Bayesian Nonparametric Dynamic State Space Modeling with Circular Latent States

Satyaki Mazumder\(^1\) and Sourabh Bhattacharya\(^2\)

\(^1\)Indian Institute of Science Education and Research, Kolkata
\(^2\)Indian Statistical Institute, Kolkata. Corresponding e-mail: sourabh@isical.ac.in.
Abstract

State space models are well-known for their versatility in modeling dynamic systems that arise in various scientific disciplines. Although parametric state space models are well-studied, nonparametric approaches are much less explored in comparison. In this article we propose a novel Bayesian nonparametric approach to state space modeling assuming that both the observational and evolutionary functions are unknown and are varying with time; crucially, we assume that the unknown evolutionary equation describes dynamic evolution of some latent circular random variable.

Based on appropriate kernel convolution of the standard Weiner process we model the time-varying observational and evolutionary functions as suitable Gaussian processes that take both linear and circular variables as arguments. Additionally, for the time-varying evolutionary function, we wrap the Gaussian process thus constructed around the unit circle to form an appropriate circular Gaussian process. We show that our process thus created satisfies desirable properties.

For the purpose of inference we develop an MCMC based methodology combining Gibbs sampling and Metropolis-Hastings algorithms. Applications to a simulated data set and a real ozone data set demonstrated quite encouraging performances of our model and methodologies.

Keywords: Circular random variable; Kernel convolution; Markov Chain Monte Carlo; State-space model; Weiner process; Wrapped Gaussian process.
## Contents

1 Introduction
   1.1 Flexibility of state space models .................................. 3
   1.2 Need for nonparametric approaches to state space models .......... 3
   1.3 A brief discussion on state space models with circular states ........ 4
   1.4 A brief overview of the contributions and organisation of this paper .... 5

2 Gaussian process based dynamic state space model with circular latent states
   2.1 Gaussian and wrapped Gaussian process representations of the observational and evolutionary functions ......................... 8
   2.2 Bayesian hierarchical structure of our nonparametric model based on circular latent states ........................................ 9
   2.3 Prior specifications ..................................................... 10

3 Look-up table approach to representing the distribution of the latent circular time series
   3.1 Joint distribution of of the latent circular variables induced by the look-up table 13
   3.2 Advantages of the look-up table approach .............................. 15

4 Simulation study
   4.1 True model ............................................................... 16
   4.2 Choices of prior parameters and the grid $G_z$ .......................... 16
   4.3 Brief discussion related to impropriety of the posteriors of some unknowns and the remedy ................................. 17
# 4.4 MCMC details ................................................................. 18
# 4.5 Results of our simulation study ........................................... 18

## 5 Real data analysis ............................................................. 22
5.1 A brief description of the data set ........................................ 22
5.2 Prior choices ..................................................................... 22
5.3 MCMC implementation ...................................................... 23
5.4 Results of our real data analysis ......................................... 23

## 6 Discussion and conclusion .................................................. 26

## A Appendix ............................................................. 28
A.1 Gaussian process on linear and angular component and its properties . . . . . . 28
A.2 Covariance structure of our Gaussian process ................................................. 29

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

S-1.1 Mean square continuity: .................................................. 31
S-1.2 Mean square differentiability ........................................... 32

**S-2 MCMC-based inference** .................................................. 34
S-2.1 Full conditional distributions ........................................... 36
S-2.1.1 Updating $\beta_f$ by Gibbs steps .................................. 37
S-2.1.2 Updating $\beta_g$ ......................................................... 38
S-2.1.3 Updating $\sigma_f^2$ and $\sigma_g^2$ ...................................... 39
S-2.1.4 Updating $\sigma_\epsilon^2$ .................................................. 40
S-2.1.5 Updating $\sigma_\eta^2$ .................................................... 40
1 Introduction

1.1 Flexibility of state space models

The versatility of state space models is clearly reflected from their utility in multifarious disciplines such as engineering, finance, medicine, ecology, statistics, etc. One reason for such widespread use of state space models is their inherent flexibility which allows modeling complex dynamic systems through the underlying latent states associated with an “evolutionary equation” and an “observational equation” that corresponds to the observed dynamic data. That most established time series models admit appropriate state space representations (see, for example, Durbin and Koopman (2001), Shumway and Stoffer (2011)) is vindication of the enormous flexibility of state space models.

1.2 Need for nonparametric approaches to state space models

Till date, most of the research on state space models have adhered to the parametric set-up, assuming known forms (either linear or non-linear) of the observational and evolutionary functions. The parametric approaches also explicitly specify that the latent states satisfy the
Markov property. These assumptions are not unquestionable, particularly in complex, real situations. Recently Ghosh et al. (2014) considered a Bayesian nonparametric approach to state space modeling, assuming that these time-varying functional forms are unknown, which they modeled by Gaussian processes. As a byproduct of their nonparametric approach based on Gaussian processes, it turned out that the latent states have a non-Markov, non-Gaussian, nonparametric distribution with complex dependence structure, which is suitable for modeling complex, realistic, dynamic systems. See Ghosh et al. (2014) for a detailed discussion on the advantages of their nonparametric approach. As argued in Ghosh et al. (2014), particularly while modeling complex scientific phenomena with the state space approach, nonparametric approaches are preferable because validation of parametric approaches require data on both observed time series and the latent states, but clearly data on the latter are not available.

1.3 A brief discussion on state space models with circular states

The approach of Ghosh et al. (2014) assumes that the latent states are linear random variables. However, in reality, there may be strong evidences that the observed time series data depends upon some circular time series which may not have been recorded, thereby rendering the latter latent time series data. A case in point is the ozone level time series data; even though it is well-known that ozone level depends upon wind direction (see Jammalamadaka and Lund (2006), for example), data on wind direction are often not recorded along with ozone level. A concrete example of such a real data, on which we illustrate our model and methodologies, is provided in Section 5. Other examples (see Holzmann et al. (2006)) include time series data on wind speed (linear) and ocean current (linear) which depend upon wind direction (circular); daily peak load of pollutants (linear) and the time of day when the peak
is attained (circular); speed (linear) and direction change (circular) of movements of objects, organisms and animals, to name a few. When both the linear and circular time series data are available, Holzmann et al. (2006) consider hidden Markov models in a discrete mixture context to statistically analyse such data sets. Our aim in this article is to propose a novel nonparametric state space approach when the circular time series data are unobserved, even though they are known to affect the available linear time series data.

1.4 A brief overview of the contributions and organisation of this paper

In this paper we adopt the philosophy of Ghosh et al. (2014) in modeling the unknown observational and evolutionary functions using Gaussian processes, but since here both the functions have arguments which are linear as well as circular, and since the evolutionary function itself is circular, it is clear that the approach of Ghosh et al. (2014), which is based on linear random variables, is no longer appropriate and quite substantial methodological advancement is necessary in our case.

We introduce our Bayesian nonparametric state space model with circular latent states in Section 2. The first challenge is to define a Gaussian process taking time and angle as arguments. We construct an appropriate Gaussian process by convolving a suitable kernel with the standard Brownian motion (Weiner process). The Gaussian process so defined enjoys desirable smoothness properties; moreover, as the absolute difference between two time points tends to infinity and/or the absolute difference between two angles tend to $\pi/2$ indicating orthogonality, the Gaussian process based covariance tends to zero as it should be. We provide these technical details in the Appendix. We provide further details of our Gaussian process with respect to continuity and smoothness properties in the supplement.
Mazumder and Bhattacharya (2014b), whose sections, figures and tables have the prefix “S-” when referred to in this paper. The Gaussian process that we create is an appropriate model for the time-varying observational function, but to model the evolutionary function which is circular in nature, we convert this Gaussian process into a wrapped Gaussian process so that it becomes a well-defined circular process.

To obtain the joint distribution of the latent states, in Section 3 we employ the “look-up table” approach of Bhattacharya (2007) (see also Ghosh et al. (2014)), but substantially modified for our circular set-up, which will play an important role in our MCMC based Bayesian inference.

In Section 4 we illustrate our model and methodologies with a simulation study, where we simulate the data set from a highly non-linear dynamic model, but fit our nonparametric model, pretending that the data-generating mechanism is unknown. Our experiment shows that even in this highly challenging situation our method successfully captures the entire set of true latent states and future observations in terms of coverage associated with 95% credible regions. In Section 5 we demonstrate the performance of our dynamic nonparametric model in the case of a real time series data set comprising the level of ozone present in the atmosphere; we obtain quite encouraging results, particularly in terms of forecasting.
2 Gaussian process based dynamic state space model with circular latent states

We introduce our proposed state space model as follows: For \( t = 1, 2, \ldots, T \),

\[
y_t = f(t, x_t) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_\epsilon),
\]

\[
x_t = \{g(t, x_{t-1}) + \eta_t\} \mod 2\pi, \quad \eta_t \sim N(0, \sigma^2_\eta),
\]

where \( \{y_t; t = 1, \ldots, T\} \) is the time series observed on the real line; \( \{x_t; t = 0, 1, \ldots, T\} \) are the latent circular states; \( f(\cdot, \cdot) \) is the unknown observational function taking values on the real line, and \( g(\cdot, \cdot) \) is the unknown evolutionary function with values on the circular manifold. In [2], \( \mod 2\pi \) stands for the mod 2\pi operation. Note that

\[
\{g(t, x_{t-1}) + \eta_t\} \mod 2\pi = \{g(t, x_{t-1}) \mod 2\pi + \eta_t \mod 2\pi\} \mod 2\pi
\]

\[
= \{g^*(t, x_{t-1}) + \eta_t\} \mod 2\pi,
\]

where \( g^* \) is the linear counterpart of \( g \). For convenience, we shall often use representation [3]. Indeed, for obtaining the distribution of \( x_t \), we shall first obtain the distribution of the linear random variable \( g^*(t, x_{t-1}) + \eta_t \) and then apply the mod 2\pi operation to \( g^*(t, x_{t-1}) + \eta_t \) to compute the distribution of the circular variable \( x_t \).

