Nonparametric Nonstationary Modeling of Spatio-Temporal Data Through State Space Approach

Suman Guha* and Sourabh Bhattacharya†

May 27, 2014

Abstract

Recently, there has been an upsurge of the number of articles on spatio-temporal modeling in statistical journals. Many of them focus on building good nonstationary spatio-temporal models. In this article, we introduce a state space based nonparametric nonstationary model for the analysis of spatio-temporal data. We consider that there are some fixed spatial locations (generally called the monitoring sites) and that the data have been observed at those locations over a period of time. To model the data we assume that the data generating process is driven by some latent spatio-temporal process, which itself is evolving with time in some unknown way. We model this evolutionary transformation via compositions of a Gaussian process and also model the unknown functional dependence between the data generating process and the latent spatio-temporal process (observational transformation) by another Gaussian process. We investigate this model in detail, explore the covariance structure and formulate a fully Bayesian method for inference and prediction. Finally, we apply our nonparametric model on two simulated data sets and a real data set and establish its effectiveness.

Keywords: Observational equation; Evolutionary equation; Gaussian process; State-space model; MCMC; Gibbs sampler; Posterior predictive distribution.

AMS 2000 Subject Classification: Primary 62M20, 62M30; Secondary 60G15.

1 Introduction

Spatio-temporal modeling has received much attention in recent years. Particularly, the rise in global temperature being a major environmental concern, scientists are now taking keen interest in the study of the dynamics of such climatic spatio-temporal processes\[19, 28, 37, 39, 41, 46\]. Another closely related class of spatio-temporal processes, that are also of much importance to climatologists, are daily rainfall and precipitation (like mist, snowfall etc.) across a region. Apart from climatology, many important spatio-temporal processes are also associated with different subfields of environmental and ecological science. To mention a few, studies on ground level concentration of ozone, SO$_2$, NO$_2$ and PM, species distribution over a region, change in land usage pattern over time, etc. Other than these areas, spatio-temporal models are also useful in geostatistics,
hydrology, astrophysics, social science, archaeology, systems biology, and many more. So, knowledge of spatio-temporal modeling is becoming increasingly necessary for better understanding of a broad range of subjects. Hence, it is of no wonder that practitioners from distant fields are working together to develop new and effective spatio-temporal models. In what follows, we review some such existing models, discuss some issues related to them and then propose a novel nonparametric model. Then we devote the rest of the article to the development and exploration of the proposed model. Finally, some simulated and real data analysis results are presented which shows that the proposed model is particularly useful for analysis of nonstationary spatial time series data. The proofs of all our mathematical results are deferred to the Appendix.

2 Existing approaches for spatio-temporal data

The term spatio-temporal data covers many different types of data. But here we consider only one specific type of spatio-temporal data. We consider that there are some arbitrary spatial locations \( s_1, s_2, s_3, \cdots, s_n \) and that the data have been observed at each of those spatial locations at times \( t = 1, 2, 3, \cdots, T \). Such data sets are very common in the context of environmental science. One important example is ground level ozone data.

An early approach for analysis of this type of data was based on stationary Gaussian process. But this approach to modeling did not turn out to be adequate. Since time differs intrinsically from space in that time moves only forward, while there may not be any preferred direction in space, time cannot be treated simply as an additional co-ordinate attached with the spatial co-ordinates. Another drawback of this approach is stationarity which is seldom satisfied by real, physical processes. So, alternative approaches to modeling spatio-temporal data were clearly necessary.

Briggs (1968)\[5\] proposed a simple method for constructing spatio-temporal models, based on a purely spatial model. Later, Cox and Isham (1988)\[7\] adopted a similar approach for modeling rainfall pattern. Roughly, their idea was to start with a purely spatial stationary covariance kernel \( C_s(h) \), where \( h \in \mathbb{R}^2 \) and construct a spatio-temporal stationary covariance kernel \( C(h, u) = EC_s(h - uv) \), where \( v \in \mathbb{R}^2 \) is a random velocity vector and \( u \) represents the time lag. In case the velocity vector is nonrandom, this model reduces to the classical frozen field model (see page 428 of \[20\]), which are useful for modeling environmental processes that are under the influence of prevailing winds or ocean currents. Again, this model suffers from the oversimplified assumption of stationarity.

Sampson and Guttorp (1992)\[38\] proposed a nonparametric nonstationary model for analyzing spatial data, based on spatial deformation. Later, they extended it to the spatio-temporal setup\[6, 23\]. Schmidt and O’Hagan (2003)\[40\] gave a Bayesian formulation of the model in the spatial setup. One basic problem with this approach is the requirement of replicates of the data, which is rarely available in practice. Only very recently, Anderes and Stein (2008)\[2\] came up with some idea on estimating such deformation model from a single realization. Still, they considered only the spatial setup and implementing the model in real data problems requires solving a large and quite difficult optimization problem. (See \[43\]).

A completely different approach to modeling spatio-temporal data is the convolution approach proposed by Higdon et al. (1999)\[26\]. Initially, they proposed it for purely spatial processes and later extended it to the spatio-temporal setup\[24, 25\]. Several extensions and modifications of the convolution approach are adopted in \[31, 32, 48\]. The basic idea is to start with a white noise process in space (space-time) and create a nonstationary spatial (spatio-temporal) process by convolving it with a spatially varying kernel. A different type of kernel convolution based non stationary spatial process was formed by Fuentes and Smith (2001)\[18\]. They convolved a family of spatially varying locally stationary processes with a spatial kernel to build a nonstationary process.
A more physically motivated approach to modeling the kind of spatio-temporal data considered here, is based on the dynamic linear model (DLM) introduced by West and Harrison (1997)\cite{47}. Such models are called dynamic spatio-temporal (DSTM) models. Stroud et al. (2001)\cite{42} proposed a dynamic spatio-temporal model for the analysis of tropical rainfall and Atlantic ocean temperature. Huerta et al. (2004)\cite{27} proposed a seasonal dynamic spatio-temporal model for the analysis of ozone levels. A more flexible dynamic spatio-temporal model with spatially varying state vector is proposed in Banerjee et al. (2005)\cite{4}. Several other authors studied these models in the context of dimension reductions etc.\cite{30}. Although DSTMs are used to incorporate the covariate information in spatio-temporal modeling, these models can also be used in the absence of covariates. The basic idea behind the dynamic approach is to model the observed spatio-temporal process as a linear state space model, where the coefficients associated with different covariates constitute the state vector.

Direct construction of nonstationary covariance functions is yet another approach. While Paciorek and Schervish (2006)\cite{34} directly construct a nonstationary covariance function in the spatial setup, Fuentes (2001, 2002)\cite{14,17} proposed a spectral method.

All the above mentioned approaches mainly focus on achieving nonstationarity. But nonseparability in space-time is also another important aspect in the context of space time modeling. Roughly, nonseparability in space-time means that space effect and time effect interact in the formation of the process; to rephrase, space effect depends on time effect and vice versa. Although this is a highly realistic assumption, most of the spatio-temporal models existing in the literature assume separability. Some works on nonseparable models can be found in \cite{8,22,15,16}. Among them, Cressie and Huang (1999)\cite{8} and Gneiting (2002)\cite{22} considered nonseparability in the stationary setup while the others have developed spatio-temporal processes, which are both nonseparable and nonstationary.

Apart from the above mentioned approaches there are many more works in the context of spatio-temporal modeling. We attempted to provide only a broad overview and it is by no way exhaustive.

3 Our proposed spatio-temporal process

First recall, that there are some arbitrary spatial locations $s_1, s_2, s_3 \ldots, s_n$ and the data $y(s_i, t)$ have been observed at those spatial locations at times $t = 1, 2, 3, \ldots, T$. We assume that the observed spatio-temporal process $Y(s, t)$ is driven by an unobserved spatio-temporal process $X(s, t)$ which itself is evolving in time in some unknown way. So, we model them via the state space approach but instead of considering any known form of observational and evolutionary equations we allow them to be unknown.

We propose a nonparametric state space based model, which can even capture the nonlinear evolution flexibly. We refer to it as a nonparametric state space based spatio-temporal model. Our model has the following form:

\begin{align}
Y(s, t) &= f(X(s, t)) + \epsilon(s, t), \\
X(s, t) &= g(X(s, t - 1)) + \eta(s, t);
\end{align}

where $s \in \mathbb{R}^2$ and $t \in \{1, 2, 3, \ldots\}; X(\cdot, 0)$ is a spatial Gaussian process with appropriate parameters; $\epsilon(\cdot, t)$ and $\eta(\cdot, t)$ are temporally independent and identically distributed spatial Gaussian processes, and $g(\cdot)$ and $f(\cdot)$ are Gaussian processes on $\mathbb{R}$. They are all independent of each other.

To elaborate, let us consider a hierarchical break up of our model. Suppose, that the observed process depends on the unobserved process through some unknown continuous function $f(\cdot)$. Then, for the purpose
of Bayesian inference we must put some prior on \( f(\cdot) \). Hence, we put a Gaussian process prior on \( f(\cdot) \). Similarly we put a Gaussian process prior on \( g(\cdot) \). This approach is able to capture arbitrary functional dependence between the observed and the latent variables, hence the term nonparametric.

Finally, to complete the model specification, we need to describe the parameters associated with the Gaussian processes. We assume that the Gaussian process \( f(x) \) has mean function of the form \( \beta_0 f + \beta_1 f x \) (where \( \beta_0, \beta_1 \) are suitable parameters) and has covariance kernel of the form \( c_f(x_1, x_2) = \gamma(\|x_1 - x_2\|) \), where \( \gamma \) is a positive definite function. It consists of parameters that determine the smoothness of the sample path of \( f(\cdot) \). Moreover \( \gamma \) is such that the centered Gaussian process associated with it has continuous sample paths. Thus, we consider isotropic covariance functions with suitable smoothness properties. Typical examples of \( \gamma \) are exponential, powered exponential, Gaussian, Matérn etc. (see, for example, Table 2.1 of [3] for other examples of such covariance kernels). Similarly, parameters \( \beta_0 g, \beta_1 g \) and \( c_g(x_1, x_2) \) are associated with the Gaussian process \( g(x) \).

The zero mean Gaussian processes associated with the noise variables have covariance kernels \( c_\epsilon(s, s') \) and \( c_\eta(s, s') \) which are also of the form as discussed above.

Regarding the Gaussian process associated with \( X(\cdot, 0) \), we assume a mean process of the form \( \mu_0(\cdot) \) and isotropic covariance kernel \( c_0(\cdot, \cdot) \). For convenience we introduce separate symbols for the mean vector and the covariance matrix associated with \( (X(s_1, 0), X(s_2, 0), \cdots X(s_n, 0)) \) where \( s_1, s_2, s_3, \cdots, s_n \) are the spatial locations where the data are observed. We denote them by \( \mu_0 \) and \( \Sigma_0 \) respectively.

From (1) and (2) it is not difficult to see that our model boils down to simple DSTM with no covariate if the process means associated with the Gaussian processes \( f(\cdot) \) and \( g(\cdot) \) are given by \( \beta_0 f = \beta_0 g = 0, \beta_1 f = \beta_1 g = 1 \), and the process variance (denoted by \( \sigma_f^2 \) and \( \sigma_g^2 \)) become 0. The model equations (1) and (2) then reduce to the following forms

\[
Y(s, t) = X(s, t) + \epsilon(s, t), \tag{3}
\]

\[
X(s, t) = X(s, t - 1) + \eta(s, t); \tag{4}
\]

where \( s \in \mathbb{R}^2 \) and \( t \in \{1, 2, 3, \ldots\} \). Although we develop our Gaussian process based spatio-temporal model for equi-spaced time points, simple modification of this model can handle non equi-spaced time points as well. However, for the sake of simplicity and brevity in this article we will confine ourselves only within the framework of equi-spaced time points.

### 3.1 Some measurability and existential issues

Before we proceed to explore the properties of our model, we need to ensure that a family of valid spatio-temporal stochastic processes is induced by the proposed model. Only then physical processes can be modeled by it. In general such issues are almost always trivially satisfied and so, never discussed in detail. But in this case we need to show that \( f(X(s_1, t)), f(X(s_2, t)), \cdots, f(X(s_n, t)) \) are jointly measurable for any \( n \) and any set of spatial locations \( s_1, s_2, \cdots, s_n \), and this is not a trivial problem. The difficulty is that when \( f \) and \( X \) both are random, \( f(X) \) need not be a measurable or valid random variable. It is the sample path continuity of \( f(\cdot) \), which compels \( f(X(s, t)) \) to be measurable. The following theorem establishes this mathematically.

**Theorem 3.1.** The model defines a family of valid (measurable) spatio-temporal processes on \( \mathbb{R}^2 \times \mathbb{Z}^+ \).

Once it is established that the proposed model induces a family of valid spatio-temporal processes, we look into some important properties like the joint distributions of state variables and observed variables, covariance structure of the observed process etc. An important point is that although we develop our model considering the space is \( \mathbb{R}^2 \) all the results go through for \( \mathbb{R}^d \) (\( d > 2 \)) as well.
3.2 Joint distribution of the variables

It is of importance to derive the joint distribution of the observed variables. As our model is based on an implicit hierarchical structure, the joint distribution of the observed variables is non-Gaussian. But before going into the details, we need to derive the joint distribution of the state variables, which will also be required for our MCMC based posterior inference.

Theorem 3.2. Suppose that the spatio-temporal process is observed at locations $s_1, s_2, s_3 \cdots, s_n$ for times $t = 1, 2, 3, \cdots, T$. Then the joint distribution of the state variables is non-Gaussian and has the pdf

$$
\frac{1}{(2\pi)^{\frac{nT}{2}} |\Sigma_0|^{\frac{1}{2}}} \exp \left[ -\frac{1}{2} \begin{pmatrix} x(s_1, 0) - \mu_{01} \\ x(s_2, 0) - \mu_{02} \\ \vdots \\ x(s_n, 0) - \mu_{0n} \end{pmatrix}' \Sigma_0^{-1} \begin{pmatrix} x(s_1, 0) - \mu_{01} \\ x(s_2, 0) - \mu_{02} \\ \vdots \\ x(s_n, 0) - \mu_{0n} \end{pmatrix} \right]
$$

where $\mu_0 = (\mu_{01}, \mu_{02}, \cdots, \mu_{0n})'$ and $\Sigma_0$ are already defined to be the mean vector and the covariance matrix of $(X(s_1, 0), X(s_2, 0), \ldots, X(s_n, 0))$, and

$$
\Sigma = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \bigotimes \Sigma_{\eta} + \Sigma
$$

where the elements of $\Sigma_{\eta}$ are obtained from the purely spatial covariance function $c_\eta$ and the elements of $\Sigma$ are obtained from the covariance function $c_\eta$ in the following way

the $(i, j)$th entry of $\Sigma_{\eta}$ is $c_\eta(s_i, s_j)$ and the $((t_1 - 1)n + i, (t_2 - 1)n + j)$th entry of $\Sigma$ is $c_\eta(x(s_i, t_1 - 1), x(s_j, t_2 - 1))$ where $1 \leq t_1, t_2 \leq T$ and $1 \leq i, j \leq n$.

