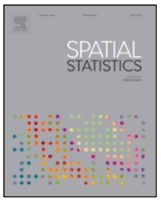
Spatial Statistics 55 (2023) 100751

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



A kernel-enriched order-dependent nonparametric spatio-temporal process

Moumita Das [a,*], Sourabh Bhattacharya [b]

[a] Basque Center For Applied Mathematics, Spain
[b] Indian Statistical Institute, Kolkata, India

ARTICLE INFO

Article history:
Received 20 April 2022
Received in revised form 13 April 2023
Accepted 13 April 2023
Available online 26 April 2023

Keywords:
Kernel convolution
Nonstationary
Nonparametric
Order based dependent Dirichlet process
Spatio-temporal data
Transdimensional Transformation based Markov Chain Monte Carlo

ABSTRACT

Spatio-temporal processes are necessary modeling tools for various environmental, biological, and geographical problems. The underlying model is commonly considered to be parametric and to be a Gaussian process. Additionally, the covariance function is expected to be stationary and separable. This structure need not always be realistic. Moreover, attempts have been made to construct nonparametric processes of neither stationary nor separable covariance functions. Nevertheless, as we elucidate, some desirable and necessary spatio-temporal properties are not guaranteed by the existing approaches, thus, calling for further innovative ideas. In this article, using kernel convolution of order-based dependent Dirichlet process, we construct a novel spatio-temporal model. We show that this satisfies desirable properties and includes the stationary, separable, parametric processes as special cases. Our resultant posterior distribution is variable dimensional, which we attack using Transdimensional Transformation based Markov Chain Monte Carlo, which can update all the variables and change dimensions using deterministic transformations of a random variable drawn from some arbitrary density defined on relevant support. We demonstrate our model's performance on simulated and real data sets. In all situations, the findings are highly encouraging.

* Corresponding author.
E-mail addresses: moumita.statistics@gmail.com (M. Das), bhsourabh@gmail.com (S. Bhattacharya).

https://doi.org/10.1016/j.spasta.2023.100751
2211-6753/

1. Introduction

Recent years have witnessed considerable research on spatial and spatio-temporal modeling. The major inferential objectives of spatio-temporal modeling are to predict a plausible value at some point in space and time, forecasting the future value at some location, and to make inferences about the parameters of the spatio-temporal processes. A model must account for the given process's spatio-temporal dependence structure. It is a common practice to assume that the underlying spatial or spatio-temporal process is stationary and isotropic Gaussian process, as it facilitates prediction. In particular, the geostatistical method of kriging assumes a Gaussian process structure for the unknown spatial or spatio-temporal field. It focuses on calculating the optimal linear predictor of the field. Researchers generally assume a stationary, often isotropic, covariance function when performing kriging. The covariance of responses at any two locations is assumed to be a function of the separation vector or of the distance between locations but not a function of the actual locations. Researchers often estimate the parameters of an isotropic covariance function from the semivariogram, the estimation of which is based on the squared differences between the responses as a function of the distance between locations. The standard kriging approach allows one to estimate a smooth spatial field flexibly, with no pre-specified parametric stochastic model for the data. However, these approaches have several drawbacks. The most important is that the actual covariance structure may not be stationary. This is because local influences may affect the random process's correlation structure. For instance, orographic effects influence the atmospheric transport of pollutants and result in a correlation structure that depends on different spatial locations (Guttorp and Sampson, 1994). If one is modeling an environmental variable across the United States, the field will likely be much smoother in the topographically-challenged Great Plains than in the Rocky Mountains. This is manifested as different covariance structures in those two regions. Assuming a stationary covariance structure will result in over-smoothing the field in the mountains and under-smoothing the field in great plains (Paciorek, 2003).

Realizing the limitations of stationary parametric processes (almost invariably Gaussian processes), researchers have come up with many novel ideas for constructing nonstationary and/or nonparametric processes. The first significant work in the framework of nonstationary parametric processes is by Sampson and Guttorp (1992), who proposed an approach based on spatial deformation. This work is followed up by Damian et al. (2001) and Schmidt and O'Hagan (2003), providing the corresponding Bayesian generalizations. Nonstationarity has been induced in parametric space–time models by Haas (1995), by proposing a moving window regression residual kriging. A similar approach has been proposed by Nott and Dunsmuir (2002). Higdon (1998) (see also Higdon et al. (1999), Higdon (2001)) proposed a kernel convolution approach for inducing nonstationarity in Gaussian processes. Similar approaches are also proposed by Fuentes and Smith (2001) and Fuentes (2002). Approaches that attempt to model the underlying process as nonparametric, in addition to modeling the covariance structure as nonstationary, are more recent in comparison, the approach of Gelfand et al. (2005) based on Dirichlet processes (see, for example, Ferguson (1973), Ferguson (1974)) being the first in this regard; see Duan et al. (2007) for a generalization. Duan et al. (2009) use stochastic differential equations to construct a nonstationary, non-Gaussian process. We discuss these proposals in some detail in Section 1.1.

Fuentes and Reich (2013) proposed a nonparametric nonstationary model based on kernel processes mixing. In their study, they showed that their proposed model outperformed all other models for several types of simulation designs (Stationary Gaussian, Nonstationary Gaussian, Stationary Non-Gaussian, Nonstationary Non-Gaussian). They illustrated their model with application to the monthly average values of ammonium and nitrate at 209 monitoring stations in the US. Their proposed nonstationary non-Gaussian model reduced the root mean square error (RMSE) by 24% for ammonium and 18% for nitrate when compared to the nonstationary Gaussian approach. RMSE is also reduced compared to stationary Gaussian and stationary non-Gaussian approaches, although the gain is more moderate in these cases.