Both the observational and the evolutionary functions have arguments \( t \), which is linear in nature, and \( x \), which is angular. The linear argument has been brought in to ensure that the functions are time-varying; see [Ghosh et al. 2014] for the details.
2.1 Gaussian and wrapped Gaussian process representations of the observational and evolutionary functions

Following Ghosh et al. (2014) we consider Gaussian and wrapped Gaussian processes to model \( f \) and \( g \) respectively; for this purpose we first construct appropriate Gaussian processes for \( f \) and \( g^* \) by convolving a suitable kernel with the standard Weiner process. The details are provided in Appendix A.1. Once we build such Gaussian processes, we can convert that for modeling \( g \) into a wrapped Gaussian process with the mod 2\( \pi \) operation applied to \( g^* \). However, since our evolutionary equation given by (2) involves the error term \( \eta_t \), we will need to compute the distribution of \( g^*(\cdot, \cdot) + \eta_t \) before applying the mod 2\( \pi \) operation.

In the Gaussian process construction detailed in Appendix A.1 we assume the mean functions to be of forms \( \mu_f(\cdot, \cdot) = h(\cdot, \cdot)'\beta_f \) and \( \mu_g(\cdot, \cdot) = h(\cdot, \cdot)'\beta_g \), where \( h(t, z) = (1, t, \tan(z))' \); here \( z \) is an angular quantity. As regards the covariance structure it turned out that for any fixed \((t_1, z_1) \) and \((t_2, z_2) \), where \( t_1, t_2 \) are linear quantities and \( z_1, z_2 \) are angular quantities, the forms of the covariances are given by

\[
c_f((t_1, z_1), (t_2, z_2)) = \exp\{-\sigma_f^4(t_1 - t_2)^2\} \cos(|z_1 - z_2|)
\]

and

\[
c_g((t_1, z_1), (t_2, z_2)) = \exp\{-\sigma_g^4(t_1 - t_2)^2\} \cos(|z_1 - z_2|).
\]

A very attractive property of our Gaussian process is that whenever \(|\theta_1 - \theta_2| = \pi/2\), implying orthogonality of two directions, the covariance becomes 0, the difference in time notwithstanding. Obviously, as the time difference tends to infinity, then also the covariance tends to zero. That desired continuity and smoothness properties hold for our Gaussian process are proved in Section S-1 of the supplement. Thus, the Gaussian process we constructed seems to have quite reasonable features that are desirable in our linear-circular context.
2.2 Bayesian hierarchical structure of our nonparametric model based on circular latent states

Our model admits the following hierarchical representation, which is analogous to that provided in Ghosh et al. (2014) but with the mod $2\pi$ operation applied to the distributions associated with the circular variables:

\[
\begin{align*}
[y_t|f, \theta_f, x_t] &\sim N\left(f(t, x_t), \sigma^2_f\right); \ t = 1, \ldots, T, \\
[x_t|g, \theta_g, x_{t-1}] &\sim N\left(g^*(t, x_{t-1}), \sigma^2_\eta\right) [2\pi]; \ t = 1, \ldots, T, \\
[x_0] &\sim N\left(\mu_0, \sigma^2_{x_0}\right) [2\pi], \\
[f(\cdot, \cdot)|\theta_f] &\sim GP\left(h(\cdot, \cdot), \beta_f, \sigma^2_f c_f(\cdot, \cdot)\right), \\
g(\cdot, \cdot)|\theta_g] &\sim GP\left(h(\cdot, \cdot), \beta_g, \sigma^2_g c_g(\cdot, \cdot)\right) [2\pi], \\
[\beta_f, \sigma^2_f, \beta_g, \sigma^2_g, \sigma^2_\epsilon, \sigma^2_\eta] &= [\beta_f, \sigma^2_f][\beta_g, \sigma^2_g][\sigma^2_\epsilon, \sigma^2_\eta].
\end{align*}
\]

In the above, GP stands for “Gaussian Process”. Integrating out $f(\cdot, \cdot)$ from the above hierarchical structure we obtain that given $x_1, \ldots, x_T$, $D_T = (y_1, \ldots, y_T)'$ has the multivariate normal distribution of dimension $T$ with mean

\[
\mu_{yt} = H_{D_T} \beta_f
\]

and covariance matrix

\[
\Sigma_{yt} = \sigma^2_f A_f + \sigma^2_\epsilon I_T,
\]

with $H'_{D_T} = (h(1, x_1), \ldots, h(T, x_T))$ and the $(i, j)$-th element of $A_f$ being $c_f((i, x_i), (j, x_j))$.

For obtaining the joint distribution of the latent circular state variables, we consider the
“look-up” table approach, but before introducing this, which we discuss in details in Section 3, in the next section we provide details regarding the prior distributions of the parameters associated with the above hierarchical structure.

### 2.3 Prior specifications

We assume the following prior distributions.

\[
[x_0] \sim \text{von Mises}(\mu_0, \sigma_0^2) \quad (12)
\]

\[
[\sigma^2_\epsilon] \propto (\sigma^2_\epsilon)^{\left(-\frac{\alpha_\epsilon+2}{2}\right)} \exp \left\{ -\frac{\gamma_\epsilon}{2\sigma^2_\epsilon} \right\} ; \quad \alpha_\epsilon, \gamma_\epsilon > 0 \quad (13)
\]

\[
[\sigma^2_\eta] \propto (\sigma^2_\eta)^{\left(-\frac{\alpha_\eta+2}{2}\right)} \exp \left\{ -\frac{\gamma_\eta}{2\sigma^2_\eta} \right\} ; \quad \alpha_\eta, \gamma_\eta > 0 \quad (14)
\]

\[
[\sigma^2_g] \propto (\sigma^2_g)^{\left(-\frac{\alpha_g+2}{2}\right)} \exp \left\{ -\frac{\gamma_g}{2\sigma^2_g} \right\} ; \quad \alpha_g, \gamma_g > 0 \quad (15)
\]

\[
[\sigma^2_f] \propto (\sigma^2_f)^{\left(-\frac{\alpha_f+2}{2}\right)} \exp \left\{ -\frac{\gamma_f}{2\sigma^2_f} \right\} ; \quad \alpha_f, \gamma_f > 0 \quad (16)
\]

\[
[\beta_f] \sim N(\beta_{f,0}, \Sigma_{\beta_f,0}) \quad (17)
\]

\[
[\beta_g] \sim N(\beta_{g,0}, \Sigma_{\beta_g,0}) \quad (18)
\]

It is assumed that all the prior parameters are known and the choice of the prior parameters are discussed in Sections 4 and 5.
3 Look-up table approach to representing the distribution of the latent circular time series

For obtaining the joint distribution of the latent circular variables, we employ the look-up table approach of Bhattacharya (2007) and Ghosh et al. (2014), but because of the circular nature of the latent states, appropriate modifications are necessary. For the sake of simplicity of illustration, for the time being let us assume that $\eta_t = 0$ for all $t$.

We consider a set of grid points in the interval $[0, 2\pi]$; let this set be denoted by $G_z = \{z_1, \ldots, z_n\}$. Let $D_z = (g^*(1, z_1), \ldots, g^*(n, z_n))$. Note that $D_z$ has a joint multivariate normal distribution of dimension $n$ with mean vector

$$E[D_z | \beta_g, \sigma_g^2] = H_{D_z} \beta_g,$$

and covariance matrix

$$V[D_z | \beta_g, \sigma_g^2] = \sigma_g^2 A_{g,D_z},$$

where $H'_{D_z} = (h(1, z_1), \ldots, h(n, z_n))$ and the $(i, j)$-th element of $A_{g,D_z}$ is $c_g(z_i, z_j)$. The conditional distribution of $D_z$ given $(x_0, g^*(1, x_0)), \beta_g$ and $\sigma_g^2$ is an $n$-variate normal with mean vector

$$E[D_z | \beta_g, \sigma_g^2, x_0, g^*(1, x_0)] = H_{D_z} \beta_g + s_{g,D_z}((1, x_0))(g^*(1, x_0) - h(1, x_0)' \beta_g)$$

and conditional variance

$$\text{Var}[D_z | \beta_g, \sigma_g^2, x_0, g^*(1, x_0)] = \sigma_g^2 (A_{g,D_z} - s_{g,D_z}((1, x_0))(s_{g,D_z}((1, x_0)))').$$
The conditional distribution of $g^*(t, x_{t-1})$ given $D_z$ and $x_{t-1}$ is a normal distribution with mean

$$E[g^*(t, x_{t-1})|D_z, x_{t-1}, \beta_g, \sigma^2_g] = h(t, x_{t-1})'\beta_g + s_{g,D_z}((t, x_{t-1}))'A^{-1}_{g,D_z}(D_z - H_{D_z}\beta_g)$$  \hspace{1cm} (23)

and variance

$$\text{Var}[g^*(t, x_{t-1})|D_z, x_{t-1}, \beta_g, \sigma^2_g] = \sigma^2_g \left(1 - (s_{g,D_z}(t, x_{t-1}))'A^{-1}_{g,D_z}s_{g,D_z}(t, x_{t-1})\right).$$ \hspace{1cm} (24)

With the above distributional details, our procedure of representing the circular latent states in terms of the auxiliary random vector $D_z$, conditional on $\beta_g$ and $\sigma^2_g$, can be described as follows.

