Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Hierarchical spatially varying coefficient and temporal dynamic process models using spTDyn

K. Shuvo Bakar, Philip Kokic & Huidong Jin

To cite this article: K. Shuvo Bakar, Philip Kokic & Huidong Jin (2016) Hierarchical spatially varying coefficient and temporal dynamic process models using spTDyn, Journal of Statistical Computation and Simulation, 86:4, 820-840, DOI: 10.1080/00949655.2015.1038267

To link to this article: https://doi.org/10.1080/00949655.2015.1038267

Published online: 08 May 2015.

Journal of Statistical Computation and Simulation, 2016
Vol. 86, No. 4, 820–840, http://dx.doi.org/10.1080/00949655.2015.1038267

Hierarchical spatially varying coefficient and temporal dynamic process models using `spTDyn`

K. Shuvo Bakar[a*], Philip Kokic[b,c] and Huidong Jin[b,d]

[a]*Department of Statistics, Yale University, New Haven, CT, USA;* [b]*CSIRO Digital Productivity Flagship, Canberra, Australia;* [c]*NIASRA, University of Wollongong, Wollongong, Australia;* [d]*CECS, Australian National University, Canberra, Australia*

(*Received 13 February 2015; accepted 2 April 2015*)

Bayesian hierarchical spatio-temporal models are becoming increasingly important due to the increasing availability of space-time data in various domains. In this paper we develop a user friendly `R` package, `spTDyn`, for spatio-temporal modelling. It can be used to fit models with spatially varying and temporally dynamic coefficients. The former is used for modelling the spatially varying impact of explanatory variables on the response caused by spatial misalignment. This issue can arise when the covariates only vary over time, or when they are measured over a grid and hence do not match the locations of the response point-level data. The latter is to examine the temporally varying impact of explanatory variables in space-time data due, for example, to seasonality or other time-varying effects. The `spTDyn` package uses Markov chain Monte Carlo sampling written in `C`, which makes computations highly efficient, and the interface is written in `R` making these sophisticated modelling techniques easily accessible to statistical analysts. The models and software, and their advantages, are illustrated using temperature and ozone space-time data.

Keywords: Bayesian modelling; spatially varying coefficient; spatio-temporal dynamic linear model; Gibbs sampling

1. Introduction

Scientific understanding of environmental/genetic/biological/epidemiologic processes often requires making inferences over a regular or irregular spatial region with long or short time domains.[1–4] In this paper we illustrate two different types of problem-specific Bayesian hierarchical spatio-temporal models. The first is (1) a spatially varying coefficient process model designed to address the issues of spatial misalignment and non-spatially varying covariates. The second type of model is (2) a spatio-temporal dynamic linear model (DLM) that can capture the dynamic nature of the data. The main contribution of this paper is the development of user friendly software in `R` for a general spatially varying and spatio-temporal DLMs.

The work intends to fill a statistical modelling software gap. There is currently no easily accessible software package available that can tackle spatially varying coefficient process models, especially using a Bayesian hierarchical approach, even though there have been substantial developments in spatio-temporal modelling software recently. For example, some publicly

*Corresponding author. Email: shuvo.bakar@yale.edu

available software, including WinBUGS,[5] OpenBUGS [6] and JAGS [7] have been applied successfully to a variety of problems, see, for example, Arab and Courter.[8] They normally require end-users to specify a model using a description language, which makes them accessible only to well-trained statistical modellers, instead of the much broader `R` user community. In addition, as they often were not implemented in `C/C++` (OpenBUGS is implemented in `Component Pascal` and JAGS in `Java` programming language), they are computationally less efficient. A recent extension of the package `spBayes` [9] incorporates dynamic space-time regression models. The model structure used in `spBayes` includes a random-walk term for the spatial random process that differs considerably to the spatio-temporal DLM, we present in this paper (see details in [10]). Moreover, `spBayes` does not support any temporal predictions for the dynamic models they considered, while temporal predictions are often required in real applications.[4,11] Another `R` package `spate` [12,13] can also model spatio-temporal DLMs. However, it uses a stochastic partial differential equation (SPDE) approach to obtain inference of the model parameters, and it does not allow models with spatially varying parameters. The package `spTimer` [14] can fit, predict and forecast spatio-temporal data using three different types of hierarchical spatio-temporal models. However, it does not have the capability of fitting spatially varying and spatio-temporal DLMs. Motivated by these needs, we develop the `spTDyn` with mainly coding in `C` to incorporate these models within the flexibility of the Bayesian paradigm.

Spatially varying coefficient process models are often useful when the covariate measurements are fixed over space. This is common in environmental applications. For example, environmental indices such as the Southern Oscillation Index (SOI) varies over time only,[15] and its impact may vary across space. This kind of models can also be used to downscale,[1] where output from gridded numerical deterministic models is calibrated with the point-level observation data. This type of calibration is also known as a change of support problem.[2] There are other approaches to solve this point-to-area problem, for example, block kriging [16] and Bayesian melding.[17] However, in this paper we limit our approach to spatially varying coefficient models.[1,3] Recently, spatially varying models have been used in different application areas including biological, genetic and epidemiological applications.[18–20]

DLMs [21] are popular for modelling data with seasonal variations. Stroud et al. [22] introduced a spatio-temporal version of the DLMs and later Huerta et al. [23] incorporated seasonal effect with application to ground-level ozone concentrations. Another form of DLM with a dynamic spatio-temporal process is described in [24]. A review of different types of approaches can be found in [25]. Recently, a number of practical applications have been undertaken using different versions of the spatio-temporal DLM, for example, Dou et al.,[26] Ghosh et al.,[27] Mahmoudian and Mohammadzadeh,[28] Sahu and Bakar [29] and Vanem et al.[4] The `spTDyn` software described in this paper makes such modelling techniques easily available to broader researchers and data analysts.

In climate applications, statistical models have frequently been used to downscale data from physical models.[30] Due to the dynamic and seasonal nature of these data, state-space models and related techniques have also been used.[31–33] However, there has been little, if any, application of spatio-temporal dynamic models to climate data. Kokic et al. [11] have recently also highlighted the need for spatio-temporal DLMs in order to achieve greater spatial and temporal coherence in climate projections. The `spTDyn` package presented in this paper clearly paves the way for spatio-temporally coherent projections.

The general structure of the spatially varying and spatio-temporal DLMs discussed in this paper have important applications to different types of spatio-temporal data. The `spTDyn` package uses the Markov chain Monte Carlo (MCMC) technique to obtain model inferences and it provides different choices of hyper-parameters for the prior distributions. The underlying code for these models is written using the low-level language `C` to facilitate fast computation and efficient storage capacity.

This paper presents the spatially varying coefficient models and the spatio-temporal DLMs separately along with several practical examples. We illustrate the spatially varying model with monthly average maximum temperature data over 2000–2009 from the Murray-Darling basin, a large geographical area in the interior of south-eastern Australia. The spatio-temporal DLM is demonstrated with daily maximum 8-hour average ground-level ozone concentration data in the state of New York for 62 days in the months of July and August, 2006. The rational for using monthly maximum temperature data and daily ozone data are discussed later in this paper. We also undertook a simulation study to validate the underlying code of the models in `spTDyn`, but for reasons of brevity omit the details from this paper. We provide detail `R` code in the supplementary file.

The rest of the paper is organized as follows. Section 2 provides the model specifications for the spatially varying and spatio-temporal DLMs used for analysing the data sets. In Section 3 we provide practical analysis and results, where we also include exploratory analysis for the maximum temperature and ozone data sets. Finally, we provide discussion and conclusions in Section 4.

2. Model specification

In this section we describe the general structure of the spatially varying coefficient process models and the spatio-temporal DLMs in detail. We illustrate the Bayesian hierarchical framework for the models with prior distributions. The hierarchical structure described in this paper consists of three stages:

- Stage 1: Data model [data|process, data parameters].
- Stage 2: Process model [process|process parameters].
- Stage 3: Parameter model [data and process parameters].

More details can be found in [2,34–36]. The spatial and temporal predictive distributions for both models are also discussed in this section. Further details on the full conditional distributions are given in Appendices 1 and 2, respectively.

2.1. *Spatially varying coefficient process models*

Let $Y_l(\mathbf{s}_i, t)$ denote the observed point referenced data at site $\mathbf{s}_i$, $i = 1, \ldots, n$ with two units of time where l denotes the longer unit, for example, year, $l = 1, \ldots, r$, and t denotes the shorter unit, for example, day, $t = 1, \ldots, T$. In the first data model stage, we write

$$Y_l(\mathbf{s}_i, t) = O_l(\mathbf{s}_i, t) + \epsilon_l(\mathbf{s}_i, t), \tag{1}$$

where $O_l(\mathbf{s}_i, t)$ is the true underlying process of the model with white noise effect $\epsilon_l(\mathbf{s}_i, t)$. The term $\epsilon_l(\mathbf{s}_i, t)$ follows a normal distribution with mean zero and variance σ_ϵ^2. In the second process model stage, we write the mean process with spatially varying coefficients as

$$O_l(\mathbf{s}_i, t) = \sum_{k=1}^{p} x_{kl}(\mathbf{s}_i, t)\alpha_k + \sum_{j=1}^{q} z_{jl}(\mathbf{s}_i, t)\beta_j(\mathbf{s}_i) + \eta_l(\mathbf{s}_i, t), \tag{2}$$

where $\eta_l(\mathbf{s}_i, t)$ is the spatially correlated error process. The term $x_{kl}(\mathbf{s}_i, t)$ is the kth regressor at time t and site $\mathbf{s}_i$ with non-spatially varying coefficient α_k, $k = 1, \ldots, p$. The term $z_{jl}(\mathbf{s}_i, t)$ is the jth regressor in the same or another set of regressors, $j = 1, \ldots, q$, at time t with spatially varying coefficient $\beta_j(\mathbf{s}_i)$ at site $\mathbf{s}_i$. Note that this variable can also be treated as a grid level output $z_{jl}(\mathbf{s}', t)$ to address the change of support problem of the grid and point-level output and can be written as

$$z_{jl}(\mathbf{s}', t) = \frac{1}{|\mathbf{s}'|} \int_{\mathbf{s}} y_l(\mathbf{s}, t)\, \mathrm{d}\mathbf{s},$$