Although the above density function has physical resemblance with a Gaussian density, the involvement of $x(s_i, t)$ in $\Sigma$ renders it non-Gaussian. In the extreme case when the process variance $c_\eta(0, 0) (= \sigma_\eta^2)$ of the Gaussian process $g(\cdot)$ is 0, $\Sigma$ becomes a block diagonal matrix with identical blocks and the joint density becomes Gaussian. In the formation of $\Sigma$ the component $I_{T \times T} \bigotimes \Sigma_{\eta}$ corresponds to linear evolution and Gaussianity whereas the component $\Sigma$ corresponds to nonlinear evolution and non-Gaussianity.

Moreover, it is clear from the form of the density function that the temporal aspect is imposed through both the location function and the scale function associated with the latent process, making it more flexible.

The interesting property that the joint density of the observed variables is also non-Gaussian, follows from non-Gaussianity of the joint distribution of the latent states. We have the following theorem in this regard.

Theorem 3.3. Suppose that the spatio-temporal process is being observed at locations $s_1, s_2, s_3 \cdots, s_n$ for times $t = 1, 2, 3, \cdots, T$. Then the following hold true:

(a) The joint distribution of the observed variables is a Gaussian mixture and has the following density
Theorem 3.3. (a) The observed spatio-temporal process \( Y(s, t) \) is light tailed in the sense that all the coordinate variables have finite variance.
(b) The covariance function \( c_g(s, t, (s^*, t^*)) \) of the observed spatio-temporal process is nonstationary and nonseparable for any \( s, s^*, t, t^* \).

Although we are able to show that our method yields nonstationary and nonseparable covariance function, it has an obvious drawback that no closed form expression for the covariance function is available. This problem, however exists in many other approaches including the convolution models, DSTM models etc., where a closed form covariance function is available only in very few special cases. Actually, the direct construction of nonstationary covariance function is the only approach that yields a closed form covariance function.

Moreover, from the prediction point of view, this is not a serious problem since what we need are the posterior predictive distributions at un-monitored locations at arbitrary points of time, which do not require closed form expressions of covariance functions.

However, we still attempted to derive some closed form expression, and are partially successful in the sense that when our model is approximately linear (in some suitable sense to be described later), the covariance function is approximately a geometric function of time lag.
3.3.1 Approximate form of the covariance function

Here, in Theorem 3.5 we show that if the process variances \((\sigma^2_g)\) and \((\sigma^2_f)\) are small then under some additional assumptions, the covariance function is approximately a geometric function of time lag. This result is mainly of theoretical interest and has nothing to do with the application of our model.

**Theorem 3.5.** Assume that \(|\beta_{1g}| < 1\). Then for given \(\epsilon'' > 0\) arbitrarily small, \(\exists \delta > 0\) such that for \(0 < \sigma^2_g, \sigma^2_f < \delta\) the covariance between \(Y(s, t)\) and \(Y(s^*, t^*)\), denoted by \(c_{y}(s, t), (s^*, t^*)\) is of following form

\[
\beta_{1g}^{t-t^*} \left[ c_0(s, s^*) + \left( \frac{1 - \beta_{1g}^{2(t^*+1)}}{1 - \beta_{1g}^2} \right) c_{\eta}(s, s^*) \right] - \epsilon'' \leq c_y((s, t), (s^*, t^*)) \\
\leq \beta_{1g}^{t-t^*} \left[ c_0(s, s^*) + \left( \frac{1 - \beta_{1g}^{2(t^*+1)}}{1 - \beta_{1g}^2} \right) c_{\eta}(s, s^*) \right] + \epsilon''.
\]

3.4 Sample path properties

Until now we have explored the finite dimensional distributions and properties of the spatio-temporal process \(Y(s, t)\). But finite dimensional properties alone are not sufficient to characterize any arbitrary general stochastic process. Two stochastic processes with completely different sample path behaviour may have identical finite dimensional distributions and properties. We demonstrate this through the following simple example (see also [?]):

Let us consider two spatio-temporal processes \(Y(s, t)(\omega)\) and \(Y^*(s, t)(\omega)\) defined on the same probability space \(\Omega = [0, 1]^2\) in the following way: \(Y(s, t)(\omega) = 0\) for all \(s, t, \omega\), and \(Y^*(s, t)(\omega) = 1\) for all \(s = \omega\) and \(= 0\) otherwise. Then one can show that for any fixed \(t\), \(Y(s, t)(\omega)\) has continuous sample path with probability 1 whereas \(Y^*(s, t)(\omega)\) has discontinuous sample path with probability 1.

Now, we study the path properties of the proposed spatio-temporal process. The first part of the following theorem states that the observed spatio-temporal process has continuous sample path almost surely and in the second part it says that under additional smoothness assumptions regarding the covariance functions, the observed spatio-temporal process will have smooth sample paths almost surely.

**Theorem 3.6.** (a) The spatio-temporal process \(Y(s, t)\) has continuous sample paths with probability 1.

(b) Assume that the covariance functions \(c_f(\cdot, \cdot), c_g(\cdot, \cdot), c_e(\cdot, \cdot), c_{\eta}(\cdot, \cdot), c_0(\cdot, \cdot)\) satisfy the additional smoothness assumptions that centered Gaussian processes with these covariance functions will have almost surely \(k\) times differentiable sample paths. Then the spatio-temporal process \(Y(s, t)\) will also have almost surely \(k\) times differentiable sample paths.

The first part of the above theorem states that almost always the spatial surface induced by our model at any time point \(t\) is continuous. So, unless the spatial surface is extremely irregular, any spatio-temporal data can be modeled reasonably adequately by our proposed spatio-temporal process. A stronger statement, however, is made in the second part of the theorem. It says that if the covariance functions \(c_f(\cdot, \cdot), c_g(\cdot, \cdot), c_e(\cdot, \cdot), c_{\eta}(\cdot, \cdot), c_0(\cdot, \cdot)\) are sufficiently regular (smooth) then the sample paths of the process \(Y(s, t)\) are almost always regular (smooth) and their degrees of smoothness depend on the degree of smoothness of the covariance functions. As an illustration consider the situation when all the covariance function belong to the Matérn family whose
smoothenes parameter is $\nu$. Then the sample paths of the centered Gaussian processes with the respective covariance functions will be almost surely $|\nu - 1|$ times differentiable, where, for any $x$, $|x|$ denotes the smallest integer greater than or equal to $x$. So, the spatio-temporal process $Y(s,t)$ will also have almost surely $|\nu - 1|$ times differentiable sample paths. Hence, when we have clear evidence from the data or prior knowledge regarding the degree of smoothness of the spatio-temporal process we should choose all the covariance functions $c_f(\cdot, \cdot), c_g(\cdot, \cdot), c_t(\cdot, \cdot), c_\eta(\cdot, \cdot), c_0(\cdot, \cdot)$ from the Matérn family with specific value of $\nu$.

4 Model fitting and prediction

Now we illustrate how our model can be used to make prediction at new spatio-temporal coordinates. Firstly, let us specify the prior structure. We consider bivariate Gaussian priors for each of $(\beta_{0g}, \beta_{1g})$ and $(\beta_{0f}, \beta_{1f})$. Although any reasonable isotropic covariance kernel that satisfy the mild regularity conditions already mentioned in Section 3 can be used in our model, for the sake of simplicity we consider the squared exponential covariance kernel with the following representation $c(u, v) = \sigma_f^2 e^{-\lambda ||u-v||^2}$. Since, we have five covariance kernels $c_f(\cdot, \cdot), c_g(\cdot, \cdot), c_t(\cdot, \cdot), c_\eta(\cdot, \cdot), c_0(\cdot, \cdot)$ we need to put prior on five scale parameters $\sigma_f^2, \sigma_g^2, \sigma_t^2, \sigma_\eta^2, \sigma_0^2$ and five smoothness parameters $\lambda_f, \lambda_g, \lambda_t, \lambda_\eta, \lambda_0$. We consider lognormal priors for all of them. Also, we have taken $N(0, I)$ prior for the vector parameter $\mu_0$. All the priors considered above are mutually independent.

With the prior specification as above and the conditional densities $|y|_x, \tilde{\theta}$ and $|x|_\tilde{\theta}$ being explicitly available ($\tilde{\theta}$ denotes the vector of all the parameters) we design a Gibbs sampler with Gaussian full conditionals for $\mu_0$, $\beta_{0g}, \beta_{1g}$ and $\beta_{0f}, \beta_{1f}$ and Metropolis steps for scale parameters $\sigma_f^2, \sigma_g^2, \sigma_t^2, \sigma_\eta^2, \sigma_0^2$ and smoothness parameters $\lambda_f, \lambda_g, \lambda_t, \lambda_\eta, \lambda_0$. We update the latent state vector $x_{nT}$ using Transformation based Markov Chain Monte Carlo (TMCMC) introduced by Dutta and Bhattacharya (2014) [12]. In particular, we use the additive transformation which has been shown by Dutta and Bhattacharya (2014)[12] to require less number of “moves types” compared to other valid transformations.

The idea of TMCMC is very simple yet a very powerful one. Here we briefly illustrate the idea of additive TMCMC by contrasting it with the traditional Random Walk Metropolis (RWM) approach, assuming that we wish to update all the variables simultaneously. Suppose that we want to simulate from the conditional distribution $|x_{nT}|_\tilde{\theta}$ using the RWM approach. Then we have to simulate $nT$ independent Gaussian random variables $E_{nT} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_{nT})$; assuming that the current state of the Markov chain is $x^{(i)}_{nT}$, we accept the new state $x^{(i)}_{nT} + E_{nT}$ with probability $\min\left\{1, \frac{|x^{(i)}_{nT} + E_{nT}|_\tilde{\theta}}{|x^{(i)}_{nT}|_\tilde{\theta}}\right\}$. However if $nT$ is large then this acceptance probability will tend to be extremely small. Hence the RWM chain sticks to a particular state for very long time and therefore the convergence to the posterior $|x_{nT}|_\tilde{\theta}$ is very slow. What additive TMCMC does is simulate only one $\epsilon > 0$ from some arbitrary distribution left-truncated at zero, and then form the $nT$ dimensional vector $E_{nT}$ setting the $k$-th element independently to $-\epsilon$ with probability $p_k$ and $+\epsilon$ with probability $1 - p_k$. For our applications we set $p_k = 1/2$ for $k = 1, 2, \ldots, nT$. Then we accept the new state $x^{(i)}_{nT} + E_{nT}$ with acceptance probability $\min\left\{1, \frac{|x^{(i)}_{nT} + E_{nT}|_\tilde{\theta}}{|x^{(i)}_{nT}|_\tilde{\theta}}\right\}$. Dutta and Bhattacharya (2014)[12], Dey and Bhattacharya (2014a)[9], Dey and Bhattacharya (2014b)[10] provide details of many advantages of TMCMC (in particular, additive TMCMC) as compared to traditional MCMC (in particular, RWM). However, in our setup, we have not used TMCMC directly. Instead, we simulated $T$ independent $\epsilon_t$ each to update $n$ latent observations corresponding to time $t$. This may be called a block TMCMC and it improves mixing significantly over ordinary TMCMC. In our model dimension of the state vector is large and updating it using usual RWM
would have been very inefficient. TMCMC saves us from such pitfall.

Using the above sampling-based approach we study the posterior distribution of the unknown quantities and make inferences regarding the parameters. But our main goal is to predict \( y(s^*, t^*) \) at some new spatio-temporal coordinate \((s^*, t^*)\). So, we augment \( x(s^*, t^*) \) with \( \{x(s_i, t)\} \) where \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \). The conditional distribution of \( [y|x, x(s^*, t^*), \theta] \) remains the same as \( [y|x, \theta] \) and posterior simulation is done exactly the same way as in the case of model-fitting. Once the post burn-in posterior samples \( (x^{(B)}(s^*, t^*), \tilde{\theta}(B)), (x^{(B+1)}(s^*, t^*), \tilde{\theta}(B+1)), \ldots \) from \([x(s^*, t^*), \tilde{\theta}|y]\) are available, we simulate \( y^{(B)}(s^*, t^*), y^{(B+1)}(s^*, t^*), \ldots \) from

\[
N \left( \beta_{0f}^{(j)} + \beta_{1f}^{(j)} x^{(j)}(s^*, t^*), \Sigma_{12}^{(j)} \left( \Sigma_{22}^{(j)} \right)^{-1} V^{(j)} \right)
\]

\[
\left[ \sigma_f^{(j)} \right]^2 + \left( \sigma_{\epsilon}^{(j)} \right)^2 - \Sigma_{12}^{(j)} \left( \Sigma_{22}^{(j)} \right)^{-1} \Sigma_{21}^{(j)} \right)
\]

where \( j = B, B + 1, \ldots \) and \( \{\beta_{0f}^{(j)}, \beta_{1f}^{(j)}, \sigma_f^{(j)}, \sigma_{\epsilon}^{(j)}\} \) are post burn-in posterior samples for the respective parameters. Here \( \Sigma_{12}^{(j)} \) is the covariance between \( x^{(j)}(s^*, t^*) \) and the vector \( \{x^{(j)}(s_i, t)\} \) and \( \Sigma_{22}^{(j)} \) is the covariance matrix of the vector \( \{x^{(j)}(s_i, t)\} \). The vector \( V^{(j)} \) consists of \( y(s_i, t) - \beta_{0f}^{(j)} - \beta_{1f}^{(j)} x^{(j)}(s_i, t) \) where \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \). These \( y^{(B)}(s^*, t^*), y^{(B+1)}(s^*, t^*), \ldots \) are samples from the posterior predictive distribution \([y(s^*, t^*)|y]\) which is used for prediction at the new spatio-temporal coordinate \((s^*, t^*)\).

5 Simulation and real data study

5.1 Illustration via a simulation example

We randomly sample 15 points from a square of length 2 and generate spatial time series of length 20 at those points using our model. The true parameter values for simulation are \( \lambda_f = 4.3, \lambda_g = 2.4, \lambda_e = \lambda_\eta = 6.25, \lambda_0 = 4.0 \). We simulate the vector \( \mu_0 \) from \( N(0, I) \).

We consider the prior structure as mentioned in Section 4. We use independent diffused normal priors with mean 0 and variance 1000 for each of \( \beta_{0g}, \beta_{1g}, \beta_{0f}, \beta_{1f} \). For the vector \( \mu_0 \) we consider \( N(0, I) \) prior.

