

Supplement to “Bayesian Nonparametric Dynamic State Space Modeling with Circular Latent States”

Satyaki Mazumder*and Sourabh Bhattacharya†

Throughout, we refer to our main paper Mazumder and Bhattacharya (2014) as MB.

S-1 Smoothness properties of our Gaussian process with linear-circular arguments

Here we assume that $\mu(t, \theta)$ is twice differentiable with respect to t and θ , and that the derivatives are bounded. Formally, we assume that $\frac{\partial^2 \mu(t, \theta)}{\partial t^2}$, $\frac{\partial^2 \mu(t, \theta)}{\partial \theta^2}$, $\frac{\partial^2 \mu(t, \theta)}{\partial t \partial \theta}$ ($= \frac{\partial^2 \mu(t, \theta)}{\partial \theta \partial t}$) exist and are bounded. We denote the covariance function $\sigma^2 \exp\{-\sigma^4 |t_1 - t_2|^2\} \cos(|\theta_1 - \theta_2|)$ (where $\sigma^2 = \frac{\psi^{-1}}{2}$) by $K(|t_1 - t_2|, |\theta_1 - \theta_2|)$.

*Indian Institute of Science Education and Research, Kolkata

†Indian Statistical Institute, Kolkata

S-1.1 Mean square continuity:

1. With respect to time t

$$\begin{aligned} & E[X(t+h, \theta) - X(t, \theta)]^2 \\ &= E[X(t+h, \theta)]^2 + E[X(t, \theta)]^2 - 2E[X(t+h, \theta)X(t, \theta)] \\ &= K(0, 0) + K(0, 0) - 2K(h, 0) \\ &= 2(K(0, 0) - K(h, 0)) \end{aligned}$$

Now as $h \rightarrow 0$, $E[X(t+h, \theta) - X(t, \theta)]^2 \rightarrow 0$ because of the fact that $K(h, 0)$ is continuous in h .

2. With respect to angle θ :

$$\begin{aligned} & E[X(t, \theta + \alpha) - X(t, \theta)]^2 \\ &= E[X(t, \theta + \alpha)]^2 + E[X(t, \theta)]^2 - 2E[X(t, \theta + \alpha)X(t, \theta)] \\ &= K(0, 0) + K(0, 0) - 2K(0, \alpha) \\ &= 2(K(0, 0) - K(0, \alpha)) \end{aligned}$$

Now as $\alpha \rightarrow 0$, $E[X(t, \theta + \alpha) - X(t, \theta)]^2 \rightarrow 0$ because of the fact that $K(0, \alpha)$ is continuous in α .

3. With respect to time t and angle θ :

$$E[X(t+h, \theta + \alpha) - X(t, \theta)]^2$$

$$\begin{aligned}
&= E[X(t+h, \theta+\alpha)]^2 + E[X(t, \theta)]^2 - 2E[X(t+h, \theta+\alpha)X(t, \theta)] \\
&= K(0,0) + K(0,0) - 2K(h, \alpha) \\
&= 2(K(0,0) - K(h, \alpha))
\end{aligned}$$

Now as $(h, \alpha) \rightarrow (0,0)$ then $E[X(t+h, \theta+\alpha) - X(t, \theta)]^2 \rightarrow 0$ because of the fact that $K(h, \alpha)$ is continuous in h and α .

S-1.2 Mean square differentiability

A process $X(\mathbf{u})$, $\mathbf{u} \in \mathbf{R}^d$, is said to be *Mean Square Differentiable* at \mathbf{u}_0 if for any direction \mathbf{p} there exists a process $L_{\mathbf{u}_0}(\mathbf{p})$, linear in \mathbf{p} , such that

$$X(\mathbf{u}_0 + \mathbf{p}) = X(\mathbf{u}_0) + L_{\mathbf{u}_0}(\mathbf{p}) + R(\mathbf{u}_0, \mathbf{p}),$$

where $\mathbf{p} \in \mathbf{R}^d$, and $R(\mathbf{u}_0, \mathbf{p})$ satisfies the following

$$\frac{R(\mathbf{u}_0, \mathbf{p})}{\|\mathbf{p}\|} \rightarrow 0, \text{ in } L^2,$$

with $\|\cdot\|$ being the usual Euclidean norm (for details see Banerjee and Gelfand (2003)).

However, we have $t \in \mathbb{R}^+$ and $\theta \in [0, 2\pi]$, so we can not directly apply the definition of mean square differentiability that is appropriate for \mathbb{R}^d . For our purpose we define a new metric on time and angular space as

$$d(t_1, t_2, \theta_1, \theta_2) = |t_1 - t_2| + |\theta_1 - \theta_2|,$$

(recall that we have used the angular distance as a metric on the angular space to represent the covariance as a function of distance in time and angle). Note that $d(\cdot, \cdot, \cdot, \cdot)$ satisfies all the three criteria for being a metric, that is,

1. $d(t_1, t_2, \theta_1, \theta_2) \geq 0$
2. $d(t_1, t_2, \theta_1, \theta_2) = 0$ iff $t_1 = t_2, \theta_1 = \theta_2$
3. $d(t_1, t_3, \theta_1, \theta_3) \leq [|t_1 - t_2| + |\theta_1 - \theta_2|] + [|t_2 - t_3| + |\theta_1 - \theta_2|]$
 $= d(t_1, t_2, \theta_1, \theta_2) + d(t_2, t_3, \theta_2, \theta_3)$

