.M. et al., chap. Bayesian Statistics 8, pp. 149–174.
- Gelfand, A. E., Banerjee, S., and Gamerman, D. (2005), "Spatial process modelling for univariate and multivariate dynamic spatial data," *Environmetrics*, 16, 465–479.
- Gelfand, A. E. and Ghosh, S. K. (1998), "Model Choice: A Minimum Posterior Predictive Loss Approach," *Biometrika*, 85, 1–11.
- Gelfand, A. E., Kim, H. J., Sirmans, C. F., and Banerjee, S. (2003), "Spatial modeling with spatially varying coefficient processes," *Journal of the American Statistical Association*, 98, 387–396.
- Gelfand, A. E. and Sahu, S. (2009), "Combining monitoring data and computer model output in assessing environmental exposure," in *The Oxford Handbook of Applied Bayesian Analysis*, eds. O'Hagan, A. and West, M., Oxford University Press, Oxford, vol. 19, pp. 482–510.
- Gelfand, A. E. and Smith, A. F. (1990), "Sampling-based approaches to calculating marginal densities," *Journal of the American statistical association*, 85, 398–409.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014), *Bayesian data analysis*, vol. 3, Chapman & Hall/CRC Boca Raton, FL, USA.
- Gelman, A. and Rubin, D. B. (1992), "Inference from iterative simulation using multiple sequences," *Statistical Science*, 7, 457–472.
- Ghosh, S. K., Bhave, P. V., Davis, J. M., and Lee, H. (2010), "Spatio-Temporal Analysis of Total Nitrate Concentrations Using Dynamic Statistical Models," *Journal of the American Statistical Association*, 105, 538–551.
- Gotway, C. A. and Young, L. J. (2007), "A Geostatistical Approach to Linking Geographically Aggregated Data From Different Sources," *Journal of Computational and Graphical Statistics*, 16, 115–135.
- Guhaniyogi, R., Finley, A. O., Banerjee, S., and Gelfand, A. E. (2011), "Adaptive Gaussian predictive process models for large spatial datasets," *Environmetrics*, 22, 997–1007.
- Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57, 97–109.

- Hossain, M. M. and Lawson, A. B. (2009), "Approximate methods in Bayesian point process spatial models," *Computational statistics & data analysis*, 53, 2831–2842.
- Huerta, G., Sanso, B., and Stroud, J. R. (2004), "A spatiotemporal model for Mexico City ozone levels," *Journal of the Royal Statistical Society Series C*, 53, 231–248.
- Kang, E. L., Cressie, N., and Sain, S. R. (2012), "Combining outputs from the North American Regional Climate Change Assessment Program by using a Bayesian hierarchical model," *Journal of the Royal Statistical Society Series C-Applied Statistics*, 61, 291–313.
- Lasinio, G. J., Mastrantonio, G., and Pollice, A. (2013), "Discussing the big n problem," *Statistical Methods & Applications*, 22, 97–112.
- Lawson, A. B., Choi, J., Cai, B., Hossain, M., Kirby, R. S., and Liu, J. (2012), "Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data," *Journal of agricultural, biological, and environmental statistics*, 1–25.
- Lee, D. (2011), "A comparison of conditional autoregressive models used in Bayesian disease mapping," *Spatial and Spatio-temporal Epidemiology*, 2, 79–89.
- Liu, J. S. and Chen, R. (1998), "Sequential Monte Carlo methods for dynamic systems," *Journal of the American statistical association*, 93, 1032–1044.
- Lopiano, K. K., Young, L. J., and Gotway, C. A. (2011), "A comparison of errors in variables methods for use in regression models with spatially misaligned data," *Statistical methods in medical research*, 20, 29–47.
- Matérn, B. (1986), *Spatial Variation*, Berlin: Springer-Verlag, 2nd ed.
- Matheron, G. (1963), "Principles of geostatistics," *Economic geology*, 58, 1246–1266.
- R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Roberts, G. O., Gelman, A., Gilks, W. R., et al. (1997), "Weak convergence and optimal scaling of random walk Metropolis algorithms," *The annals of applied probability*, 7, 110–120.
- Rouhani, S. and Myers, D. E. (1990), "Problems in space-time kriging of geohydrological data," *Mathematical Geology*, 22, 611–623.
- Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the royal statistical society: Series b (statistical methodology)*, 71, 319–392.
- Sahu, S. K. and Bakar, K. S. (2012a), "A Comparison of Bayesian Models for Daily Ozone Concentration Levels," *Statistical Methodology*, 9, 144–157.

- (2012b), “Hierarchical Bayesian Autoregressive Models for Large Space Time Data with Applications to Ozone Concentration Modelling,” *Applied Stochastic Models in Business and Industry*, 28, 395–415.
- Sahu, S. K., Bakar, K. S., and Awang, N. (2015), “Bayesian forecasting using spatiotemporal models with applications to ozone concentration levels in the Eastern United States,” *Geometry Driven Statistics*, 121, 260.
- Sahu, S. K., Gelfand, A. E., and Holland, D. M. (2007), “High-resolution space–time ozone modeling for assessing trends,” *Journal of the American Statistical Association*, 102, 1221–1234.
- Schmidt, A. M. and O’Hagan, A. (2003), “Bayesian inference for non-stationary spatial covariance structure via spatial deformations,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 743–758.
- Sigrist, F., Kunsch, H. R., and Stahel, W. A. (2013), “A dynamic nonstationary spatio-temporal model for short term prediction of precipitation,” *Annals of Applied Statistics*, 6, 1452–1477.
- Stein, M. L. (2005), “Statistical methods for regular monitoring data,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 667–687.
- Strauss, D. J. (1975), “A model for clustering,” *Biometrika*, 467–475.
- Stroud, J. R., Muller, P., and Sanso, B. (2001), “Dynamic models for spatiotemporal data,” *Journal of the Royal Statistical Society B*, 63, 673–689.
- Szpiro, A. A., Sheppard, L., and Lumley, T. (2011), “Efficient measurement error correction with spatially misaligned data,” *Biostatistics*, 12, 610–623.
- West, M. and Harrison, J. (1997), *Bayesian forecasting and dynamic models*, New York: Springer, 2nd ed.
- Zhou, Y., Johansen, A. M., and Aston, J. A. (2016), “Toward Automatic Model Comparison: An Adaptive Sequential Monte Carlo Approach,” *Journal of Computational and Graphical Statistics*, 25, 701–726.

*Received: April 4, 2017*

*Accepted: June 2, 2017*