Griffin and Steel (2006) (henceforth GS) proposed the novel order-based dependent Dirichlet processes (ODDP). They introduced a framework for nonparametric modeling with dependence on continuous covariates. Dependence is induced through relevant weights utilizing similarities in the

covariate information. Each weight is a transformation of independently and identically distributed (*iid*) random variables. GS derived an ordering π of these random variables at each covariate value such that distributions for similar covariate values are associated with similar orderings and thus will be close. These orderings combined with the Poisson point process, give a simple analytical expression for the correlation function of the distributions, which ensures that if two points are similar in the covariate space, they will get a higher correlation compared to the points that are not. Furthermore, when the distance between two points is large enough in the covariate space, the correlation approaches zero. In spatial/spatio-temporal context, it translates into the fact that when two observations are widely separated in space/space–time, the model based correlations tend to zero. Nevertheless, the ODDP process suffers from the limitation of being stationary.

Preserving all the desirable properties of the correlation function of ODDP, we attempt to incorporate further flexibility in our spatial/temporal/spatio-temporal model in terms of nonstationarity and nonseparability through our proposed kernel convolution-based methodology. Specifically, we propose a new class of spatial/temporal/spatio-temporal models that is nonparametric, nonstationary, nonseparable, and such that the correlation tends to zero if either of spatial and temporal distance tends to infinity. All these properties are desirable in real data scenarios.

In the context of real spatio-temporal data set on particulate matter, we show that empirical correlations tend to zero as the spatio-temporal lag increase. Nonstationarity of this data set is also inferred in a separate paper by Roy and Bhattacharya (2020). Furthermore, another data set on sea surface temperature used in Bhattacharya (2021) also exhibits lagged correlations tending to zero with increasing spatio-temporal lags despite nonstationarity. Moreover, these data sets are far from Gaussianity, as simple quantile–quantile plots indicate. Hence, this paper attempts to create a class of realistic stochastic processes to address these desirable properties in real-life data sets. In order to provide additional motivation, we compared our analysis to one of the competent existing models, namely, Fuentes and Reich (2013). Their proposed model is nonstationary, and nonparametric but does not guarantee those correlations fall to zero with increasing lag. Our model better captures the correct correlation structure, especially when the actual correlation is close to zero for nonstationary models. When considering all of these characteristics, our approach may be worthwhile in modeling nonstationary spatio-temporal data.

The rest of our paper is structured as follows. In Section 1.1, we provide a brief overview of the existing approaches to construction of nonstationary, nonseparable space–time processes in both parametric and nonparametric frameworks, arguing that not all desirable properties are necessarily accounted for in these approaches. Such issues necessitate development of new approaches to the construction of nonstationary, nonparametric, nonseparable space–time models. In Section 2.1 we introduce our proposed space–time model based on kernel convolution of ODDP, and show that it satisfies the properties that the existing models do not guarantee. We investigate our model's continuity and smoothness properties in Section 2.3. Since our proposed model involves an infinite random series, one needs to either truncate the series or assume a random number of summands and adopt variable dimensional Markov Chain Monte Carlo (MCMC) approaches for model fitting. Although we adopt the latter framework for our applications and implement the recently developed Transdimensional Transformation based Markov Chain Monte Carlo (TTMCMC) (Das and Bhattacharya, 2019b) for simulating from our variable dimensional model, for the sake of completeness, we also investigate the truncation approach. Indeed, in Section 2.4, we consider the difference between the prior predictive models with and without truncation of the infinite random series, providing a bound that depends upon the truncation parameter. Thus, the truncation parameter can be chosen so that the bound falls below any desired level. Section 2.5 discusses the choice of suitable kernels and prior distributions. Furthermore, the choice of the spatio-temporal domain is relevant for the computational purpose. We describe the joint posterior distribution associated with our model and briefly discuss TTMCMC in Section 2.6. We detail a simulation study illustrating the performance of our model and comparison with Fuentes and Reich (2013) in Section 3.1. Indeed, the model of Fuentes and Reich (2013), despite being very different from our ideas, comes closest to our model conceptually among the existing models. In Section 3.4 we consider applying our ideas to two real data sets: a spatial ozone data set and a spatio-temporal data set on particulate matters. Finally, we summarize our contributions and provide concluding remarks in Section 4.

Proofs of our results and requisite details of TTMCMC, particularly in the context of our spatio-temporal model, and details regarding the generation of the data for the simulation experiment, are provided in the supplement Das and Bhattacharya (2019a), whose sections and algorithms have the prefix "S-" when referred to in this paper.

1.1. Overview of other available nonstationary approaches

1.1.1. Parametric approaches

The deformation approaches of Sampson and Guttorp (1992), Damian et al. (2001), and Schmidt and O'Hagan (2003) are based on Gaussian processes. In these approaches replications of the data are necessary, which the authors relate to temporal independence of the data. This also means that space–time data cannot be modeled using these approaches, unless all the temporal dependence can be captured through a trend term in the mean structure. Moreover, in the deformation-based approaches model based theoretical correlations between random observations separated by large enough distances need not necessarily tend to zero. Letting $Y(\boldsymbol{s}, t)$ denote the response at spatial location $\boldsymbol{s}$ and time t, Sampson and Guttorp (1992) deal with the variogram of the following form:

$$\text{Var}(Y(\boldsymbol{s}_1, t) - Y(\boldsymbol{s}_2, t)) = f(\|\boldsymbol{d}(\boldsymbol{s}_1, t) - \boldsymbol{d}(\boldsymbol{s}_2, t)\|), \tag{1.1}$$

for any $\boldsymbol{s}_1, \boldsymbol{s}_2, t$, where f is an appropriate monotone function and $\boldsymbol{d}$ is a one-to-one nonlinear mapping. The technique of Sampson and Guttorp (1992) involves appropriately approximating f by $\hat{f}$ using the multidimensional scaling method, and obtaining a configuration of points $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ in a "deformed" space, where the process is assumed isotropic. Then using thin-plate splines, a nonlinear approximation of $\boldsymbol{d}$, which we denote by $\hat{\boldsymbol{d}}$, is determined such that $\hat{\boldsymbol{d}}(\boldsymbol{s}_i) \approx \boldsymbol{u}_i$, for $i = 1, \ldots, n$. Bayesian versions of the key idea have been described in Damian et al. (2001), who use random thin-plate splines and Schmidt and O'Hagan (2003), who use Gaussian process to implement the nonlinear transformation $\boldsymbol{d}$. Rather than estimating f nonparametrically, both specify a parametric functional form from a valid class of such monotone functions.