With the help of this new metric in time and angular space we define *Mean Square Differentiability* in time and circular domain as

Definition 1 A process $X(t, \theta)$ is said to be **Mean Square Differentiable** in L^2 sense at (t_0, θ_0) if for any direction (h, α) there exists a process $L_{t_0, \theta_0}(h, \alpha)$, linear in h, α , such that

$$X(t_0 + h, \theta_0 + \alpha) = X(t_0, \theta_0) + L_{t_0, \theta_0}(h, \alpha) + R(t_0, \theta_0, h, \alpha),$$

where $R(t_0, \theta_0, h, \alpha)$ satisfies the following condition

$$\frac{R(t_0, \theta_0, h, \alpha)}{d(h, 0, \alpha, 0)} \rightarrow 0, \text{ in } L^2 \text{ as } d(h, 0, \alpha, 0) \rightarrow 0.$$

In our case, since our covariance function $K(|t_1 - t_2|, |\theta_1 - \theta_2|)$ has partial derivatives of all orders, the partial derivative processes of all orders exist with covariance structures given by partial derivatives of our covariance function; see Section 2.2 of Adler (1981) for details. In fact, the partial derivative processes are all Gaussian processes, and hence, they

are bounded in L^2 .

Hence, we can apply Taylor series expansion to obtain a linear function $L_{\mathbf{u}_0}(\mathbf{p})$. The following calculation will make the things clear. Following the multivariate Taylor series expansion (using our new metric) we have

$$X(t_0 + h, \theta_0 + \alpha) = X(t_0, \theta_0) + h \left. \frac{\partial}{\partial t} X(t, \theta) \right|_{t=t_0, \theta=\theta_0} + \alpha \left. \frac{\partial}{\partial \theta} X(t, \theta) \right|_{t=t_0, \theta=\theta_0} + R(t_0, \theta_0, h, \alpha),$$

where $|R(t_0, \theta_0, h, \alpha)| \leq M^* d^2(h, 0, \alpha, 0)$, with $M^* = \max \left\{ \left| \frac{\partial^2 X(t, \theta)}{\partial t^2} \right|, \left| \frac{\partial^2 X(t, \theta)}{\partial t \partial \theta} \right|, \left| \frac{\partial^2 X(t, \theta)}{\partial \theta \partial t} \right|, \left| \frac{\partial^2 X(t, \theta)}{\partial \theta^2} \right| \right\}$ (using the analogy with multivariate Taylor series expansion in \mathbf{R}^d , recall that in the case of \mathbf{R}^d , $R(\mathbf{u}_0, \mathbf{p}) \leq M^* \|\mathbf{p}\|^2$).

Since each of the partial derivative processes is bounded in L^2 , it is obvious that M^* is also bounded in L^2 . Mean square differentiability of our kernel convolved Gaussian process thus follows.

S-2 MCMC-based inference

In our MCMC-based inference we include the problem of forecasting y_{T+1} , given the observed data set \mathbf{D}_T . The posterior predictive distribution of y_{T+1} given \mathbf{D}_T is given by

$$\begin{aligned} [y_{T+1} | \mathbf{D}_T] &= \int [y_{T+1} | \mathbf{D}_T, x_0, \dots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_f^2, \sigma_g^2] \\ &\quad \times [x_0, \dots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_g^2, \sigma_f^2 | \mathbf{D}_T] \\ &\quad d\boldsymbol{\beta}_f d\boldsymbol{\beta}_g d\sigma_\epsilon^2 d\sigma_\eta^2 d\sigma_g^2 d\sigma_f^2 dx_0 \dots dx_{T+1}. \end{aligned} \tag{1}$$

Thus, once we have a sample realization from the joint posterior

$[x_0, \dots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_g^2, \sigma_f^2 | \mathbf{D}_T]$, we can generate a realization from $[y_{T+1} | \mathbf{D}_T]$ by simply simulating from $[y_{T+1} | \mathbf{D}_T, x_0, \dots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_g^2, \sigma_f^2]$, conditional on the realization obtained from the former joint posterior. Observe that the conditional distribution $[y_{T+1} = f(T+1, x_{T+1}) + \epsilon_{T+1} | \mathbf{D}_T, x_0, \dots, x_{T+1}, \boldsymbol{\beta}_f, \sigma_\epsilon^2, \sigma_f^2]$ is normal with mean

$$\mu_{y_{T+1}} = \mathbf{h}(T+1, x_{T+1})' \boldsymbol{\beta}_f + \mathbf{s}_{f, D_T}(T+1, x_{T+1})' \mathbf{A}_{f, D_T}^{-1} (\mathbf{D}_T - \mathbf{H}_{D_T} \boldsymbol{\beta}_f) \quad (2)$$

and variance

$$\sigma_{y_{T+1}}^2 = \sigma_\epsilon^2 + \sigma_f^2 (1 - (\mathbf{s}_{f, D_T}(T+1, x_{T+1}))' \mathbf{A}_{f, D_T}^{-1} \mathbf{s}_{f, D_T}(T+1, x_{T+1})). \quad (3)$$