1. $x_0 \sim \pi^*$, where $\pi^*$ is some appropriate prior distribution on the unit circle.

2. Given $x_0, \beta_g$ and $\sigma^2_g$, $x_1 = g^*(1, x_0) [2\pi] = g(1, x_0)$, where $g^*(1, x_0)$ has a normal distribution with mean $h(1, x_0)'\beta_g$ and variance $\sigma^2_g$.

3. Given $x_0, x_1, \beta_g$ and $\sigma^2_g$, $[D_z|x_0, g^*(1, x_0), \beta_g, \sigma^2_g]$ is a multivariate normal distribution with mean (21) and covariance matrix (22).

4. For $t = 2, 3, \ldots$, $x_t^* \sim [g^*(t, x_{t-1})|D_z, x_{t-1}, \beta_g, \sigma^2_g]$ which is a normal distribution with mean and variance given by (23) and (24) respectively; $x_t$ is related to $x_t^*$ via $x_t = x_t^*[2\pi]$.
3.1 Joint distribution of of the latent circular variables induced by the look-up table

Using the look-up table approach the joint distribution of \((D_z, x_0, x_1, x_2, \ldots, x_T, x_{T+1})\) given \(\beta_g, \sigma_{\eta}^2\) and \(\sigma_g^2\) is as follows:

\[
\begin{align*}
[x_0, x_1, \ldots, x_{T+1}, D_z | \beta_g, \sigma_{\eta}^2, \sigma_g^2] &= [x_0][x_1 = \{g^*(1, x_0) + \eta_1\} [2\pi][x_0, \sigma_{\eta}^2, \sigma_g^2][D_z | x_0, g^*(1, x_0), \\
& \quad \beta_g, \sigma_g^2] \times \prod_{t=1}^{T} [x_{t+1} = \{g^*((t+1), x_t) + \eta_{t+1}\} [2\pi][\beta_g, \sigma_g^2, \\
& \quad D_z, x_t, \sigma_{\eta}^2].
\end{align*}
\]

In the above, \([x_0] \sim \pi^*\) is a prior distribution on the unit circle, \([x_1 = \{g^*(1, x_0) + \eta_1\} [2\pi][x_0, \beta_g, \sigma_{\eta}^2, \sigma_g^2]\) follows a wrapped normal distribution, derived from \([x_1^* = g^*(1, x_0) + \eta_1|x_0, \beta_g, \sigma_{\eta}^2, \sigma_g^2]\), which is a normal distribution with mean \(\mu_g(1, x_0) = h(1, x_0)'\beta_g\) and variance \(\sigma_g^2 + \sigma_{\eta}^2\).

As already noted in Section 3, \([D_z | x_0, g^*(1, x_0), \beta_g, \sigma_g^2]\) is multivariate normal with mean and covariance matrix given by (21) and (22) respectively, and finally the conditional distribution \([x_{t+1} = \{g^*((t+1), x_t) + \eta_{t+1}\} [2\pi][\beta_g, \sigma_g^2, D_z, x_t, \sigma_{\eta}^2]\) is again a wrapped normal distribution derived from \([x_{t+1}^* = g^*((t+1), x_t) + \eta_{t+1}|\beta_g, \sigma_g^2, D_z, x_t, \sigma_{\eta}^2]\), which is a normal distribution with mean \(\mu_{x_t}\) given by (23) and variance

\[
\sigma_{x_t}^2 = \sigma_{\eta}^2 + \sigma_g^2 (1 - (s_{g,D_z}(t, x_{t-1}))'A_{g,D_z}^{-1}s_{g,D_z}(t, x_{t-1})).
\]

(26)

For explicit derivations of the conditional distributions associated with (25) it is necessary to bring in some more auxiliary variables. To be specific, note that \(x_t^* = x_t + 2\pi K_t\), where \(K_t = [x_t^*/2\pi]\), where, for any \(u\), \([u]\) denotes the greatest integer not exceeding \(u\). Note that
for each $t$, $K_t$ can take values in the set $\{\cdots, -2, -1, 0, 1, 2, \cdots\}$. Here we view the wrapped number $K_t$ as a random variable; see also Ravindran and Ghosh (2011).

Note that $x_1$ given $(g^*(1, x_0), \beta_g, \sigma^2_\eta, \sigma^2_g, K_1)$ has the following distribution:

$$[x_1 | g^*(1, x_0), \beta_g, \sigma^2_\eta, \sigma^2_g, K_1] = \frac{1}{\sqrt{2\pi} \sigma_\eta} \exp \left( -\frac{1}{2\sigma_\eta^2} (x_1 + 2\pi K_1 - g^*(1, x_0))^2 \right) I_{[0, 2\pi]}(x_1)$$

and the distribution of $K_1$ given $(g^*(1, x_0), \sigma^2_\eta)$ is

$$[K_1 | g^*(1, x_0), \sigma^2_\eta] = \Phi \left( \frac{2\pi (K_1 + 1) - g^*(1, x_0)}{\sigma_\eta} \right) - \Phi \left( \frac{2\pi K_1 - g^*(1, x_0)}{\sigma_\eta} \right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

Similarly, for $t = 2, \ldots, T + 1$, the distributions of $x_t$ given $(\beta_g, \sigma^2_\eta, \sigma^2_g, D_z, x_{t-1}, K_t)$ and $K_t$ given $(\beta_g, \sigma^2_\eta, \sigma^2_g, D_z, x_{t-1})$, respectively, are

$$[x_t | \beta_g, \sigma^2_\eta, \sigma^2_g, D_z, x_{t-1}, K_t] = \frac{1}{\sqrt{2\pi} \sigma_{x_t}} \exp \left( -\frac{1}{2\sigma_{x_t}^2} (x_t + 2\pi K_t - \mu_{x_t})^2 \right) I_{[0, 2\pi]}(x_t)$$

and

$$[K_t | \beta_g, \sigma^2_\eta, \sigma^2_g, D_z, x_{t-1}] = \Phi \left( \frac{2\pi (K_t + 1) - \mu_{x_t}}{\sigma_{x_t}} \right) - \Phi \left( \frac{2\pi K_t - \mu_{x_t}}{\sigma_{x_t}} \right),$$

where $\mu_{x_t}$ and $\sigma_{x_t}$ are given by (23) and (26), respectively.

Thus, using the conditionals (27), (28), (29) and (30), the joint distribution of the latent circular variables, conditional on $\beta_g, \sigma^2_\eta$ and $\sigma^2_g$ can be represented as

$$[x_0, x_1, \ldots, x_{T+1} | \beta_g, \sigma^2_\eta, \sigma^2_g]$$
\[ = \sum_{K_1, \ldots, K_{t+1}} \int [x_0, x_1, \ldots, x_{T+1}, D_z, K_1, \ldots, K_{T+1}] \beta_g, \sigma^2_\eta, \sigma^2_g dD_z \]

\[ = \sum_{K_1, \ldots, K_{t+1}} \int [x_0][D_z|x_0, g^*(1, x_0), \beta_g, \sigma^2_g][g^*(1, x_0)|x_0, \beta_g, \sigma^2_g] \]

\[ \times [x_1|x_0, g^*(1, x_0), K_1, \sigma^2_\eta, \sigma^2_g][K_1|x_0, g^*(1, x_0), \sigma^2_\eta, \sigma^2_g] \]

\[ \times \prod_{t=1}^{T} [x_{t+1}|\beta_g, \sigma^2_g, D_z, x_t, \sigma^2_\eta, K_{t+1}][K_{t+1}|\beta_g, \sigma^2_\eta, \sigma^2_g, D_z, x_t] dg^*(1, x_0)dD_z. \]  

(31)

### 3.2 Advantages of the look-up table approach

Ghosh et al. (2014) provide ample details on the accuracy of the look-up table approach. In particular, they prove a theorem on the accuracy of the approximation of the distribution of the latent states using the look-up table, show that the joint distribution of the latent states is non-Markovian, even though conditionally on \( D_z \), the latent states have a Markov structure. Quite importantly, Ghosh et al. (2014) point out that this approach leads to great computational savings and remarkable numerical stability of the associated MCMC algorithm thanks to the fact that \( A_{g,D_z} \) needs to be inverted only once, even before beginning the MCMC simulations, and that the set of grid points \( G_z \) can be chosen so that \( A_{g,D_z} \) is invertible. These advantages clearly remain valid even in our circular set-up.
4 Simulation study

4.1 True model

We now illustrate the performance of our model and methodologies using a simulation study. For this purpose we simulate a set of observations of size 101 from the following nonlinear dynamic model:

\[ y_t = \tan^2(\theta_t)/20 + v_t; \]

\[ \tan\left(\frac{\theta_t - \pi}{2}\right) = \alpha \tan\left(\frac{\theta_{t-1} - \pi}{2}\right) + \beta \tan\left(\frac{\theta_{t-1} - \pi}{2}\right) + \gamma \cos(1.2(t - 1)) + u_t, \]

for \( t = 1, \ldots, 101 \), where \( u_t \) and \( v_t \) are normally distributed with means zero and variances \( \sigma^2_\eta \) and \( \sigma^2_\epsilon \). We set the values of \( \alpha, \beta \) and \( \gamma \) to be 0.05, 0.1 and 0.2, respectively; we fix the values of both \( \sigma^2_\eta \) and \( \sigma^2_\epsilon \) at 0.1. We consider the first 100 observations of \( y_t \) as known, and set aside the last observation for the purpose of forecasting.