As is clear, since large differences $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ does not imply that $\|\boldsymbol{d}(\boldsymbol{s}_1) - \boldsymbol{d}(\boldsymbol{s}_2)\|$ is also large, the model based correlations between two observations widely separated need not necessarily tend to zero, in either of the aforementioned deformation-based approaches.

The kernel convolution approaches of Higdon et al. (1999), Higdon (2001), and Fuentes and Smith (2001) overcome some of the difficulties of the deformation approach. In these approaches data replication is not necessary, and for appropriate choices of the kernel, stationarity, nonstationarity, separability, and nonseparability can be achieved with respect to spatio-temporal data. In the approach of Higdon et al. (1999) and Higdon (2001),

$$Y(\boldsymbol{x}) = \int K(\boldsymbol{x}, \boldsymbol{u}) Z(\boldsymbol{u}) d\boldsymbol{u}, \tag{1.2}$$

where K is a kernel function and $Z(\cdot)$ is a white noise process. Then the covariance between $Y(\boldsymbol{x}_1)$ and $Y(\boldsymbol{x}_2)$ is given by

$$C(\boldsymbol{x}_1, \boldsymbol{x}_2) = \int K(\boldsymbol{x}_1, \boldsymbol{u}) K(\boldsymbol{x}_2, \boldsymbol{u}) d\boldsymbol{u}. \tag{1.3}$$

In general, this does not depend upon $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ only through $\boldsymbol{x}_1 - \boldsymbol{x}_2$, thus achieving nonstationarity. However, it is clear from the covariance structure (1.3) that $C(\boldsymbol{x}_1, \boldsymbol{x}_2)$ does not generally tend to zero as $d = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \to \infty$, except under seperability, with the additional assumption of isotropy with respect to either space or time.

The approach of Fuentes and Smith (2001), which may be thought of as a generalization of Higdon et al. (1999), comes close towards solving the problem of zero covariance in the limit with large enough separation between observations, but still unfortunately fails to achieve this desirable property.

A nonstationary process has been constructed by Chang et al. (2011), by representing the underlying process as a linear combination of basis functions and stationary Gaussian processes. This approach also does not guarantee that the correlation tends to zero if $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \to \infty$. For other available parametric approaches to nonstationarity we refer to the references provided in Chang et al. (2011).

1.1.2. Nonparametric approaches

Gelfand et al. (2005) seem to be the first to propose a nonstationary, noparametric Bayesian model based on Dirichlet process mixing. They represent the random field $\boldsymbol{Y}_D = \{Y(\boldsymbol{x}); \boldsymbol{x} \in D\}$ as $\sum_{\ell=1}^{\infty} w_\ell \delta_{\boldsymbol{\theta}_{\ell,D}}$, where $\boldsymbol{\theta}_{\ell,D} = \{\theta_\ell(\boldsymbol{x}); \boldsymbol{x} \in D\}$ are realizations from a specified stationary Gaussian process, which we denote as $\boldsymbol{G}_0$, $w_1 = V_1$, $w_\ell = V_\ell \prod_{r=1}^{\ell-1}(1-V_r)$ for $\ell \geq 2$, where $V_r \overset{iid}{\sim} Beta(1,\alpha)$; $r = 1, 2, \ldots$. Thus, a random process $\boldsymbol{G}$ is induced on the space of processes of $\boldsymbol{Y}_D$ with $\boldsymbol{G}_0$ being the "central" process. Gelfand et al. (2005) assume the space–time data $\boldsymbol{Y}_t = (Y(\boldsymbol{s}_1, t), \ldots, Y(\boldsymbol{s}_n, t))'$ to be time-independent for $t = 1, \ldots, T$, which is the same assumption of data replication used in the deformation-based approaches. The temporal-independence assumption allows Gelfand et al. (2005) to model the data as follows: for $t = 1, \ldots, T$, $\boldsymbol{Y}_t \overset{iid}{\sim} \boldsymbol{G}^{(n)}$ and $\boldsymbol{G}^{(n)} \sim DP(\boldsymbol{G}_0^{(n)})$, where $\boldsymbol{G}^{(n)}$ and $\boldsymbol{G}_0^{(n)}$ denote the n-variate distributions corresponding to the processes $\boldsymbol{G}$ and $\boldsymbol{G}_0$. The development leads to the following covariance structure: for any $\boldsymbol{s}_1, \boldsymbol{s}_2, t$,

$$\mathrm{Cov}(Y(\boldsymbol{s}_1, t), Y(\boldsymbol{s}_2, t) \mid \boldsymbol{G}) = \sum_{\ell=1}^{\infty} w_\ell \theta_\ell(\boldsymbol{s}_1)\theta_\ell(\boldsymbol{s}_2) - \left\{\sum_{\ell=1}^{\infty} w_\ell \theta_\ell(\boldsymbol{s}_1)\right\}\left\{\sum_{\ell=1}^{\infty} w_\ell \theta_\ell(\boldsymbol{s}_2)\right\}, \tag{1.4}$$

which is nonstationary. However, marginalized over $\boldsymbol{G}$, the covariance between $Y(\boldsymbol{s}_1, t)$ and $Y(\boldsymbol{s}_2, t)$ turns out to be stationary. Since, in Gelfand et al. (2005), the Bayesian inference of the data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ proceeds by integrating out $\boldsymbol{G}^{(n)}$, the entire flavor of nonstationarity is lost. Also, given $\boldsymbol{G}$, (1.4) is nonstationary but does not necessarily converge to zero if $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \to \infty$.