Using the auxiliary variables K_1, \dots, K_{T+1} , the posterior distribution of the latent circular variables and the other parameters can be represented as

$$\begin{aligned} & [x_0, x_1, \dots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_g^2, \sigma_f^2 | \mathbf{D}_T] \\ &= \sum_{K_1, \dots, K_{T+1}} \int [x_0, x_1, \dots, x_T, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_g^2, \sigma_f^2, g^*(1, x_0), \mathbf{D}_z, K_1, \dots, K_T, K_{T+1} | \mathbf{D}_T] \\ & \quad \times dg^*(1, x_0) d\mathbf{D}_z \\ &\propto \sum_{K_1, \dots, K_{T+1}} \int [x_0, x_1, \dots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_g^2, \sigma_f^2, g^*(1, x_0), \mathbf{D}_z, K_1, \dots, K_T, K_{T+1}, \mathbf{D}_T] \\ & \quad \times dg^*(1, x_0) d\mathbf{D}_z \\ &= \sum_{K_1, \dots, K_{T+1}} \int [\boldsymbol{\beta}_f][\boldsymbol{\beta}_g][\sigma_\epsilon^2][\sigma_\eta^2][\sigma_g^2][\sigma_f^2][x_0][g^*(1, x_0) | x_0, \boldsymbol{\beta}_g, \sigma_g^2][\mathbf{D}_z | g^*(1, x_0), x_0, \boldsymbol{\beta}_g, \sigma_g^2] \\ & \quad [x_1 | g^*(1, x_0), \sigma_\eta^2, K_1][K_1 | g^*(1, x_0), \sigma_\eta^2][\mathbf{D}_T | x_1, \dots, x_T, \boldsymbol{\beta}_f, \sigma_\epsilon^2, \sigma_f^2] \end{aligned}$$

$$\prod_{t=2}^{T+1} [x_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \sigma_g^2, \mathbf{D}_z, x_{t-1}, K_t] \prod_{t=2}^{T+1} [K_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \sigma_g^2, \mathbf{D}_z, x_{t-1}] dg^*(1, x_0) d\mathbf{D}_z. \quad (4)$$

In order to obtain MCMC samples from $[x_0, x_1, \dots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_g^2, \sigma_f^2 | \mathbf{D}_T]$, we first carry out MCMC simulations from the joint posterior which is proportional to integrand (4). Ignoring $g^*(1, x_0)$, \mathbf{D}_z and K_1, \dots, K_{T+1} in these MCMC simulations and storing the realizations associated with the remaining parameters yield the desired samples.

S-2.1 Full conditional distributions

Here we provide the full conditional distributions of the unknowns. In what follows, we shall express $[g^*(1, x_0) | x_0, \boldsymbol{\beta}_g, \sigma_g^2] [\mathbf{D}_z | g^*(1, x_0), x_0, \boldsymbol{\beta}_g, \sigma_g^2]$ as $[\mathbf{D}_z, g^*(1, x_0) | x_0, \boldsymbol{\beta}_g, \sigma_g^2]$.

$$[\boldsymbol{\beta}_f | \dots] \propto [\boldsymbol{\beta}_f] [\mathbf{D}_T | x_1, \dots, x_T, \boldsymbol{\beta}_f, \sigma_\epsilon^2] \quad (5)$$

$$[\boldsymbol{\beta}_g | \dots] \propto [\boldsymbol{\beta}_g] [\mathbf{D}_z, g^*(1, x_0) | x_0, \boldsymbol{\beta}_g, \sigma_g^2] \prod_{t=2}^{T+1} [x_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \sigma_g^2, \mathbf{D}_z, x_{t-1}, K_t] \prod_{t=2}^{T+1} [K_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \sigma_g^2, \mathbf{D}_z, x_{t-1}] \quad (6)$$

$$[\sigma_\epsilon^2 | \dots] \propto [\sigma_\epsilon^2] [\mathbf{D}_T | x_1, \dots, x_T, \boldsymbol{\beta}_f, \sigma_\epsilon^2] \quad (7)$$

$$[\sigma_f^2 | \dots] \propto [\sigma_f^2] [\mathbf{D}_T | x_1, \dots, x_T, \boldsymbol{\beta}_f, \sigma_f^2] \quad (8)$$

$$[\sigma_\eta^2 | \dots] \propto [\sigma_\eta^2] [x_1 | g^*(1, x_0), \sigma_\eta^2, K_1] [K_1 | g^*(1, x_0), \sigma_\eta^2] \prod_{t=2}^{T+1} [x_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \sigma_g^2, \mathbf{D}_z, x_{t-1}, K_t] \prod_{t=2}^{T+1} [K_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \sigma_g^2, \mathbf{D}_z, x_{t-1}] \quad (9)$$

$$[\sigma_g^2 | \dots] \propto [\sigma_g^2] [\mathbf{D}_z, g^*(1, x_0) | x_0, \boldsymbol{\beta}_g, \sigma_g^2] \prod_{t=2}^{T+1} [x_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \sigma_g^2, \mathbf{D}_z, x_{t-1}, K_t] \prod_{t=2}^{T+1} [K_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \sigma_g^2, \mathbf{D}_z, x_{t-1}]$$