For the scale parameters \( \sigma_g^2, \sigma_\eta^2 \) we used lognormal priors with location parameter 0 and scale parameter 0.7. For the scale parameters \( \sigma_f^2, \sigma_{\epsilon}^2, \sigma_0^2 \) associated respectively with the error processes and the spatial process at time 0 we used lognormal priors with location parameter 3 and scale parameter 0.1. The choice of the hyper parameters associated with \( \sigma_f^2, \sigma_{\epsilon}^2 \) are made based on the variance of the data. We used lognormal priors also for the smoothness parameters. However these lognormal priors are concentrated near 0. This somewhat concentrated prior provides safeguard against ill-conditioning of the relevant covariance matrices during MCMC iterations — indeed, free movement of the smoothness parameters in their respective parameter spaces for MCMC purposes can make the covariance matrices \( \Sigma \) and \( \Sigma_{f,\epsilon} \) drastically ill-conditioned. In making the above selection of hyper parameters we have extensively used MCMC pilot runs based on a smaller subset of the entire dataset. Since, due to smaller sizes of the data sets the pilot runs were much faster, we were able to experiment with many such pilot MCMC runs with different combinations of the hyper parameters. We chose that combination with smallest deviation of the posterior predictive median surface from the true data surface in terms of total absolute deviation. Then we perform our final MCMC computation based on the entire dataset using those selected values of the hyper parameters. In a standard laptop we have run 200,000 iterations at a rate almost 70,000 iterations per day. We have taken first 100,000 iterations as burn-in and the post burn-in 100,000 iterations have been used for posterior inference. Convergence of the MCMC chains are checked using different starting points. Visual monitoring suggests that the convergence is satisfactory.
Apart from the above set of priors we also experimented with some non informative priors but the results turned out to be only negligibly different, indicating some degree of robustness with respect to the prior choices.

We obtain the leave-one-out posterior predictive distribution for the observed data at each of the 300 space-time coordinates. Out of the 300 data points, at 292 space-time coordinates the observed data fell within the 95\% Bayesian prediction intervals of their respective posterior predictive distributions. The average length of the 95\% Bayesian prediction interval being 20.25 indicates that the results are encouraging.

Although we have used different sets of parameter values to conduct this simulation experiment we report only one study for space constraint. The results are shown in Figures 1(a)-1(f).
Figure 1: Leave-one-out posterior predictive performance of our model when the data are generated from our own model.
A brief of Figure 1 follows. Panel (a) displays the leave-one-out 95% Bayesian prediction interval for all of the 15 spatial locations at time point $t = 7$. The upper and lower quantiles (which form the intervals) are interpolated to form two surfaces. The middle surface is obtained by interpolating the leave-one-out point predictions (which is the median in this case) at the monitoring sites $s_1, s_2, \ldots, s_{15}$ for time point $t = 7$. Panel (b) displays once again the leave-one-out 95% Bayesian prediction interval for all of the 15 spatial locations but this time the middle surface is interpolated through the truly observed data (the truly observed data being represented by black dots). Panel (c) and Panel (d) display similar plots as (a) and (b) but for time point $t = 14$. Panels (e) and (f) show the leave-one-out 95% Bayesian prediction intervals for the time series at spatial locations $s_6$ and $s_9$ respectively. The starred curve represents the curve interpolated through the truly observed data (the observed data itself being represented by star) and the plain curves represent respectively the upper, middle and lower quantiles associated with the posterior predictive distributions.

In other words, Figures 1(a)-1(f) demonstrate the leave-one-out posterior predictive performance of our model. While figures 1(b) and 1(d) show that our model performs satisfactorily in terms of interval prediction at purely spatial level, the similarity of the median surface in figure 1(a) with the true data surface in figure 1(b) indicates that our model performs well also in terms of point prediction. The same conclusion can be drawn by observing figures 1(c) and 1(d). Figures 1(e) and 1(f) confirm that our model is also successful in terms of prediction at the temporal level.
Figure 2: Panels (a)-(f) display the 95% Bayesian credible intervals for the correlation function for some fixed combination of \((s_i, t_j), (s'_i, t'_j)\). The coverage region is denoted by the horizontal black line and the true value is denoted by the vertical black line.

While figures 1(a)-1(f) display the performance of our model in terms of prediction, figures 2(a)-2(f) focus on the issue of capturing the true, underlying correlation structure. The figures display the posterior distribution of the correlation function \(\rho((s_i, t_j), (s'_i, t'_j))\) for different choices of \(s_i, t_j, s'_i, t'_j\). The true value and the 95% Bayesian credible interval are also depicted in the figures. As is evident from the diagrams, the true values lie well within the respective 95% Bayesian credible intervals, vindicating reasonable performance of our model in terms of capturing the true correlation structure.

Hence, in summary, this simulation study demonstrates the effectiveness of our model for the purpose of prediction as well as for learning about the true correlation structure.

5.2 Simulation from a nonlinear non-Gaussian spatial state space model

In the previous section we described how our model performs when the data is simulated from our own model. But, now we describe a simulation study where the data is generated from a nonlinear non-Gaussian state space model which is completely different from our model. Let us first describe our data generation method.
We consider a square of length 2 and generate 20 random locations inside it. Then we simulate a spatial time series of length 20 at those points using the following nonlinear state space model:

\[
Y(s, t) = -4.1 + 0.7X(s, t) + \epsilon(s, t),
\]

\[
X(s, t) = -1.1 + 0.5X(s, t-1) + 3\sin\left(\frac{\pi}{4}X(s, t-1)\right) - 5\sin\left(\frac{\pi}{5}X(s, t-1)\right) + \eta(s, t),
\]

\[(X(s_1, 0), X(s_2, 0), \cdots, X(s_{20}, 0)) \sim N(0, \Sigma_0);
\]

where \(s \in \mathbb{R}^2\) and \(t \in \{1, 2, 3, \ldots\}. \epsilon(\cdot, t)\) and \(\eta(\cdot, t)\) are temporally independent and identically distributed spatial Gaussian processes with squared exponential covariance kernel. The corresponding scale parameters are \(\sigma_\epsilon = 3.0, \sigma_\eta = 3.9\) and the smoothness parameters are \(\lambda_\epsilon = 6.25, \lambda_\eta = 6.25\). The covariance matrix \(\Sigma_0\) is also generated by a squared exponential covariance kernel with scale and smoothness parameters 3.8 and 4.0 respectively.

Figure 3: Panel (a) shows the distribution of the spatial coordinates where the data have been generated. Panel (b) displays the plot of the evolutionary transformation \(-1.1 + 0.5x + 3\sin\left(\frac{\pi}{4}x\right) - 5\sin\left(\frac{\pi}{5}x\right)\) over the range of \(X(s_i, t)\).

As is evident from the above figures the evolutionary transformation \(-1.1+0.5x+3\sin\left(\frac{\pi}{4}x\right)-5\sin\left(\frac{\pi}{5}x\right)\) is highly nonlinear over the range of \(X(s_i, t)\), which serves our purpose. Our prior specification is very similar to the previous simulation study, where we have taken diffused normal priors for the location parameters and lognormal priors for the scale and smoothness parameters. We have taken sufficiently large burn-in and convergence of the chain is confirmed via visual monitoring. The computation is a bit time consuming with almost 60,000 iterations per day in a standard laptop. This is however natural for dynamic models, as they are so high-dimensional. In the following set of figures we describe how our model performs in terms of prediction.
Figure 4: Leave-one-out posterior predictive performance of our model when the data are generated from a highly non-linear, parametric model.
A brief description of Section 4 follows. Panel (a) displays the leave-one-out 95% Bayesian prediction interval for all of the 20 spatial locations at time point \( t = 3 \). The upper and lower quantiles (which form the intervals) are interpolated to form two surfaces. The middle surface is obtained by interpolating the leave-one-out point predictions (which is the median in this case) at the monitoring sites \( s_1, s_2, \ldots, s_{20} \) for time point \( t = 3 \). Panel (b) displays once again the leave-one-out 95% Bayesian prediction interval for all of the 20 spatial locations but this time the middle surface is interpolated through the truly observed data (the truly observed data being represented by black dots). Panel (c) and Panel (d) display plots similar as (a) and (b) but for time point \( t = 11 \). Panels (e) and (f) depict the leave-one-out 95% Bayesian prediction interval for the time series at spatial locations \( s_2 \) and \( s_8 \), respectively. The starred curve represents the curve interpolated through the truly observed data (the observed data itself being represented by star) and the plain curves represent respectively the upper, middle and lower quantiles associated with the posterior predictive distributions.

From the above set of figures it is evident that our model performs very well even though the data arose from a completely different model.

5.3 \( SO_2 \) precipitation over Great Britain and France

Air pollution over large geographical regions is a topic of wide range of studies involving statistics and other disciplines. Among them statistical modeling of pollution caused by \( SO_2 \) draws considerable attention. Here we consider a \( SO_2 \) precipitation dataset over Great Britain and France. The dataset consists of monthly measurement of sulphur dioxide precipitation taken at 16 monitoring stations spread over Great Britain and France, starting from April 1999 to January 2001. This dataset is a part of data collected through the European monitoring and evaluation programme (EMEP) which co-ordinates the monitoring of airborne pollution over Europe. Further information is available at \texttt{http://www.emep.int}, the website from which we have obtained the data set.

First we consider some exploratory analysis based on the dataset. It is important to note that instead of direct measurement what we have is the measurement of natural logarithm of \( SO_2 \) precipitation for these stations.
We plot the station-wise time series in a single figure and we see clear seasonal component and a very mild decreasing trend in most of the time series plots shown in Panel (b) of Figure 5. Then we estimate the trend and seasonal component of each of the time series separately using simple descriptive techniques and detrend and deseasonalize them. Finally, we model the residual spatio-temporal data using our method. The prior structure used in the real data analysis is similar to that of the simulation study. We used diffused bivariate normal priors for $(\beta_0, \beta_1, g, \beta_0, \beta_1, f)$ and lognormal priors for all the other parameters.

Since the monitoring stations are spread over a large geographical region and distance stretches horizontally as latitude increases, the use of simple longitude and latitude as spatial coordinates would not be appropriate. The Lambert (or Schmidt) projection addresses this problem by preserving the area. This projection is defined by the transformation of longitude and latitude, expressed in radians as $\psi$ and $\phi$, to the new co-ordinate system $s = (2 \sin(\pi/4 - \phi^2) \sin(\psi), -2 \sin(\pi/4 - \phi^2) \cos(\psi))$. However, with respect to the temporal coordinate we simply take one month as one unit of time.

We implemented an MCMC chain with sufficiently large burn-in; visual inspection suggests satisfactory convergence. Based on the MCMC samples we calculated leave-one-out posterior predictive distributions of $\ln SO_2$ at each of the monitoring stations for all the months ranging from April 1999 to January 2001. We calculated the 95% Bayesian prediction interval associated with the leave-one-out posterior predictive distribution for each of the 352 space time coordinates. More than 95% of the data fell within the respective prediction interval.

In Figures 6(a)-6(c) we provide the 95% Bayesian prediction intervals associated with the leave-one-out posterior predictive distributions for time series of length 22 for different spatial locations. These three plots show that our model performs satisfactorily also in terms of prediction at purely temporal level.

Similarly, in Figure 7(a)-7(i) we display the contour plots of the leave-one-out posterior point prediction surface, along with the contour plot of the true data surfaces for different months. The contour plots clearly...
show that our model performs very well in terms of point prediction. We also display the interval prediction through the combined plot of the interpolating surfaces of the upper quantiles, the true data and the lower quantiles, for the corresponding months. Overall, this shows that our model performs well in terms of both point prediction and interval prediction at purely spatial level.

Figure 6: Panel (a) shows the time series of the true data, the corresponding point predictions, the upper and lower quantiles (form the prediction bands) at a particular spatial location. The starred line represents the true data and the other middle curve denotes the point prediction. Panel (b) and Panel (c) display similar time series plots for two more monitoring sites.
Figure 7: Panels (a), (b), (c) show the contourplots of surfaces interpolated through leave-one-out point predictions for July 1999, April 2000 and January 2001 respectively. Panels (d), (e), (f) show the contourplots of surfaces interpolated through the true data for July 1999, April 2000 and January 2001 respectively. Panels (g), (h), (i) display the true data surfaces and the corresponding upper and lower quantile surfaces for the same three months, to demonstrate the uncertainty in prediction graphically. The red triangles indicate the observed values. The black dots in all the figures indicate the locations of the monitoring stations on the map.
One point to be noted here is that we have fitted our model to the residual space-time dataset (after detrending and deseasonalizing the original dataset), not to the original ln $SO_2$ measurements. So, while doing posterior prediction (or leave-one-out posterior prediction) at a space-time location, we needed to add back the trend and seasonal components to obtain the correct predictions. Only then we can compare the posterior prediction and the true value at any particular space-time location.

6 Possible extensions

A very important aspect of such space-time data is covariate information. Although in this article, we only consider model without any covariate information, it can be easily incorporated in our setup. In that case our model will be of the following form:

$$
Y(s, t) = f(X(s, t), Z(s, t)) + \epsilon(s, t),
$$

(8)

$$
X(s, t) = g(X(s, t - 1)) + \eta(s, t),
$$

(9)

where $Z(s, t)$ is the covariate process and $f(\cdot, \cdot)$ is now a Gaussian processes on $\mathbb{R}^2$. The rest of the theory remains the same as before. In fact, we can have $k$ different covariate processes $Z_1(s, t), Z_2(s, t), \cdots, Z_k(s, t)$, in which case $f(\cdot, \cdot, \cdots, \cdot)$ will be a Gaussian process on $\mathbb{R}^{k+1}$.

Another direction for extension is time varying nonparametric state space model. In that case we let the evolutionary and observational transformations vary with time, so that our model assumes the form:

$$
Y(s, t) = f_t(X(s, t)) + \epsilon(s, t),
$$

(10)

$$
X(s, t) = g_t(X(s, t - 1)) + \eta(s, t);
$$

(11)

Such modification is particularly useful when the time interval on which we are observing the data is very wide so that it is unlikely for the functions $f$ and $g$ to remain invariant with respect to time. In the context of purely temporal nonparametric state space models, [21] consider time-varying functions $f_t(\cdot)$ and $g_t(\cdot)$, which they re-write as $f(\cdot, t)$ and $g(\cdot, t)$, respectively. In other words, they consider the time component as an argument of their random functions $f$ and $g$, which are modeled by Gaussian processes in the usual manner. Such ideas can be easily adopted in our spatio-temporal model. Currently, we are working on these extensions.

