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 $h \to 0$, $E[X(t+h,\theta)-X(t,\theta)]^2 \to 0$ 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 $\alpha \to 0$, $E[X(t,\theta+\alpha)-X(t,\theta)]^2 \to 0$ 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 $(h,\, \alpha) \to$ (0,0) then $E[X(t+h,\theta+\alpha) - X(t,\theta)]^2 \to 0$ 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

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

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

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

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

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

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 $\boldsymbol{D}_T$. The posterior predictive distribution of y_{T+1} given $\boldsymbol{D}_T$ is given by

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

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

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

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

and variance

$$\sigma^2_{y_{T+1}} = \sigma^2_\epsilon + \sigma^2_f \left(1 - (\boldsymbol{s}_{f,D_T}(T+1, x_{T+1}))'\boldsymbol{A}^{-1}_{f,D_T}\boldsymbol{s}_{f,D_T}(T+1, x_{T+1})\right). \tag{3}$$

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

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

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

In order to obtain MCMC samples from $[x_0,x_1,\ldots,x_{T+1},\boldsymbol{\beta}_f,\boldsymbol{\beta}_g,\sigma^2_\epsilon,\sigma^2_\eta,\sigma^2_g,\sigma^2_f|\boldsymbol{D}_T]$, we first carry out MCMC simulations from the joint posterior which is proportional to integrand (4). Ignoring $g^*(1,x_0)$, D_z and $K_1,\ldots,K_{T+1}$ 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 $[g^*(1,x_0)|x_0,\boldsymbol{\beta}_g,\sigma^2_g][\boldsymbol{D}_z|g^*(1,x_0),x_0,\boldsymbol{\beta}_g,\sigma^2_g]$ as $[\boldsymbol{D}_z,g^*(1,x_0)|x_0,\boldsymbol{\beta}_g,\sigma^2_g]$.

$$[\boldsymbol{\beta}_f|\cdots]\propto[\boldsymbol{\beta}_f][\boldsymbol{D}_T|x_1,\ldots,x_T,\boldsymbol{\beta}_f,\sigma^2_\epsilon] \tag{5}$$

$$\begin{aligned}[\boldsymbol{\beta}_g|\cdots]\propto{}&[\boldsymbol{\beta}_g][\boldsymbol{D}_z,g^*(1,x_0)|x_0,\boldsymbol{\beta}_g,\sigma^2_g]\prod_{t=2}^{T+1}[x_t|\boldsymbol{\beta}_g,\sigma^2_\eta,\sigma^2_g,\boldsymbol{D}_z,x_{t-1},K_t]\prod_{t=2}^{T+1}\big[K_t|\boldsymbol{\beta}_g,\sigma^2_\eta\\&\sigma^2_g,\boldsymbol{D}_z,x_{t-1}\big]\end{aligned} \tag{6}$$

$$[\sigma^2_\epsilon|\cdots]\propto[\sigma^2_\epsilon][\boldsymbol{D}_T|x_1,\ldots,x_T,\boldsymbol{\beta}_f,\sigma^2_\epsilon] \tag{7}$$

$$[\sigma^2_f|\cdots]\propto[\sigma^2_f][\boldsymbol{D}_T|x_1,\ldots,x_T,\boldsymbol{\beta}_f,\sigma^2_f] \tag{8}$$

$$\begin{aligned}[\sigma^2_\eta|\cdots]\propto{}&[\sigma^2_\eta][x_1|g^*(1,x_0),\sigma^2_\eta,K_1][K_1|g^*(1,x_0),\sigma^2_\eta]\prod_{t=2}^{T+1}[x_t|\boldsymbol{\beta}_g,\sigma^2_\eta,\sigma^2_g,\boldsymbol{D}_z,x_{t-1},K_t]\\&\prod_{t=2}^{T+1}[K_t|\boldsymbol{\beta}_g,\sigma^2_g,\sigma^2_\eta,\boldsymbol{D}_z,x_{t-1}]\end{aligned} \tag{9}$$