where $\mathbf{s}'$ is the grid-level data and $\mathbf{s}$ are corresponding point-level observations inside the grid $\mathbf{s}'$. Let $\mathbf{Y}_{lt} = (Y_l(\mathbf{s}_1, t), \ldots, Y_l(\mathbf{s}_n, t))'$, $\mathbf{O}_{lt} = (O_l(\mathbf{s}_1, t), \ldots, O_l(\mathbf{s}_n, t))'$, $\boldsymbol{\epsilon}_{lt} = (\epsilon_l(\mathbf{s}_1, t), \ldots, \epsilon_l(\mathbf{s}_n, t))'$ and $\boldsymbol{\eta}_{lt} = (\eta_l(\mathbf{s}_1, t), \ldots, \eta_l(\mathbf{s}_n, t))'$. We write Equations (1), (2) in vector and matrix notations as

$$\mathbf{Y}_{lt} = \mathbf{O}_{lt} + \boldsymbol{\epsilon}_{lt}, \tag{3}$$

$$\mathbf{O}_{lt} = \mathbf{X}_{lt}\boldsymbol{\alpha} + \sum_{j=1}^{q} \mathbf{Z}_{jlt}\boldsymbol{\beta}_j(\mathbf{s}) + \boldsymbol{\eta}_{lt}, \tag{4}$$

where $\mathbf{X}_{lt}$ represents the $n \times p$ design matrix in the model defined as

$$\mathbf{X}_{lt} = \begin{bmatrix} x_{1l}(\mathbf{s}_1, t) & \ldots & x_{pl}(\mathbf{s}_1, t) \\ \vdots & \ldots & \vdots \\ x_{1l}(\mathbf{s}_n, t) & \ldots & x_{pl}(\mathbf{s}_n, t) \end{bmatrix},$$

and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ is the $p \times 1$ vector of non-spatially varying parameters corresponding to $\mathbf{X}_{lt}$. The term $\boldsymbol{\eta}_{lt} = (\eta_t(\mathbf{s}_1, t), \ldots, \eta_t(\mathbf{s}_n, t))'$ is the spatially correlated error follows a Gaussian process as $\boldsymbol{\eta}_t \sim GP(\mathbf{0}, \Sigma_\eta)$. Here, $\Sigma_\eta = \sigma_\eta^2 \kappa(\phi, \nu)$ is the covariance and $\kappa(\phi, \nu)$ is the spatial correlation that follows Handcock and Stein,[37] Handcock and Wallis [38] and Matérn [39] family defined as

$$\kappa(\mathbf{s}_i, \mathbf{s}_j; \phi, \nu) = \frac{1}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu}\|\mathbf{s}_i - \mathbf{s}_j\|\phi)^\nu K_\nu(2\sqrt{\nu}\|\mathbf{s}_i - \mathbf{s}_j\|\phi), \quad \phi > 0, \nu > 0, \tag{5}$$

where $\Gamma(\nu)$ is the standard gamma function, K_ν is the modified Bessel function of second kind with order ν, and $\|\mathbf{s}_i - \mathbf{s}_j\|$ is the distance between sites $\mathbf{s}_i$ and $\mathbf{s}_j$. The parameter ϕ controls the rate of decay of the correlation as the distance $\|\mathbf{s}_i - \mathbf{s}_j\|$ increases and the parameter ν controls smoothness of the random field.[40,41]

Finally, the term $\mathbf{Z}_{jlt}$ is a $n \times n$ diagonal matrix of another set of covariates for each $j = 1, \ldots, q$; defined as

$$\mathbf{Z}_{jlt} = \begin{bmatrix} z_{jl}(\mathbf{s}_1, t) & 0 & \ldots & 0 \\ 0 & z_{jl}(\mathbf{s}_2, t) & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & z_{jl}(\mathbf{s}_n, t) \end{bmatrix}.$$

[/math1][/math1] This corresponds to the spatially varying $n \times 1$ coefficients $\boldsymbol{\beta}_j(\mathbf{s}) = (\beta(\mathbf{s}_1), \ldots, \beta_j(\mathbf{s}_n))'$, $j = 1, \ldots, q$. We also assume $\boldsymbol{\beta}_j(\mathbf{s}) \sim GP(\mathbf{0}, \Sigma_\beta)$, $\Sigma_\beta = \sigma_\beta^2 \kappa(\phi, \nu)$ same for each covariate at the third parameter stage. Here, σ_β^2 is the hyper-parameter for $\boldsymbol{\beta}_j(\mathbf{s})$ whose prior distribution is detailed below.

2.1.1. *Bayesian framework*

Let $N = nrT$, $\mathbf{y}^*$ denote all missing, and $\mathbf{y}$ denote the non-missing data for $i = 1, \ldots, n$, $t = 1, \ldots, T$ and $l = 1, \ldots, r$. Again, let $\boldsymbol{\theta}$ denote all unknown parameters in the model defined in

Equations (3) and (4). We write the joint posterior distribution as

$$\pi(\boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y}) \propto (\sigma_\epsilon^2)^{-N/2} \exp\left[-\frac{1}{2\sigma_\epsilon^2}\sum_{l=1}^{r}\sum_{t=1}^{T}((\mathbf{Y}_{lt}-\mathbf{O}_{lt})'(\mathbf{Y}_{lt}-\mathbf{O}_{lt}))\right]$$
$$\times |\Sigma_\eta|^{-N/2} \exp\left[-\frac{1}{2\sigma_\eta^2}\sum_{l=1}^{r}\sum_{t=1}^{T}((\mathbf{O}_{lt}-\boldsymbol{\nu}_{lt})'S^{-1}(\mathbf{O}_{lt}-\boldsymbol{\nu}_{lt}))\right]$$
$$\times \prod_{j=1}^{q}(\sigma_\beta^2)^{-n/2} \exp\left[-\frac{1}{2\sigma_\beta^2}\boldsymbol{\beta}_j(\mathbf{s})'S^{-1}\boldsymbol{\beta}_j(\mathbf{s})\right] \times \pi(\sigma_\epsilon^2)\pi(\sigma_\eta^2)\pi(\sigma_\beta^2)\pi(\boldsymbol{\alpha})\pi(\phi), \quad (6)$$

where $\boldsymbol{\nu}_{lt} = \mathbf{X}_{lt}\boldsymbol{\alpha} + \sum_{j=1}^{q}\mathbf{Z}_{jlt}\boldsymbol{\beta}_j(\mathbf{s})$, $S = \kappa(\phi, \nu)$; and $\pi(\sigma_\epsilon^2)$, $\pi(\sigma_\eta^2)$ and $\pi(\phi)$ are the prior distributions for the variance and smooth parameters of the models follows Inverse Gamma distribution with shape and rate hyper-parameters a and b. Hence, we obtain the mean of the distribution as a/b and variance as a/b^2. The hyper-prior distribution for $\pi(\sigma_\beta^2)$ is also Inverse Gamma with parameters a and b. We also assume prior distribution for $\boldsymbol{\alpha}$ follows $\pi(\boldsymbol{\alpha}) \sim N(\mu_\alpha \mathbf{1}_p, \sigma_\alpha^2 \mathbf{I}_{p\times p})$. We consider a proper prior specification for all prior distributions, see details in Section 3.1.1. In Appendix 1, we provide the full conditional distributions of the spatially varying model in detail for the MCMC computation.

2.1.2. *Spatial predictive distribution*

Suppose a prediction to location $\mathbf{s}_0$ at time t and l is required. We can easily get the predictive distribution of $Y_l(\mathbf{s}_0, t)$ from the distribution:

$$Y_l(\mathbf{s}_0, t) \sim N(O_l(\mathbf{s}_0, t), \sigma_\epsilon^2),$$

where we obtain $O_l(\mathbf{s}_0, t)$ sequentially as

$$O_l(\mathbf{s}_0, t) = \sum_{k=1}^{p} x_{kl}(\mathbf{s}_0, t)\alpha_k + \sum_{j=1}^{q} z_{jl}(\mathbf{s}_0, t)\beta_j(\mathbf{s}_0) + \eta_l(\mathbf{s}_0, t)$$

Thus, the posterior spatial predictive distribution of $Y_l(\mathbf{s}_0, t)$ is obtained as

$$\pi(Y_l(\mathbf{s}_0, t)|\mathbf{y}) = \int \pi(Y_l(\mathbf{s}_0, t)|O_l(\mathbf{s}_0, t), \sigma_\epsilon^2) \times \pi(O_l(\mathbf{s}_0, t)|\boldsymbol{\beta}_j(\mathbf{s}_0), \boldsymbol{\theta}, \mathbf{y}^*)$$
$$\times \pi(\boldsymbol{\beta}_j(\mathbf{s}_0)|\boldsymbol{\theta}, \mathbf{y}^*) \times \pi(\boldsymbol{\theta}|\mathbf{y})\, \mathrm{d}O_l(\mathbf{s}_0, t)\, \mathrm{d}\boldsymbol{\beta}_j(\mathbf{s}_0)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\mathbf{y}^*, \quad (7)$$

where $\boldsymbol{\beta}_j(\mathbf{s}_0) = (\beta_1(\mathbf{s}_0), \ldots, \beta_q(\mathbf{s}_0))'$ represents all values for q variables defined in Equation (2). Now, we obtain $\pi(\boldsymbol{\beta}_j(\mathbf{s}_0)|\boldsymbol{\theta})$ from the following distribution for each j as

$$\begin{pmatrix}\beta_j(\mathbf{s}_0)\\ \boldsymbol{\beta}_j(\mathbf{s})\end{pmatrix} \sim N\left[\begin{pmatrix}0\\ \mathbf{0}\end{pmatrix}, \sigma_\beta^2\begin{pmatrix}1 & S_{12}\\ S_{21} & S\end{pmatrix}\right],$$

where S_{12} is $1 \times n$ vector with the ith entry as $\kappa(\mathbf{s}_i, \mathbf{s}_0; .)$ and $S_{12} = S_{21}'$. Hence, we obtain

$$\beta_j(\mathbf{s}_0)|\boldsymbol{\theta} \sim N(S_{12}S^{-1}\boldsymbol{\beta}_j(\mathbf{s}), \sigma_\beta^2(1 - S_{12}S^{-1}S_{21})). \quad (8)$$