$$\sigma_\eta^2, \mathbf{D}_z, x_{t-1}] \quad (10)$$

$$[x_0 | \cdots] \propto [x_0] [\mathbf{D}_z, g^*(1, x_0) | x_0, \boldsymbol{\beta}_g, \sigma_g^2] \quad (11)$$

$$[g^*(1, x_0) | \cdots] \propto [g^*(1, x_0) | x_0, \boldsymbol{\beta}_g, \sigma_g^2] [\mathbf{D}_z | g^*(1, x_0), x_0, \boldsymbol{\beta}_g, \sigma_g^2] [x_1 | g^*(1, x_0), x_0, \sigma_\eta^2, K_1] \\ [K_1 | g^*(1, x_0), \sigma_\eta^2] \quad (12)$$

$$[\mathbf{D}_z | \cdots] \propto [\mathbf{D}_z | g^*(1, x_0), x_0, \boldsymbol{\beta}_g, \sigma_g^2] \prod_{t=2}^{T+1} [x_t | \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \mathbf{D}_z, x_{t-1}, K_t] \prod_{t=2}^{T+1} [K_t | \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \\ \mathbf{D}_z, x_{t-1}] \quad (13)$$

$$[x_1 | \cdots] \propto [x_1 | g^*(1, x_0), \sigma_\eta^2] [\mathbf{D}_T | x_1, \dots, x_T, \boldsymbol{\beta}_f, \sigma_\epsilon^2]$$

$$[x_2 | \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \mathbf{D}_z, x_1, K_2] [K_2 | \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \mathbf{D}_z, x_1] \quad (14)$$

$$[x_{T+1} | \cdots] \propto [x_{T+1} | \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \mathbf{D}_z, x_T, K_{T+1}] \quad (15)$$

$$[x_{t+1} | \cdots] \propto [x_{t+1} | \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \mathbf{D}_z, x_t] [x_{t+2} | \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \mathbf{D}_z, x_{t+1}, K_{t+2}] [K_{t+2} | \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \\ \mathbf{D}_z, x_{t+1}] [\mathbf{D}_T | x_1, \dots, x_T, \boldsymbol{\beta}_f, \sigma_\epsilon^2], \quad t = 1, \dots, T-1 \quad (16)$$

Finally, we write down the full conditional distribution of K_t , for $t = 1, \dots, T+1$, as

$$[K_1 | \cdots] \propto [K_1 | g^*(1, x_0), \sigma_\eta^2] [x_1 | g^*(1, x_0), \boldsymbol{\beta}_g, \sigma_g^2, K_1] \quad (17)$$

$$[K_t | \cdots] \propto [x_t | \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \mathbf{D}_z, x_{t-1}, K_t] [K_t | \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \mathbf{D}_z, x_{t-1}], \quad t = 2, \dots, T+1. \quad (18)$$

S-2.1.1 Updating $\boldsymbol{\beta}_f$ by Gibbs steps

The full conditional of $\boldsymbol{\beta}_f$ is a multivariate normal distribution with mean

$$E[\boldsymbol{\beta}_f | \cdots] = \{\mathbf{H}'_{D_T} (\sigma_f^2 \mathbf{A}_{f,D_T} + \sigma_\epsilon^2 I)^{-1} \mathbf{H}_{D_T} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{f,0}}\}^{-1} \\ \times \{\mathbf{H}'_{D_T} (\sigma_f^2 \mathbf{A}_{f,D_T} + \sigma_\epsilon^2 I)^{-1} \mathbf{D}_T + \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{f,0}}^{-1} \boldsymbol{\beta}_{f,0}\} \quad (19)$$

and variance

$$V[\boldsymbol{\beta}_f | \dots] = \{\mathbf{H}'_{D_T}(\sigma_f^2 \mathbf{A}_{f,D_T} + \sigma_\epsilon^2 I)^{-1} \mathbf{H}_{D_T} + \boldsymbol{\Sigma}_{\beta_{f,0}}\}^{-1}. \quad (20)$$

S-2.1.2 Updating $\boldsymbol{\beta}_g$

We first explicitly write down the right hand side of (6).

$$\begin{aligned} & [\boldsymbol{\beta}_g][\mathbf{D}_z, g^*(1, x_0) | x_0, \boldsymbol{\beta}_g] \prod_{t=2}^{T+1} [x_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \mathbf{D}_z, x_{t-1}, K_t] \prod_{t=2}^{T+1} [K_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \mathbf{D}_z, x_{t-1}] \\ & \propto \exp \left(-\frac{1}{2} (\boldsymbol{\beta}_g - \boldsymbol{\beta}_{g,0})' \boldsymbol{\Sigma}_{\beta_{g,0}}^{-1} (\boldsymbol{\beta}_g - \boldsymbol{\beta}_{g,0}) \right) \\ & \exp \left(-\frac{1}{2} [(\mathbf{D}_z, g^*)' - (\mathbf{H}_{D_z} \boldsymbol{\beta}_g, \mathbf{h}'(1, x_0))]' \mathbf{A}_{D_z, g^*(1, x_0)}^{-1} [(\mathbf{D}_z, g^*)' - (\mathbf{H}_{D_z} \boldsymbol{\beta}_g, \mathbf{h}'(1, x_0))]' \right) \\ & \exp \left\{ -\sum_{i=2}^{T+1} \frac{1}{2\sigma_{x_t}^2} (x_t + 2\pi K_t - \mu_{x_t})^2 \right\} \prod_{t=2}^{T+1} I_{[0, 2\pi]}(x_t) \end{aligned} \quad (21)$$