7 Discussion and concluding remarks

One common problem in the context of space-time data is missing observations. Malfunction of monitoring device, inexperienced handling etc., may lead to such problem. So, any good spatio-temporal model should take care of missing data. In our case it is very simple. We augment the latent variable (say $x(s^*, t^*)$) corresponding to the missing observation with $\{x(s_i, t)\}$, $i = 1, \cdots, n$ and $t = 1, \cdots, T$. Then we can predict the missing observation $y(s^*, t^*)$ in the same way we make prediction at a new spatio-temporal coordinate (see Section 4). Thus, prediction at a new spatio-temporal co-ordinate and missing data are taken care of simultaneously in our model.

Although we have developed a new nonstationary model, we have not done any comparative study with respect to other existing nonstationary models (mentioned in section 2). Unfortunately, these nonstationary models are built on entirely different philosophies and hence are not comparable. For example, the nonstationary model developed in [25] depends heavily on the choice of kernel and may outperform the nonstationary
model developed in [23] depending on whether some particular kernel is being used. In fact, Dou and Zidek (2010) [11], while comparing two different methods for analyzing hourly ozone data expressed similar view (see page 1209). Similarly, our model, which is built on the idea of state space models is not comparable at all with the nonstationary models that are based on the ideas of kernel convolution or deformation. In fact, our model is not even comparable with DSTM, another model based on the idea of state space. The reason behind it is that although our model reduces to the no-covariate DSTM model if \( \beta_0 = \beta_g = 0, \beta_1 = \beta_1 = 1, \sigma_f^2 = \sigma_g^2 = 0 \), in spirit it is entirely different from DSTM models, which are generally used to utilize covariate information in improving the prediction. Indeed, our objective is to utilize unobserved information to enhance prediction. Both of these models fall in the category of state space models, but the motivation behind them are quite different. DSTM models are much more like spatio-temporal regression modeling where the linear regression structure varies with time (varying coefficient linear regression).

**Acknowledgments**

The research of the first author is fully funded by CSIR SPM Fellowship, Govt. of India. The authors thank European monitoring and evaluation programme (EMEP) for the \( SO_2 \) dataset. They also thank Moumita Das for many fruitful discussions on this article.

**Appendix**

Before proving the theorems let us make a notational clarification. The notations \( X|Y, x|y \) and \( X = x|Y = y \) are equivalent and throughout this section they will denote the value of conditional pdf of \( X \) given \( Y = y \) at \( X = x \).

**Proof of Theorem 3.1.** For this proof we assume that there exist continuous modifications of the Gaussian processes that we consider, that is, there exist processes with sample paths that are continuous everywhere, not just almost everywhere, and that our Gaussian processes equals such processes with probability one (see, for example, see [?] for details). Existence of such continuous modifications are guaranteed under the correlation structure that we consider for our Gaussian processes.

Let us first notice that, \( \exists \) a probability space \( (\Omega, \mathcal{F}, P) \) such that

\[
(g(x), X(s_1, 0), \ldots, X(s_n, 0)) : (\Omega, \mathcal{F}) \to (C(\mathbb{R}), \mathcal{A}) \bigotimes_{\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)}
\]

where \( C(\mathbb{R}) \) is the space of all real valued continuous functions on \( \mathbb{R} \) and \( \mathcal{A} \) is the Borel sigma field obtained from the topology of compact convergence on the space \( C(\mathbb{R}) \). Such a joint measurability result need not hold unless \( g(x) \) and \( (X(s_1, 0), \ldots, X(s_n, 0)) \) are independent. Now, we will show that \( g(X(s_1, 0)) \) is a measurable real valued function or a proper random variable. To show this, first see that \( g(X(s_1, 0)) : \Omega \to \mathbb{R} \) can be written as \( T(g(\cdot), X(s_1, 0)) \) where \( T : C(\mathbb{R}) \bigotimes \mathbb{R} \to \mathbb{R} \) is a transformation such that, \( T(g, x) = g(x) \) where \( g \) is a real valued continuous function on \( \mathbb{R} \) and \( x \) is a real number.

**Lemma 7.1.** \( T : C(\mathbb{R}) \bigotimes \mathbb{R} \to \mathbb{R} \) is a continuous transformation where the topology associated with \( C(\mathbb{R}) \) is the topology of compact convergence and the topology associated with \( \mathbb{R} \) is the usual Euclidean distance based topology on real numbers.
Proof of Lemma 7.4: Let us consider the metric \( d(g, g') = \sum_{i=1}^{\infty} \frac{1}{2^i} 1 + \sup_{x \in [-i, i]} |g(x) - g'(x)| \). This metric induces the topology of compact convergence on the space \( C(\mathbb{R}) \). To prove continuity of \( T \) one needs to show that \( d(g_n, g) \to 0 \) and \( |x_n - x| \to 0 \) \( \Rightarrow \) \( |T(g_n, x_n) - T(g, x)| \to 0 \).

Let us assume that \( d(g_n, g) \to 0 \) and \( |x_n - x| \to 0 \). So, \( \exists N_0 \) and \( j_0 \) such that \( \forall n \geq N_0, x_n \in [-j_0, j_0] \).

Now, \( \frac{1}{2^{j_0}} \sup_{x \in [-j_0, j_0]} |g_n(x) - g(x)| \leq d(g_n, g) \Rightarrow \sup_{x \in [-j_0, j_0]} |g_n(x) - g(x)| \to 0 \)

and \( |g(x_n) - g(x)| \to 0 \) because \( g \) is continuous.

But,

\[
|g_n(x_n) - g(x)| \leq |g_n(x_n) - g(x_n)| + |g(x_n) - g(x)|
\]

So, \( \forall n \geq N_0, |g_n(x_n) - g(x)| \leq \sup_{x \in [-j_0, j_0]} |g_n(x) - g(x)| + |g(x_n) - g(x)| \)

The RHS goes to 0 as \( n \to \infty \). Hence, \( |T(g_n, x_n) - T(g, x)| = |g_n(x_n) - g(x)| \to 0 \).

Once continuity of \( T \) is proved, note that \( T^{-1}(U) \), for any open set \( U \subseteq \mathbb{R} \), is an open set in the product topology on the space \( C(\mathbb{R}) \times \mathbb{R} \). Hence, \( T^{-1}(U) \) belongs to the Borel sigma field generated by this product topology which is in this case equivalent to the product sigma field \( \mathcal{A} \otimes \mathcal{B}(\mathbb{R}) \) associated with \( (C(\mathbb{R}), \mathcal{A}) \otimes (\mathbb{R}, \mathcal{B}(\mathbb{R})) \). This equivalence holds because both of the spaces \( C(\mathbb{R}) \) and \( \mathbb{R} \) are separable. But \( (g(x), X(s_1, 0)) \) is measurable with respect to \( (\Omega, \mathcal{F}) \) and \( (C(\mathbb{R}), \mathcal{A}) \otimes (\mathbb{R}, \mathcal{B}(\mathbb{R})) \). Hence, the inverse image of \( T^{-1}(U) \) with respect to \( (g(x), X(s_1, 0)) \) is in \( \mathcal{F} \). So, the inverse image of any open set \( U \subseteq \mathbb{R} \) with respect to \( g(X(s_1, 0)) \) is in \( \mathcal{F} \). This proves the measurability of \( g(X(s_1, 0)) \).

Following exactly same argument as above we can further prove that \( g(X(s_2, 0)), \ldots, g(X(s_n, 0)) \) are jointly measurable. Now, as \( \eta(s, t) \) is independent of \( g(x) \) and \( (X(s_1, 0), \ldots, X(s_n, 0)) \), we have the joint measurability of \( (X(s_1, 1), \ldots, X(s_n, 1)) \) (Sec 7). Infact, we can prove that \( (g(x), X(s_1, 1), \ldots, X(s_n, 1)) \) are jointly measurable. To do it we consider \( T' : C(\mathbb{R}) \otimes \mathbb{R} \to C(\mathbb{R}) \otimes \mathbb{R} \) such that \( T'(g, x) = (g(x), g(x)) \) where \( g \) is a real valued continuous function on \( \mathbb{R} \) and \( x \) is a real number. Then similarly as in case of \( T \) we can prove that \( T' \) is also a continuous map which immediately implies that \( (g, g(X(s_1, 0))) \) are jointly measurable. Then \( \eta(s, t) \) being independent of \( g(x) \) and \( (X(s_1, 0), \ldots, X(s_n, 0)) \), implies the joint measurability of \( (g(x), X(s_1, 1), \ldots, X(s_n, 1)) \). Hence, starting with the joint measurability of \( (g(x), X(s_1, 0), \ldots, X(s_n, 0)) \) we prove the joint measurability of \( (X(s_1, 1), \ldots, X(s_n, 1)) \) and \( (g(x), X(s_1, 1), \ldots, X(s_n, 1)) \). Similarly, if we start with joint measurability of \( (g(x), X(s_1, 1), \ldots, X(s_n, 1)) \) we can prove the joint measurability of \( (X(s_1, 2), \ldots, X(s_n, 2)) \). Thus, the joint measurability of the whole collection of state variables \( X(s_i, t) \) \( \forall \ i = 1, 2, \ldots, n; t = 0, 1, \ldots, T \) is mathematically established.

Now, to prove joint measurability of the collection of observed variables \( Y(s_i, t) \) \( \forall \ i = 1, 2, \ldots, n; t = 1, \ldots, T \) recall the observational equation (1). Since, \( f(x) \) takes values in \( (C(\mathbb{R}), \mathcal{A}) \) just as \( g(x) \), and also since \( \epsilon(s, t) \) is independent of \( f(s) \) just as \( \eta(s, t) \) is independent of \( g(x) \), all the previous arguments go through in this case and joint measurability of \( Y(s_i, t) \) \( \forall \ i = 1, 2, \ldots, n; t = 0, 1, \ldots, T \) is established.

Finally, it remains to show that a valid spatio-temporal process is induced by this model. But it is immediate from the application of Kolmogorov consistency theorem. The consistency conditions of the theorem are trivially satisfied by our construction and hence the result follows. \( \square \)
Alternative proof of Theorem 3.1. In the previous proof, we needed to assume the continuous modification of the underlying Gaussian process. Here we consider a proof where the underlying Gaussian process need not be continuous. In fact, the alternative proof that we now provide is valid even if the underlying process admits at most countable number of discontinuities.

Note that it is possible to represent any stochastic process \( \{ Z(s); s \in T \} \), for fixed \( s \) as a random variable \( \omega \mapsto Z(s, \omega) \), where \( \omega \in \Omega; \Omega \) being the set of all functions from \( T \) into \( \mathbb{R} \). Also, fixing \( \omega \in \Omega \), the function \( s \mapsto Z(s, \omega); s \in T \), represents a path of \( Z(s); s \in T \). Indeed, we can identify \( \omega \) with the function \( s \mapsto Z(s, \omega) \) from \( T \) to \( \mathbb{R} \); see, for example, [33], for a lucid discussion. This latter identification will be convenient for our purpose, and we adopt this for proving our result on measurability.

Note that the \( \sigma \)-algebra \( \mathcal{F} \) associated with \( Z \) is generated by sets of the form

\[
\{ \omega : \omega(s_i) \in B_1, \omega(s_2) \in B_2, \ldots, \omega(s_k) \in B_k \},
\]

where \( B_i \subset \mathbb{R}; i = 1, \ldots, k \), are Borel sets in \( \mathbb{R} \).

In our case, the Gaussian process \( g(\cdot) \) can be identified with \( g(x)(\omega_1) = \omega_1(x) \), for any fixed \( x \in \mathbb{R} \) and \( \omega_1 \in \Omega_1 \), where \( \Omega_1 \) is the set of all functions from \( \mathbb{R} \) to \( \mathbb{R} \). The initial Gaussian process \( X(\cdot, 0) \) can be identified with \( X(s, 0)(\omega_2) = \omega_2(s) \), where \( s \in \mathbb{R}^d (d \geq 2) \) and \( \omega_2 \in \Omega_2 \). Here \( \Omega_2 \) is the set of all functions from \( \mathbb{R}^d \) to \( \mathbb{R} \). Let \( \mathcal{F}_1 \) and \( \mathcal{F}_2 \) be the Borel \( \sigma \)-fields associated with \( \Omega_1 \) and \( \Omega_2 \), respectively.

We first show that the composition of \( g(\cdot) \) with \( X(\cdot, 0) \), given by \( g(X(s, 0)) \) is a measurable random variable for any \( s \) is a measurable function. Since \( g \) and \( X(\cdot, 0) \) are independent, we need to consider the product space \( \Omega_1 \otimes \Omega_2 \), and noting that \( g(X(s, 0)(\omega_2))(\omega_1) = \omega_1(\omega_2(s)) \), where \( (\omega_1, \omega_2) \in \Omega_1 \otimes \Omega_2 \), need to show that sets of the form

\[
A(s_1, \ldots, s_k) = \{ (\omega_1, \omega_2) : \omega_1(\omega_2(s_1)) \in B_1, \omega_1(\omega_2(s_2)) \in B_2, \ldots, \omega_1(\omega_2(s_k)) \in B_k \},
\]

where \( B_i \subset \mathbb{R}; i = 1, \ldots, k \), are Borel sets in \( \mathbb{R} \), are in \( \mathcal{F}_1 \otimes \mathcal{F}_2 \), the product Borel \( \sigma \)-field associated with \( \Omega_1 \otimes \Omega_2 \). For our purpose, we let \( B_i \) be of the form \([a_i, b_i] \) for real values \( a_i < b_i \).

Now, suppose that \( (\omega_1, \omega_2) \in A(s_1, \ldots, s_k) \). Then \( \omega_1(\omega_2(s_i)) \), \( i = 1, \ldots, k \), which implies that \( \omega_2(s_i) \) is at most a countable union of sets of the form \([a_j^{(i)}, b_j^{(i)}]; j \in D_i \), where \( D_i \) is a countable set of indices. Also, it holds that \( \omega_1(x^*) \in [a_i, b_i] \); \( \forall x^* \in \mathbb{Q} \cap \bigcup_{j \in D_i} [a_j^{(i)}, b_j^{(i)}] \). Here \( \mathbb{Q} \) is the countable set of rationals in \( \mathbb{R} \).