4.2 Choices of prior parameters and the grid \( G_z \)

In this experiment we consider a tri-variate normal prior distribution for \( \beta_f \) with mean \((0, 0, 0)'\) and the identity matrix as the covariance matrix. For \( \beta_g \) we choose a tri-variate normal prior with mean vector \((2.5, 0.04, 1.0)'\) and covariance matrix being the identity matrix once again. Choices of these prior parameters ensured adequate mixing of our MCMC algorithm.

Observe that in [2] of our proposed model, in the mean function of the underlying Gaussian process \( g \), the third component of \( \beta_g \) and \( \tan(x_{t-1}) \) are multiplied, so they suffer
an identifiability problem. To counter this problem, we set the third component of \( \beta_g \) to be 1, throughout the experiment.

For \( \sigma_\epsilon \) and \( \sigma_f \) we consider inverse gamma priors with parameters \((4.01, 0.005 \times 5.01)\) and \((4.01, 0.1 \times 5.01)\), respectively, so that the mode of \( \sigma_\epsilon \) is 0.005 and that of \( \sigma_f \) is 0.1, respectively. We choose the first parameter of the inverse gamma distribution to be equal to 4.01 so that the variance is 200 times the square of the mean of the inverse gamma distribution, which are in this case 0.012 and 0.25, respectively. The choices of second prior parameters for \( \sigma_\epsilon \) and \( \sigma_f \) yielded adequate mixing of our MCMC algorithm.

Finally we divide the interval \([0, 2\pi]\) into 100 sub-intervals and choose one point from each of the sub-intervals; these values constitute the second component of the two dimensional grid \( G_z \). For the first component of \( G_z \), we select a random number uniformly from each of the 100 subintervals \([i, i+1], i = 0, \ldots, 99\).

### 4.3 Brief discussion related to impropriety of the posteriors of some unknowns and the remedy

An interesting feature associated with our model is the impropriety of the posteriors of \( \sigma_g \), \( \sigma_\eta \) and \( K_1, \ldots, K_{T+1} \), when they are all allowed to be random. In a nutshell, for any value of \( K_t \), exactly the same value of the circular variable \( x_t \) is obtained by the mod \( 2\pi \) operation applied to \( x_t^* = x_t + 2\pi K_t \). Thus, given \( x_t \), it is not possible to constrain \( K_t \) unless both \( \sigma_g \) and \( \sigma_\eta \) are bounded. Boundedness of \( \sigma_g \) and \( \sigma_\eta \) would ensure that \( x_t^* \) has finite variance, which would imply finite variability of \( K_t \).

Since it is unclear how to select a bounded prior for \( \sigma_g \) and \( \sigma_\eta \), we obtain the maximum likelihood estimates (MLEs) of these variances and plug in these estimates in our model. To obtain the MLEs, we implemented the simulated annealing methodology (see, for example,
Robert and Casella (2004), Liu (2001)) where at each iteration we proposed new values of these variances, then integrated out all the other parameters using averages of Monte Carlo simulations, given the proposed values of $\sigma_g$ and $\sigma_\eta$, so that we obtain the integrated likelihood given the proposed variances; then we calculated the acceptance ratio, and finally decreased the temperature parameter of our simulated annealing algorithm before proceeding to the next iteration. The MLEs turned out to be $\hat{\sigma}_g = 0.54$ and $\hat{\sigma}_\eta = 0.44$.

4.4 MCMC details

As detailed in Section S-2 of the supplement, our MCMC algorithm updates some parameters using Gibbs steps, and the remaining using random walk Metropolis-Hastings steps. To update $\sigma_\epsilon$ and $\sigma_f$ we implemented normal random walk with variance 0.05; $x_0$ is updated using vonMises distribution with $\kappa = 3.0$, and for updating $x_t$ a mixture of two vonMises distributions with $\kappa = 0.5$ and $\kappa = 3.0$ is used for $t = 1, \ldots, T$. The wrapping variables $K_t; t = 1, \ldots, T$, are updated using the discrete normal random walk with variance 1.0. All these choices are made very painstakingly after carefully adjudging the mixing properties of many pilot MCMC runs. The rest of the parameters are updated using Gibbs steps, as detailed in Section S-2 of the supplement.

With the above choices of the prior parameters and $G_2$, and with the above MCMC updating procedure of the parameters, we performed 1,40,000 MCMC simulations with a burn-in period consisting of the first 70,000 iterations.

4.5 Results of our simulation study

The posterior densities of the components of $\beta_f$ are provided in Figure 1. Figure 2 displays the posterior densities of the first two components of $\beta_g$, and the posterior density of $\sigma_f$. 

18
Figure 3 depicts the posterior density of the $\sigma_e$ and $x_{101}$. The horizontal bold black lines denote the 95% credible interval credible intervals and the vertical lines denote the true values. Observe that the true values in each of the cases fall well within the 95% credible intervals.

The 95% credible intervals of the latent time series $x_1, \ldots, x_T$ is displayed in Figure 4. It is worth observing that the entire set of true latent circular data $\{x_1, \ldots, x_T\}$ falls well within the respective 95% credible regions.

Figure 5 depicts the posterior predictive density corresponding to $y_{101}$; the true value is well within the 95% credible intervals of the predictive density.

Thus, our model performs quite encouragingly, in spite of the true model being highly non-linear and assumed to be unknown. As a result, we expect our model to perform adequately in general situations.

Figure 1: Posterior densities of the three components of $\beta_f$. 

19
Figure 2: Posterior densities of the first and second components of $\beta_g$ and the posterior density of $\sigma_f$.

Figure 3: Posterior densities of $\sigma_\epsilon$ and the $x_{101}$. 
Figure 4: 95% credible intervals of the time series $x_1, \ldots, x_T$, black line denotes the true time series.

Figure 5: Posterior predictive density corresponding to $y_{101}$, where the vertical line denotes the true value and the bold black horizontal line denotes the 95% credible interval.
5 Real data analysis

5.1 A brief description of the data set

We now apply our model and methodologies to a real data set obtained from the website [http://www.esrl.noaa.gov/gmd/grad/neubrew/OmiDataTimeSeries.jsp](http://www.esrl.noaa.gov/gmd/grad/neubrew/OmiDataTimeSeries.jsp). The data concerns the ozone level present in the atmosphere at a particular location and at a particular year. For our analysis we select a location with latitude 40.125 and longitude 105.238. We collected 101 observations starting from January 1 of 2013. Although it is expected that the ozone level present in the atmosphere depends upon the direction of wind flow (see Jammalamadaka and Lund [2006]), the data on the direction of wind flow is not available at that particular location and time. Therefore, we expect that our general, nonparametric model and the associated methods will be quite useful in this situation. We retain 100 observations for our analysis and keep aside the last observation for the purpose of prediction. Before applying our model and methods, we first de-trend the data-set.

5.2 Prior choices

We chose the prior parameters so as to obtain reasonable prediction of the future observation (set aside as \(y_{101}\)), and to obtain adequate mixing of our MCMC algorithm. As such we specify the prior means of \(\beta_f\) and \(\beta_g\) to be \((0, 0, 0)’\) and \((1, 1, 1)’\), respectively. The prior covariance matrix for \(\beta_f\) has been chosen to be an identity matrix of order 3 \(\times\) 3 and for \(\beta_g\) it has been taken to be a diagonal matrix of order 3 \(\times\) 3, with diagonal elements 0.01, 0.01 and 0.05, respectively. Following the discussion in Section 4, we fixed the third component \(\beta_g\) at 1 throughout the experiment to avoid identifiability issues. The shape parameters for \(\sigma_e\) and \(\sigma_f\) in the respective inverse gamma prior distributions are chosen to

22
be 4.01 and the scale parameters are chosen to be 0.01 × 5.01 and 0.001 × 5.01, respectively, so that the prior modes for \( \sigma_e \) and \( \sigma_f \) are 0.01 and 0.001, respectively. The choice of the first parameter of inverse gamma is justified in Section 4.

The MLEs of \( \sigma_h \) and \( \sigma_g \), obtained by the simulated annealing method discussed in Section 4.3, are 0.9794 and 0.2950, respectively.

5.3 MCMC implementation

With these choices of prior parameters we implement our MCMC algorithm detailed in Section S-2 of the supplement with the random walk scales chosen on the basis of informal trial and error method associated with many pilot runs of our MCMC algorithm. Our final MCMC run is based on 65000 iterations of MCMC, of which we discarded the first 25000 iterations as the burn-in period.

5.4 Results of our real data analysis

The posterior densities of the three components of \( \beta_f \) and the two components of \( \beta_g \) are provided in Figures 6 and 7, respectively. The posterior densities of \( \sigma_e \) and \( \sigma_f \) are shown in Figure 8. Figure 9 presents the posterior 95% credible intervals of the latent circular process, along with the posterior median passing through the middle. Finally the posterior predictive density corresponding to \( y_{101} \) is provided in Figure 10. Here the thin vertical line denotes the true value of the 101-th observation and the thick line represents the 95% credible region of the posterior predictive density. The true value falls well within the 95% credible interval, which shows how well our model and the prior distributions of the parameters succeed in describing the uncertainty present in the data (see, for instance, Box and Tiao (1973) and Bickel and Doksum (2007)).
Figure 6: Posterior densities of the three components of $\beta_f$ for the ozone data.