Duan et al. (2007) attempt to generalize the model of Gelfand et al. (2005) by specifying $\boldsymbol{G}$ as

$$Pr\{Y(\boldsymbol{x}_1) \in A_1, \ldots, Y(\boldsymbol{x}_n) \in A_n\} = \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} p_{i_1,\ldots,i_n} \delta_{\theta_{i_1}(\boldsymbol{x}_1)}(A_1) \cdots \delta_{\theta_{i_n}(\boldsymbol{x}_n)}(A_n), \tag{1.5}$$

where $\boldsymbol{\theta}_j$'s are *iid* $\boldsymbol{G}_0$ as in Gelfand et al. (2005), and $\{p_{i_1,\ldots,i_n} \geq 0 : \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} p_{i_1,\ldots,i_n} = 1\}$ determines the site-specific joint selection probabilities, which also must satisfies simple constraints to ensure consistency. The resulting conditional covariance (conditional on $\boldsymbol{G}$) and the marginal covariance are somewhat modified versions of those of Gelfand et al. (2005), but now even the marginal covariance is nonstationary. By choosing $\boldsymbol{G}_0$ to be an isotropic Gaussian process it can be ensured that the marginal covariance tends to zero as two observations are widely separated, but the same cannot be ensured for the conditional covariance. Moreover, replications of the data is necessary even for this generalized version of Gelfand et al. (2005), and modeling temporal dependence is precluded as before. A methodology very similar to that of Duan et al. (2007) is proposed in Petrone et al. (2009).

Although the aforementioned approaches are temporally independent, Kottas et al. (2007) have considered a first order autoregressive setup to model temporal dependence as a simple parametric temporal extension of the temporally independent model proposed in Gelfand et al. (2005).

A nonstationary, nonseparable non-Gaussian spatiotemporal process has been constructed by Duan et al. (2009) using discretized versions of stochastic differential equations, but again, the correlations between largely separated observations do not necessarily tend to zero under their model. Also, stationarity or separability cannot be derived as special cases of this approach.

A flexible approach using kernel convolution of Lévy random measures has been detailed in Wolpert et al. (2011), but even this approach does not guarantee that correlations tend to zero for largely separated distances for arbitrarily chosen kernels.

An univariate and multivariate nonparametric spatial model based on kernel process mixing has been proposed by Fuentes and Reich (2013) (henceforth FR). In this work, the idea of stick-breaking prior of Sethuraman (1994) was extended to a spatial set up. A different, unknown distribution was assigned to each location, with a series of space-dependent kernel functions that have a space-varying bandwidth parameter. Essentially, the Beta-distributed sequence $\{V_r : r = 1, 2, \ldots\}$ in the stick-breaking construction of the traditional Dirichlet process are multiplied with a sequence of space-dependent kernels $\{K_r(\boldsymbol{s}) : r = 1, 2, \ldots\}$, and the $\boldsymbol{G}_0$-distributed sequence is replaced with an isotropic Gaussian process with nonstationary variance. The kernel functions attempt to impose a natural ranking for the different mixture components based on distances of locations to knots, which

seems to be an alternative way to mimic the role of the orderings imposed in GS. As the bandwidths of the kernels tend to zero uniformly, the covariance conditional on $\{V_r : r = 1, 2, \ldots\}$ tends to the isotropic covariance of the underlying Gaussian process. Marginally, the covariance structure, albeit nonstationary, need not yield zero covariance even if the distance between the locations tend to infinity. Moreover, this idea has been considered only for spatial modeling. Although it is simple to extend the method to spatio-temporal situations, enforcing separability is needed, which does not seem to be as straightforward.

Nonstationarity in both space and time is researched far less thoroughly than it is for modeling nonstationarity in space, which has a large body of literature (Shand and Li, 2017). Recently Shand and Li (2017) proposed a nonstationary space time model using dimension expansion technique. Compared to the vast literature on continuous nonstationary spatio temporal processes, there are very few methods available to model non-smooth covariance structures over the space or both space–time (Guttorp et al., 2013). Among them, Kim et al. (2005) developed a method based on a Bayesian approach to Voronoi tessellation. Since our approach hinges upon the idea of GS, and smoothness properties of the ODDP depends on the order generating process, it is discontinuous in nature. We will discuss in details the smoothness properties of our model in Section 2.3. Another possible source of nonstationarity is the local influence of some covariates on the spatial process of interest. Recently, there have been some proposals in the literature that account for covariate information in the covariance structure of spatial and spatio-temporal processes; see, for example, Reich et al. (2011), Schmidt et al. (2011), Neto et al. (2014), Ingebrigtsen et al. (2014), Risser and Calder (2015), Gilani et al. (2016), Risser et al. (2019). Since in our model we introduce dependence via the ODDP, where weights in the Sethuraman representation are dependent on the covariate information, we can efficiently incorporate the local influence of covariate information into our model. The covariate information can also be incorporated in the kernel that we convolve the ODDP with.

More recent works on spatial and spatio-temporal models focus on the large data context, and most of them are Gaussian process models. Spatial random effects model (Cressie and Johannesson, 2008; Katzfuss, 2013; Nychka et al., 2015) and predictive process models (Banerjee et al., 2008; Ren and Banerjee, 2013) were proposed to make the computation feasible in modeling nonstationary spatial data. For big spatial and spatio temporal data, Banerjee (2017) discusses methods based on Gaussian processes with a focus on low rank based models and methods based on sparse covariance matrices. The potential statistical inefficiency of low rank models for spatial interpolation was illustrated (Stein, 2014).

Very recently Bhattacharya (2021) introduce a new spatial/spatio-temporal model built upon Lévy processes. Another model based on Hamiltonian equations of physics has been introduced by Mazumder et al. (2022). The goal of all the models the authors and their collaborators put out is to create a class of nonstationary spatial/spatio-temporal models that are computationally feasible and have desirable properties. In the next section we introduce our idea based on kernel convolution of ODDP and show that it overcomes the issues faced by the traditional approaches to construction of flexible space–time models.