If necessary, we can envisage a countable set \( D^* \) consisting of points of discontinuities of \( \omega_1 \). If \( \xi \) is a point of discontinuity, then \( \omega_1(\xi) \) may be only the left limit of particular sequence \( \{ \omega_1(\xi_m^0); m = 1, 2, \ldots \} \) or only the right limit of a particular sequence \( \{ \omega_1(\xi_m); m = 1, 2, \ldots \} \), or \( \omega_1(\xi) \) may be an isolated point, not reachable by sequences of the above forms. It follows that \( (\omega_1, \omega_2) \) must lie in

\[
A^*(s_1, \ldots, s_k) = \bigcap_{i=1}^k \left\{ (\omega_1, \omega_2) : \omega_1(x) \in [a_i, b_i] \; \forall x \in \left( \mathbb{Q} \cap \bigcup_{j \in D_i} [a_j^{(i)}, b_j^{(i)}] \right) \cup D^*, \omega_2(s_i) \in [a_j^{(i)}, b_j^{(i)}] \right\}.
\]

Now, if \( (\omega_1, \omega_2) \in A^*(s_1, \ldots, s_k) \), then, noting that for any point \( x \in \bigcup_{j \in D_i} [a_j^{(i)}, b_j^{(i)}] \) of \( \omega_1 \), \( \omega_1(x) = \lim_{m \to \infty} \omega_1(\xi_m) \), where \( \{ \xi_m; m = 1, 2, \ldots \} \in \mathbb{Q} \cap \bigcup_{j \in D_i} [a_j^{(i)}, b_j^{(i)}] \), it is easily seen that \( (\omega_1, \omega_2) \in A(s_1, \ldots, s_k) \).

Hence, \( A(s_1, \ldots, s_k) = A^*(s_1, \ldots, s_k) \).

Now observe that \( A^*(s_1, \ldots, s_k) \) is a finite intersection of countable union of measurable sets; hence, \( A^*(s_1, \ldots, s_k) \) is itself a measurable set. In other words, we have proved that \( g(X(\cdot, 0)) \) is measurable. Now, as \( \eta(\cdot, t) \) is independent of \( g(\cdot) \) and \( X(\cdot, 0) \), it follows from (2) that \( X(\cdot, 1) \) is measurable.
To prove measurability of $X(\cdot, 2)$, note that
\[
X(s, 2) = g(X(s, 1)) + \eta(s, 2)
= g(g(X(s, 0)) + \eta(s, 1)) + \eta(s, 2).
\tag{12}
\]

The process $\eta(\cdot, 1)$ requires introduction an extra sample space $\Omega_3$, so that we can identify $\eta(s, 1)(\omega_3)$ as $\omega_3(s)$. With this, we can represent $g(g(X(s, 0)) + \eta(s, 1))$ of $\Omega_3$ as $\omega_1(\omega_2(s)) + \omega_3(s)$.

Now, $\omega_1(\omega_2(s)) + \omega_3(s) \in [a_i, b_i]$ implies that $\omega_1(\omega_2(s)) + \omega_3(s) \in \bigcup_{j \in D_i} [a_i^{(j)}, b_i^{(j)}]$. If $\omega_1(\omega_2(s)) + \omega_3(s) \in [a_i^{(k)}, b_i^{(k)}]$ for some $k \in D_i$, the the set of solutions is
\[
\{ \omega_1(\omega_2(s)) \in [a_i^{(k)} - r, b_i^{(k)} - r] : w_3(s) = r \},
\tag{13}
\]
where $\omega_1(\omega_2(s)) \in [a_i^{(k)} - r, b_i^{(k)} - r]$ implies, as before, that $\omega_2(s)$ belongs to a countable union of measurable sets in $\mathbb{R}$. Although the set of (13) is an uncountable union, following the technique used for proving measurability of $g(X(\cdot, 0))$, we will intersect the set by $\mathbb{Q}$, the (countable) set of rationals in $\mathbb{R}$; this will render the intersection a countable set. The proof of measurability then follows similarly as before.

Proceeding likewise, we can prove measurability of $X(\cdot, t)$ is measurable for $t = 2, 3, \ldots$. Proceeding exactly in the same way, we can also prove that $Y(\cdot, t)$; $t = 1, 2, \ldots, T$ are measurable. Moreover, it can be easily seen that the same methods employed for proving the above results on measurability can be extended in a straightforward (albeit notationally cumbersome) manner to prove that the sets of the forms
\[
\{X(s_i, t_i) \in [a_i, b_i] : i = 1, \ldots, k\}
\quad \text{and} \quad
\{Y(s_i, t_i) \in [a_i, b_i] : i = 1, \ldots, k\}
\]
are also measurable. Furthermore, it can be easily verified that $X$ and $Y$ satisfy Kolmogorov’s consistency criteria. In other words, $X$ and $Y$ are well-defined stochastic processes in both space and time.

**Proof of Theorem 3.2.** Let us first observe that conditional on $g(x)$ our latent process satisfies the Markov property. That is,
\[
[(x(s_1, t), \ldots, x(s_n, t)) \mid (g(x(s_1, t-1)), \ldots, g(x(s_n, t-1))), (x(s_1, t-1),
\ldots, x(s_n, t-1)), (x(s_1, t-2), \ldots, x(s_n, t-2)), \ldots, (x(s_1, 0), \ldots, x(s_n, 0))]
\]
\[
= [(x(s_1, t), \ldots, x(s_n, t)) \mid (g(x(s_1, t-1)), \ldots, g(x(s_n, t-1))), (x(s_1, t-1), \ldots, x(s_n, t-1))]
\]
\[
\sim \frac{1}{|\Sigma_\eta|^2} \exp \left[ -\frac{1}{2} \begin{pmatrix}
(x(s_1, t) - g(x(s_1, t-1)) \cr
x(s_2, t) - g(x(s_2, t-1)) \cr
\vdots \cr
x(s_n, t) - g(x(s_n, t-1))
\end{pmatrix}^\top \Sigma_\eta^{-1} \begin{pmatrix}
(x(s_1, t) - g(x(s_1, t-1)) \cr
x(s_2, t) - g(x(s_2, t-1)) \cr
\vdots \cr
x(s_n, t) - g(x(s_n, t-1))
\end{pmatrix} \right],
\]

where $[x \mid y]$ denotes the conditional density of $X$ at $x$ given $Y = y$. Now, let us represent $g(x(s_i, t-1))$ by $u(i, t)$ for all $i = 1, \ldots, n$ and $t = 1, 2, \ldots, T$. Then repeatedly using the Markov property we have following
\[
[x(s_1, T), \ldots, x(s_n, T), \ldots, x(s_1, 0), \ldots, x(s_n, 0) \mid g(x(s_1, T-1)), \ldots, g(x(s_n, T-1)), \ldots, g(x(s_1, 0)), \ldots, g(x(s_n, 0))]\]
\[
\sim [x(s_1, T), \ldots, x(s_n, T) \mid g(x(s_1, T - 1)), \ldots, g(x(s_n, T - 1)), x(s_1, T - 1), \ldots, x(s_n, T - 1)] \times \\
\ldots \times [x(s_1, 1), \ldots, x(s_n, 1) \mid g(x(s_1, 0)), \ldots, g(x(s_n, 0)), x(s_1, 0), \ldots, x(s_n, 0)] \\
\times [x(s_1, 0), \ldots, x(s_n, 0)]
\]

\[
\sim \frac{1}{(2\pi)^{n(T+1)/2} |\Sigma|^T} \prod_{t=1}^{T} \exp \left[ -\frac{1}{2} \begin{pmatrix} (x(s_1, t) - u(1, t))^T \\
(x(s_2, t) - u(2, t)) \\
\vdots \\
(x(s_n, t) - u(n, t)) \end{pmatrix} \Sigma_{\eta}^{-1} \begin{pmatrix} (x(s_1, t) - u(1, t)) \\
(x(s_2, t) - u(2, t)) \\
\vdots \\
(x(s_n, t) - u(n, t)) \end{pmatrix} \right]
\]

\[
\times \frac{1}{(2\pi)^{n/2} |\Sigma_0|^T} \exp \left[ -\frac{1}{2} \begin{pmatrix} (x(s_1, 0) - \mu_0)_1 \\
(x(s_2, 0) - \mu_0)_2 \\
\vdots \\
(x(s_n, 0) - \mu_0)_n \end{pmatrix} \Sigma_0^{-1} \begin{pmatrix} (x(s_1, 0) - \mu_0)_1 \\
(x(s_2, 0) - \mu_0)_2 \\
\vdots \\
(x(s_n, 0) - \mu_0)_n \end{pmatrix} \right]
\]

But, this is the joint density of the state variables conditioned on \( g(x) \). To obtain the joint density of the state variables one needs to marginalize it with respect to the Gaussian process \( g(\cdot) \). After marginalization, the joint density takes the following form:

\[
\frac{1}{(2\pi)^{n(T+1)/2} |\Sigma|^T} \prod_{t=1}^{T} \exp \left[ -\frac{1}{2} \begin{pmatrix} (x(s_1, 0) - \mu_0)_1 \\
(x(s_2, 0) - \mu_0)_2 \\
\vdots \\
(x(s_n, 0) - \mu_0)_n \end{pmatrix} \Sigma_0^{-1} \begin{pmatrix} (x(s_1, 0) - \mu_0)_1 \\
(x(s_2, 0) - \mu_0)_2 \\
\vdots \\
(x(s_n, 0) - \mu_0)_n \end{pmatrix} \right]
\]

\[
\times \int_{\mathbb{R}^{nT}} \prod_{t=1}^{T} \exp \left[ -\frac{1}{2} \begin{pmatrix} (x(s_1, t) - u(1, t)) \\
(x(s_2, t) - u(2, t)) \\
\vdots \\
(x(s_n, t) - u(n, t)) \end{pmatrix} \Sigma_{\eta}^{-1} \begin{pmatrix} (x(s_1, t) - u(1, t)) \\
(x(s_2, t) - u(2, t)) \\
\vdots \\
(x(s_n, t) - u(n, t)) \end{pmatrix} \right] \frac{1}{(2\pi)^{2T/2} |\Sigma|^T} \exp \left[ -\frac{1}{2} \begin{pmatrix} u(1, 1) - \beta_0 - \beta_1 g(x(s_1, 0)) \\
u(2, 1) - \beta_0 - \beta_1 g(x(s_2, 0)) \\
\vdots \\
u(n, T) - \beta_0 - \beta_1 g(x(s_n, T - 1)) \end{pmatrix} \Sigma^{-1} \begin{pmatrix} u(1, 1) - \beta_0 - \beta_1 g(x(s_1, 0)) \\
u(2, 1) - \beta_0 - \beta_1 g(x(s_2, 0)) \\
\vdots \\
u(n, T) - \beta_0 - \beta_1 g(x(s_n, T - 1)) \end{pmatrix} \right] \, du
\]

where \( \Sigma \) is as in (3.2). This is nothing but a convolution of two \( \mathbb{R}^{nT} \) dimensional Gaussian densities, one with mean vector \( 0 \) and covariance matrix \( I_{T \times T} \otimes \Sigma_0 \) and the other one with mean vector \( (\beta_0 + \beta_1 g(x(s_1, 0)), \ldots, \beta_0 + \beta_1 g(x(s_n, T - 1)))' \) and covariance matrix \( \Sigma \).
Hence, the integral boils down to

\[
\frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{|\Sigma_0|^{\frac{1}{2}}} \exp\left[ -\frac{1}{2} \begin{pmatrix} (x(s_1, 0) - \mu_{01})' & (x(s_2, 0) - \mu_{02})' & \cdots & (x(s_n, 0) - \mu_{0n})' \end{pmatrix} \Sigma_0^{-1} \begin{pmatrix} (x(s_1, 0) - \mu_{01}) \\ (x(s_2, 0) - \mu_{02}) \\ \vdots \\ (x(s_n, 0) - \mu_{0n}) \end{pmatrix} \right] \frac{1}{(2\pi)^{\frac{nT}{2}}} \frac{1}{|\tilde{\Sigma}|^{\frac{1}{2}}} \times \exp\left[ -\frac{1}{2} \begin{pmatrix} x(s_1, 1) - \beta_{0g} - \beta_{1g}x(s_1, 0) \\ x(s_2, 1) - \beta_{0g} - \beta_{1g}x(s_2, 0) \\ \vdots \\ x(s_n, T) - \beta_{0g} - \beta_{1g}x(s_n, T - 1) \end{pmatrix} \tilde{\Sigma}^{-1} \begin{pmatrix} x(s_1, 1) - \beta_{0g} - \beta_{1g}x(s_1, 0) \\ x(s_2, 1) - \beta_{0g} - \beta_{1g}x(s_2, 0) \\ \vdots \\ x(s_n, T) - \beta_{0g} - \beta_{1g}x(s_n, T - 1) \end{pmatrix} \right],
\]

where \(\tilde{\Sigma}\) is as in (3.2).

\(\square\)

**Proof of Theorem 3.3.** First, see that for fixed \(x(s_i, t_1)\), \(Y(s_i, t_1)\) is distributed as a Gaussian with mean \(\beta_{0f} + \beta_{1f}x(s_i, t_1)\) and variance \(\sigma_f^2 + \sigma_e^2\) where \(\sigma_f^2\) and \(\sigma_e^2\) are respectively the process variance associated with the isotropic Gaussian processes \(\epsilon(\cdot, t)\) and \(f(x)\) (see (3) and (1)). Now, see that for fixed \(x(s_i, t_1)\) and \(x(s_j, t_2)\), \(f(x(s_i, t_1))\) and \(f(x(s_j, t_2))\) have covariance \(c_f(x(s_i, t_1), x(s_j, t_2))\). Also, \(\epsilon(\cdot, t_1)\) and \(\epsilon(\cdot, t_2)\) are mutually independent spatial Gaussian processes for \(t_1 \neq t_2\). Hence, conditional on state variables the covariance between \(Y(s_i, t_1)\) and \(Y(s_j, t_2)\) is \(c_f(x(s_i, t_1), x(s_j, t_2)) + c_e(s_i, s_j)\delta(t_1 - t_2)\). Here \(\delta(\cdot)\) is the delta function i.e. \(\delta(t) = 1\) for \(t = 0\) and \(= 0\) otherwise.

So, the joint density of the observed variables, which is denoted by \([y(s_1, 1), y(s_2, 1), \ldots, y(s_n, T)]\), is given by

\[
[y(s_1, 1), y(s_2, 1), \cdots, y(s_n, T)] = 
\int_{\mathbb{R}^{nT}} [y(s_1, 1), y(s_2, 1), \cdots, y(s_n, T) \mid x(s_1, 1), x(s_2, 1), \cdots, x(s_n, T)] [x(s_1, 1), x(s_2, 1), \cdots, x(s_n, T)] dx
\]

Hence, part \((a)\) follows.