We obtain $O_l(\mathbf{s}_0, t)$ from the joint distribution as

$$\begin{pmatrix} O_l(\mathbf{s}_0, t) \\ \mathbf{O}_{lt} \end{pmatrix} \sim N\left[\begin{pmatrix} \sum_{k=1}^{p} x_{kl}(\mathbf{s}_0, t)\alpha_k + \sum_{j=1}^{q} z_{jl}(\mathbf{s}_0, t)\beta_j(\mathbf{s}_0) \\ \mathbf{X}_{lt}\boldsymbol{\alpha} + \sum_{j=1}^{q} \mathbf{Z}_{jlt}\boldsymbol{\beta}_j(\mathbf{s}) \end{pmatrix}, \sigma_\eta^2 \begin{pmatrix} 1 & S_{12} \\ S_{21} & S \end{pmatrix}\right].$$

Thus, the predictive conditional distribution for $O_l(\mathbf{s}_0, t)$ is:

$$O_l(\mathbf{s}_0, t)|\boldsymbol{\beta}_j(\mathbf{s}_0), \boldsymbol{\theta} \sim N(\zeta, \Lambda), \tag{9}$$

where $\Lambda = \sigma_\eta^2(1 - S_{12}S^{-1}S_{21})$ and

$$\zeta = \sum_{k=1}^{p} x_{kl}(\mathbf{s}_0, t)\alpha_k + \sum_{j=1}^{q} z_{jl}(\mathbf{s}_0, t)\boldsymbol{\beta}_j(\mathbf{s}_0) + S_{12}S^{-1}\left(\mathbf{O}_{lt} - \mathbf{X}_{lt}\boldsymbol{\alpha} - \sum_{j=1}^{q} \mathbf{Z}_{jlt}\boldsymbol{\beta}_j(\mathbf{s})\right).$$

Now, we use the MCMC samples obtained from the posterior to get the samples of predictive distributions by using simple step-by-step algorithm.

2.1.3. *Temporal predictive distribution*

To obtain one-step ahead temporal prediction or forecast we consider two particular cases; forecasting (1) in the observed spatial locations $\mathbf{s}_i$, $i = 1, \ldots, n$ and (2) in the unknown spatial location $\mathbf{s}_0$.

Suppose we want to forecast at time $T+1$ in locations $\mathbf{s}_i$ denoted by $Y_l(\mathbf{s}_i, T+1)$. We can easily obtain the posterior temporal predictive distribution of $Y_l(\mathbf{s}_i, T+1)$ from $O_l(\mathbf{s}_i, T+1)$ with variance σ_ϵ^2, and the term $O_l(\mathbf{s}_i, T+1)$ is calculated as:

$$O_l(\mathbf{s}_i, T+1) = \sum_{k=1}^{p} x_{kl}(\mathbf{s}_i, T+1)\alpha_k + \sum_{j=1}^{q} z_{jl}(\mathbf{s}_i, T+1)\beta_j(\mathbf{s}) + \eta_l(\mathbf{s}_i, T+1).$$

Now for forecasting in a new location $\mathbf{s}_0$ at time $T+1$ we define the observation as $Y_l(\mathbf{s}_0, T+1)$. We already obtain the forecast values $Y_l(\mathbf{s}_i, T+1)$ and hence obtain the forecast distribution in location $\mathbf{s}_0$ using spatial predictive distribution described in Section 2.1.2. Finally, we use a simple algorithm to obtain MCMC samples for the temporal predictive distributions.

2.2. *Spatio-temporal DLMs*

For dynamic temporally varying coefficient process model, at the second process model stage, the true process $O_l(\mathbf{s}_i, t)$ in Equation (1) is written as

$$O_l(\mathbf{s}_i, t) = \sum_{k=1}^{p} x_{kl}(\mathbf{s}_i, t)\alpha_k + \sum_{j=1}^{u} z_{jl}(\mathbf{s}_i, t)\beta_j(t) + \eta_l(\mathbf{s}_i, t), \tag{10}$$

where $x_{kl}(\mathbf{s}_i, t)$, α_k and $\eta_l(\mathbf{s}_i, t)$ are defined in Section 2.1. The term $\beta_j(t)$ is the dynamic parameter that varies over time for each covariate $j, j = 1, \ldots, u$. We model $\beta_j(t)$ further in the process stage using state-space structure as

$$\beta_j(t) = \rho_j\beta_j(t-1) + \delta_j(t), \tag{11}$$

where $0 \leq \rho_j \leq 1$ is the autoregressive parameter. Note that the software `spTDyn` developed to fit this model has flexibility on choosing the parameter ρ_j. For example, user has the option to estimate ρ_j (treating this as unknown) or to fix the value of $\rho_j = 1$ to construct a random-walk model, see details in Section 3.2.1.

The term $\delta_j(t)$ is the dynamic error follows $N(0, \sigma_\delta^2)$, with fixed variance σ_δ^2 over time. We also define $\beta_j(0)$ at the initial time point that follows $N(0, \sigma_0^2)$. Let $\boldsymbol{\beta}_t = (\beta_1(t), \ldots, \beta_u(t))'$ and $\boldsymbol{\delta}_t = (\delta_1(t), \ldots, \delta_u(t))'$, hence, using vector and matrix notations we write the models as

$$\mathbf{Y}_{lt} = \mathbf{O}_{lt} + \boldsymbol{\epsilon}_{lt}, \tag{12}$$

$$\mathbf{O}_{lt} = \mathbf{X}_{lt}\boldsymbol{\alpha} + \mathbf{Z}_{lt}\boldsymbol{\beta}_t + \boldsymbol{\eta}_{lt}, \tag{13}$$

$$\boldsymbol{\beta}_t = \mathbf{G}\boldsymbol{\beta}_{t-1} + \boldsymbol{\delta}_t, \tag{14}$$

where $\boldsymbol{\delta}_t \sim GP(\mathbf{0}, \sigma_\delta^2 \mathbf{I}_u)$, $\mathbf{I}_u$ is the $u \times u$ identity matrix. Unlike $\mathbf{Z}_{lt}$ defined in Section 2.1 for the spatially varying model, we define $\mathbf{Z}_{lt}$ as

$$\mathbf{Z}_{lt} = \begin{bmatrix} z_{1l}(\mathbf{s}_1, t) & \ldots & z_{ul}(\mathbf{s}_1, t) \\ \vdots & \ldots & \vdots \\ z_{1l}(\mathbf{s}_n, t) & \ldots & z_{ul}(\mathbf{s}_n, t) \end{bmatrix}.$$

The random effect $\boldsymbol{\eta}_{lt} \sim GP(\mathbf{0}, \Sigma_\eta)$ is described in Section 2.1 with spatial covariance follows Matérn family defined in Equation (5). Finally, the term $\mathbf{G}$ is a $u \times u$ diagonal matrix defined as

$$\mathbf{G} = \begin{bmatrix} \rho_1 & 0 & \ldots & 0 \\ 0 & \rho_2 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & \rho_u \end{bmatrix}.$$

2.2.1. *Bayesian framework*

Let $\boldsymbol{\theta}$ denote all the parameters in the model defined in Equations (12)–(14); and $\boldsymbol{\nu}_{lt} = \mathbf{X}_{lt}\boldsymbol{\alpha} + \mathbf{Z}_{lt}\boldsymbol{\beta}_t$. We write the joint posterior distribution for the spatio-temporal DLM as

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y}) \propto\ & (\sigma_\epsilon^2)^{-N/2} \exp\left[-\frac{1}{2\sigma_\epsilon^2}\sum_{l=1}^{r}\sum_{t=1}^{T}(\mathbf{Y}_{lt} - \mathbf{O}_{lt})'(\mathbf{Y}_{lt} - \mathbf{O}_{lt})\right] \\ & \times |\Sigma_\eta|^{-N/2} \exp\left[-\frac{1}{2\sigma_\eta^2}\sum_{l=1}^{r}\sum_{t=1}^{T}(\mathbf{O}_{lt} - \boldsymbol{\nu}_{lt})' S^{-1}(\mathbf{O}_{lt} - \boldsymbol{\nu}_{lt})\right] \\ & \times (\sigma_\delta^2)^{-uT/2} \exp\left[-\frac{1}{2\sigma_\delta^2}\sum_{t=1}^{T}(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1})'(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1})\right] \\ & \times \pi(\sigma_\epsilon^2)\pi(\sigma_\eta^2)\pi(\sigma_\delta^2)\pi(\sigma_0^2)\pi(\boldsymbol{\alpha})\pi(\phi) \times \prod_{t=1}^{T}\pi(\beta_t) \times \prod_{j=1}^{u}\pi(\rho_j). \end{aligned} \tag{15}$$

We define the prior distribution for $\rho_j \sim N(0, \sigma_\rho^2)$. For rest of the model parameters, we define similar prior distributions discussed in Section 2.1. We chose hyper-parameters of the prior distributions in a way to obtain proper prior specifications, see details in Section 3.2.1. The full conditional distributions obtained from Equation (15) are given in Appendix 2.