2. Methods

2.1. Kernel convolution of ODDP

Before introducing our proposal, We will give a brief introduction of ODDP. ODDP hinges upon the idea of Dirichlet process prior introduced by Ferguson (1973). Dirichlet process are overwhelmingly used as the prior for the unknown distribution. However, the Dirichlet process cannot be used to establish the relationship between covariates and unknown distribution if one must model the data with covariates.We briefly introduce the concept of stick-breaking representation of Dirichlet process proposed by Sethuraman (1994) before delving deeper into the introduction to ODDP.

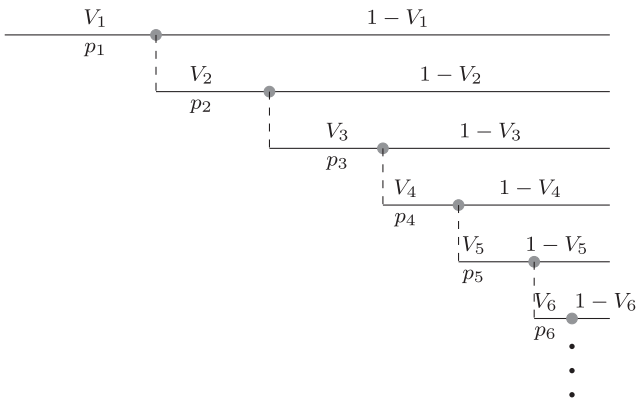


Fig. 2.1. Stick Breaking Representation of Dirichlet Process.

2.2. Stick breaking representation of Dirichlet process

$$G \overset{\mathcal{D}}{=} \sum_{i=1}^{\infty} p_i \delta_{\boldsymbol{\psi}_i}, \tag{2.1}$$

where

$$p_i = V_i \prod_{j<i} (1 - V_j). \tag{2.2}$$

$$V_i \overset{iid}{\sim} Beta(1, \alpha)$$

$$\boldsymbol{\psi}_i \overset{iid}{\sim} G_0$$

This representation of Dirichlet process is known as a "stick-breaking" representation since it involves breaking off successive "sticks" of length p_i from a single stick of unit length ($\sum_{i=1}^{\infty} p_i = 1$) (Fig : 2.1).

2.2.1. Overview of ODDP

In order to induce spatial dependence between observations at different locations GS modifies the nonparametric stick-breaking construction of Sethuraman (1994) in the following way:

for each point $\boldsymbol{x} \in D$, where D is some specified domain, they define the distribution:

$$G_{\boldsymbol{x}} \overset{\mathcal{D}}{=} \sum_{i=1}^{\infty} p_i(\boldsymbol{x}) \delta_{\boldsymbol{\psi}_{\pi_i(\boldsymbol{x})}}, \tag{2.3}$$

where

$$p_i(\boldsymbol{x}) = V_{\pi_i(\boldsymbol{x})} \prod_{j<i} (1 - V_{\pi_j(\boldsymbol{x})}). \tag{2.4}$$

In (2.3) and (2.4), $\boldsymbol{\pi}(\boldsymbol{x}) = (\pi_1(\boldsymbol{x}), \pi_2(\boldsymbol{x}), \ldots)$ denotes the ordering at $\boldsymbol{x}$, where $\pi_i(\boldsymbol{x}) \in \{1, 2, \ldots\}$ and $\pi_i(\boldsymbol{x}) = \pi_j(\boldsymbol{x})$ if and only if $i = j$. For $j = 1, 2, \ldots$, the parameters $\boldsymbol{\psi}_j \overset{iid}{\sim} G_0$, where G_0 is some specified parametric centering distribution, and $V_j \overset{iid}{\sim} Beta(1, \alpha)$, where $\alpha > 0$ is a specified parameter. The process associated with specification (2.1) is the ODDP. Clearly, if $\pi_i(\boldsymbol{x}) = i$ for each $\boldsymbol{x}$ and i, then the Dirichlet process (DP) results at all locations.

GS constructs $\boldsymbol{\pi}(\boldsymbol{x})$ in a way such that it is associated with the realization of a point process. Specifically, they consider a stationary Poisson process Φ and a sequence of sets $U(\boldsymbol{x})$ for $\boldsymbol{x} \in D$, the

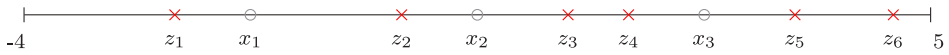


Fig. 2.2. Three covariate values x_1, x_2, x_3 with a section of point process in the $U(x) = (-4, 5)$.

latter determining the relevant region for the ordering purpose. In the case of only spatial problems, if $\boldsymbol{x} \in D \subset \mathbb{R}^d$, for $d \geq 1$, then GS suggests $U(\boldsymbol{x}) = D$ for all $\boldsymbol{x} \in D$ as a suitable construction of $U(\boldsymbol{x})$. For time series problems they suggest $D = \mathbb{R}$ and $U(x) = (-\infty, x]$. When $\boldsymbol{x} = (\boldsymbol{s}', t)'$, that is, when $\boldsymbol{x}$ consists of both spatial and temporal co-ordinates, for our modeling purpose, we use $U(\boldsymbol{x}) = D \times (-\infty, t]$.

Letting $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots\}$ denote a realization of the stationary Poisson point process, the ordering $\boldsymbol{\pi}(\boldsymbol{x})$ is chosen to satisfy $\|\boldsymbol{x} - \boldsymbol{z}_{\pi_1(\boldsymbol{x})}\| < \|\boldsymbol{x} - \boldsymbol{z}_{\pi_2(\boldsymbol{x})}\| < \|\boldsymbol{x} - \boldsymbol{z}_{\pi_3(\boldsymbol{x})}\| < \cdots$, where $\|\cdot\|$ is a distance measure and $\boldsymbol{z}_{\pi(\boldsymbol{x})} \in \Phi \cap U(\boldsymbol{x})$. Thus, although the set of probabilities $\{p_i(\boldsymbol{x}); i = 1, 2, \ldots\}$ remains the same for all locations, they are randomly permuted. This random permutation, in turn, induces spatial dependence. To explain using a simple example, x_1, x_2 is spatially closer than x_1, x_3 in Fig. 2.2. The stick-breaking representation of ODDP suggests a higher degree of weight similarity between x_1 and x_2 than between x_1 and x_3, as shown in Fig. 2.3. Because the weights are similar, there is a greater dependence between x_1 and x_2 than between x_1 and x_3.