Figure 7: Posterior densities of the first two components of $\beta_g$ for the ozone data.

Figure 8: Posterior densities of $\sigma_f$ and $\sigma_e$ for the ozone data.
Figure 9: 95% credible intervals and the median for the latent process of the ozone data.

Figure 10: Posterior predictive density of the 101-th observation for the ozone data. The thick horizontal line denotes the 95% credible interval and the vertical line denotes the true value.
6 Discussion and conclusion

In this paper we have proposed a novel nonparametric dynamic state space model where the latent process is in the circular manifold. We assumed that both the observational and the evolutionary functions are time-varying, but have unknown functional forms, which we model nonparametrically via appropriate Gaussian processes. For this purpose we derived a suitable Gaussian processes with both linear and circular arguments using kernel convolution.

Previously, some research has been carried out regarding Gaussian process with circular argument; see, for example, Dufour and Roy (1976), Gneiting (1998). However, most of the previous works considered the circular variable as the only argument. The main issue with the procedure of Dufour and Roy (1976) is that the covariance function turns out to be an infinite sum, and therefore, one has to approximate the infinite sum with proper truncation while applying to data. Hence, the question of error of approximation lurks in their procedure. Gneiting (1998) provided sufficient conditions under which any correlation function on the real line can be treated as a correlation function on circles. For that purpose Gneiting (1998) had to bound the argument of the correlation function on a finite interval, and therefore, the correlation can not tend to zero for the underlying Gaussian process.

The kernel convolution method has been used in Shafie et al. (2003), although they derived the Gaussian process on two linear arguments, one in \( \mathbb{R} \) (real line) and the other in \( \mathbb{R}^+ \) (positive part of the real line). Adler (1981) and Adler and Taylor (2007) dealt with Gaussian processes on manifolds in great details. However, they focused on Gaussian processes with arguments only on single manifold. Here we mention that although we also use the kernel convolution approach to forming appropriate Gaussian processes, our case is substantially different in that our Gaussian process construction is based on both linear and circular arguments. Moreover, we have chosen our kernel appropriately such that the
Gaussian process satisfies all desirable smoothness properties. The most elegant property of our Gaussian process is that the covariance function becomes 0 whenever $|\theta_1 - \theta_2| = \pi/2$. This implies that whenever two angular observations have orthogonal directions, their correlation turns out to be 0 irrespective of the difference in time. Obviously, we also have shown that as $|t_1 - t_2| \to \infty$ then the covariance function tends to 0, that is, as the difference in time goes to $\infty$, the correlation goes to 0.

The main aim of our research is to predict single or multiple future observations given the dynamic data at hand. That is, considering the Bayesian paradigm, our main objective is to obtain posterior predictive distributions. To achieve the posterior predictive distributions, appropriate MCMC simulation techniques needed to be devised. The main MCMC challenge for this model is to simulate the complete latent process; aided by the look up table concept of Bhattacharya (2007) (see also Ghosh et al. (2014)), appropriately adapted to suit the circular context, we could create an MCMC algorithm that has demonstrated very reasonable performances in both simulated and real data situations.

Our model and methods are applied to a simulated data where the data is generated from a highly nonlinear model, which is completely different from our own model. This simulation is done purposefully to demonstrate that our method is applicable to any nonlinear dynamic model where the latent process is in the circular manifold. It is also successfully shown that the future observation is well within the 95% credible region of the posterior predictive density. Quite importantly, the complete latent process fell well within the 95% credible intervals. The encouraging results are expected to provide any practitioner with some degree of latitude in applying our model in any practical context.

Finally, to demonstrate the performance of our model in practice we applied our model to the level of ozone present in the atmosphere for a particular location over a period of
time consisting of hundred observations. Even in this real example, our ideas yielded quite encouraging results. In particular, our posterior predictive density for the set-aside “future” observation successfully captured the true, set-aside value within the 95% credible interval.

In this paper we assumed that the observations $y_t$, $t = 1, \ldots, T$, are in $\mathbb{R}$. However, it is straightforward to extend our theory to $\mathbb{R}^p$ by suitably adjusting the kernel convolution technique. Even the technique can be extended to cases where the latent $x_t$, $t = 1, \ldots, T$, are also multidimensional. To keep the size of the paper reasonable we skip the multivariate part for this paper.

Acknowledgement

The authors are thankful to Moumita Das for very useful discussions.

A Appendix

A.1 Gaussian process on linear and angular component and its properties

To define a Gaussian process on linear and angular component we use the well known kernel convolution method. Let $k$ be any $d$-dimensional kernel such that

$$\int k^2(t) \, dt < \infty.$$ 

Here we choose two kernels as follows (in case $d = 1$)
\[ k_1(t) = \frac{1}{\nu} \pi^{-1/4} e^{-\frac{1}{2\nu^2} t^2}, \]

where \( \psi > 0 \), and
\[ k_2(t) = \pi^{-1/2} \cos(t) I(0 \leq t \leq \pi), \]
a trigonometric kernel. Based on above two choices of the kernel we propose a new Gaussian process for time and angle as arguments as follows.

\[
X(t, \theta) = \mu(t, \theta) + \left( \int_{-\infty}^{\infty} \psi^{-1/4} e^{-\frac{1}{2\nu^2} (y-t)^2} dW(y) \right) \left( \int_{0}^{\pi} \pi^{-1/2} \cos(u - \theta) dW(u) \right)
= \mu(t, \theta) + \psi^{-3/4} \int_{-\infty}^{\infty} \int_{0}^{\pi} e^{-\frac{1}{2\nu^2} (y-t)^2} \cos(u - \theta) dW(u) dW(y),
\]
where \( \mu(t, \theta) \) is the mean of the process which may depend on time \( t \) and angle \( \theta \) (as in our case mean is assumed to be of the form \( h(\cdot, \cdot)^T \beta \), with \( h(t, \theta)' = (1, t, \tan(\theta)) \); \( W(\cdot) \) is the one dimensional standard Wiener process. Next, we determine the structure of the covariance of our Gaussian process thus constructed.

### A.2 Covariance structure of our Gaussian process

With these separable kernels we calculate the covariance function of \( X(t_1, \theta_1) \) and \( X(t_2, \theta_2) \) for fixed \((t_1, \theta_1)\) and \((t_2, \theta_2)\) as

\[
\text{cov}(X(t_1, \theta_1), X(t_2, \theta_2)) = \psi^{-2} \pi^{-3/4} E \left\{ \left( \int_{-\infty}^{\infty} \int_{0}^{\pi} e^{-\frac{1}{2\nu^2} (y-t_1)^2} \cos(u - \theta_1) dW(u) dW(y) \right) \right. \\
\left. \left( \int_{-\infty}^{\infty} \int_{0}^{\pi} e^{-\frac{1}{2\nu^2} (y-t_2)^2} \cos(u - \theta_2) dW(u) dW(y) \right) \right\}
= \psi^{-2} \pi^{-3/4} \int_{-\infty}^{\infty} e^{-\frac{1}{2\nu^2} (y-t_1)^2} (y-t_1)^2 dW(y) \int_{0}^{\pi} \cos(u - \theta_1) \cos(u - \theta_2) du
\]
\[
\frac{1}{2} \psi^{-2} - \frac{4}{\pi} e^{-\frac{1}{\psi} (t_1^2 + t_2^2)} \int_\infty^{-\infty} e^{-\frac{1}{\psi} (y^2 - y(t_1 + t_2))} dy
\]

\[
\int_0^\pi \left[ \cos((-\theta_1 - \theta_2)) + \cos(2u - (\theta_1 + \theta_2)) \right] du
\]

\[
= \frac{1}{2} \psi^{-2} - \frac{4}{\pi} e^{-\frac{1}{\psi} (t_1^2 + t_2^2)} \left\{ \int_\infty^{-\infty} e^{-\frac{1}{\psi} (y - \frac{1}{\psi} t_1 + \frac{1}{\psi} t_2)^2} dy \right\}
\]

\[
\left\{ \pi \cos(|\theta_1 - \theta_2|) + \int_0^\pi \cos(2u - (\theta_1 + \theta_2)) du \right\}
\]

\[
= \frac{1}{2} \psi^{-2} - \frac{4}{\pi} e^{-\frac{1}{\psi} (t_1^2 + t_2^2)} \pi \cos(|\theta_1 - \theta_2|)
\]

\[
= \frac{1}{\psi^{-1} e^{-\frac{1}{\psi} (t_1^2 + t_2^2)} \pi \cos(|\theta_1 - \theta_2|)}
\]

\[
= \sigma^2 \exp\{-\sigma^4 |t_1 - t_2|^2\} \cos(|\theta_1 - \theta_2|),
\]

where \(\sigma^2 = \frac{\psi^{-1}}{2}\).