Observe that the denominator of $[x_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \mathbf{D}_z, x_{t-1}, K_t]$ cancels with the density of $[K_t | \boldsymbol{\beta}_g, \sigma_\eta^2, \mathbf{D}_z, x_{t-1}]$ for each $t = 2, \dots, T+1$. Also we note that the indicator function does not involve $\boldsymbol{\beta}_g$ for all $t = 2, \dots, T+1$. Therefore, after simplifying the exponent terms and ignoring the indicator function we can write

$$[\boldsymbol{\beta}_g | \dots] \propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}_g - \mu_{\beta_g})' \boldsymbol{\Sigma}_{\beta_g}^{-1} (\boldsymbol{\beta}_g - \mu_{\beta_g}) \right\}, \quad (22)$$

where

$$\begin{aligned}
\mu_{\beta_g} = E[\beta_g | \dots] &= \left\{ \Sigma_{\beta_g,0}^{-1} + \frac{1}{\sigma_g^2} [\mathbf{H}'_{D_z}, \mathbf{h}(1, x_0)] \mathbf{A}_{D_z, g^*(1, x_0)}^{-1} [\mathbf{H}'_{D_z}, \mathbf{h}(1, x_0)]' \right. \\
&+ \sum_{t=1}^T \frac{(\mathbf{H}'_{D_z} \mathbf{A}_{g, D_z}^{-1} \mathbf{s}_{g, D_z}(t+1, x_t) - \mathbf{h}(t+1, x_t)) (\mathbf{H}'_{D_z} \mathbf{A}_{g, D_z}^{-1} \mathbf{s}_{g, D_z}(t+1, x_t) - \mathbf{h}(t+1, x_t))'}{\sigma_{x_t}^2} \left. \right\}^{-1} \\
&\left\{ \Sigma_{\beta_g,0}^{-1} \beta_{g,0} + \frac{1}{\sigma_g^2} [\mathbf{H}'_{D_z}, \mathbf{h}(1, x_0)] \mathbf{A}_{D_z, g^*(1, x_0)}^{-1} [\mathbf{D}_z, g^*(1, x_0)] \right. \\
&+ \sum_{t=1}^T \frac{(x_{t+1} + 2\pi K_{t+1} - \mathbf{s}_{g, D_z}(t+1, x_t)' \mathbf{A}_{g, D_z}^{-1} \mathbf{D}_z) (\mathbf{h}(t+1, x_t) - \mathbf{H}'_{D_z} \mathbf{A}_{g, D_z}^{-1} \mathbf{s}_{g, D_z}(t+1, x_t))}{\sigma_{x_t}^2} \left. \right\} \\
&\tag{23}
\end{aligned}$$

and

$$\begin{aligned}
\Sigma_{\beta_g} = V[\beta_g | \dots] &= \left\{ \Sigma_{\beta_g,0}^{-1} + \frac{1}{\sigma_g^2} [\mathbf{H}'_{D_z}, \mathbf{h}(1, x_0)] \mathbf{A}_{D_z, g^*(1, x_0)}^{-1} [\mathbf{H}'_{D_z}, \mathbf{h}(1, x_0)]' \right. \\
&+ \sum_{t=1}^T \frac{(\mathbf{H}'_{D_z} \mathbf{A}_{g, D_z}^{-1} \mathbf{s}_{g, D_z}(t+1, x_t) - \mathbf{h}(t+1, x_t)) (\mathbf{H}'_{D_z} \mathbf{A}_{g, D_z}^{-1} \mathbf{s}_{g, D_z}(t+1, x_t) - \mathbf{h}(t+1, x_t))'}{\sigma_{x_t}^2} \left. \right\}. \\
&\tag{24}
\end{aligned}$$

Hence $[\beta_g | \dots]$ follows a four-variate normal distribution with mean and variance μ_{β_g} and Σ_{β_g} , respectively, and therefore, we update β_g using Gibbs sampling.

S-2.1.3 Updating σ_f^2 and σ_g^2

The mathematical form of the full conditional distributions of σ_f^2 and σ_g^2 are not tractable, so we update σ_f^2 and σ_g^2 by random walk Metropolis-Hastings steps.

S-2.1.4 Updating σ_ϵ^2

The mathematical form of the full conditional distribution of σ_ϵ^2 is not tractable, so we update σ_ϵ^2 by a random walk Metropolis-Hastings step.