2.2.2. *Spatial predictive distribution*

For predicting at location $\mathbf{s}_0$ at time t and l, we write the spatial predictive distribution of $Y_l(\mathbf{s}_0, t)$ as:

$$\pi(Y_l(\mathbf{s}_0,t)|\mathbf{y}) = \int \pi(Y_l(\mathbf{s}_0,t)|O_l(\mathbf{s}_0,t),\sigma_\epsilon^2)\pi(O_l(\mathbf{s}_0,t)|\boldsymbol{\theta},\mathbf{y}^*)\pi(\boldsymbol{\theta}|\mathbf{y})\,\mathrm{d}O_l(\mathbf{s}_0,t)\,\mathrm{d}\boldsymbol{\theta}\,d\mathbf{y}^*. \quad (16)$$

Here, we get $O_l(\mathbf{s}_0, t)|\boldsymbol{\theta}$ from the following joint distribution and use posterior MCMC samples to get the predictive distributions.

$$\begin{pmatrix} O_l(\mathbf{s}_0,t) \\ \mathbf{O}_{lt} \end{pmatrix} \sim N\left[\begin{pmatrix} \sum_{k=1}^{p} x_{kl}(\mathbf{s}_0,t)\alpha_k + \sum_{j=1}^{u} z_{jl}(\mathbf{s}_0,t)\beta_j(t) \\ \mathbf{X}_{lt}\boldsymbol{\alpha} + \mathbf{Z}_{lt}\boldsymbol{\beta}_t \end{pmatrix}, \sigma_\eta^2 \begin{pmatrix} 1 & S_{12} \\ S_{21} & S \end{pmatrix}\right].$$

2.2.3. *Temporal predictive distribution*

Similar to Section 2.1, we calculate one-step ahead temporal prediction for the spatio-temporal DLMs. We obtain forecast in the fitted location $Y_l(\mathbf{s}_i, T+1)$ from the true process $O_l(\mathbf{s}_i, T+1)$, where

$$O_l(\mathbf{s}_i, T+1) = \sum_{k=1}^{p} x_{kl}(\mathbf{s}_i, T+1)\alpha_k + \sum_{j=1}^{u} z_{jl}(\mathbf{s}_i, T+1)\beta_j(T+1) + \eta_l(\mathbf{s}_i, T+1),$$

and $\beta_j(T+1)$ is sampled following Equation (14). For the unobserved location $\mathbf{s}_0$, we get forecast distribution using spatial predictive distribution of the spatio-temporal DLM discussed in Section 2.2.2 at time $T+1$.

2.3. ***Model selection***

In this paper, we use the predictive model choice criteria (PMCC) for comparing models, which is suitable for models with normally distributed errors.[42,43] The PMCC is based on posterior predictive distribution and derived from a minimum posterior predictive loss approach. The first term of the PMCC is a goodness-of-fit term while the second is a penalty term for model complexity. The model with the smallest value of PMCC is preferred among competing models. Thus, to be selected a model must strike a good balance between goodness of fit and model complexity. It is defined as

$$\mathrm{PMCC} = \sum_{i=1}^{n}\sum_{l=1}^{r}\sum_{t=1}^{T} E(Y_l(\mathbf{s}_i,t)_{\mathrm{rep}} - y_l(\mathbf{s}_i,t))^2 + \sum_{i=1}^{n} \mathrm{Var}(Y_l(\mathbf{s}_i,t)_{\mathrm{rep}}). \quad (17)$$

where $Y_l(\mathbf{s}_i, t)_{\mathrm{rep}}$ is a future replicate of the data $Y_l(\mathbf{s}_i, t)$. The package `spTDyn` automatically provides PMCC values for the fitted models. The package also provides a number of validation statistics, for example, mean-squared error (MSE), root-mean-squared error, mean absolute error (MAE), mean absolute percentage error (MAPE), bias, relative bias and relative mean separation. For detailed definitions of these criteria, please see Bakar and Sahu.[14]

3. **Applications**

In this section we analyse two practical data sets using spatially varying coefficient models and the spatio-temporal DLMs. We illustrate the implementations of the `spTDyn` package for both models. We then compare the results with a Gaussian process spatio-temporal model,[2] which do not involve any spatially or temporally varying dynamic parameters. For this we use the package `spTimer` that has the ability to fit and predict data using the Gaussian process spatio-temporal models.[14] We also performed a simulation study to validate the code for the models. However, for brevity we omit the results. Detail `R` code for the simulation study and real-life data examples are given in the supplementary file.

3.1. *Temperature data*

In this paper, we model mean monthly maximum temperature (in °C) data (denote as `Tmax`) using the spatially varying coefficient process models. The data are obtained from 12 monitoring locations in the Murray-Darling basin of Australia from 2000 to 2009 (Figure 1). The data set contains 1440 data points for maximum temperature, among them 6.25% are missing. For fitting the model, we use nine locations and the remaining three locations are set aside for spatial validation. We also set aside the last year's data set for temporal validation.

We use grid-level maximum temperature data (denoted as `grid` in the data set) obtained from the NCEP as a covariate in the model. The NCEP reanalysis data [44] are output from a blending of high quality observed climate data and a dynamic climate model. These data are available historically over a 2.5 × 2.5 degree global grid for a range of climatic variables including rainfall and temperature.

The grid-level NCEP data are spatially misaligned with the point-level measurement locations. As discussed in Section 1, we use the spatially varying coefficient process model to tackle this spatial misalignment. We also use the Southern Oscillation Index (SOI) as a predictor of maximum temperatures in south-east Australia.[15] These values are measured in a way that assumes constant spatial effect; however, their effects may vary over space. We include this variable in our model to obtain its spatial effect through spatially varying coefficients.

Table 1 provides summary statistics for maximum temperature and explanatory variables used for the spatially varying model. The maximum temperature varies from 12.10°C to 40.20°C

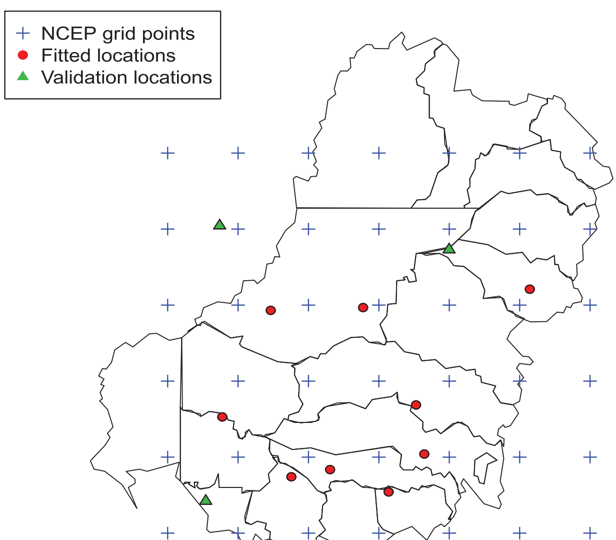

Figure 1. Map of the Murray-Darling basin, Australia. The temperature monitoring locations and National Centers for Environmental Prediction (NCEP) grid locations are superimposed. Fitted and validation locations are also superimposed on the map.

Table 1. Summary statistics for the variables used in the spatially varying coefficient process models.

Variables		Minimum	Mean	Maximum
Predictor	Tmax (°C)	12.10	25.09	40.20
Predictand	SOI (index)	– 29.10	0.01	22.40
	grid (°C)	8.71	24.74	43.28

in our region of study, whereas the grid output for maximum temperature varies from 8.71°C to 43.28°C. The SOI has average index value as 0.01 with minimum and maximum values of −29.1 and 22.4, respectively. The explanatory variables used in this paper have no missing observations.

The NCEP grid data are stored in a 4-dimensional data matrix structure in `R`, where the first two dimensions of the matrix are defined as the longitude and latitude positions. In this paper, we define the last two dimensions of the matrix as month and year, respectively. In the `spTDyn` package, we also create a function `ObsGridData` to prepare the NCEP data for modelling purposes. Note that the maximum temperature data set modelled in Section 3.1.1 using the spatially varying coefficient process model includes grid-level information for the temperature covariate a priori.

3.1.1. *Analysis for the Australian maximum temperature data*

We modified the static downscaler model structure described in [1] and define the true process $O_l(\mathbf{s}_i, t)$ of the spatially varying coefficient process model stated in Equation (1) as

$$O_l(\mathbf{s}_i, t) = \alpha + \sum_{j=1}^{2} z_{jl}(\mathbf{s}_i, t)\tilde{\beta}_j + \eta_l(\mathbf{s}_i, t), \tag{18}$$

where

$$\tilde{\beta}_j = \beta_{j0} + \beta_j(\mathbf{s}_i), \quad j = 1, 2. \tag{19}$$

The spatially varying model defined by Equations (18) and (19) is a special case of the model defined at Equation (2). To fit the spatially varying coefficient model defined above using `spTDyn`, one needs to set the argument `sp` in the `formula` statement. For example,

```
R> formula = tmax soi + sp(soi) + grid + sp(grid)
```

allows `spTDyn` to identify variables ‘SOI’ and ‘grid’ as having spatially varying coefficients $\beta_j(\mathbf{s}_i)$, $j = 1, 2$. Note that the `formula` argument also uses the intercept term α and the non-spatially varying effects β_{j0}, $j = 1, 2$ for the covariates defined in model (19). It is also possible to omit the intercept term from the model using ‘-1’ in the `formula` argument. Thus, a typical code example in `spTDyn` for the spatially varying coefficient models is

```
R> post.sp <- GibbsDyn(tmax soi+sp(soi)+grid+sp(grid), data=AUSdataFit,
+ coords= lon+lat, priors=NULL, initials=NULL, nItr=13000, nBurn=3000,
+ spatial.decay=decay(distribution=Gamm(2,1),tuning=0.06))
```

A number of remarks are in order. The argument `coords` requires the spatial coordinate points of the data set. This can be supplied using an $n \times 2$ matrix or data frame of longitude (`lon`) and latitude (`lat`) axes or by using a formula argument $\sim$`lon`$+$`lat`. Note that for supplying the `coords` using formula argument, one should make sure that the `data` contain the corresponding variables for spatial coordinates.

Users can specify the values for the hyper-parameters of the prior distributions using the argument `priors`. The argument `initials` can be used to provide initial values of the parameters for the MCMC chain. However, suitable defaults are provided in `spTDyn` by writing `NULL`. For example, `priors = NULL` indicates that model parameters uses a conjugate prior with a large variance value like 10^{10}. For `initials = NULL`, the default initial values are $\sigma_\eta^2 = 0.1$, $\sigma_\epsilon^2 = 0.01$. For the spatial decay parameter, ϕ the default is set as $-\log(0.05)/d_{\max} \approx 3/d_{\max}$, where $d_{\max}$ is the maximum distance calculated from the coordinates of the model fitting locations.[45] The initial values for the regression parameters are obtained by fitting a simple linear model using the `lm` function in `R`.