Supplementary Material

Throughout, we refer to our main paper Mazumder and Bhattacharya (2014a) 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 \(\theta\), and that the derivatives are bounded. Formally, we assume that \(\frac{\partial^2 \mu(t, \theta)}{\partial t^2}, \frac{\partial^2 \mu(t, \theta)}{\partial \theta^2}, \frac{\partial^2 \mu(t, \theta)}{\partial \theta \partial t} (= \frac{\partial^2 \mu(t, \theta)}{\partial t \partial \theta})\) 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 \(K(|t_1 - t_2|, |\theta_1 - \theta_2|)\).
S-1.1 Mean square continuity:

1. With respect to time $t$

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

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

$$= K(0, 0) + K(0, 0) - 2K(h, 0)$$

$$= 2(K(0, 0) - K(h, 0))$$

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 $\theta$:

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

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

$$= K(0, 0) + K(0, 0) - 2K(0, \alpha)$$

$$= 2(K(0, 0) - K(0, \alpha))$$

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 $\alpha$.

3. With respect to time $t$ and angle $\theta$:

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

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 \(\alpha\).

### S-1.2 Mean square differentiability

A process \(X(u), u \in \mathbb{R}^d\), is said to be Mean Square Differentiable at \(u_0\) if for any direction \(p\) there exists a process \(L_{u_0}(p)\), linear in \(p\), such that

\[ X(u_0 + p) = X(u_0) + L_{u_0}(p) + R(u_0, p), \]

where \(p \in \mathbb{R}^d\), and \(R(u_0, p)\) satisfies the following

\[ \frac{R(u_0, p)}{||p||} \to 0, \text{ in } L^2, \]

with \(|| \cdot ||\) being the usual Euclidean norm (for details see \[\text{Banerjee and Gelfand (2003)}\]).

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
$(t_0, \theta_0)$ if for any direction $(h, \alpha)$ there exists a process $L_{t_0, \theta_0}(h, \alpha)$, linear in $h, \alpha$, 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 $L_{u_0}(p)$. The following calculation will make the things clear. Following the multivariate Taylor series expansion (using our new metric) we have

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

where $|R(t_0, \theta_0, h, \alpha)| \leq M^* d^2(h, 0, \alpha, 0)$, with $M^* = \max\left\{ \left| \frac{\partial^2 X(t, \theta)}{\partial t^2} \right|, \left| \frac{\partial^2 X(t, \theta)}{\partial t \partial \theta} \right|, \left| \frac{\partial^2 X(t, \theta)}{\partial \theta^2} \right| \right\}$ (using the analogy with multivariate Taylor series expansion in $\mathbb{R}^d$, recall that in the case of $\mathbb{R}^d$, $R(u_0, p) \leq M^* ||p||^2$).

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 $D_T$. The posterior predictive distribution of $y_{T+1}$ given $D_T$ is given by

$$[y_{T+1}|D_T] = \int [y_{T+1}|D_T, x_0, \ldots, x_{T+1}, \beta_f, \beta_g, \sigma_e^2, \sigma_y^2, \sigma_f^2, \sigma_g^2] \times [x_0, \ldots, x_{T+1}, \beta_f, \beta_g, \sigma_e^2, \sigma_y^2, \sigma_f^2, \sigma_g^2|D_T] \times d\beta_f d\beta_g d\sigma_e^2 d\sigma_y^2 d\sigma_f^2 d\sigma_g^2 dx_0 \ldots dx_{T+1}. \quad (32)$$

Thus, once we have a sample realization from the joint posterior
\[ x_0, \ldots, x_{T+1}, \beta_f, \beta_g, \sigma_{\epsilon}^2, \sigma_{\eta}^2, \sigma_{g}^2, \sigma_f^2 | D_T \], we can generate a realization from \( y_{T+1} | D_T \) by simply simulating from \( y_{T+1} | D_T, x_0, \ldots, x_{T+1}, \beta_f, \beta_g, \sigma_{\epsilon}^2, \sigma_{\eta}^2, \sigma_f^2, \sigma_g^2 \), conditional on the realization obtained from the former joint posterior. Observe that the conditional distribution \( y_{T+1} = f(T+1, x_{T+1}) + \epsilon_{T+1} | D_T, x_0, \ldots, x_{T+1}, \beta_f, \beta_g, \sigma_{\epsilon}^2, \sigma_f^2 \) is normal with mean

\[
\mu_{y_{T+1}} = h(T+1, x_{T+1})' \beta_f + s_{f,D_T}(T+1, x_{T+1})' A_{f,D_T}^{-1} (D_T - H_D T \beta_f)
\]

and variance

\[
\sigma_{y_{T+1}}^2 = \sigma_{\epsilon}^2 + \sigma_f^2 (1 - (s_{f,D_T}(T+1, x_{T+1}))' A_{f,D_T}^{-1} s_{f,D_T}(T+1, x_{T+1})).
\]

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

\[
[x_0, x_1, \ldots, x_{T+1}, \beta_f, \beta_g, \sigma_{\epsilon}^2, \sigma_{\eta}^2, \sigma_g^2, \sigma_f^2 | D_T] \\
= \sum_{K_1, \ldots, K_{T+1}} \int [x_0, x_1, \ldots, x_T, x_{T+1}, \beta_f, \beta_g, \sigma_{\epsilon}^2, \sigma_{\eta}^2, \sigma_g^2, \sigma_f^2, g^*(1, x_0), D_z, K_1, \ldots, K_T, K_{T+1} | D_T] \\
\times dg^*(1, x_0) dD_z \\
\times dK_1(1, x_0, \beta_g, \sigma_g^2) \\
\times dK_1(1, x_0, \eta, \sigma_{\eta}^2) \\
\times dK_1(1, x_0, x_T, \beta_f, \sigma_f^2) \\
= \sum_{K_1, \ldots, K_{T+1}} \int [\beta_f][\beta_g][\sigma_{\epsilon}^2][\sigma_{\eta}^2][\sigma_g^2][\sigma_f^2][x_0][g^*(1, x_0)][x_0, \beta_g, \sigma_g^2][D_z][g^*(1, x_0), x_0, \beta_g, \sigma_g^2] \\
[x_1 | g^*(1, x_0), \sigma_{\eta}^2, K_1][K_1 | g^*(1, x_0), \sigma_{\eta}^2][D_T | x_1, \ldots, x_T, \beta_f, \sigma_{\epsilon}^2, \sigma_f^2]
\]
Here we provide the full conditional distributions of the unknowns. In what follows, we shall  
express \( g^*(1,x_0) \mid x_0, \beta_g, \varphi_g^2 \mid D_z \mid g^*(1,x_0), x_0, \beta_g, \sigma_g^2 \) as \( D_z, g^*(1,x_0) \mid x_0, \beta_g, \sigma_g^2 \).

\[
\prod_{t=2}^{T+1} \left[ x_t \mid \beta_g, \sigma_y^2, \sigma_g^2, D_z, x_{t-1}, K_t \right] \prod_{t=2}^{T+1} \left[ K_t \mid \beta_g, \sigma_y^2, \sigma_g^2, D_z, x_{t-1} \right] \text{d}g^*(1,x_0) \text{d}D_z . \tag{35}
\]

In order to obtain MCMC samples from \([x_0, x_1, \ldots, x_{T+1}, \beta_f, \beta_g, \sigma_y^2, \sigma_g^2, \sigma_f^2, \sigma_f^2 | D_T]\), we first carry out MCMC simulations from the joint posterior which is proportional to integrand (35). 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 \([g^*(1,x_0) \mid x_0, \beta_g, \varphi_g^2 \mid D_z \mid g^*(1,x_0), x_0, \beta_g, \sigma_g^2] \) as \([D_z, g^*(1,x_0) \mid x_0, \beta_g, \sigma_g^2] \).
\[
\sigma_y^2, D_z, x_{t-1}) \\
[x_0 \cdots] \propto [x_0|D_z, g^*(1, x_0)|x_0, \beta_g, \sigma_g^2] \\
g^*(1, x_0) \cdots \propto [g^*(1, x_0)|x_0, \beta_g, \sigma_g^2][D_z|g^*(1, x_0), x_0, \beta_g, \sigma_g^2][x_1|g^*(1, x_0), x_0, \sigma_q^2, K_1] \\
 [K_1|g^*(1, x_0), \sigma_q^2] \\
[D_z \cdots] \propto [D_z|g^*(1, x_0), x_0, \beta_g, \sigma_g^2, \sigma_q^2] \prod_{t=2}^{T+1} [x_t|\beta_g, \sigma_g^2, \sigma_q^2, D_z, x_{t-1}, K_t] \prod_{t=2}^{T+1} [K_t|\beta_g, \sigma_g^2, \sigma_q^2, \sigma_y^2] \\
[D_z, x_{t-1}] \\
[x_1 \cdots] \propto [x_1|g^*(1, x_0), \sigma_q^2][D_T|x_1, \ldots, x_T, \beta_f, \sigma_f^2] \\
 [x_2|\beta_g, \sigma_g^2, \sigma_q^2, D_z, x_1, K_2][K_2|\beta_g, \sigma_g^2, \sigma_q^2, D_z, x_1] \\
 [x_{T+1} \cdots] \propto [x_{T+1}|\beta_g, \sigma_g^2, \sigma_q^2, D_z, x_T, K_{T+1}] \\
 [x_{t+1} \cdots] \propto [x_{t+1}|\beta_g, \sigma_g^2, \sigma_q^2, D_z, x_t, x_{t+2}|\beta_g, \sigma_g^2, \sigma_q^2, D_z, x_{t+1}, K_{t+2}][K_{t+2}|\beta_g, \sigma_g^2, \sigma_q^2, \sigma_y^2, D_z, x_{t+1}] [D_T|x_1, \ldots, x_T, \beta_f, \sigma_f^2], 
 t = 1, \ldots, T - 1
\]