Assuming a homogeneous Poisson point process with intensity λ, ODDP is characterized by G_0, α, and λ. We express dependence of ODDP on these parameters by ODDP$(\alpha G_0, \lambda)$.

Assuming that data $\{y_1, \ldots, y_n\}$ are available at sites $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, GS embed the ODDP in a hierarchical Bayesian model:

$$\begin{aligned} y_i &\sim g_{\boldsymbol{\psi}_i(\cdot)} \\ \boldsymbol{\psi}_i &\sim G_{\boldsymbol{x}_i} \text{ and} \\ G_{\boldsymbol{x}_i} &\sim \text{ODDP}(\alpha G_0, \lambda). \end{aligned}$$

Note that the same theory can be extended to space–time situations with $\boldsymbol{x} = (\boldsymbol{s}', t)'$, where $\boldsymbol{s}$ stands for the spatial location and t stands for the time point.

Next, we introduce our proposed idea of kernel convolution of ODDP.

2.2.2. Kernel convolution of ODDP

We consider the following model for the data $\boldsymbol{Y} = \{y_1, \ldots, y_n\}$ at locations/times $\{\boldsymbol{x}_i = (\boldsymbol{s}_i', t_i)'; \ i = 1, \ldots, n\}$:

$$y_i = f(\boldsymbol{x}_i) + \epsilon_i, \tag{2.5}$$

where $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, for unknown σ^2. We represent the spatio-temporal process as noisy measurements y_i of an unknown real valued function $f(\boldsymbol{x})$. The unknown mean function $f(\cdot)$ is expressed as a convolution of ODDP $G_{\boldsymbol{x}}$ with a smoothing kernel $K(\boldsymbol{x}, \cdot)$:

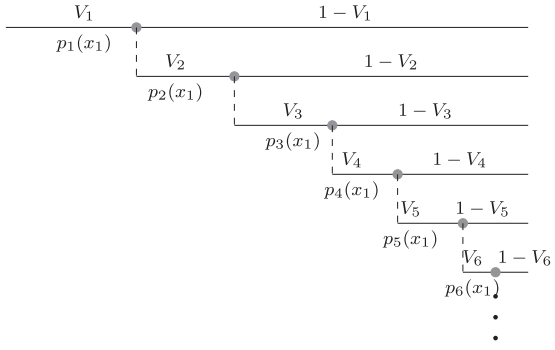
$$\begin{aligned} f(\boldsymbol{x}) &= \int K(\boldsymbol{x}, \boldsymbol{\theta}) dG_{\boldsymbol{x}}(\boldsymbol{\theta}) = \sum_{i=1}^{\infty} K(\boldsymbol{x}, \boldsymbol{\theta}_{\pi_i(\boldsymbol{x})}) p_i(\boldsymbol{x}) \ \forall \boldsymbol{x} \in D \subseteq \mathbb{R}^d, \boldsymbol{\theta} \in \mathbb{R}^d, \\ \boldsymbol{\theta} &\sim G_{\boldsymbol{x}} \text{ and} \\ G_{\boldsymbol{x}} &\sim \mathit{ODDP}(\alpha G_0, \lambda), \end{aligned} \tag{2.6}$$

d (≥ 1) being the dimension of $\boldsymbol{x}$. Note that, theoretically there is no reason to consider the same dimension for $\boldsymbol{x}$ and $\boldsymbol{\theta}$, since $K(\boldsymbol{x}, \boldsymbol{\theta})$ may be well-defined for different dimensions of $\boldsymbol{x}$ and $\boldsymbol{\theta}$. However, for our purpose, and for simplicity, we assume the dimensions to be the same.

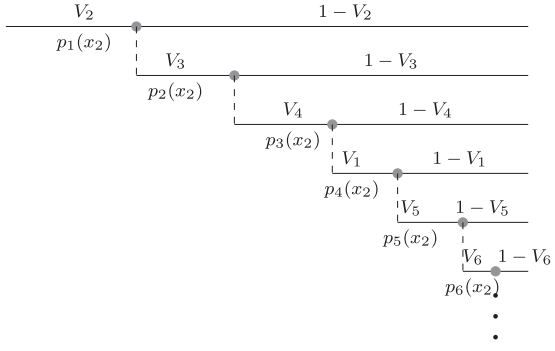
Thus, given $\boldsymbol{G}_{\boldsymbol{x}_i}$,

$$y_i \sim N\left(f(\boldsymbol{x}_i), \sigma^2\right), \tag{2.7}$$

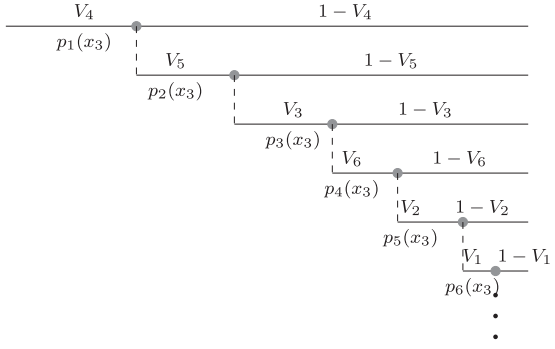
the normal distribution with mean $f(\boldsymbol{x}_i)$ of the form (2.6) and variance σ^2. Thus, given $\boldsymbol{G}_{\boldsymbol{x}_i}$ and $\boldsymbol{G}_{\boldsymbol{x}_j}$, y_i and y_j are independent.



(a)



(b)



(c)

Fig. 2.3. Stick Breaking Representation of ODDP at different covariate values a) x_1, (b) x_2, (c) x_3 : (Illustrates The higher degree of weight similarity between two nearby points in space).