For part \((b)\) note that if \(\sigma_e^2 = 0\), the conditional density

\[
[y(s_1, 1), y(s_2, 1), \cdots, y(s_n, T) \mid x(s_1, 1), x(s_2, 1), \cdots, x(s_n, T)]
\]

is Gaussian with block diagonal covariance matrix \(I_{T \times T} \otimes \Sigma_e\). On the other hand, we have already noted that if \(\sigma_e^2 = 0\), the joint density of the state variables boils down to Gaussian (see the discussion following Theorem 3.2). Let us consider only the state variables from time \(t = 1\) onwards. They jointly follow an \(nT\) dimensional Gaussian distribution. It is not difficult to see that the mean vector and the covariance matrix of the \(nT\) dimensional Gaussian distribution are of following forms:

the \(((t - 1)n + i)\) th entry of the mean vector is \(\beta_{1g}^t \mu_{0i} + \beta_{0g}^t \beta_{1g}^t - 1\)

where \(1 \leq t \leq T\)

and the \(((t_1 - 1)n + i), (t_2 - 1)n + j))\) th entry of the covariance matrix is

\(\beta_{1g}^{t_1 + t_2} \sigma_{i,j}^0 + (\beta_{1g}^{t_1 + t_2 - 2} + \beta_{1g}^{t_1 + t_2 - 4} + \cdots + \beta_{1g}^{t_1 - t_2}) c_e(s_i, s_j)\)

where \(1 \leq t_1, t_2 \leq T\) and \(1 \leq i, j \leq n\)

\(\sigma_{i,j}^0\) is the \((i, j)\) th entry of the covariance matrix \(\Sigma_0\)

26
Now, using part (a) we see that the joint distribution of \(Y(s_1, 1), Y(s_2, 1), \ldots, Y(s_n, T)\) is nothing but a convolution of two \(\mathbb{R}^{nT}\)-dimensional Gaussian densities. Hence, it is a Gaussian distribution whose mean vector has \(((t-1)n+i)\) th entry as \(\beta_{0f} + \beta_{1f} \left( \beta_{1g} \mu_0 + \beta_{0g} \frac{\rho_i^{d-1}}{\rho_{d-1}} \right)\) where \(1 \leq t \leq T\) and the \(((t-1)n+i), ((t-1)n+j)\) th entry of the covariance matrix is 

\[
\beta_{1f}^2 \left( \beta_{1g}^{t_1+t_2} \rho_{l_1,l_2} + (\beta_{1g}^{t_1} + \beta_{1g}^{t_2} - 2) + \beta_{1g}^{t_1+t_2} - 4 + \cdots + \beta_{1g}^{t_1-t_2} \right) c_\eta(s_i, s_j) + c_\epsilon(s_i, s_j) \delta(t_1 - t_2)
\]

where \(1 \leq t_1, t_2 \leq T\) and \(1 \leq i, j \leq n\). So, part (b) is proved.

**Proof of Theorem 3.4.** part (a) We first show that \(\text{Var}(X(s, t))\) is finite. Then by using the formula

\[
\text{Var}(Y(s, t)) = \text{Var}(E(Y(s, t)|X(s, t))) + \text{E}(\text{Var}(Y(s, t)|X(s, t)))
\]

\[
= \text{Var}(\beta_{0f} + \beta_{1f} X(s, t)) + \text{E}(\sigma_f^2 + \sigma_t^2)
\]

\[
= \beta_{1f}^2 \text{Var}(X(s, t)) + \sigma_f^2 + \sigma_t^2
\]

we establish that \(\text{Var}(Y(s, t))\) is finite. To show \(\text{Var}(X(s, t))\) is finite we use principle of mathematical induction, i.e. we first show that \(\text{Var}(X(s, 0))\) is finite and then we show that if \(\text{Var}(X(s, 0)), \text{Var}(X(s, 1)), \ldots, \text{Var}(X(s, t-1))\) are finite then \(\text{Var}(X(s, t))\) is finite. These two steps together compel \(\text{Var}(X(s, t))\) to be finite.

The first step is trivially shown as \(X(s, 0)\) is a Gaussian random variable. Now we show the second step of mathematical induction, that is, we show that if \(\text{Var}(X(s, 0)), \text{Var}(X(s, 1)), \ldots, \text{Var}(X(s, t-1))\) are finite then \(\text{Var}(X(s, t))\) is finite.

Now consider the following:

\[
\text{Var}(X(s, t)|X(s, t-1) = x_{t-1}, X(s, t-2) = x_{t-2}, \ldots, X(s, 0) = x_0)
\]

\[
= \text{Var}(g(X(s, t-1)) + \eta(s, t)|X(s, t-1) = x_{t-1}, X(s, t-2) = x_{t-2}, \ldots, X(s, 0) = x_0)
\]

\[
= \text{Var}(g(X(s, t-1))|X(s, t-1) = x_{t-1}, X(s, t-2) = x_{t-2}, \ldots, X(s, 0) = x_0) + \sigma_f^2
\]

\[
= \text{Var}(g(x_{t-1})g(x_{t-2}) + \eta(s, t-1) = x_{t-1}, \ldots, g(x_0) + \eta(s, 1) = x_1, X(s, 0) = x_0) + \sigma_f^2
\]

\[
= \sigma_f^2 - \Sigma_{g12}(\Sigma_{g22} + \sigma_f^2 \mathbf{I})^{-1}\Sigma_{g12} + \sigma_\eta^2
\]

[see page 16 of [30]]

where \(\Sigma_{g12}\) is the row vector \(\left( c_g(x_{t-1}, x_0), c_g(x_{t-1}, x_1), \ldots, c_g(x_{t-1}, x_{t-2}) \right)\) and \(\Sigma_{g22}\) is the variance covariance matrix

\[
\begin{pmatrix}
    c_g(x_0, x_0) & c_g(x_0, x_1) & \cdots & c_g(x_0, x_{t-2}) \\
    c_g(x_1, x_0) & c_g(x_1, x_1) & \cdots & c_g(x_1, x_{t-2}) \\
    \vdots & \vdots & \ddots & \vdots \\
    c_g(x_{t-2}, x_0) & c_g(x_{t-2}, x_1) & \cdots & c_g(x_{t-2}, x_{t-2})
\end{pmatrix}
\]

induced by covariance function \(c_g(\cdot, \cdot)\). Now, we consider \(E(\text{Var}(X(s, t)|X(s, t-1) = x_{t-1}, X(s, t-2) = x_{t-2}, \ldots, X(s, 0) = x_0))\). We want to show that this quantity is finite. But the problem is that we have to deal with the inverse of a random matrix \((\Sigma_{g22} + \sigma_f^2 \mathbf{I})\). Fortunately, the random matrix \((\Sigma_{g22} + \sigma_f^2 \mathbf{I})\) is non-negative definite (nd). Hence, \(\sigma_f^2 - \Sigma_{g12}(\Sigma_{g22} + \sigma_f^2 \mathbf{I})^{-1}\Sigma_{g12} + \sigma_\eta^2 \leq \sigma_f^2 + \sigma_\eta^2\). On the other hand, this quantity being a conditional variance is always nonnegative. So, the following inequality holds

\[
0 \leq \sigma_f^2 - \Sigma_{g12}(\Sigma_{g22} + \sigma_f^2 \mathbf{I})^{-1}\Sigma_{g12} + \sigma_\eta^2 \leq \sigma_f^2 + \sigma_\eta^2.
\]
Unfortunately, the (random) elements of the row vector \( c \) (composed of bounded random variables. So, its elements \\
are fixed numbers it is easy to see that \\
Now we show that \\
where \( Z(s) \) is the row vector \((x_1 - \beta_{g_0} - \beta_{12}x_0, x_2 - \beta_{g_0} - \beta_{12}x_1 \cdots x_{t-1} - \beta_{g_0} - \beta_{12}x_{t-2})\). We want to \\
This term is finite. Equivalently, we want to show that \\
Now we show that \( \text{Var}(E(X(s,t)|X(s,t-1) = x_{t-1}, X(s,t-2) = x_{t-2} \cdots, X(s,0) = x_0)) \) is finite. For that it is enough to show \\
so that they are bounded random variables and then \\
First we show that \( w_i(S) \) are bounded random variables. Consider the spectral decomposition of the real \\
Let us assume that \( \Sigma_{g_{22}} = U D U' \), where \( U \) is an orthogonal matrix and \( D \) is \\
Since \( U \) is a (random) orthogonal matrix its elements are bounded random variables between \(-1\) and \(1\). The (random) elements of the row vector \( \Sigma_{g_{12}}' \) are covariances induced by the isotropic covariance kernel \\
Hence, they are bounded random variables between \(-\sigma_g^2\) and \(\sigma_g^2\). Finally, the (random) elements of \\
are bounded random variables. Thus, the (random) elements of \( (D + \sigma_g^2)^{-1} \) are bounded random variables between \(0\) and \(\frac{1}{\sigma_g^2}\). Hence, the (random) row vector \( \Sigma_{g_{12}}' (\Sigma_{g_{22}} + \sigma_g^2) \), being a product of some random matrices whose elements are bounded random variables, is itself composed of bounded random variables. So, its elements \( w_i(S) \), although random, are bounded. Now, we state a crucial lemma.
Lemma 7.2. Let us assume that $X_1, X_2, \ldots, X_n$ are random variables with finite variance and $W_1, W_2, \ldots, W_n$ are bounded random variables all defined on same probability space. Then the random variables $Y = (W_1 X_1 + W_2 X_2 + \cdots + W_n X_n)$ also has finite variance.

Proof. Let us assume that $W_1, W_2, \cdots, W_n$ lie between $[-M, M]$. $E(X_i^2) \leq K$ for $i = 1, 2, \cdots, n$. Now $\text{Var}(W_i X_i) = E(\text{Var}(W_i X_i | X_i)) + \text{Var}(E(W_i X_i | X_i)) = E(X_i^2 \text{Var}(W_i | X_i)) + \text{Var}(E(W_i | X_i))$. But $\text{Var}(W_i X_i) \leq E(W_i^2 | X_i) \leq M^2$. So, $E(X_i^2 \text{Var}(W_i | X_i)) \leq M^2$. Similarly, $E(W_i | X_i)$ lies between $[-M, M]$. So, $\text{Var}(W_i X_i) \leq E(X_i^2 E(W_i | X_i))^2 \leq M^2 E(X_i^2)$. Hence, $\text{Var}(W_i X_i) \leq 2M^2 E(X_i^2)$. So,

$$|\text{Var}(Y)| = \sum_{i=1}^{n} \text{Var}(W_i X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(W_i X_i, W_j X_j) \leq \sum_{i=1}^{n} \text{Var}(W_i X_i) + 2 \sum_{1 \leq i < j \leq n} |\text{Cov}(W_i X_i, W_j X_j)| \leq \sum_{i=1}^{n} \text{Var}(W_i X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Var}^{1/2}(W_i X_i) \text{Var}^{1/2}(W_j X_j) \leq 2nM^2 K + n(n-1)2M^2 K$$

So, $Y$ has finite variance. □

Once we apply the lemma to $w_1 z_1(s) + w_2 z_2(s) + \cdots + w_t z_t(s)$ the finiteness of $\text{Var}(E(X(s, t) | X(s, t-1) = x_{t-1}, X(s, t-2) = x_{t-2}, \ldots, X(s, 0) = x_0))$ is immediate. Then by the formula $\text{Var}(X(s, t)) = E(\text{Var}(X(s, t) | X(s, t-1)) = x_{t-1}, X(s, t-2) = x_{t-2}, \ldots, X(s, 0) = x_0)) + \text{Var}(E(X(s, t) | X(s, t-1) = x_{t-1}, X(s, t-2) = x_{t-2}, \ldots, X(s, 0) = x_0))$ we get that $\text{Var}(X(s, t))$ is finite. □

part (b) Since we have already proved in part (a) that the coordinate variables of the observed spatio-temporal process have finite variances, now we can consider the covariance function associated with the process and study its properties. Let us denote the covariance between $Y(s, t)$ and $Y(s^*, t^*)$ by $c_y(s, t, s^*, t^*)$. Then

$$c_y(s, t, s^*, t^*) = E[\text{Cov}(Y(s, t), Y(s^*, t^*) | x(s, t), x(s^*, t^*))] + \text{Cov}[E(Y(s, t) | x(s, t)), E(Y(s^*, t^*) | x(s^*, t^*))] = E[c_f(X(s, t), X(s^*, t^*))] + c_\delta(s, s^*) \delta(t - t^*) + \beta_f^2 \text{Cov}[X(s, t), X(s^*, t^*)]$$

(14)

Now, the term $E[c_f(X(s, t), X(s^*, t^*))]$ will be nonstationary and hence $E[c_f(X(s + h, t + k), X(s^* + h, t^* + k))] \neq E[c_f(X(s, t), X(s^*, t^*))]$. In fact, $|X(s + h, t + k) - X(s^* + h, t^* + k)| \not= |X(s, t) - X(s^*, t^*)|$ with probability 1 because $X(s, t)$ has density with respect to Lebesgue measure and this heuristically justifies our argument. So, the covariance function $c_y(\cdot, \cdot)$ is nonstationary in both space and time.

To prove non separability, first see that $c_f(x(s, t), x(s^*, t^*))$ is non separable in space and time, because both space and time are involved in it through $x(s, t)$. Hence, $E[c_f(X(s, t), X(s^*, t^*))]$ is nonseparable and therefore $c_y(\cdot, \cdot)$ is nonseparable in space and time.

29
Proof of Theorem 3.5. First consider $\text{Cov}(X(s, t), X(s^*, t^*))$ where WLOG we assume $t > t^*$. Also, assume that $g^*(\cdot)$ is the centered Gaussian process obtained from $g(\cdot)$. Then

$$\text{Cov}(X(s, t), X(s^*, t^*)) = \text{Cov}(g(X(s, t - 1)) + \eta(s, t), X(s^*, t^*))$$
$$= \text{Cov}(\beta_0 g + \beta_1 g(X(s, t - 1)) + g^*(X(s, t - 1)) + \eta(s, t), X(s^*, t^*))$$
$$= \beta_1 g \text{Cov}(X(s, t - 1), X(s^*, t^*)) + \text{Cov}(g^*(X(s, t - 1)), X(s^*, t^*))$$

Repeatedly expanding the term in the same way we get

$$= \beta_{1g}^{t-t^*} \text{Cov}(X(s, t^*), X(s^*, t^*)) + \beta_{1g}^{t-t^*-1} \text{Cov}(g^*(X(s, t^*)), X(s^*, t^*)) + \cdots + \text{Cov}(g^*(X(s, t - 1)), X(s^*, t^*)) \quad (15)$$

Just as the previous paragraph we can further see that

$$\text{Cov}(X(s, t^*), X(s^*, t^*))$$
$$= \text{Cov}(\beta_0 g + \beta_1 g(X(s, t^* - 1)) + g^*(X(s, t^* - 1)) + \eta(s, t^*), \beta_0 g + \beta_1 g(X(s^*, t^* - 1)) + g^*(X(s^*, t^* - 1)) + \eta(s^*, t^*))$$
$$= \beta_1^2 \text{Cov}(X(s, t^* - 1), X(s^*, t^* - 1)) + \beta_{1g} \text{Cov}(X(s, t^* - 1), g^*(X(s^*, t^* - 1)) + \beta_{1g} \text{Cov}(X(s^*, t^* - 1), g^*(X(s, t^* - 1))) + \text{Cov}(g^*(X(s, t^* - 1)), g^*(X(s^*, t^* - 1))) + c_g(s, s^*) \quad (16)$$

Now we plan to show that terms of the types $\text{Cov}(g^*(X(s^*, t^* - 1)), X(s, t^* - 1))$ and $\text{Cov}(g^*(X(s, t^* - 1)), g^*(X(s^*, t^* - 1)))$ are negligible if $\sigma_g^2$ is small enough. Our next lemma proves it rigorously.