The argument `spatial.decay` is used for sampling ϕ. This can be done using the function `decay`. In this example, we use a Gamma prior distribution with shape and rate hyper-parameters $a = 2$ and $b = 1$, respectively, with mean a/b and variance a/b^2. We also need to tune the normal proposal distribution for ϕ using the argument `tuning` inside the function `decay` to obtain acceptance rate between 15 % and 40% , see Gelman et al.[46] Note that the Metropolis–Hastings algorithm is used only for ϕ as a closed form of full conditional distribution is not known. We use the Gibbs sampler for other parameters in the models, see details in Appendices 1 and 2. User can also use uniform discrete prior distribution or provide a fixed value for ϕ to proceed with an empirical Bayes approach as used by Sahu et al.[47]

The function `GibbsDyn` contains few other arguments. For example, the argument `cov.fnc` is used for defining the covariance model, we use for fitting the data. Currently, it can take `exponential`, `gaussian`, `spherical` and `matern`. Note that for the Matérn model the smoothing parameter ν is estimated using a discrete uniform distribution. The argument `distance.method` of function `GibbsDyn` is used to calculate distances between locations. It can take geodetic calculations in kilometres or in miles, it is also possible to use other distance measurements, for example, Euclidean. There is also an option for on-the-fly transformation of the data to `LOG` and `SQRT` scale, and the posterior predictive distribution will return the values on the original scale. For further details, see the `spTDyn` package manual.

We use 13,000 iterations for the MCMC run, where the first 3000 are discarded as burn-in. The user also has the flexibility to change the prior distributions and other input for the MCMC chain. It is also possible to perform the MCMC diagnostics using the `coda` package; see details in `spTDyn` for further specifications. We obtain the MCMC summary statistics using the `summary` argument:

```
> summary(post.sp)
##
# Spatially varying parameters are not included.
##
 -----------------------------------------------------
Model: GP
Call: tmax soi + sp(soi) + grid + sp(grid)
Iterations: 13000
nBurn: 3000
Acceptance rate for phi (
-----------------------------------------------------
  Goodness.of.fit Penalty PMCC
values:   1382.35 1733.04 3115.39
-----------------------------------------------------
Computation time: 22.74 - Sec.
-----------------------------------------------------
Parameters:
```

```
    Mean Median  SD Low2.5p Up97.5p
(Intercept) 1.2299 1.2001 0.4439 0.4234 2.1736
soi   0.0857 0.0856 0.0133 0.0600 0.1124
grid  0.9210 0.9218 0.0171 0.8848 0.9523
sig2eps  0.1136 0.1114 0.0233 0.0746 0.1662
sig2eta  3.4972 3.3458 0.7034 2.6020 5.2564
sig2beta 0.1109 0.1042 0.0367 0.0603 0.2043
phi   0.0023 0.0023 0.0004 0.0014 0.0031
------------------------------------------------------
```

The coefficient estimates of the terms soi and grid in the above R output are for the fixed β_{j0}, $j = 1, 2$ coefficients of the explanatory variables, as defined in Equation (19). We also observe from the summary output that, as expected, the spatial variance σ_η^2 is much higher than the nugget effect σ_ϵ^2. Furthermore, we can obtain the summary statistics for the spatially varying parameters by typing `coefficient = "spatial"` inside the summary statement:

```
R> summary(post.sp, coefficient="spatial")
```

The output of this summary statement is omitted here. Similarly, we can also plot the MCMC samples using the `plot` function in `R`.

Figure 2 shows the distribution of the spatially varying coefficients for the explanatory variables. We observe for both `soi` and `grid` variables, the parameter estimates vary by locations. This also indicates the necessity of using the spatially varying coefficient process models compared with the constant or non-spatially varying spatio-temporal models.

To obtain the predictive posterior distribution of the spatially varying coefficient process models, we set aside two validation data sets: `AUSdataPred` for spatial and `AUSdataFore` for temporal predictions. Typical code for spatial prediction in `spTDyn` is:

```
R> pred.sp<-predict(post.sp,newcoords= lon+lat,newdata=AUSdataPred)
```

It is also possible to use further burn-in of the MCMC samples using the `nBurn` argument in function `predict`. Figure 3(a) and 3 (b) shows the mean of the predictive values of the spatial parameters for the variables `soi` and `grid` that correspond to $\beta_j(\mathbf{s}_0)$, $j = 1, 2$ in Equation (8). We observe a negative spatial effect of `soi` in the northern part of the study region, whereas a positive effect is observed in the southern part. On the other hand for the grid-based NCEP

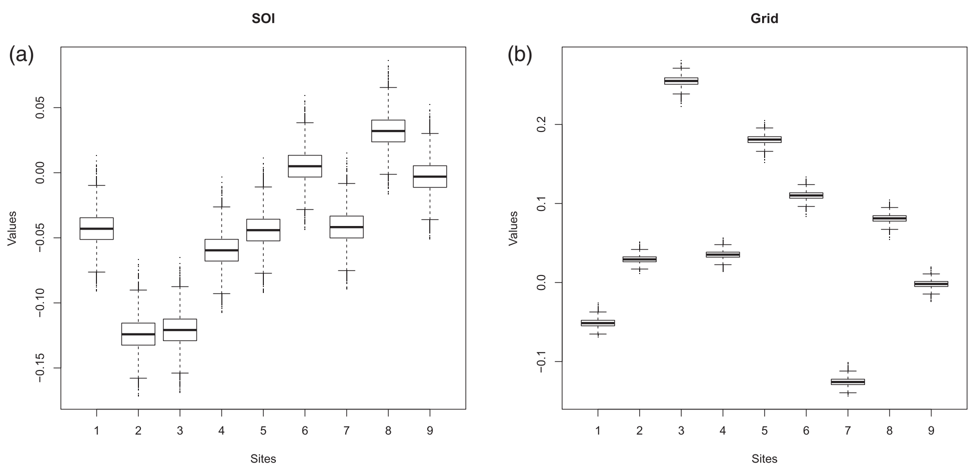

Figure 2. Boxplot of the spatially varying coefficients for the variables (a) SOI and (b) grid-level data.

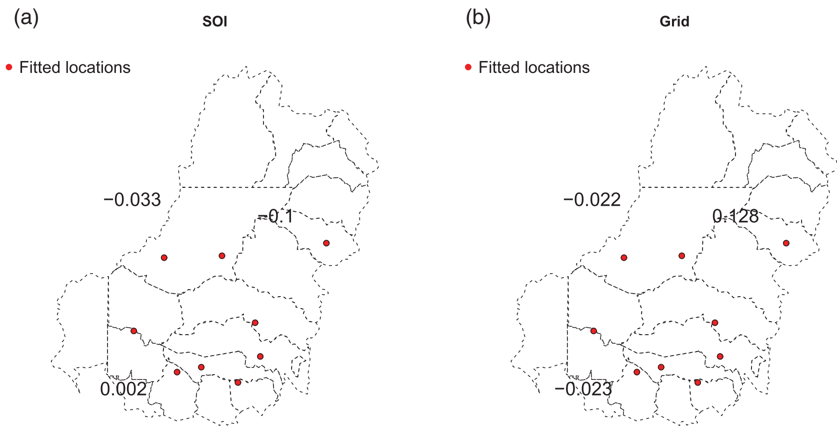

Figure 3. Estimates of the spatial coefficients in unmonitored locations for the explanatory variables (a) SOI and (b) grid-level NCEP output.

covariate, grids on the eastern part of the study region have a positive effect and western part have a negative effect on maximum temperature.

We can also obtain temporal prediction of the monthly maximum temperature using the set aside validation data `AUSdataFore` for 12 months in 2009. For example, we write a typical `spTDyn` code example as

```
R> pred.sp.f <- predict(post.sp,type="temporal",foreStep=12,
   + newcoords= lon+lat, newdata=AUSdataFore)
```

Note that we define `type = "temporal"` in `predict` function and also provide number of days ahead forecast using `foreStep = 12`.

3.1.2. *Comparisons*

A spatio-temporal Gaussian process model without spatially varying coefficients is fitted with the same data set and compared with the spatially varying coefficient process models, see the `R` code in the supplementary file. The PMCC values [42] are give in Table 2 for both models. They indicate a significantly better description for the spatially varying model compared with the spatio-temporal model.

Comparison results are also provided for the predictive performance of the models. Table 3 shows some validation criteria, for example, MSE, MAE, MAPE and bias for the spatially varying coefficient process model and the non-spatially varying model. The argument `spT.validation` of the package `spTimer` [14] provides these validation statistics. For both the spatial and temporal out-of-sample predictions, the spatially varying model shows smaller

Table 2. Predictive model choice criteria (PMCC) for spatially and non-spatially varying coefficient process spatio-temporal models. Here, Gof refers to the goodness of fit.

Coefficients of the spatio-temporal models	Gof	Penalty	PMCC
Spatially varying	1383.67	1731.60	3115.27
Fixed	2051.77	2810.19	4861.96

Table 3. Validation statistics for the spatially and non-spatially varying models for spatial and temporal predictions.

	Spatial prediction		Temporal prediction	
Criteria	Coefficients of the spatio-temporal models			
Models	Spatially varying	Fixed	Spatially varying	Fixed
MSE	5.49	9.81	19.52	38.39
MAE	1.95	2.79	3.97	5.04
MAPE (%)	8.19	11.00	14.02	16.15
BIAS	− 0.76	− 1.50	− 1.20	− 4.01

values of the various predictive accuracy criteria. Thus, we conclude that this is the preferred predictive model for the maximum temperature data set.

3.2. *Ozone data*

The daily maximum 8-hour average ozone concentration data from 28 monitoring sites in New York (NY) have previously been examined by Sahu and Bakar.[29] It is analysed in this paper using the spatio-temporal DLM, see Figure 4. A total 62 days of observations in the months of July and August in 2006 are used for analysis. Hence, we have a total 1736 data points of which 1.38% are missing. We fit the spatio-temporal DLM using data from 20 locations and the remaining 8 locations are set a side for validation purpose. We also set aside the last 5 days data for forecast validation. We use maximum 8-hour average ozone in parts per billion (ppb) as this is the measurement standard defined by the US environmental protection agency.[1] In our data set, we denote this ozone measurement as `o8hrmax`.