Since the ODDP model of GS can also be viewed as a convolution, it is important to clarify its differences with (2.6) and (2.7). Indeed, note that with respect to GS, the response data y_i has the

following distribution:

$$y_i \sim \int g_{\boldsymbol{\psi}}(\cdot) dG_{\boldsymbol{x}_i}(\boldsymbol{\psi}) = \sum_{j=1}^{\infty} g_{\boldsymbol{\psi}_{\pi_j(\boldsymbol{x}_i)}}(\cdot) p_j(\boldsymbol{x}_i). \tag{2.8}$$

Thus, under (2.8) (that is, under the model proposed by GS), for any choice of $g_{\boldsymbol{\psi}}$, y_i arises from an infinite mixture with mixture density components $g_{\boldsymbol{\psi}_{\pi_j(\boldsymbol{x}_i)}}(\cdot)$ and corresponding mixture probabilities $p_j(\boldsymbol{x}_i)$. On the other hand, our model postulates a normal distribution for y_i via (2.6) and (2.7), where the mean is the kernel convolution given by (2.6). The convolutions given by (2.6) and (2.8) also have different interpretations. The latter is a density, whereas, the former is any real-valued function. Further implications, with respect to nonstationarity, and correlation structure tending to zero with widely separated distances, are discussed following Theorem 5.

In spatio-temporal processes we have to specify the joint distribution for an uncountable number of random variables. But, in practice we observe the process at a finite number of locations only. To infer about the process, it is better to have finite moments, that ensures existence of the posterior distribution. It also facilitates the prediction of the process at an arbitrary unobserved location. The following theorem, the proof of which is presented in Section S-1 of the supplement, gives an expression of the expectation of $f(\boldsymbol{x})$.

Theorem 1. *Let* $\int |K(\boldsymbol{x}, \boldsymbol{\theta})| dG_0(\boldsymbol{\theta}) < \infty$. *Then* $\int |K(\boldsymbol{x}, \boldsymbol{\theta})| dG_{\boldsymbol{x}}(\boldsymbol{\theta}) < \infty$ *with probability one, and*

$$E(f(\boldsymbol{x})) = E \int K(\boldsymbol{x}, \boldsymbol{\theta}) dG_{\boldsymbol{x}}(\boldsymbol{\theta}) = \int K(\boldsymbol{x}, \boldsymbol{\theta}) dE G_{\boldsymbol{x}}(\boldsymbol{\theta}) = \int K(\boldsymbol{x}, \boldsymbol{\theta}) dG_0(\boldsymbol{\theta}) = E_{G_0} K(\boldsymbol{x}, \boldsymbol{\theta}).$$

Before deriving the covariance structure of $f(\cdot)$, we define the necessary notation following GS. Let

$$T(\boldsymbol{x_1}, \boldsymbol{x_2}) = \{k : \text{ there exists } i, j \text{ such that } \pi_i(\boldsymbol{x_1}) = \pi_j(\boldsymbol{x_2}) = k\}.$$

For $k \in T(\boldsymbol{x_1}, \boldsymbol{x_2})$, we further define $A_{lk} = \{\pi_j(\boldsymbol{x_l}) : j < i \text{ where } \pi_i(\boldsymbol{x_l}) = k\}$, $S_k = A_{1k} \cap A_{2k}$ and $S_k' = A_{1k} \cup A_{2k} - S_k$. Then the following theorem, the proof of which is deferred to Section S-2 of the supplement, provides an expression for the covariance structure of $f(\cdot)$, which will be our reference point for arguments regarding nonstationarity and other desirable spatial properties in comparison with the existing methods.

Theorem 2. *If* $\int |K(\boldsymbol{x}, \boldsymbol{\theta})| dG_0(\boldsymbol{\theta}) < \infty$ *and* $\int |K(\boldsymbol{x_1}, \boldsymbol{\theta}) K(\boldsymbol{x_2}, \boldsymbol{\theta})| dG_0(\boldsymbol{\theta}) < \infty$, *then for a fixed ordering at* $\boldsymbol{x}_1$ *and* $\boldsymbol{x}_2$,

$$\begin{aligned} Cov(f(\boldsymbol{x_1}), f(\boldsymbol{x_2})) =& Cov_{G_0}(K(\boldsymbol{x_1}, \boldsymbol{\theta}), K(\boldsymbol{x_2}, \boldsymbol{\theta})) \\ & \times \frac{2}{(\alpha+1)(\alpha+2)} \sum_{k \in T(\boldsymbol{x_1}, \boldsymbol{x_2})} \left(\frac{\alpha}{\alpha+2}\right)^{\#S_k} \left(\frac{\alpha}{\alpha+1}\right)^{\#S_k'}, \end{aligned} \tag{2.9}$$

where

$$Cov_{G_0}(K(\boldsymbol{x_1}, \boldsymbol{\theta}), K(\boldsymbol{x_2}, \boldsymbol{\theta})) = \int K(\boldsymbol{x_1}, \boldsymbol{\theta}) K(\boldsymbol{x_2}, \boldsymbol{\theta}) dG_0(\boldsymbol{\theta}) - E_{G_0}(K(\boldsymbol{x_1}, \boldsymbol{\theta})) E_{G_0}(K(\boldsymbol{x_2}, \boldsymbol{\theta})). \tag{2.10}$$

Corollary 3. *It follows from the above theorem that for* $i = 1, 2$, *if* $\int K^2(\boldsymbol{x}_i, \boldsymbol{\theta}) dG_0(\boldsymbol{\theta}) < \infty$, *then*

$$Var(f(\boldsymbol{x}_i)) = \frac{Var_{G_0}(K(\boldsymbol{x}_i, \boldsymbol{\theta}))}{\alpha + 1} \tag{2.11}$$

and

$$Corr(f(\boldsymbol{x_1}), f(\boldsymbol{x_2})) = Corr_{G_0}(K(\boldsymbol{x_1}, \boldsymbol{\theta}), K(\boldsymbol{x_2}, \boldsymbol{\theta})) \times Corr(G_{\boldsymbol{x}_1}, G_{\boldsymbol{x}_2}), \tag{2.12}$$

where

$$Corr(G_{\boldsymbol{x}_1}, G_{\boldsymbol{x}_2}) = \frac{2}{\alpha+2} \sum_{k \in T(\boldsymbol{x_1}, \boldsymbol{x_2})} \left(\frac{\alpha}{\alpha+2}\right)^{\#S_k} \left(\frac{\alpha}{\alpha+1}\right)^{\#S_k'}. \tag{2.13}$$