Lemma 7.3. For arbitrarily small $\epsilon > 0$, $\exists \delta > 0$ such that $\text{Cov}(g^*(X(s, t - 1)), X(s^*, t^*)) < \epsilon$ for $0 < \sigma_g^2 < \delta$.

See that it is enough to prove that $V \text{ar}(g^*(X(s, t - 1)))$ is arbitrarily small $\forall s, t$. Then Cauchy-Schwartz inequality implies $\text{Cov}^2(g^*(X(s, t - 1)), g^*(X(s^*, t^*))) \leq V \text{ar}(g^*(X(s, t - 1))V \text{ar}(g^*(X(s^*, t^*)))$ is arbitrarily small. Similarly, Cauchy-Schwartz inequality implies $\text{Cov}^2(g^*(X(s, t - 1)), \eta(s^*, t^*)) \leq V \text{ar}(g^*(X(s, t - 1))V \text{ar}(\eta(s^*, t^*))) = V \text{ar}(g^*(X(s, t - 1))) \sigma_g^2$ is arbitrarily small. Then we are done by the expansion

$$\text{Cov}(g^*(x(s, t)), x(s^*, t^*)) = \text{Cov}(g^*(x(s, t - 1)), g^*(x(s^*, t^* - 1)) + \cdots + \beta_{1g}^t \text{Cov}(g^*(x(s, t - 1)), g^*(x(s^*, 0)))$$
$$+ \text{Cov}(g^*(x(s, t - 1)), \eta(s^*, t^* - 1)) + \cdots + \beta_{1g}^t \text{Cov}(g^*(x(s, t - 1)), \eta(s^*, 0))$$

Before proceeding towards the proof we mention two results from Gaussian process (see [1] for details) that will be used subsequently.

Result 7.4 (Borell-TIS inequality). Let us assume that $g$ is an almost surely bounded centered Gaussian process on index set $T \subseteq \mathbb{R}$. Define $\sigma_T^2 = \sup_{t \in T} E(g_t^2)$.

Then $P(||g|| > s) \leq \exp\left(-\frac{(s - E||g||)^2}{2\sigma_T^2}\right)$ for $s > E(||g||)$ where $\|g\| = \sup_{t} g_t$.

Result 7.5 (Dudley’s metric entropy bound). Under the assumption of the Borel-TIS inequality,
\[
E\|g\| \leq K \int_0^{\text{diam}(T)} \sqrt{H(e)} \, de,
\]

where \( \text{diam}(T) = \sup_{s_1, s_2 \in T} d(s_1, s_2) \) is the diameter of the index set \( T \) with respect to the canonical pseudo-metric \( d \) associated with the Gaussian process \( g \) given by \( d(s_1, s_2) = \sqrt{E[(g(s_1) - g(s_2))^2]} \), and \( H(\epsilon) = \ln N(\epsilon) \) where \( N(\epsilon) \) is the minimum number of \( \epsilon \) balls required to cover the index set \( T \) with respect to the canonical pseudo-metric \( d \); \( K \) is a universal constant.

With the above two results, we are ready to prove Lemma 7.3.

**Proof of Lemma 7.3.** Consider \( \text{Var}((g^*(X(s, t - 1)))) \). Observe that

\[
\text{Var}((g^*(X(s, t - 1)))) \leq E((g^*(X(s, t - 1))))^2
\]

\[
\leq E(\sup_x |g^*(x)|^2)
\]

\[
= \int_0^\infty P(\sup_x |g^*(x)|^2 > u) \, du \quad \text{(by the tail sum formula)}
\]

\[
\leq 2 \int_0^\infty P(\sup_x g^*(x) > \sqrt{u}) \, du
\]

\[
= 2 \int_0^{L^2} P(\sup_x g^*(x) > \sqrt{u}) \, du + 2 \int_{L^2}^\infty P(\sup_x g^*(x) > \sqrt{u}) \, du
\]

(\text{where } L = \max (E(\sup_x g^*(x)), 0))

\[
\leq 2L^2 + 2 \int_{L^2}^\infty e^{-\frac{\sqrt{\pi} \cdot u}{2\sigma_g^2}} \, du \quad \text{(using Result 7.4)}
\]

Now, using the change of variable \( \sqrt{u} = z + L \) the integral \( \int_{L^2}^\infty e^{-\frac{\sqrt{\pi} \cdot u}{2\sigma_g^2}} \) can be reduced to the form

\[
\int_{L^2}^\infty e^{-\frac{\sigma_g^2 z^2}{2\sigma_g^2}} \, dz + 2L \int_{L^2}^\infty e^{-\frac{\sigma_g^2 z^2}{2\sigma_g^2}} \, dz
\]

\[
= 2\sigma_g^2 + L\sigma_g(\sqrt{2\pi})
\]

Hence, \( \text{Var}((g^*(X(s, t - 1)))) \leq 2L^2 + 4\sigma_g^2 + 2L\sigma_g(\sqrt{2\pi}) \).

But, \( 0 \leq L \leq K \int_0^{\text{diam}(T)} \sqrt{H(e)} \, de \) by Result 7.5 and it is not difficult to see that \( H(e) \) is a decreasing function of \( \sigma_g^2 \). The same is true of \( \text{diam}(T) \) when as a function of \( \sigma_g^2 \). These two facts together permit applicability of the monotone convergence theorem to yield

\[
0 \leq \lim_{\sigma_g^2 \to 0^+} L \leq \lim_{\sigma_g^2 \to 0^+} K \int_0^{\text{diam}(T)} \sqrt{H(e)} \, de \leq \lim_{\sigma_g^2 \to 0^+} K \int_0^\infty \sqrt{H(e)} \, H(\epsilon \leq \text{diam}(T)) \, de
\]

\[
\leq K \int_0^\infty \lim_{\sigma_g^2 \to 0^+} \sqrt{H(e)} \, H(\epsilon \leq \text{diam}(T)) \, de = 0.
\]

31
Similarly, using the bound repeatedly in (16), we obtain

\[ |\text{Cov}(X(s, t), X(s^*, t^*)) - \beta_{1g}^{t-t^*} \text{Cov}(X(s, t^*), X(s^*, t^*))| \leq \frac{\epsilon}{1 - |\beta_{1g}|} \]

Arguing similarly one can also show that for arbitrarily small \( \epsilon > 0, \exists \delta > 0 \) such that \( \text{Cov}(g^*(X(s, t^*-1)), g^*(X(s^*, t^*-1))) < \epsilon \) for \( 0 < \sigma_g^2 < \delta \). Moreover, see that the bound is uniform in \( s \) and \( t \). Since \( |\beta_{1g}| < 1 \), using the bound repeatedly in (15), we obtain

\[ |\text{Cov}(X(s, t), X(s^*, t^*)) - \beta_{1g}^{t-t^*} \text{Cov}(X(s, t^*), X(s^*, t^*))| \leq \frac{\epsilon}{1 - |\beta_{1g}|} \]

Similarly, using the bound repeatedly in (16), we obtain

\[ |\text{Cov}(X(s, t^*), X(s^*, t^*)) - \text{Cov}(X(s, 0), X(s^*, 0)) - \frac{1 - \beta_{1g}^{2(t^*+1)}}{1 - \beta_{1g}^2} |c_\eta(s, s^*)| | \leq \frac{\epsilon}{1 - |\beta_{1g}|} + \frac{\epsilon}{1 - |\beta_{1g}|} \]

Combining them we get

\[ |\text{Cov}(X(s, t), X(s^*, t^*)) - \beta_{1g}^{t-t^*} \text{Cov}(X(s, 0), X(s^*, 0)) - \beta_{1g}^{t-t^*} \frac{1 - \beta_{1g}^{2(t^*+1)}}{1 - \beta_{1g}^2} |c_\eta(s, s^*)| \]

\[ \leq \frac{\epsilon}{1 - |\beta_{1g}|} \left[ 1 + 3|\beta_{1g}|^{t-t^*} \right] \leq \frac{4\epsilon}{1 - |\beta_{1g}|} \]

Now plugging in this approximation in the expression for \( c_g((s, t), (s^*, t^*)) \) we get the desired result. So, Theorem 3.5 is finally proved.

**Proof of Theorem 3.6.** Part (a): From the condition of the theorem it is clear that \( \exists \) a probability space \( (\Omega, \mathcal{F}, P) \) and a set \( A \in \mathcal{F} \) such that \( P(A) = 1 \) and for \( \omega \in A \), \( X(s, 0)(\omega), \eta(s, t)(\omega), \epsilon(s, t)(\omega) \) are continuous functions in \( s \) where \( t = 1, 2, 3 \cdots \) and \( g(x)(\omega), f(x)(\omega) \) are continuous functions in \( x \). Then by the property of composition of two functions \( X(s, 1)(\omega) = g(X(s, 0)(\omega))(\omega) + \eta(s, 1)(\omega) \) is a continuous function in \( s \). Proceeding recursively, one can prove that \( X(s, t)(\omega) \) is a continuous function in \( s \) for any \( t \). Once we show \( X(s, t)(\omega) \) is a continuous function, we prove \( Y(s, t)(\omega) = f(X(s, t)(\omega))(\omega) + \epsilon(s, t)(\omega) \) is a continuous function in \( s \). So, part (a) is proved.

Part (b): Proof of part (b) follows the similar lines of the proof of part (a). Firstly, we state a simple lemma.

**Lemma 7.6.** Let us consider two real valued functions \( u(z) \) and \( v(x, y) \) such that both of them are \( k \) times differentiable. Then the composition function \( u(v(x, y)) \) is also \( k \) times differentiable.

**Proof.** Proof of this lemma is basically a generalization of chain rule for multivariate functions and can be found in advanced multivariate calculus textbooks. We give a brief sketch of the proof. First we clarify the term \( k \) times differentiable for the function \( v(x, y) \). It means all mixed partial derivatives of \( v(x, y) \) of order \( k \) exist. We prove the lemma using mathematical induction. Firstly, We show that the lemma is true for \( k = 1 \) and then we show that if the lemma is true for \( k - 1 \) then it must be true for \( k \) as well.
That the lemma is true for \( k = 1 \) easily follows from the chain rule for multivariate functions. Now we prove the second step. By the induction hypothesis the lemma is true for the \( k - 1 \) case and \( u(z) \) and \( v(x, y) \) are \( k \) times differentiable. We want to show that \( u(v(x, y)) \) is also \( k \) times differentiable. Without loss of generality, we consider the mixed partial derivative \( \frac{\partial}{\partial x^k \partial y^{k_2}} (u(v(x, y))) \) where \( k_1 + k_2 = k \) and show that it exists. Observe that the partial derivative is equivalent to \( \frac{\partial}{\partial x^k} \frac{\partial}{\partial y^{k_2 - 1}} (u'(v(x, y))(\frac{\partial}{\partial y} v(x, y))) \) provided it exists. Since, by the induction hypothesis the lemma is true for the \( k - 1 \) case and \( u'(z) \) and \( v(x, y) \) are \( k - 1 \) times differentiable, the composition of them \( u'(v(x, y)) \) is also \( k - 1 \) times differentiable. On the other hand, \( \frac{\partial}{\partial y} v(x, y) \) is also \( k - 1 \) times differentiable. So, the product of them \( u'(v(x, y))(\frac{\partial}{\partial y} v(x, y)) \) is also \( k - 1 \) times differentiable. Hence the partial derivative \( \frac{\partial}{\partial x^k \partial y^{k_2 - 1}} (u'(v(x, y))(\frac{\partial}{\partial y} v(x, y))) \) exists. Equivalently, \( \frac{\partial}{\partial x^k \partial y^{k_2}} (u(v(x, y))) \) exists. Similarly one can prove the existence of other mixed partial derivatives of \( u(v(x, y)) \) of order \( k \). Hence, by induction the proof follows.

Part (b): From the condition of the theorem it is clear that \( \exists \) a probability space \( (\Omega, \mathcal{F}, P) \) and a set \( A \in \mathcal{F} \) such that \( P(A) = 1 \) and for \( \omega \in A \), \( X(s,0)(\omega), \eta(s,t)(\omega), \epsilon(s,t)(\omega) \) are \( k \) times differentiable functions in \( s \) where \( t = 1, 2, 3, \cdots \) and \( g(x)(\omega), f(x)(\omega) \) are \( k \) times differentiable functions in \( x \). Then by the above lemma \( X(s,1)(\omega) = g(X(s,0)(\omega))(\omega) + \eta(s,1)(\omega) \) is a \( k \) times differentiable function in \( s \). The rest of the proof is exactly similar as in part (a).

References

[1] ADLER, R. J. and TAYLOR, J. E. (2007). Random Fields and Geometry. Springer. ISBN-13: 978-0-387-48112-8.

[2] ANDERES, E. B. and STEIN, M. L. (2008). Estimating deformations of isotropic Gaussian random fields on the plane. Ann. Appl. Stat., 36(2): 719-741, 2008. ISSN 0090-5364. doi: 10.1214/09053607000000893. URL [http://www.jstor.org/stable/25464644](http://www.jstor.org/stable/25464644)

[3] BANERJEE, S., CARLIN, B. P. and GELFAND, A. (2004). Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC. ISBN 1-58488-410-X.

[4] BANERJEE, S., GAMERMAN, D. and GELFAND, A. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. Environmetrics, 16(5): 465-479, 2005. ISSN 1099-095X. doi: 10.1002/env.715. URL [http://onlinelibrary.wiley.com/doi/10.1002/env.715/pdf](http://onlinelibrary.wiley.com/doi/10.1002/env.715/pdf)

[5] BRIGGS, B. H. (1968). On the analysis of moving patterns in geophysics I. Correlation analysis. J. Atmos. Terr. Phys., 30(10): 1777-1788, 1968. ISSN 0021-9169. doi: 10.1016/0021-9169(68)90097-4. URL [http://www.sciencedirect.com/science/article/pii/0021916968900974](http://www.sciencedirect.com/science/article/pii/0021916968900974)

[6] BRUNO, F., GUTTORP, P., SAMPSON, P. D. and COCCHI, D. (2009). A simple non-separable, non-stationary spatiotemporal model for ozone. Environ. Ecol. Stat., 16(4): 515-529, 2009. ISSN 1573-3009. doi: 10.1007/s10651-008-0094-8. URL [http://link.springer.com/article/10.1007/s10651-008-0094-8](http://link.springer.com/article/10.1007/s10651-008-0094-8)
COX, D. R. and ISHAM, V. (1988). A Simple Spatial-Temporal Model of Rainfall. *Proc. R. Soc. Lon. Ser. A.*, 415(1849): 317-328, 1988. ISSN 1471-2946. URL [http://www.jstor.org/stable/2398094](http://www.jstor.org/stable/2398094).