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

\[
[K_1|\cdots] \propto [K_1|g^*(1, x_0), \sigma_q^2][x_1|g^*(1, x_0), \beta_g, \sigma_g^2, K_1] \\
[K_t|\cdots] \propto [x_t|\beta_g, \sigma_q^2, D_z, x_{t-1}, K_t][K_t|\beta_g, \sigma_q^2, D_z, x_{t-1}], 
 t = 2, \ldots, T + 1.
\]

**S-2.1.1 Updating $\beta_f$ by Gibbs steps**

The full conditional of $\beta_f$ is a multivariate normal distribution with mean

\[
E[\beta_f|\cdots] = \left\{ H_{D_T'}(\sigma_f^2 A_{f,D_T} + \sigma_f^2 I)^{-1} H_{D_T} + \Sigma_{f,0} \right\}^{-1} \\
\times \left\{ H_{D_T'}(\sigma_f^2 A_{f,D_T} + \sigma_f^2 I)^{-1} D_T + \Sigma_{f,0}^{-1} \beta_{f,0} \right\}
\]

37
and variance

\[ V[\beta_f | \cdots] = \{H_{D_T}(\sigma_f^2 A_{f,D_T} + \sigma^2 \epsilon I)^{-1} H_{D_T} + \Sigma_{\beta_f,0}\}^{-1}. \] (51)

S-2.1.2 Updating \( \beta_g \)

We first explicitly write down the right hand side of (37).

\[
\begin{aligned}
[\beta_g][D_z, g^*(1, x_0)|x_0, \beta_g] & \prod_{t=2}^{T+1} [x_t|\beta_g, \sigma^2_{\eta}, D_z, x_{t-1}, K_t] \prod_{t=2}^{T+1} [K_t|\beta_g, \sigma^2_{\eta}, D_z, x_{t-1}] \\
& \propto \exp \left( -\frac{1}{2}(\beta_g - \beta_{g,0})' \Sigma_{\beta_g,0}^{-1}(\beta_g - \beta_{g,0}) \right) \\
& \exp \left( -\frac{1}{2}[(D_z, g^*)' - (H_{D_z} \beta_g, h'(1, x_0)')]' A_{D_z, g^*(1, x_0)}^{-1} [(D_z, g^*)' - (H_{D_z} \beta_g, h'(1, x_0))]' \right) \\
& \exp \left\{ -\sum_{i=2}^{T+1} \frac{1}{2\sigma^2_{x_t}} (x_t + 2\pi K_t - \mu_x)^2 \right\} \prod_{t=2}^{T+1} I_{[0,2\pi]}(x_t) \right) \right) \] (52)

Observe that the denominator of \([x_t|\beta_g, \sigma^2_{\eta}, D_z, x_{t-1}, K_t]\) cancels with the density of \([K_t|\beta_g, \sigma^2_{\eta}, D_z, x_{t-1}]\) for each \( t = 2, \ldots, T + 1 \). Also we note that the indicator function does not involve \( \beta_g \) for all \( t = 2, \ldots, T + 1 \). Therefore, after simplifying the exponent terms and ignoring the indicator function we can write

\[
[\beta_g | \cdots] \propto \exp \left\{ -\frac{1}{2}(\beta_g - \mu_{\beta_g})' \Sigma_{\beta_g}^{-1}(\beta_g - \mu_{\beta_g}) \right\}, \] (53)

where
\[ \mu_{\beta_g} = E[\beta_g | \cdots] = \left\{ \Sigma_{\beta_g,0}^{-1} + \frac{1}{\sigma_g^2} [H_{D_z}^t, h(1, x_0)] A_{D_z, g^*(1, x_0)}^{-1} [H_{D_z}^t, h(1, x_0)]^t \right\}^{-1} \\
+ \sum_{t=1}^{T} \left( \frac{H_{D_z}^t A_{g, D_z}^{-1} s_{g, D_z}(t + 1, x_t) - h(t + 1, x_t)}{\sigma_{x_t}^2} \right) \left( \frac{H_{D_z}^t A_{g, D_z}^{-1} s_{g, D_z}(t + 1, x_t) - h(t + 1, x_t)}{\sigma_{x_t}^2} \right)^t \right\}^{-1} \\
\left\{ \Sigma_{\beta_g,0}^{-1} + \frac{1}{\sigma_g^2} [H_{D_z}^t, h(1, x_0)] A_{D_z, g^*(1, x_0)}^{-1} [D_z, g^*(1, x_0)] \right\} \\
+ \sum_{t=1}^{T} \left( x_{t+1} + 2\pi K_{t+1} - s_{g, D_z}(t + 1, x_t)^t A_{g, D_z}^{-1} D_z (h(t + 1, x_t) - H_{D_z}^t A_{g, D_z}^{-1} s_{g, D_z}(t + 1, x_t)) \right) \right\} \tag{54} \]

and

\[ \Sigma_{\beta_g} = V[\beta_g | \cdots] = \left\{ \Sigma_{\beta_g,0}^{-1} + \frac{1}{\sigma_g^2} [H_{D_z}^t, h(1, x_0)] A_{D_z, g^*(1, x_0)}^{-1} [H_{D_z}^t, h(1, x_0)]^t \right\}^{-1} \\
+ \sum_{t=1}^{T} \left( \frac{H_{D_z}^t A_{g, D_z}^{-1} s_{g, D_z}(t + 1, x_t) - h(t + 1, x_t)}{\sigma_{x_t}^2} \right) \left( \frac{H_{D_z}^t A_{g, D_z}^{-1} s_{g, D_z}(t + 1, x_t) - h(t + 1, x_t)}{\sigma_{x_t}^2} \right)^t \right\}^{-1} \tag{55} \]

Hence \([\beta_g | \cdots]\) follows a tri-variate normal distribution with mean and variance \( \mu_{\beta_g} \) and \( \Sigma_{\beta_g} \), respectively, and therefore, we update \( \beta_g \) using Gibbs sampling.

**S-2.1.3 Updating \( \sigma_f^2 \) and \( \sigma_g^2 \)**

The mathematical form of the full conditional distributions of \( \sigma_f^2 \) and \( \sigma_g^2 \) are not tractable, so we update \( \sigma_f^2 \) and \( \sigma_g^2 \) by random walk Metropolis-Hastings steps.
S-2.1.4 Updating $\sigma^2$

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

S-2.1.5 Updating $\sigma^2_{\eta}$

For full conditional distribution of $\sigma^2_{\eta}$ right hand side of (40) simplifies a bit in the sense that the denominator of $[x_t|\beta_g, \sigma^2_{\eta}, D_z, x_{t-1}, K_t]$ cancels with the density of $[K_t|\beta_g, \sigma^2_{\eta}, D_z, x_{t-1}]$ for $t = 2, \ldots, T+1$, and the denominator of $[x_1|g^*(1, x_0), \beta_g, \sigma^2_{\eta}, K_1]$ cancels with the density of $[K_1|g^*(1, x_0), \beta_g, \sigma^2_{\eta}]$, which, in turn, gives the following form:

$$[\sigma^2_{\eta}] \propto [\sigma^2_{\eta}] \exp \left\{ - \sum_{i=2}^{T+1} \frac{1}{2\sigma^2_{x_t}} (x_t + 2\pi K_t - \mu_{x_t})^2 \right\} \exp \left\{ - \frac{1}{2\sigma^2_{\eta}} (x_1 + 2\pi K_1 - g^*)^2 \right\}. \quad (56)$$

However, the above equation does not have a closed form; hence, for updating $\sigma^2_{\eta}$ 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 (43), after cancelling the denominator of $[x_1|g^*(1, x_0), x_0, \beta_g, \sigma^2, K_1]$ with the density of $[K_1|g^*(1, x_0), x_0, \beta_g, \sigma^2]$, and ignoring the indicator function on $x_0$, reduces to

$$ [g^*(1, x_0) \cdots] \propto [g^*(1, x_0)|x_0, \beta_g][D_z|g^*(1, x_0), x_0, \beta_g] \exp \left\{ -\frac{1}{2\sigma^2}(x_1 + 2\pi K_1 - g^*)^2 \right\}. $$