S-2.1.5 Updating σ_η^2

For full conditional distribution of σ_η^2 right hand side of (9) simplifies a bit in the sense that the denominator of $[x_t|\boldsymbol{\beta}_g, \sigma_\eta^2, \mathbf{D}_z, x_{t-1}, K_t]$ cancels with the density of $[K_t|\boldsymbol{\beta}_g, \sigma_\eta^2, \mathbf{D}_z, x_{t-1}]$ for $t = 2, \dots, T+1$, and the denominator of $[x_1|g^*(1, x_0), \boldsymbol{\beta}_g, \sigma_\eta^2, K_1]$ cancels with the density of $[K_1|g^*(1, x_0), \boldsymbol{\beta}_g, \sigma_\eta^2]$, which, in turn, gives the following form:

$$[\sigma_\eta^2|\dots] \propto [\sigma_\eta^2] \exp \left\{ - \sum_{i=2}^{T+1} \frac{1}{2\sigma_{x_t}^2} (x_t + 2\pi K_t - \mu_{x_t})^2 \right\} \exp \left\{ - \frac{1}{2\sigma_\eta^2} (x_1 + 2\pi K_1 - g^*)^2 \right\}. \quad (25)$$

However, the above equation does not have a closed form; hence, for updating σ_η^2 as well, we use random walk Metropolis-Hastings.

S-2.1.6 Updating x_0

The full conditional distribution of x_0 is not tractable and hence again here we use random walk Metropolis-Hastings for updating x_0 . Now note that x_0 is a circular random variable, so to update $x_0^{(old)}$ to $x_0^{(new)}$ we use the vonMises distribution with location parameter $x_0^{(old)}$.

S-2.1.7 Updating $g^*(1, x_0)$

Equation (12), after cancelling the denominator of $[x_1|g^*(1, x_0), x_0, \boldsymbol{\beta}_g, \sigma_\eta^2, K_1]$ with the density of $[K_1|g^*(1, x_0), x_0, \boldsymbol{\beta}_g, \sigma_\eta^2]$, and ignoring the indicator function on x_0 , reduces to

$$[g^*(1, x_0)|\dots] \propto [g^*(1, x_0)|x_0, \boldsymbol{\beta}_g][\boldsymbol{D}_z|g^*(1, x_0), x_0, \boldsymbol{\beta}_g] \exp \left\{ -\frac{1}{2\sigma_\eta^2}(x_1 + 2\pi K_1 - g^*)^2 \right\}.$$

After further simplification the full conditional distribution of $g^*(1, x_0)$ reduces to

$$[g^*(1, x_0)|\dots] \propto \exp \left\{ -\frac{1}{2\gamma_g^2}(g^* - \nu_g)^2 \right\}, \quad (26)$$

where

$$\begin{aligned} \nu_g = E[g^*(1, x_0)|\dots] &= \left\{ \frac{1}{\sigma_\eta^2} + \frac{1}{\sigma_g^2}(1 + \boldsymbol{s}_{g,D_z}(1, x_0)' \boldsymbol{\Sigma}_{g,D_z}^{-1} \boldsymbol{s}_{g,D_z}(1, x_0)) \right\}^{-1} \\ &\quad \left\{ \frac{x_1 + 2\pi K_1}{\sigma_\eta^2} + \frac{1}{\sigma_g^2}(\boldsymbol{h}(1, x_0)' \boldsymbol{\beta}_g + \boldsymbol{s}_{g,D_z}' \boldsymbol{\Sigma}_{g,D_z}^{-1} \boldsymbol{D}_z^*) \right\} \end{aligned} \quad (27)$$

and

$$\gamma_g^2 = V[g^*(1, x_0)|\dots] = \left\{ \frac{1}{\sigma_\eta^2} + \frac{1}{\sigma_g^2}(1 + \boldsymbol{s}_{g,D_z}(1, x_0)' \boldsymbol{\Sigma}_{g,D_z}^{-1} \boldsymbol{s}_{g,D_z}(1, x_0)) \right\}, \quad (28)$$

with

$$\boldsymbol{D}_z^* = \boldsymbol{D}_z - \boldsymbol{H}_{D_z} \boldsymbol{\beta}_g + \boldsymbol{h}(1, x_0)' \boldsymbol{\beta}_g \boldsymbol{s}_{g,D_z}, \quad (29)$$

and

$$\boldsymbol{\Sigma}_{g,D_z} = \boldsymbol{A}_{g,D_z} - \boldsymbol{s}_{g,D_z}(1, x_0) \boldsymbol{s}_{g,D_z}(1, x_0)'. \quad (30)$$

Hence $[g^*|\cdots]$ follows a normal distribution with mean ν_g and variance γ_g . Therefore, we update g^* using Gibbs sampling.