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

$$\sigma_\eta^2, \boldsymbol{D}_z, x_{t-1}] \tag{10}$$

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

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

$$[K_1|g^*(1,x_0), \sigma_\eta^2] \tag{12}$$

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

$$\left.\boldsymbol{D}_z, x_{t-1}\right] \tag{13}$$

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

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

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

$$[x_{t+1}|\cdots] \propto [x_{t+1}|\boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \boldsymbol{D}_z, x_t][x_{t+2}|\boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2, \boldsymbol{D}_z, x_{t+1}, K_{t+2}] \left[K_{t+2}|\boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2,\right.$$

$$\left.\boldsymbol{D}_z, x_{t+1}\right] [\boldsymbol{D}_T|x_1, \ldots, x_T, \boldsymbol{\beta}_f, \sigma_\epsilon^2], \quad t = 1, \ldots, T-1 \tag{16}$$

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

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

$$[K_t|\cdots] \propto [x_t|\boldsymbol{\beta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_{t-1}, K_t][K_t|\boldsymbol{\beta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_{t-1}], \ t = 2, \ldots, T+1. \tag{18}$$

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

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

$$\times \{\boldsymbol{H}'_{D_T}(\sigma_f^2 \boldsymbol{A}_{f,D_T} + \sigma_\epsilon^2 I)^{-1} \boldsymbol{D}_T \ + \ \boldsymbol{\Sigma}^{-1}_{\beta_{f,0}} \boldsymbol{\beta}_{f,0}\} \tag{19}$$

and variance

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

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

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

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

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

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

where

$$\mu_{\beta_g} = E[\boldsymbol{\beta}_g|\cdots] = \Bigg\{\boldsymbol{\Sigma}^{-1}_{\beta_g,0} + \frac{1}{\sigma^2_g}[\boldsymbol{H}'_{D_z}, \boldsymbol{h}(1,x_0)]\boldsymbol{A}^{-1}_{D_z,g^*(1,x_0)}[\boldsymbol{H}'_{D_z}, \boldsymbol{h}(1,x_0)]'$$

$$+ \sum_{t=1}^{T} \frac{\left(\boldsymbol{H}'_{D_z}\boldsymbol{A}^{-1}_{g,D_z}\boldsymbol{s}_{g,D_z}(t+1,x_t) - \boldsymbol{h}(t+1,x_t)\right)\left(\boldsymbol{H}'_{D_z}\boldsymbol{A}^{-1}_{g,D_z}\boldsymbol{s}_{g,D_z}(t+1,x_t) - \boldsymbol{h}(t+1,x_t)\right)'}{\sigma^2_{x_t}}\Bigg\}^{-1}$$

$$\Bigg\{\boldsymbol{\Sigma}^{-1}_{\beta_g,0}\boldsymbol{\beta}_{g,0} + \frac{1}{\sigma^2_g}[\boldsymbol{H}'_{D_z}, \boldsymbol{h}(1,x_0)]\boldsymbol{A}^{-1}_{D_z,g^*(1,x_0)}[\boldsymbol{D}_z, g^*(1,x_0)]$$

$$+ \sum_{t=1}^{T} \frac{\left(x_{t+1} + 2\pi K_{t+1} - \boldsymbol{s}_{g,D_z}(t+1,x_t)'\boldsymbol{A}^{-1}_{g,D_z}\boldsymbol{D}_z\right)\left(\boldsymbol{h}(t+1,x_t) - \boldsymbol{H}'_{D_z}\boldsymbol{A}^{-1}_{g,D_z}\boldsymbol{s}_{g,D_z}(t+1,x_t)\right)}{\sigma^2_{x_t}}\Bigg\} \tag{23}$$

and

$$\boldsymbol{\Sigma}_{\beta_g} = V[\boldsymbol{\beta}_g|\cdots] = \Bigg\{\boldsymbol{\Sigma}^{-1}_{\beta_g,0} + \frac{1}{\sigma^2_g}[\boldsymbol{H}'_{D_z}, \boldsymbol{h}(1,x_0)]\boldsymbol{A}^{-1}_{D_z,g^*(1,x_0)}[\boldsymbol{H}'_{D_z}, \boldsymbol{h}(1,x_0)]'$$

$$+ \sum_{t=1}^{T} \frac{\left(\boldsymbol{H}'_{D_z}\boldsymbol{A}^{-1}_{g,D_z}\boldsymbol{s}_{g,D_z}(t+1,x_t) - \boldsymbol{h}(t+1,x_t)\right)\left(\boldsymbol{H}'_{D_z}\boldsymbol{A}^{-1}_{g,D_z}\boldsymbol{s}_{g,D_z}(t+1,x_t) - \boldsymbol{h}(t+1,x_t)\right)'}{\sigma^2_{x_t}}\Bigg\}. \tag{24}$$