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

$$ [g^*(1, x_0) \cdots] \propto \exp \left\{ -\frac{1}{2\gamma^2}(g^* - \nu_g)^2 \right\}, \quad (57) $$

where

$$ \nu_g = E[g^*(1, x_0) \cdots] = \left\{ \frac{1}{\sigma^2} + \frac{1}{\sigma^2}(1 + s_{g,D_z}(1, x_0)'\Sigma^{-1}_{g,D_z} s_{g,D_z}(1, x_0)) \right\}^{-1} \left\{ \frac{x_1 + 2\pi K_1}{\sigma^2} + \frac{1}{\sigma^2}(h(1, x_0)'\beta_g + s_{g,D_z}'\Sigma^{-1}_{g,D_z} D^*_z) \right\}, \quad (58) $$

and

$$ \gamma^2 = V[g^*(1, x_0) \cdots] = \left\{ \frac{1}{\sigma^2} + \frac{1}{\sigma^2}(1 + s_{g,D_z}(1, x_0)'\Sigma^{-1}_{g,D_z} s_{g,D_z}(1, x_0)) \right\}, \quad (59) $$

with

$$ D^*_z = D_z - H_{D_z}\beta_g + h(1, x_0)'\beta_g s_{g,D_z}, \quad (60) $$

and

$$ \Sigma_{g,D_z} = A_{g,D_z} - s_{g,D_z}(1, x_0)s_{g,D_z}(1, x_0)'. \quad (61) $$
Hence $[g^*|\cdots]$ follows a normal distribution with mean $\nu_g$ and variance $\gamma_g$. 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 $[x_t|\beta_g, \sigma_g^2, D_z, x_{t-1}, K_t]$ cancels with the density of $[K_t|\beta_g, \sigma_g^2, D_z, x_{t-1}]$ for each $t = 2, \ldots, T + 1$. After simplification it turns out that the full conditional distribution of $D_z$ is an $n$-variate normal with mean

\[
E(D_z|\cdots) = \left\{ \frac{\Sigma_{g,D_z}}{\sigma_g^2} + A_{g,D_z}^{-1} \left( \sum_{t=1}^{T} s_{g,D_z}(t+1,x_t)s'_{g,D_z}(t+1,x_t) \right) A_{g,D_z}^{-1} \right\}^{-1} \times \left\{ \frac{\Sigma_{g,D_z} \mu_{g,D_z}}{\sigma_g^2} + A_{g,D_z}^{-1} \right\}
\]

\[
\sum_{t=1}^{T} s_{g,D_z}(t+1,x_t) \left\{ x_{t+1} + 2\pi K_{t+1} - \beta'_g(h(1,t+1,x_{t+1}) - H'_{D_z} A_{g,D_z}^{-1} s_{g,D_z}(t+1,x_{t+1})) \right\} \sigma_{x_t}^2
\]

and covariance matrix

\[
V(D_z|\cdots) = \left\{ \frac{\Sigma_{g,D_z}}{\sigma_g^2} + A_{g,D_z}^{-1} \left( \sum_{t=1}^{T} s_{g,D_z}(t+1,x_t)s'_{g,D_z}(t+1,x_t) \right) A_{g,D_z}^{-1} \right\}^{-1}.
\]

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 $[45]$ as follows:
\[
[x_1|\cdots] \propto \frac{1}{\sqrt{2\pi \sigma_\eta}} \exp \left( -\frac{1}{2\sigma_\eta^2} (x_1 + 2\pi K_1 - g^*)^2 \right) I_{[0,2\pi]}(x_1) \\
\exp \left\{ -\frac{1}{2} (D_T - \mu_{y_t})' \Sigma_{y_t}^{-1} (D_T - \mu_{y_t}) \right\} \\
\frac{1}{\sqrt{2\pi \sigma_{x_2}}} \exp \left( -\frac{1}{2\sigma_{x_2}^2} (x_2 + 2\pi K_2 - \mu_{x_2})^2 \right),
\]

(64)

where \(\mu_{y_t}\) and \(\Sigma_{y_t}\) are given by (10) and (11) of MB. Here we note that the denominator of \([x_2|\beta_g, \sigma_\eta^2, D_z, x_1, K_2]\) cancels with \([K_2|\beta_g, \sigma_\eta^2, D_z, x_1]\). 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

\[
[x_1|\cdots] \propto \frac{1}{\sqrt{2\pi \sigma_\eta}} \exp \left( -\frac{1}{2\sigma_\eta^2} (x_1 + 2\pi K_1 - g^*)^2 \right) I_{[0,2\pi]}(x_1) \\
\exp \left\{ -\frac{1}{2} (D_T - \mu_{y_t})' \Sigma_{y_t}^{-1} (D_T - \mu_{y_t}) \right\} \\
\frac{1}{\sqrt{2\pi \sigma_{x_2}}} \exp \left( -\frac{1}{2\sigma_{x_2}^2} (x_2 + 2\pi K_2 - \mu_{x_2})^2 \right),
\]

(65)

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

\[
[x_{t+1} \mid \cdots] \propto \frac{1}{\sqrt{2\pi}\sigma_{x_{t+1}}} \exp \left( -\frac{1}{2\sigma_{x_{t+1}}^2} (x_{t+1} + 2\pi K_{t+1} - \mu_{x_{t+1}})^2 \right) I_{[0,2\pi]}(x_{t+1})
\]

\[
\times \left( \Phi \left( \frac{2\pi (K_{t+1} + 1) - \mu_{x_{t+1}}}{\sigma_{x_{t+1}}} \right) - \Phi \left( \frac{2\pi K_{t+1} - \mu_{x_{t+1}}}{\sigma_{x_{t+1}}} \right) \right)
\]

\[
\times \frac{1}{\sqrt{2\pi}\sigma_{x_{t+2}}} \exp \left( -\frac{1}{2\sigma_{x_{t+2}}^2} (x_{t+2} + 2\pi K_{t+2} - \mu_{x_{t+2}})^2 \right)
\]

\[
\exp \left\{ -\frac{1}{2} (D_T - \mu_{y_T})' \Sigma_{y_T}^{-1} (D_T - \mu_{y_T}) \right\}. \tag{66}
\]

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 (66) as

\[
[x_{t+1} \mid \cdots] \propto \frac{1}{\sqrt{2\pi}\sigma_{x_{t+1}}} \exp \left( -\frac{1}{2\sigma_{x_{t+1}}^2} (x_{t+1} + 2\pi K_{t+1} - \mu_{x_{t+1}})^2 \right) I_{[0,2\pi]}(x_{t+1})
\]

\[
\times \frac{1}{\sqrt{2\pi}\sigma_{x_{t+2}}} \exp \left( -\frac{1}{2\sigma_{x_{t+2}}^2} (x_{t+2} + 2\pi K_{t+2} - \mu_{x_{t+2}})^2 \right)
\]

\[
\exp \left\{ -\frac{1}{2} (D_T - \mu_{y_T})' \Sigma_{y_T}^{-1} (D_T - \mu_{y_T}) \right\}. \tag{67}
\]

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

$$\mu_{x_{T+1}} = h(1, x_T)' \beta_g + s_{g,D_z}(T + 1, x_T)' A_{g,D_z}^{-1} (D_z - H_{D,1} \beta_g)$$

(68)

and

$$\sigma^2_{x_{T+1}} = \sigma^2_\eta + \sigma^2_g \{1 - s_{g,D_z}(T + 1, x_T)' A_{g,D_z}^{-1} s_{g,D_z}(T + 1, x_T)\}.$$ (69)

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

$$[K_1| \cdots] \propto \frac{1}{\sqrt{2\pi \sigma_\eta}} \exp \left(-\frac{1}{2\sigma^2_\eta} (x_1 + 2\pi K_1 - g^*)^2\right) I_{\{\ldots, -1, 0, 1, \ldots\}}(K_1),$$ (70)

and similarly the full conditional distribution of $K_t$ becomes

$$[K_t| \cdots] \propto \frac{1}{\sqrt{2\pi \sigma_{x_t}}} \exp \left(-\frac{1}{2\sigma^2_{x_t}} (x_t + 2\pi K_t - \mu_{x_t})^2\right) I_{\{\ldots, -1, 0, 1, \ldots\}}(K_t),$$ (71)

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.

Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*. Springer, Boston.

Banerjee, S. and Gelfand, A. E. (2003). On Smoothness Properties of Spatial Processes. *Journal of Multivariate Analysis*, 84, 85–100.

Bhattacharya, S. (2007). A Simulation Approach to Bayesian Emulation of Complex Dynamic Computer Models. *Bayesian Analysis*, 2, 783–816.

Bickel, J. P. and Doksum, A. K. (2007). *Mathematical Statistics*. Number 2nd Edition in Volume I. Pearson Prentice Hall.

Box, E. P. G. and Tiao, C. G. (1973). *Bayesian Inference in Statistical Analysis*. Addison Wesley Publishing Co.

Dufour, M. J. and Roy, R. (1976). On Spectral Estimation for a Homogeneous Random Process on the Circle. *Stochastic Processes and their Applications*, 4, 107–120.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.

Ghosh, A., Mukhopadhyay, S., Roy, S., and Bhattacharya, S. (2014). Bayesian Inference in Nonparametric Dynamic State Space Models. *Statistical Methodology*, 21, 35–48.

Gneiting, T. (1998). Simple Tests for the Validity of Correlation Function Models on the Circle. *Statistics and Probability Letters*, 39, 119–122.
Holzmann, H., Munk, A., Suster, M., and Zucchini, W. (2006). Hidden Markov Models for Circular and Linear-Circular Time Series. *Environmental and Ecological Statistics*, **13**, 325–347.

Jammalamadaka, R. S. and Lund, J. U. (2006). The Effect of Wind Direction on Ozone Levels: A Case Study. *Environmental and Ecological Statistics*, **13**, 287–298.

Liu, J. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, New York.

Mazumder, S. and Bhattacharya, S. (2014a). Bayesian Nonparametric Dynamic State-Space Modeling with Circular Latent States. Submitted.

Mazumder, S. and Bhattacharya, S. (2014b). Supplement to “Bayesian Nonparametric Dynamic State-Space Modeling with Circular Latent States”. Submitted.

Ravindran, P. and Ghosh, S. (2011). Bayesian Analysis of Circular Data Using Wrapped Distributions. *Journal of Statistical Theory and Practice*, **4**, 1–20.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.

Shafie, K., Siegmund, D., Sigal, B., and Worsley, K. J. (2003). Rotation Space Random Fields with an Application to fMRI Data. *Annals of Statistics*, **31**, 1732–1771.

Shumway, R. H. and Stoffer, D. S. (2011). *Time Series Analysis and Its Applications*. Springer-Verlag, New York.