S-2.1.8 Updating D_z

Here also we observe that in the full conditional distribution of D_z , the denominator of $[x_t|\beta_g, \sigma_\eta^2, D_z, x_{t-1}, K_t]$ cancels with the density of $[K_t|\beta_g, \sigma_\eta^2, D_z, x_{t-1}]$ for each $t = 2, \dots, T+1$. After simplification it turns out that the full conditional distribution of D_z is an n -variate normal with mean

$$E(D_z|\cdots) = \left\{ \frac{\Sigma_{g,D_z}^{-1}}{\sigma_g^2} + \mathbf{A}_{g,D_z}^{-1} \left(\sum_{t=1}^T \frac{s_{g,D_z}(t+1, x_t) s'_{g,D_z}(t+1, x_t)}{\sigma_{x_t}^2} \right) \mathbf{A}_{g,D_z}^{-1} \right\}^{-1} \\ \times \left\{ \frac{\Sigma_{g,D_z}^{-1} \boldsymbol{\mu}_{g,D_z}}{\sigma_g^2} + \mathbf{A}_{g,D_z}^{-1} \sum_{t=1}^T \frac{s_{g,D_z}(t+1, x_t) \{x_{t+1} + 2\pi K_{t+1} - \beta'_g(\mathbf{h}(1, t+1, x_t) - \mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} s_{g,D_z}(t+1, x_t))\}}{\sigma_{x_t}^2} \right\} \quad (31)$$

and covariance matrix

$$V(D_z|\cdots) = \left\{ \frac{\Sigma_{g,D_z}^{-1}}{\sigma_g^2} + \mathbf{A}_{g,D_z}^{-1} \left(\sum_{t=1}^T \frac{s_{g,D_z}(t+1, x_t) s'_{g,D_z}(t+1, x_t)}{\sigma_{x_t}^2} \right) \mathbf{A}_{g,D_z}^{-1} \right\}^{-1}. \quad (32)$$

Therefore, we update D_z using Gibbs sampling.

S-2.1.9 Updating x_1

For the full conditional distribution of x_1 we write down the complete expression of (14) as follows:

$$\begin{aligned}
[x_1 | \dots] &\propto \frac{\frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left(-\frac{1}{2\sigma_\eta^2}(x_1 + 2\pi K_1 - g^*)^2\right) I_{[0,2\pi]}(x_1)}{\Phi\left(\frac{2\pi(K_1+1)-g^*}{\sigma_\eta}\right) - \Phi\left(\frac{2\pi K_1-g^*}{\sigma_\eta}\right)} \\
&\quad \exp\left\{-\frac{1}{2}(\mathbf{D}_T - \boldsymbol{\mu}_{y_t})' \boldsymbol{\Sigma}_{y_t}^{-1}(\mathbf{D}_T - \boldsymbol{\mu}_{y_t})\right\} \\
&\quad \frac{1}{\sqrt{2\pi}\sigma_{x_2}} \exp\left(-\frac{1}{2\sigma_{x_2}^2}(x_2 + 2\pi K_2 - \mu_{x_2})^2\right), \tag{33}
\end{aligned}$$

where $\boldsymbol{\mu}_{y_t}$ and $\boldsymbol{\Sigma}_{y_t}$ are given by (10) and (11) of MB. Here we note that the denominator of $[x_2 | \boldsymbol{\beta}_g, \sigma_\eta^2, \mathbf{D}_z, x_1, K_2]$ cancels with $[K_2 | \boldsymbol{\beta}_g, \sigma_\eta^2, \mathbf{D}_z, x_1]$. Also we ignore the indicator term associated with x_2 . We note that the term $\Phi\left(\frac{2\pi(K_1+1)-g^*}{\sigma_\eta}\right) - \Phi\left(\frac{2\pi K_1-g^*}{\sigma_\eta}\right)$ does not involve x_1 . Hence ignoring $\Phi\left(\frac{2\pi(K_1+1)-g^*}{\sigma_\eta}\right) - \Phi\left(\frac{2\pi K_1-g^*}{\sigma_\eta}\right)$ we get

$$\begin{aligned}
[x_1 | \dots] &\propto \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left(-\frac{1}{2\sigma_\eta^2}(x_1 + 2\pi K_1 - g^*)^2\right) I_{[0,2\pi]}(x_1) \\
&\quad \exp\left\{-\frac{1}{2}(\mathbf{D}_T - \boldsymbol{\mu}_{y_t})' \boldsymbol{\Sigma}_{y_t}^{-1}(\mathbf{D}_T - \boldsymbol{\mu}_{y_t})\right\} \\
&\quad \frac{1}{\sqrt{2\pi}\sigma_{x_2}} \exp\left(-\frac{1}{2\sigma_{x_2}^2}(x_2 + 2\pi K_2 - \mu_{x_2})^2\right), \tag{34}
\end{aligned}$$

However, it is not possible to get a closed form expression of $[x_1 | \dots]$, so we update it by random walk Metropolis-Hastings.

S-2.1.10 Updating x_{t+1} , $t = 1, \dots, T-1$

For x_{t+1} we have the same structure as for x_1 , except for some changes in the parameters.