We consider the state of NY because the ozone monitoring network in this state represents a typical practical situations, that is a cluster of a few sites in and around a big city and a moderate number of other sites, situated large distances apart and covering a vast region. We only

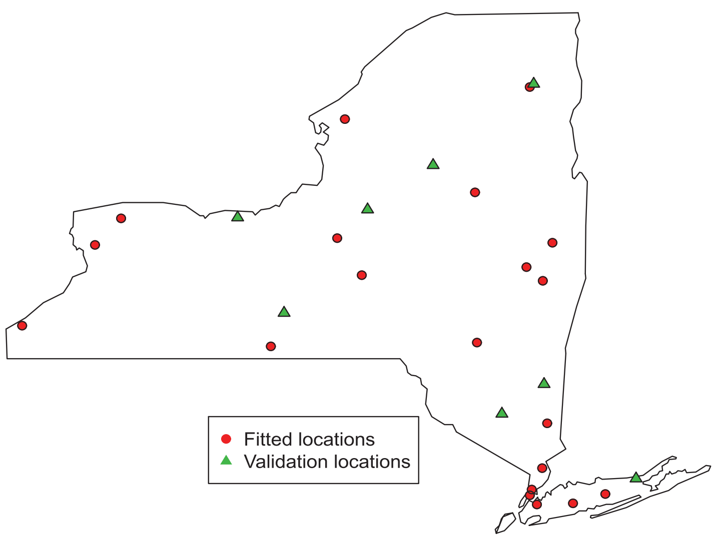

Figure 4. The ozone monitoring locations in the state of New York. Fitted and validation locations are also superimposed.

Table 4. Summary statistics for the variables used in the spatio-temporal DLMs.

Variables		Minimum	Mean	Maximum
Predictor	o8hrmax	12.38	47.82	107.00
Predictand	cMAXTMP (°C)	18.16	27.54	37.06

consider July and August as this is the high ozone season in the USA; detailed explanation of the data set are given in [29]. To illustrate the spatio-temporal DLM, we use the most important and statistically significant covariate, maximum temperature (`cMAXTMP` in °C) for modelling ozone data.[14,29,48,49] Ozone levels in NY for the study time period range from 12.38 parts per billion (ppb) to 107 ppb with an average of 47.82 ppb (Table 4). Maximum temperature ranges from 18.16°C to 37.06°C. Furthermore, the ozone level has a diurnal pattern that reflects the one day increase in ozone level at a measurement point and a decrease in the following day related to the change of maximum temperature in that area. These data are modelled in Section 3.2.1 using spatio-temporal DLM. Note that we denote the maximum temperature as '`Tmax`' in `AUSdata` and '`cMAXTMP`' in `NYdata` for the spatially varying and the spatio-temporal DLMs, respectively; see Tables 1 and 4.

3.2.1. *Analysis for the NY ozone data*

The daily maximum 8-hour average ozone level is modelled using the spatio-temporal DLM that described in Section 2.2. A summary description of the variables are also given in Section 3.2. We model ozone levels with dynamic covariate effect of maximum temperature (`cMAXTMP`). In particular, the following model specification of the spatio-temporal DLM defined in Equations (12)–(14) is used.

$$O_l(\mathbf{s}_i, t) = z_l(\mathbf{s}_i, t)\beta(t) + \eta_l(\mathbf{s}_i, t), \tag{20}$$

$$\beta(t) = \beta(t-1) + \delta(t). \tag{21}$$

Note that in the observation equation (20) we only use the temporal dynamics for maximum temperature, and in the state equation (21) we use a pure random-walk model by assuming $\rho = 1$, that is, defined in Equation (11). Also note that Equations (20) and (21) is a particular case of the spatio-temporal DLMs in Equations (12)–(14).

The argument `tp` is used in `formula` of the `GibbsDyn` function to specify which covariates are dynamic.

```
R> formula <- o8hrmax tp(cMAXTMP) - 1
```

Note that we omit the intercept term from the analysis by using '-1' in the `formula` argument. To analyse ground-level ozone, we fix ρ at 1 by setting the initial value of ρ in argument `initials` to one.

```
R> initials <- initials(rhotp=1)
```

Users of `spTDyn` can also get estimates of the auto-regressive parameter ρ defined in Equation (11) using the argument `rhotp = 0`. For NY ozone data, we estimated its median as 0.96. However, we use a random-walk model for illustrative purpose.

Typical R code to fit the ozone data with a spatio-temporal DLM and to produce MCMC summary statistics for the model parameters is

```
R> post.tp <-GibbsDyn(o8hrmax tp(cMAXTMP)-1, coords=Longitude+Latitude,
+ data=NYdataFit,initials=initials(rhotp=1), scale.transform="SQRT",
+ spatial.decay=decay(distribution=Gamm(2,1),tuning=0.01))
```

Note that we use an on-the-fly square root transformation for ozone data as it stabilises the variability, see details in [29,50,51]. By default `spTDyn` considers a proper prior specification, where the hyper-parameters for the Gamma distribution is consider as $a = 1$ and $b = 2$; in addition the variance of all normal prior distributions are considered as 10^{10} with mean zero. The function `summary(post.tp)` will give summary statistics of the model parameters, and to obtain summary statistics of the temporal dynamic parameters of the model we include the argument `coefficients = "temporal"` inside the function `summary`.

Figure 5 shows the 95% credible interval plots of the dynamic maximum temperature coefficient we used in the model. We observe that the estimates vary over time and indicate a dynamic changing pattern, which justifies the use of spatio-temporal DLM for the NY data. Similar to the spatially varying model, for prediction we use the S3 class function `predict`. The default choice is spatial prediction. For a temporal prediction (or forecast) set the argument `type = "temporal"` inside the `predict` function. Detailed code to do prediction is given in the supplementary file.

For the inference of this model and the one in Section 3.1.1, `spTDyn` took about 19.05 and 22.74 s, respectively, to run 13,000 iteration on a laptop with a 2.8 GHz processor, which could be dozens of times faster than `OpenBUGS` or `JAGS` for these medium-sized data sets.

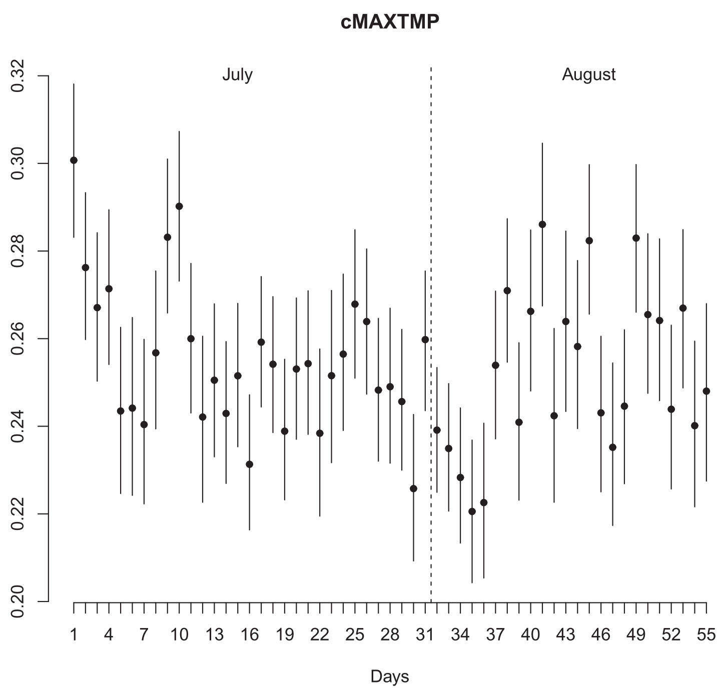

Figure 5. Boxplot of the temporally varying coefficients for the covariate maximum temperature (cMAXTMP).

Table 5. Predictive model choice criteria (PMCC) for temporally and non-temporally varying models. Here, Gof refers to the goodness of fit.

Coefficients of the spatio-temporal models	Gof	Penalty	PMCC
Dynamic	98.41	426.05	524.46
Fixed	165.78	553.24	719.02

Table 6. Validation statistics for the temporally and non-temporally varying models for spatial and temporal predictions.

Criteria	Spatial prediction		Temporal prediction	
	Coefficients of the spatio-temporal models			
Models	Spatio-temporal DLM	Fixed	Spatio-temporal DLM	Fixed
MSE	61.32	68.97	60.58	71.25
MAE	6.21	6.62	6.01	6.67
MAPE (%)	13.62	14.65	19.71	21.35
BIAS	1.06	1.82	− 0.55	− 1.57

3.2.2. *Comparisons*

A non-temporally varying Gaussian process spatio-temporal model is fitted using the same covariate `cMAXTMP` of the NY ozone data to compare its performance with the spatio-temporal DLM. Another model that we can consider for comparison purpose is an auto-regressive model, where temporal dynamics are captured through true underlying process. However, we omit this analysis in this paper as this has been done by Sahu and Bakar.[29]

From Table 5, we observe that the PMCC value for the spatio-temporal DLM is much smaller than for the spatio-temporal models, which demonstrates the superior model fitting performance of the DLM. This result is further confirmed by the validation statistics for the hold out data sets, see Table 6. We observe for both spatial and temporal predictions the spatio-temporal DLM provides lower, that is, better, MSE. Other validation criteria, for example, MAE, MAPE and BIAS also support the same comparison conclusion.

4. Discussion and conclusion

This paper develops the `spTDyn` package in `R` for Bayesian hierarchical spatially varying coefficient process models and for spatio-temporal DLMs. The MCMC sampling algorithm is used to draw inferences for the model parameters. Associated methods and code are developed for producing spatial and temporal predictive distributions for the general model structures considered in the paper. Several practical applications of the software to real data sets were considered. It was illustrated that the proposed models can have superior predictive performance compared with the spatio-temporal models in [2,40].

The underlying code of `spTDyn` is written in `C` to facilitate fast computation and efficient storage capacity, while the front end is written in `R` making the sophisticated modelling techniques easily accessible to statistical analysts. The `spTDyn` package gives researchers the flexibility to develop different types of spatially varying coefficient process models and spatio-temporal DLMs as required by the nature of data. For example, the specific models used for analysing maximum temperature and ozone concentration data are complex in structure, but they are both special cases of the general model structure given in Section 2. The software also allows the user to choose from a wide variety of different prior distributions for the model parameters.