CRESSIE, N. and HUANG, H. C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.*, 94(448): 1330-1339, 1999. ISSN 1537-274X. doi: 10.1080/01621459.1999.10473885. URL [http://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1999.10473885](http://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1999.10473885).

DEY, K. K., and BHATTACHARYA, S. (2014a). On Optimal Scaling of Additive Transformation Based Markov Chain Monte Carlo. URL [http://arxiv.org/abs/1307.1446](http://arxiv.org/abs/1307.1446).

DEY, K. K., and BHATTACHARYA, S. (2014b). On Geometric Ergodicity of Additive and Multiplicative Transformation Based Markov Chain Monte Carlo in High Dimensions. URL [http://arxiv.org/pdf/1312.0915.pdf](http://arxiv.org/pdf/1312.0915.pdf).

DOU, Y., LE, N. D., and ZIDEK, V. (2010). Modeling hourly ozone concentration fields. *Ann. Appl. Stat.*, 4(3): 1183-1213, 2010. ISSN 1932-6157. doi: 10.1214/09-AOAS318. URL [http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdfview_1&handle=euclid.aoas/1287409369](http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdfview_1&handle=euclid.aoas/1287409369).

DUTTA, S. and BHATTACHARYA, S. (2014). Markov chain monte carlo based on deterministic transformations. *Statistical Methodology*, 16: 100-116, 2014. ISSN 1572-3127. URL [http://www.sciencedirect.com/science/article/pii/S1572312713000683](http://www.sciencedirect.com/science/article/pii/S1572312713000683).

FONSECA, T. C. O. and STEEL, M. F. J. (2011). Non-Gaussian spatiotemporal modelling through scale mixing. *Biometrika*, 98(4): 761-774, 2011. ISSN 1464-3510. doi: 10.1093/biomet/asr047. URL [http://biomet.oxfordjournals.org/content/98/4/761.abstract](http://biomet.oxfordjournals.org/content/98/4/761.abstract).

FUENTES, M. (2001). A high frequency kriging approach for non-stationary environmental processes. *Environmetrics*, 12(5): 469-483, 2001. ISSN 1099-095X. doi: 10.1002/env.473. URL [http://onlinelibrary.wiley.com/doi/10.1002/env.473/pdf](http://onlinelibrary.wiley.com/doi/10.1002/env.473/pdf).

FUENTES, M., CHEN, L., and DAVIS, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics*, 19(5): 487-507, 2008. ISSN 1099-095X. doi: 10.1002/env.891. URL [http://onlinelibrary.wiley.com/doi/10.1002/env.891/pdf](http://onlinelibrary.wiley.com/doi/10.1002/env.891/pdf).

FUENTES, M., CHEN, L., DAVIS, J. M. and LACKMANN, G. (2003). A new class of nonseparable and nonstationary covariance models for wind fields. Technical Report. URL [http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA515693](http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA515693).

FUENTES, M. (2002). Spectral methods for non stationary spatial processes. *Biometrika*, 89(1): 197-210, 2002. ISSN 1464-3510. doi: 10.1093/biomet/89.1.197. URL [http://biomet.oxfordjournals.org/content/89/1/197.full.pdf+html](http://biomet.oxfordjournals.org/content/89/1/197.full.pdf+html).

FUENTES, M. and SMITH, R. L. (2001). A new class of nonstationary spatial models. Technical Report, North Carolina State University (2001). doi: 10.1.1.7.6794. URL [http://www.stat.unc.edu/postscript/rs/nonstat.pdf](http://www.stat.unc.edu/postscript/rs/nonstat.pdf).
[19] FURRER, R. and SAIN, S. R. (2009). Spatial model fitting for large datasets with applications to climate and microarray problems. *Stat. Comput.*, 19(2): 113-128, 2009. ISSN 1573-1375. doi: 10.1007/s11222-008-9075-x. URL [http://www.springerlink.com/content/203124p861356383/](http://www.springerlink.com/content/203124p861356383/)

[20] GELFAND, A. E., DIGGLE, P. J., FUENTES, M., and GUTTORP, P. (2010). *Handbook of Spatial Statistics*. CRC Press. ISBN 978-1-4200-7287-7. URL [http://www.crcpress.com/product/isbn/9781420072877](http://www.crcpress.com/product/isbn/9781420072877).

[21] GHOSH, A., MUKHOPADHYAY, S., ROY, S., and BHATTACHARYA, S. (2014). Bayesian Inference in Nonparametric Dynamic State-Space Models. *Statistical Methodology*, 21: 35-48, 2014. doi: http://dx.doi.org/10.1016/j.stamet.2014.02.004. URL [http://www.sciencedirect.com/science/article/pii/S1572312714000197](http://www.sciencedirect.com/science/article/pii/S1572312714000197).

[22] GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.*, 97(458): 590-600, 2002. ISSN 1537-274X. doi: 10.1198/016214502760047113. URL [http://www.jstor.org/stable/3085674?origin=JSTOR-pdf](http://www.jstor.org/stable/3085674?origin=JSTOR-pdf).

[23] GUTTORP, P., MEIRING, W. and SAMPSON, P. D. (1994). A space-time analysis of ground-level ozone data. *Environmetrics*, 5(3): 241-254, 1994. ISSN 1099-095X. doi: 10.1002/env.3170050305. URL [http://onlinelibrary.wiley.com/doi/10.1002/env.3170050305/pdf](http://onlinelibrary.wiley.com/doi/10.1002/env.3170050305/pdf).

[24] HIGDON, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environ. Ecol. Stat.*, 5(2): 173-190, 1998. ISSN 1352-8505. doi: 10.1023/A:1009666805688. URL [http://www.springerlink.com/content/g3v4441h61qu1v60/fulltext.pdf](http://www.springerlink.com/content/g3v4441h61qu1v60/fulltext.pdf).

[25] HIGDON, D. (2002). Space and space-time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues*. Springer-Verlag. ISBN 978-1-85233-294-5.

[26] HIGDON, D., SWALL, J. and KERN, J. (1999). Nonstationary spatial modeling. *Bayesian Statistics 6*. Oxford:Clarendon. ISBN 978-0-1985-0485-6.

[27] HUERTA, G., SANSÓ, B. and STROUD, J. R. (2004)). A spatiotemporal model for Mexico City ozone levels. *Appl. Statist.*, 53(2): 231-248, 2004. ISSN 1467-9876. doi: 10.1046/j.1467-9876.2003.01100.x. URL [http://onlinelibrary.wiley.com/doi/10.1046/j.1467-9876.2003.01100.x/pdf](http://onlinelibrary.wiley.com/doi/10.1046/j.1467-9876.2003.01100.x/pdf).

[28] JUN, M., KNUTTI, R. and NYCHKA, D. W. (2008). Spatial analysis to quantify numerical model bias and dependence: How many climate models are there? *J. Amer. Statist. Assoc.*, 103(483): 934-947, 2008. ISSN 1537-274X. doi: 10.1198/016214507000001265. URL [http://www.tandfonline.com/doi/abs/10.1198/016214507000001265](http://www.tandfonline.com/doi/abs/10.1198/016214507000001265).

[29] KIM, H. M. and MALLICK, B. K. (2004). A Bayesian prediction using the skew Gaussian distribution. *J. Stat. Plan. Infer.*, 120(1-2): 85-101, 2004. ISSN 0378-3758. doi: 10.1016/S0378-3758(02)00501-3. URL [http://www.sciencedirect.com/science/article/pii/S0378375802005013](http://www.sciencedirect.com/science/article/pii/S0378375802005013).
[30] LOPES, H. F., SALAZAR, E. and GAMERMAN, D. (2008). Spatial Dynamic Factor Analysis. *Bayesian Anal.*, 3(4): 759-792, 2008. ISSN 1931-6690. doi: 10.1214/08-BA329. URL [http://ba.stat.cmu.edu/journal/2008/vol03/issue04/lopes.pdf](http://ba.stat.cmu.edu/journal/2008/vol03/issue04/lopes.pdf)

[31] MAJUMDAR, A. and GELFAND, A. E. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Math. Geol.*, 39(2): 225-245, 2007. ISSN 1573-8868. doi: 10.1007/s11004-006-9072-6. URL [http://link.springer.com/article/10.1007/s11004-006-9072-6#page-1](http://link.springer.com/article/10.1007/s11004-006-9072-6#page-1)

[32] MAJUMDAR, A., PAUL, D. and BAUTISTA, D. (2010). A generalized convolution model for multivariate nonstationary spatial processes. *Stat. Sinica.*, 20(2): 675-695, 2010. ISSN 1017-0405. URL [http://www3.stat.sinica.edu.tw/statistica/oldpdf/A20n211.pdf](http://www3.stat.sinica.edu.tw/statistica/oldpdf/A20n211.pdf)

[33] ØKSENDAL, B. (2000). *Stochastic Differential Equations. 5th Edition*. Springer-Verlag. Hiedelberg, New York.

[34] PACIOREK, C. J. and SCHERVISH, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5): 483-506, 2006. ISSN 1099-095X. doi: 10.1002/env.785. URL [http://onlinelibrary.wiley.com/doi/10.1002/env.785/pdf](http://onlinelibrary.wiley.com/doi/10.1002/env.785/pdf)

[35] PALACIOS, M. B. and STEEL, M. F. J. (2006). Non-Gaussian Bayesian Geostatistical Modeling. *J. Amer. Statist. Assoc.*, 101(474): 604-618, 2006. ISSN 1537-274X. doi: 10.1198/016214505000001195. URL [http://www.tandfonline.com/doi/abs/10.1198/016214505000001195#.UjrXM0nJXQo](http://www.tandfonline.com/doi/abs/10.1198/016214505000001195#.UjrXM0nJXQo)

[36] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press. ISBN 0-262-18253-X.

[37] SAIN, S. R., FURRER, R. and CRESSIE, N. (2011). A spatial analysis of multivariate output from regional climate models. *Ann. Appl. Stat.*, 5(1): 150-175, 2011. ISSN 1932-6157. doi: 10.1214/10-AOAS369. URL [http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aaos/1300715186](http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aaos/1300715186)

[38] SAMPSON, P. D., and GUTTORP, P. (1992). Nonparametric Estimation of Nonstationary Spatial Covariance Structure. *J. Amer. Statist. Assoc.*, 87(417): 108-119, 1992. ISSN 1537-274X. doi: 10.1080/01621459.1992.10475181. URL [http://www.jstor.org/stable/2290458](http://www.jstor.org/stable/2290458)

[39] SANG, H., JUN, M. and HUANG, J. Z. (2011). Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors. *Ann. Appl. Stat.*, 5(4): 2519-2548, 2011. ISSN 1932-6157. doi: 10.1214/11-AOAS478. URL [http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aaos/1324399605](http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aaos/1324399605)

[40] SCHMIDT, A. M., and O’HAGAN, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *J. Roy. Statist. Soc. Ser. B*, 65(3): 743-758, 2003. ISSN 1467-9868. doi: 10.1111/1467-9868.00413. URL [http://www.jstor.org/stable/3647549](http://www.jstor.org/stable/3647549)

[41] SMITH, R. L., TEBALDI, C., NYCHKA, D. and MEARNS, L. O. (2009). Bayesian modeling of uncertainty in ensembles of climate models. *J. Amer. Statist. Assoc.*, 104(485): 97-116, 2009. ISSN 1537-274X. doi: 10.1198/jasa.2009.0007. URL [http://amstat.tandfonline.com/doi/abs/10.1198/jasa.2009.0007](http://amstat.tandfonline.com/doi/abs/10.1198/jasa.2009.0007)
[42] STROUD, J. R., MÜLLER, P. and SANSÓ, B. (2001). Dynamic models for spatiotemporal data. *J. Roy. Statist. Soc. Ser. B*, 63(4): 673-689, 2001. ISSN 1467-9868. doi: 10.1111/1467-9868.00305. URL [http://www.jstor.org/stable/2680659](http://www.jstor.org/stable/2680659).

[43] MEIRING, W., MONESTIEZ, P., SAMPSON, D. P. and GUTTORP, P. (1997). Developments in the modelling of nonstationary spatial covariance structure from space-time monitoring data. *Geostatistics Wollongong ’96*. Springer. ISBN 978-0-7923-4496-4.

[44] OLIVEIRA, V. D., KEDEM, B. and SHORT, D. A. (1997). Bayesian Prediction of Transformed Gaussian Random Fields. *J. Amer. Statist. Assoc.*, 92(440): 1422-1433, 1997. ISSN 1537-274X. doi: 10.1080/01621459.1997.10473663. URL [http://www.tandfonline.com/doi/abs/10.1080/01621459.1997.10473663#.UjrV0EnJXQo](http://www.tandfonline.com/doi/abs/10.1080/01621459.1997.10473663#.UjrV0EnJXQo).

[45] TEBALDI, C. and KNUTTI, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 365: 2053-2075, 2007. ISSN 1471-2962. doi: 10.1098/rsta.2007.2076. URL [http://rsta.royalsocietypublishing.org/content/365/1857/2053](http://rsta.royalsocietypublishing.org/content/365/1857/2053).

[46] TEBALDI, C. and SANSÓ, B. (2009). Joint projections of temperature and precipitation change from multiple climate models: A hierarchical Bayesian approach. *J. Roy. Statist. Soc. Ser. A*, 172(1): 83-106, 2009. ISSN 1467-985X. doi: 10.1111/j.1467-985X.2008.00545.x. URL [http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2008.00545.x/abstract](http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2008.00545.x/abstract).

[47] WEST, M. and HARRISON, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag. ISBN 978-0-387-94725-9. doi: 10.1007/b98971. URL [http://www.springerlink.com/content/978-0-387-94725-9](http://www.springerlink.com/content/978-0-387-94725-9).

[48] ZHU, Z. and WU, Y. (2012). Estimation and prediction of a class of convolution based spatial non-stationary models for large spatial data. *J. Comput. Graph. Stat.*, 19(1): 74-95, 2012. ISSN 1537-2715. doi: 10.1198/jcgs.2009.07123. URL [http://amstat.tandfonline.com/doi/pdf/10.1198/jcgs.2009.07123](http://amstat.tandfonline.com/doi/pdf/10.1198/jcgs.2009.07123).