To be precise, the full conditional distribution can be explicitly written as

$$\begin{aligned}
[x_{t+1} | \dots] &\propto \frac{\frac{1}{\sqrt{2\pi}\sigma_{x_{t+1}}} \exp\left(-\frac{1}{2\sigma_{x_{t+1}}^2}(x_{t+1} + 2\pi K_{t+1} - \mu_{x_{t+1}})^2\right) I_{[0,2\pi]}(x_{t+1})}{\Phi\left(\frac{2\pi(K_{t+1}+1)-\mu_{x_{t+1}}}{\sigma_{x_{t+1}}}\right) - \Phi\left(\frac{2\pi K_{t+1}-\mu_{x_{t+1}}}{\sigma_{x_{t+1}}}\right)} \\
&\quad \frac{1}{\sqrt{2\pi}\sigma_{x_{t+2}}} \exp\left(-\frac{1}{2\sigma_{x_{t+2}}^2}(x_{t+2} + 2\pi K_{t+2} - \mu_{x_{t+2}})^2\right) \\
&\quad \exp\left\{-\frac{1}{2}(\mathbf{D}_T - \boldsymbol{\mu}_{y_t})' \boldsymbol{\Sigma}_{y_t}^{-1}(\mathbf{D}_T - \boldsymbol{\mu}_{y_t})\right\}. \tag{35}
\end{aligned}$$

We note here that $\Phi\left(\frac{2\pi(K_{t+1}+1)-\mu_{x_{t+1}}}{\sigma_{x_{t+1}}}\right) - \Phi\left(\frac{2\pi K_{t+1}-\mu_{x_{t+1}}}{\sigma_{x_{t+1}}}\right)$ does not involve x_{t+1} because $\mu_{x_{t+1}}$ and $\sigma_{x_{t+1}}$ depend on x_t , not on x_{t+1} , and hence we can ignore the term $\Phi\left(\frac{2\pi(K_{t+1}+1)-\mu_{x_{t+1}}}{\sigma_{x_{t+1}}}\right) - \Phi\left(\frac{2\pi K_{t+1}-\mu_{x_{t+1}}}{\sigma_{x_{t+1}}}\right)$ and rewrite (35) as

$$\begin{aligned}
[x_{t+1} | \dots] &\propto \frac{1}{\sqrt{2\pi}\sigma_{x_{t+1}}} \exp\left(-\frac{1}{2\sigma_{x_{t+1}}^2}(x_{t+1} + 2\pi K_{t+1} - \mu_{x_{t+1}})^2\right) I_{[0,2\pi]}(x_{t+1}) \\
&\quad \frac{1}{\sqrt{2\pi}\sigma_{x_{t+2}}} \exp\left(-\frac{1}{2\sigma_{x_{t+2}}^2}(x_{t+2} + 2\pi K_{t+2} - \mu_{x_{t+2}})^2\right) \\
&\quad \exp\left\{-\frac{1}{2}(\mathbf{D}_T - \boldsymbol{\mu}_{y_t})' \boldsymbol{\Sigma}_{y_t}^{-1}(\mathbf{D}_T - \boldsymbol{\mu}_{y_t})\right\}. \tag{36}
\end{aligned}$$

Here also the expression of the full conditional distribution of x_{t+1} is not tractable. So, we adopt random walk Metropolis-Hastings to update x_{t+1} , for $t = 1, \dots, T$.

S-2.1.11 Updating x_{T+1}

The full conditional distribution of x_{T+1} has probability density function of the form (29) of MB with parameters

$$\mu_{x_{T+1}} = \mathbf{h}(1, x_T)' \boldsymbol{\beta}_g + \mathbf{s}_{g,D_z}(T+1, x_T)' \mathbf{A}_{g,D_z}^{-1} (\mathbf{D}_z - \mathbf{H}_{D_z} \boldsymbol{\beta}_g) \quad (37)$$

and

$$\sigma_{x_{T+1}}^2 = \sigma_\eta^2 + \sigma_g^2 \{1 - \mathbf{s}_{g,D_z}(T+1, x_T)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(T+1, x_T)\}. \quad (38)$$

We note here that given all unknowns except x_{T+1} , $x_{T+1} + 2\pi K_{T+1}$ follows a truncated normal distribution with left side truncation at $2\pi K_{T+1}$ and right side truncation at $2\pi(K_{T+1} + 1)$ (K_{T+1} is constant in this case). Hence we update $x_{T+1} + 2\pi K_{T+1}$ using Gibbs sampling and then subtract $2\pi K_{T+1}$ from it to update x_{T+1} .

S-2.1.12 Updating K_t , $t = 1, \dots, T+1$

The full conditional distribution of K_1 reduces to the following form

$$[K_1 | \dots] \propto \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left(-\frac{1}{2\sigma_\eta^2}(x_1 + 2\pi K_1 - g^*)^2\right) I_{\{\dots, -1, 0, 1, \dots\}}(K_1), \quad (39)$$

and similarly the full conditional distribution of K_t becomes

$$[K_t | \dots] \propto \frac{1}{\sqrt{2\pi}\sigma_{x_t}} \exp\left(-\frac{1}{2\sigma_{x_t}^2}(x_t + 2\pi K_t - \mu_{x_t})^2\right) I_{\{\dots, -1, 0, 1, \dots\}}(K_t), \quad (40)$$

for $t = 2, \dots, T+1$. We update K_t , for $t = 1, \dots, K+1$, by random walk Metropolis-Hastings.

References

- Adler, R. J. (1981). *The Geometry of Random Fields*. Wiley, London.
- Banerjee, S. and Gelfand, A. E. (2003). On Smoothness Properties of Spatial Processes. *Journal of Multivariate Analysis*, **84**, 85–100.
- Mazumder, S. and Bhattacharya, S. (2014). Bayesian Nonparametric Dynamic State-Space Modeling with Circular Latent States. Submitted.