In future enhancements, the spatially varying coefficient process model and the spatio-temporal DLM can potentially be improved in several ways. For example, it could be extended to handle a large number of spatial locations. This is the well-known big-n problem, and can be addressed with row-rank approximation or spectral methods.[2,52,53] It should also be possible to develop an additive form of the model that incorporates both spatially varying and temporal dynamic coefficient strategies, and which includes a non-separable space-time covariance structure.[2] In addition, an extension of these models to non-Gaussian first stage models or areal data can also be considered.[8] Another possible extension is to include other Bayesian model selection criteria, such as deviance information criterion (DIC).[54]

Acknowledgements

The authors are grateful to the anonymous reviewers for their constructive comments. They would like to acknowledge the advice provided by Dr Mark Howden, Dr Steve Crimp and Mr Garry Hopwood of CSIRO. The authors also wish to acknowledge the support of the Agriculture Flagship, CSIRO, for funding this research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Note

1. See, http://www.epa.gov/region1/airquality/avg8hr.html

References

[1] Berrocal VJ, Gelfand AE, Holland DM. A spatio-temporal downscaler for output from numerical models. J Agric Biol Environ Stat. 2010;15(2):176–197.
[2] Cressie NAC, Wikle CK. Statistics for spatio-temporal data. New York: Wiley; 2011.
[3] Gelfand AE, Kim H-J, Sirmans CF, Banerjee S. Spatial modeling with spatially varying coefficient processes. J Amer Statist Assoc. 2003;98:387–396.
[4] Vanem E, Huseby AB, Natvig B. Bayesian hierarchical spatio-temporal modelling of trends and future projections in the ocean wave climate with a CO_2 regression component. Environ Ecol Stat. 2014;21(2):189–220.
[5] Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput. 2000;10(4):325–337.
[6] Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. Stat Med. 2009;28(25):3049–3067.
[7] Plummer M, et al. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vol. 124. Vienna; 2003. p. 125.
[8] Arab A, Courter JR. Spatio-temporal trend analysis of spring arrival data for migratory birds. Commun Stat – Simul Comput. http://dx.doi.org/10.1080/03610918.2013.809100
[9] Finley AO, Banerjee S, Carlin BP. `spBayes`: univariate and multivariate spatial modeling. `R` package version 0.3-8. 2013.
[10] Finley AO, Banerjee S, Gelfand AE. Bayesian dynamic modeling for large space-time datasets using Gaussian predictive processes. J Geograph Syst. 2012;14(1):29–47.
[11] Kokic P, Jin H, Crimp S. Improved point scale climate projections using a block bootstrap simulation and quantile matching method. Climate Dyn. 2013;41(3–4):853–866.
[12] Sigrist F, Künsch HR, Stahel WA. spate: an R package for statistical modeling with SPDE based spatio-temporal Gaussian processes. Technical report, Seminar for Statistics, ETH Zurich; 2012.
[13] Sigrist F, Kunsch HR, Stahel WA. A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. Ann Appl Stat. 2013;6:1452–1477.
[14] Bakar KS, Sahu SK. spTimer: Spatio-temporal Bayesian modelling using `r`. J Statist Softw. 2015;63(15):1–32.
[15] Power S, Tseitkin F, Mehta V, Lavery B, Torok S, Holbrook N. Decadal climate variability in australia during the twentieth century. Int J Climatol. 1999;19(2):169–184.
[16] Gotway CA, Young LJ. A geostatistical approach to linking geographically aggregated data from different sources. J Comput Graph Stat. 2007;16:115–135.

[17] Fuentes M, Raftery AE. Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. Biometrics. 2005;61:36–45.
[18] Arab A, Jackson MC, Kongoli C. Modelling the effects of weather and climate on malaria distributions in West Africa. Malaria J. 2014;13(126):1.
[19] Cai B, Lawson AB, Hossain MM, Choi J, Kirby RS, Liu J. Bayesian semiparametric model with spatially-temporally varying coefficients selection. Stat Med. 2013;32(21):3670–3685.
[20] Jensen SJ, Park J, Braunstein AF, McAuliffe J. Bayesian hierarchical modeling of the HIV evolutionary response to therapy. J Amer Statist Assoc. 2013;108(508):1230–1242.
[21] West M, Harrison J. Bayesian forecasting and dynamic models. 2nd ed. New York: Springer; 1997.
[22] Stroud JR, Muller P, Sanso B. Dynamic models for spatiotemporal data. J R Statist Soc B. 2001;63(4):673–689.
[23] Huerta G, Sanso B, Stroud JR. A spatiotemporal model for mexico city ozone levels. J R Statist Soc Ser C. 2004;53(2):231–248.
[24] Gelfand AE, Banerjee S, Gamerman D. Spatial process modelling for univariate and multivariate dynamic spatial data. Environmetrics. 2005;16(5):465–479.
[25] Gamerman D, Salazar E, Reis E. Dynamic Gaussian process priors, with applications to the analysis of space-time data (with discussion). In: Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM, West M, editors. Bayesian statistics 8. Oxford: Oxford University Press; 2007. p. 149–174.
[26] Dou Y, Le ND, Zidek JV. Modeling hourly ozone concentration fields. Ann Appl Stat. 2010;4(3):1183–1213.
[27] Ghosh SK, Bhave PV, Davis JM, Lee H. Spatio-temporal analysis of total nitrate concentrations using dynamic statistical models. J Amer Statist Assoc. 2010;105(490):538–551.
[28] Mahmoudian B, Mohammadzadeh M. A spatio-temporal dynamic regression model for extreme wind speeds. Extremes. 2014;17(2):221–245.
[29] Sahu SK, Bakar KS. A comparison of Bayesian models for daily ozone concentration levels. Statist Methodol. 2012;9(1):144–157.
[30] Fowler HJ, Blenkinsop S, Tebaldi C. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. Int J Climatol. 2007;27:1547–1578.
[31] Greene AM, Robertson AW, Kirshner S. Analysis of indian monsoon daily rainfall on subseasonal to multidecadal time scales using a hidden Markov model. Q J R Meteorol Soc. 2008;134:875–887.
[32] Kokic P, Crimp S, Howden M. Forecasting climate variables using a mixed-effect state-space model. Environmetrics. 2011;22:409–419.
[33] Shumway RH, Stoffer DS. Time series analysis and its applications: with R examples. 3rd ed. New York: Springer; 2010.
[34] Arab A, Hooten MB, Wikle CK. Hierarchical spatial models. In: Shekar S, Xiong H, editors. Encyclopedia of GIS. New York: Springer US; 2008. p. 425–431.
[35] Berliner LM. Hierarchical Bayesian time series models. In: Hanson KM, Silver RN, editors. Proceedings of the fifteenth international workshop on maximum entropy and Bayesian methods, Santa Fe, New Mexico, USA; 1996. p. 15–22.
[36] Gelfand AE. Hierarchical modeling for spatial data problems. Spatial Stat. 2012;1:30–39.
[37] Handcock MS, Stein ML. A Bayesian analysis of kriging. Technometrics. 1993;35:403–410.
[38] Handcock MS, Wallis JR. An approach to statistical spatial-temporal modelling of meteorological fields. J Amer Statist Assoc. 1994;89:368–390.
[39] Matérn B. Spatial variation. 2nd ed. Berlin: Springer-Verlag; 1986.
[40] Banerjee S, Carlin BP, Gelfand AE. Hierarchical modeling and analysis for spatial data. Boca Raton, FL: Chapman & Hall/CRC; 2004.
[41] Cressie NAC. Statistics for Spatial Data. New York: Wiley; 1993.
[42] Gelfand AE, Ghosh SK. Model choice: a minimum posterior predictive loss approach. Biometrika. 1998;85(1): 1–11.
[43] Laud P, Ibrahim J. Predictive model selection. J R Statist Soc. Ser B (Methodol). 1995;57(1):247–262.
[44] Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J. The NCEP/NCAR 40-year reanalysis project. Bull Amer Meteorol Soc. 1996;77(3):437–471.
[45] Finley AO, Banerjee S, Carlin BP. spBayes: an r package for univariate and multivariate hierarchical point-referenced spatial models. J Statist Softw. 2007;19(4):1–24.
[46] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. 2nd ed Boca Raton, FL: Chapman & Hall/CRC; 2004.
[47] Sahu SK, Gelfand AE, Holland DM. High-resolution spacetime ozone modeling for assessing trends. J Amer Statisti Assoc. 2007;102(480):61–86.1221–1234.
[48] Davis J, Cox W, Reff A, Dolwick P. A comparison of CMAQ-based and observation-based statistical models relating ozone to meteorological parameters. Atmos Environ. 2011;45(20):3481–3487.
[49] Steiner AL, Davis AJ, Sillman S, Owen RC, Michalak AM, Fiore AM. Observed suppression of ozone formation at extremely high temperatures due to chemical and biophysical feedbacks. Proc Nat Acad Sci. 2010;107(46):19685–19690.
[50] Sahu SK, Gelfand AE, Holland DM. Spatio-temporal modeling of fine particulate matter. J Agric Biol Environ Stat. 2006;11:61–86.
[51] Sahu SK, Yip S, Holland DM. Improved space-time forecasting of next day ozone concentrations in the eastern U.S. Atmos Environ. 2009;43:494–501.

[52] Banerjee S, Gelfand AE, Finley AO, Sang H. Gaussian predictive process models for large spatial data sets. J R Statist Soc B. 2008;70:825–848.
[53] Sahu SK, Bakar KS. Hierarchical Bayesian autoregressive models for large space time data with applications to ozone concentration modelling. Appl Stoch Models Bus Ind. 2012;28:395–415.
[54] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. J R Statist Soc: Ser B (Statist Methodol). 2002;64(4):583–639.