Hence $[\boldsymbol{\beta}_g|\cdots]$ follows a four-variate normal distribution with mean and variance μ_{β_g} and $\boldsymbol{\Sigma}_{\beta_g}$, respectively, and therefore, we update β_g using Gibbs sampling.

S-2.1.3 Updating σ^2_f and σ^2_g

The mathematical form of the full conditional distributions of σ^2_f and σ^2_g are not tractable, so we update σ^2_f and σ^2_g 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 $[x_t|\boldsymbol{\beta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_{t-1}, K_t]$ cancels with the density of $[K_t|\boldsymbol{\beta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_{t-1}]$ for $t = 2, \ldots, T+1$, 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:

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

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

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

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

$$[g^*(1, x_0) | \cdots] \propto \exp\left\{-\frac{1}{2\gamma^2_g}(g^* - \nu_g)^2\right\}, \tag{26}$$

where

$$\nu_g = E[g^*(1, x_0) | \cdots] = \left\{\frac{1}{\sigma^2_\eta} + \frac{1}{\sigma^2_g}(1 + \boldsymbol{s}_{g,D_z}(1, x_0)' \boldsymbol{\Sigma}^{-1}_{g,D_z} \boldsymbol{s}_{g,D_z}(1, x_0))\right\}^{-1}$$
$$\left\{\frac{x_1 + 2\pi K_1}{\sigma^2_\eta} + \frac{1}{\sigma^2_g}(\boldsymbol{h}(1, x_0)' \boldsymbol{\beta}_g + \boldsymbol{s}'_{g,D_z} \boldsymbol{\Sigma}^{-1}_{g,D_z} \boldsymbol{D}^*_z)\right\} \tag{27}$$

and

$$\gamma^2_g = V[g^*(1, x_0) | \cdots] = \left\{\frac{1}{\sigma^2_\eta} + \frac{1}{\sigma^2_g}(1 + \boldsymbol{s}_{g,D_z}(1, x_0)' \boldsymbol{\Sigma}^{-1}_{g,D_z} \boldsymbol{s}_{g,D_z}(1, x_0))\right\}, \tag{28}$$

with

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

and

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

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 $\boldsymbol{D}_z$

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

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

and covariance matrix

$$V(\boldsymbol{D}_z|\cdots) = \left\{ \frac{\boldsymbol{\Sigma}^{-1}_{g,D_z}}{\sigma^2_g} + \boldsymbol{A}^{-1}_{g,D_z} \left(\sum_{t=1}^{T} \frac{s_{g,D_z}(t+1,x_t) s'_{g,D_z}(t+1,x_t)}{\sigma^2_{x_t}} \right) \boldsymbol{A}^{-1}_{g,D_z} \right\}^{-1}. \tag{32}$$

Therefore, we update $\boldsymbol{D}_z$ 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:

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

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 $[x_2|\boldsymbol{\beta}_g,\sigma_\eta^2,\boldsymbol{D}_z,x_1,K_2]$ cancels with $[K_2|\boldsymbol{\beta}_g,\sigma_\eta^2,\boldsymbol{D}_z,x_1]$. 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

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

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

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

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

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

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

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

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 $t = 1, \ldots, T$.

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

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

and

$$\sigma^2_{x_{T+1}} = \sigma^2_\eta + \sigma^2_g\{1 - \boldsymbol{s}_{g,D_z}(T+1, x_T)'\boldsymbol{A}^{-1}_{g,D_z}\boldsymbol{s}_{g,D_z}(T+1, x_T)\}. \tag{38}$$

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 $K_t,\ t = 1, \ldots, T+1$

The full conditional distribution of K_1 reduces to the following form

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

and similarly the full conditional distribution of K_t becomes

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

for $t = 2, \ldots, T+1$. We update K_t, for $t = 1, \ldots, K+1$, 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.