Appendix 1. Conditional distributions: spatially varying process model

- Using the kernel of the joint posterior distribution of Equation (6), we obtain the conditional distributions of the variance parameters $\sigma_\epsilon^2, \sigma_\eta^2, \sigma_\beta^2$:

$$\pi\left(\frac{1}{\sigma_\epsilon^2}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y}\right) \sim \text{Gam}\left(\frac{N}{2}+a, b+\frac{1}{2}\sum_{l=1}^{r}\sum_{t=1}^{T}(\mathbf{Y}_{lt}-\mathbf{O}_{lt})'(\mathbf{Y}_{lt}-\mathbf{O}_{lt})\right),$$

$$\pi\left(\frac{1}{\sigma_\eta^2}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y}\right) \sim \text{Gam}\left(\frac{N}{2}+a, b+\frac{1}{2}\sum_{l=1}^{r}\sum_{t=1}^{T}(\mathbf{O}_{lt}-\boldsymbol{\nu}_{lt})'S^{-1}(\mathbf{O}_{lt}-\boldsymbol{\nu}_{lt})\right),$$

$$\pi\left(\frac{1}{\sigma_\beta^2}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y}\right) \sim \text{Gam}\left(\frac{nq}{2}+a, b+\frac{1}{2}\sum_{j=1}^{q}\boldsymbol{\beta}_j(\mathbf{s})'S^{-1}\boldsymbol{\beta}_j(\mathbf{s})\right),$$

  where $\boldsymbol{\theta}_{(-1)}$ represents all parameters in the models except the parameter for which we are obtaining the distribution.
- We can get the conditional distribution of $\boldsymbol{\alpha}$ is defined as $\pi(\boldsymbol{\alpha}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y})$ and distributed as $N(\Delta\chi, \Delta)$, where

$$\Delta^{-1} = \sum_{l=1}^{r}\sum_{t=1}^{T}\mathbf{X}_{lt}'\Sigma_\eta^{-1}\mathbf{X}_{lt} + \frac{1}{\sigma_\alpha^2}\mathbf{I}_{p\times p},$$

$$\chi = \sum_{l=1}^{r}\sum_{t=1}^{T}\mathbf{X}_{lt}'\Sigma_\eta^{-1}\left(\mathbf{O}_{lt} - \sum_{j=1}^{q}\mathbf{Z}_{jlt}\boldsymbol{\beta}_j(\mathbf{s})\right) + \frac{\mu_\alpha}{\sigma_\alpha^2}\mathbf{1}_p.$$

- Similarly, we can obtain the conditional distribution of $\boldsymbol{\beta}_j(\mathbf{s})$ as $N(\Delta_{jlt}\chi_{jlt}, \Delta_{jlt})$, where

$$\Delta_{jlt}^{-1} = \sum_{l=1}^{r}\sum_{t=1}^{T}\mathbf{Z}_{jlt}'\Sigma_\eta^{-1}\mathbf{Z}_{jlt} + \Sigma_\beta^{-1},$$

$$\chi_{jlt} = \sum_{l=1}^{r}\sum_{t=1}^{T}\mathbf{Z}_{jlt}'\Sigma_\eta^{-1}\left(\mathbf{O}_{lt} - \mathbf{X}_{lt}\boldsymbol{\alpha} - \sum_{k\neq j}\mathbf{Z}_{klt}\boldsymbol{\beta}_k(\mathbf{s})\right).$$

- From Equation (6), we obtain the conditional distribution $\pi(\mathbf{O}_{lt}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y})$, that follows $N(\Delta_{lt}\chi_{lt}, \Delta_{lt})$, where

$$\Delta_{lt}^{-1} = \frac{1}{\sigma_\epsilon^2}\mathbf{I}_{n\times n} + \Sigma_\eta^{-1},$$

$$\chi_{lt} = \frac{1}{\sigma_\epsilon^2}\mathbf{Y}_{lt} + \Sigma_\eta^{-1}\left(\mathbf{X}_{lt}\boldsymbol{\alpha} + \sum_{j=1}^{q}\mathbf{Z}_{jlt}\boldsymbol{\beta}_j(\mathbf{s})\right),$$

- Full conditional distribution of spatial decay parameter ϕ is not available and hence obtained from the kernel of the joint posterior distribution given in Equation (6).
- A missing observation $Y_l^*(\mathbf{s}_i, t)$ has full conditional distribution $N(O_l(\mathbf{s}_i, t), \sigma_\epsilon^2)$.

Appendix 2. Conditional distributions: temporally varying process model

- Using the joint posterior distribution kernel in Equation (15), we obtain the conditional distributions of the variance parameters $\sigma_\epsilon^2, \sigma_\eta^2, \sigma_\delta^2, \sigma_0^2$:

$$\pi\left(\frac{1}{\sigma_\epsilon^2}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y}\right) \sim \text{Gam}\left(\frac{N}{2}+a, b+\frac{1}{2}\sum_{l=1}^{r}\sum_{t=1}^{T}(\mathbf{Y}_{lt}-\mathbf{O}_{lt})'(\mathbf{Y}_{lt}-\mathbf{O}_{lt})\right),$$

$$\pi\left(\frac{1}{\sigma_\eta^2}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y}\right) \sim \text{Gam}\left(\frac{N}{2} + a, b + \frac{1}{2}\sum_{l=1}^{r}\sum_{t=1}^{T}(\mathbf{O}_{lt} - \boldsymbol{\nu}_{lt})' S^{-1}(\mathbf{O}_{lt} - \boldsymbol{\nu}_{lt})\right),$$

$$\pi\left(\frac{1}{\sigma_\delta^2}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y}\right) \sim \text{Gam}\left(\frac{uT}{2} + a, b + \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1})'(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1})\right),$$

$$\pi\left(\frac{1}{\sigma_0^2}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y}\right) \sim \text{Gam}\left(\frac{u}{2} + a, b + \frac{1}{2}\boldsymbol{\beta}_0'\boldsymbol{\beta}_0\right),$$

where $\boldsymbol{\theta}_{(-1)}$ represents all parameters in the models except the parameter for which we are obtaining the distribution. The term $\boldsymbol{\nu}_{lt} = \mathbf{X}_{lt}\boldsymbol{\alpha} + \mathbf{Z}_{lt}\boldsymbol{\beta}_t$.

- The conditional distribution of $\boldsymbol{\alpha}$ is defined as $\pi(\boldsymbol{\alpha}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y})$ and distributed as $N(\Delta\chi, \Delta)$, where

$$\Delta^{-1} = \sum_{l=1}^{r}\sum_{t=1}^{T}\mathbf{X}_{lt}'\Sigma_\eta^{-1}\mathbf{X}_lt + \frac{1}{\sigma_\alpha^2}\mathbf{I}_{p\times p},$$

$$\chi = \sum_{l=1}^{r}\sum_{t=1}^{T}\mathbf{X}_{lt}'\Sigma_\eta^{-1}(\mathbf{O}_{lt} - \mathbf{Z}_{lt}\boldsymbol{\beta}_t) + \frac{\mu_\alpha}{\sigma_\alpha^2}\mathbf{1}_p.$$

- We can obtain the conditional distribution of $\boldsymbol{\beta}_t$ at time $1 \le t < T$ as $N(\Delta_t\chi_t, \Delta_t)$, where

$$\Delta_t^{-1} = \sum_{l=1}^{r}\mathbf{Z}_{lt}'\Sigma_\eta^{-1}\mathbf{Z}_{lt} + \frac{1}{\sigma_\delta^2}(\mathbf{I}_u + \mathbf{G}'\mathbf{G}),$$

$$\chi_t = \sum_{l=1}^{r}\mathbf{Z}_{lt}'\Sigma_\eta^{-1}(\mathbf{O}_{lt} - \mathbf{X}_{lt}\boldsymbol{\alpha}) + \frac{1}{\sigma_\delta^2}\mathbf{G}(\boldsymbol{\beta}_{t-1} + \boldsymbol{\beta}_{t+1}).$$

At time $t = T$, the conditional distribution can be written as $N(\Delta_t\chi_t, \Delta_t)$, where

$$\Delta_t^{-1} = \sum_{l=1}^{r}\mathbf{Z}_{lt}'\Sigma_\eta^{-1}\mathbf{Z}_{lt} + \frac{1}{\sigma_\delta^2}\mathbf{1}_u$$

$$\chi_t = \sum_{l=1}^{r}\mathbf{Z}_{lt}'\Sigma_\eta^{-1}(\mathbf{O}_{lt} - \mathbf{X}_{lt}\boldsymbol{\alpha}) + \frac{1}{\sigma_\delta^2}\mathbf{G}\boldsymbol{\beta}_{t-1}.$$

At $t = 0$, we obtain the conditional distribution of $\boldsymbol{\gamma}_0$ as $N(\Delta_0\chi_0, \Delta_0)$, where

$$\Delta_0^{-1} = \frac{1}{\sigma_\delta^2}\mathbf{G}'\mathbf{G} + \frac{1}{\sigma_0^2}\mathbf{I}_{u\times u},$$

$$\chi_0 = \frac{1}{\sigma_\delta^2}\mathbf{G}\boldsymbol{\beta}_1.$$

- The $r \times r$ matrix $\mathbf{G}$ is diagonal with elements $\rho_k, k = 1, \ldots, u$. We obtain the conditional distribution as $N(\Delta_k\chi_k, \Delta_k)$, where

$$\Delta_k^{-1} = \frac{1}{\sigma_\delta^2}\sum_{t=1}^{T}\beta_{kt-1}^2 + \frac{1}{\sigma_\rho^2},$$

$$\chi_k = \frac{1}{\sigma_\delta^2}\sum_{t=1}^{T}\beta_{kt-1}\beta_{kt} + \frac{\mu_\rho}{\sigma_\rho^2}.$$

- From Equation (15), we obtain the conditional distribution $\pi(\mathbf{O}_{lt}|\boldsymbol{\theta}_{(-1)}, \mathbf{y}^*, \mathbf{y})$, that follows $N(\Delta_{lt}\chi_{lt}, \Delta_{lt})$, where

$$\Delta_{lt}^{-1} = \frac{1}{\sigma_\epsilon^2}\mathbf{I}_{n\times n} + \Sigma_\eta^{-1},$$

$$\chi_{lt} = \frac{1}{\sigma_\epsilon^2}\mathbf{Y}_{lt} + \Sigma_\eta^{-1}\left(\mathbf{X}_{lt}\boldsymbol{\alpha} + \mathbf{Z}_{lt}\boldsymbol{\beta}_t\right).$$

- Full conditional distribution of spatial decay parameter ϕ is not available and hence obtained from the kernel of the joint posterior distribution given in Equation (15).
- A missing observation $Y_l^*(\mathbf{s}_i, t)$ has full conditional distribution $N(O_l(\mathbf{s}_i, t), \sigma_\epsilon^2)$.