The expression for the correlation in (2.13) has been obtained by GS. The above results provide an expression for the correlation conditional on a fixed ordering. To obtain the unconditional correlation it is necessary to marginalize the conditional correlation over the point process Φ. Following GS we also modify the notation as follows: we now let $T(\boldsymbol{x}_1, \boldsymbol{x}_2) = \Phi \cap U(\boldsymbol{x}_1) \cap U(\boldsymbol{x}_2)$, $A_{\ell k} = A_\ell(\boldsymbol{z}_k)$, where $A_\ell(\boldsymbol{z}) = \{\boldsymbol{w} \in \Phi \cap U(\boldsymbol{x}_\ell) : \|\boldsymbol{w} - \boldsymbol{x}_\ell\| < \|\boldsymbol{z} - \boldsymbol{x}_\ell\|\}$, for $\boldsymbol{z} \in \Phi \cap U(\boldsymbol{x}_\ell)$. As already mentioned in Section 2.2.1, when $\boldsymbol{x} = (\boldsymbol{s}', t)'$, we define $U(\boldsymbol{x}) = D \times (-\infty, t]$.

Also, for $\boldsymbol{z} \in T(\boldsymbol{x}_1, \boldsymbol{x}_2)$, we let $S(\boldsymbol{z}) = A_1(\boldsymbol{z}) \cap A_2(\boldsymbol{z})$ and $S'(\boldsymbol{z}) = A_1(\boldsymbol{z}) \cup A_2(\boldsymbol{z}) - S(\boldsymbol{z})$, which imply that $S(\boldsymbol{z}) = \{\boldsymbol{w} \in T(\boldsymbol{x}_1, \boldsymbol{x}_2) : \|\boldsymbol{w} - \boldsymbol{x}_1\| < \|\boldsymbol{z} - \boldsymbol{x}_1\| \text{ and } \|\boldsymbol{w} - \boldsymbol{x}_2\| < \|\boldsymbol{z} - \boldsymbol{x}_2\|\}$.

We further define, as in GS, $S_{-\boldsymbol{z}}(\boldsymbol{z})$ and $S'_{-\boldsymbol{z}}(\boldsymbol{z})$ to be translations of $S(\boldsymbol{z})$ and $S'(\boldsymbol{z})$, respectively, by $-\boldsymbol{z}$. Then the refined Campbell theorem yields, in the case where Φ is a stationary point process with intensity λ:

$$\begin{aligned}\mathrm{Corr}(f(\boldsymbol{x}_1), f(\boldsymbol{x}_2)) = \mathrm{Corr}_{G_0}(K(\boldsymbol{x}_1, \boldsymbol{\theta}), K(\boldsymbol{x}_2, \boldsymbol{\theta})) \\ \times \frac{2\lambda}{\alpha + 2} \int_{U(\boldsymbol{x}_1) \cap U(\boldsymbol{x}_2)} \int \left(\frac{\alpha}{\alpha + 2}\right)^{\phi_{-\boldsymbol{z}}(S_{-\boldsymbol{z}})} \left(\frac{\alpha}{\alpha + 1}\right)^{\phi_{-\boldsymbol{z}}(S'_{-\boldsymbol{z}})} P_0(d\phi) d\boldsymbol{z}. \quad (2.14)\end{aligned}$$

In (2.14), $P_0(d\phi)$ is the Palm distribution of Φ at the origin, and $\phi_{-\boldsymbol{z}}$ is the realization of Φ translated by $-\boldsymbol{z}$. Note also that the second factor of the above correlation is the unconditional correlation between $G_{\boldsymbol{x}_1}$ and $G_{\boldsymbol{x}_2}$ (see GS).

Remark 4. It is worth pointing out that unlike Gelfand et al. (2005) who obtained covariance structure conditional on the random process $\boldsymbol{G}$, in our case, the covariance structures conditional on the random measures $\boldsymbol{G}_{\boldsymbol{x}}$ are not relevant, since it follows from (2.7) and the subsequent discussion that $\mathrm{Cov}\left(y_1, y_2 | \boldsymbol{G}_{\boldsymbol{x}_1}, \boldsymbol{G}_{\boldsymbol{x}_2}\right) = 0$. Indeed, dependence among the responses is induced through dependence among $\boldsymbol{G}_{\boldsymbol{x}}$.

The following theorem, the proof of which is provided in Theorem 5 of the supplement, shows that the above correlation structure of our kernel convolution based ODDP satisfies desirable properties.

Theorem 5. *$Corr(f(\boldsymbol{x}_1), f(\boldsymbol{x}_2)) \to 1$ as $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \to 0$ and $Corr(f(\boldsymbol{x}_1), f(\boldsymbol{x}_2)) \to 0$ as $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \to \infty$.*

It is clear from the above theorem and model (2.5) that $\mathrm{Corr}(y_i, y_j) \to 0$ as $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \to \infty$.

Under a stationary Poisson process assumption for Φ, and for particular specifications of $U(\boldsymbol{x})$ mentioned in Section 2.2.1, the calculations of GS show that the second factor of (2.14) depends upon $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ only through $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$, leading to isotropy of the process. There does not seem to exist any result analogous to the refined Campbell theorem in the context of nonstationary Poisson process which might allow one to construct a nonstationary correlation structure in this case. The analytic form of the ODDP correlation structure need not be available for other constructions of $U(\boldsymbol{x})$ either. Isotropy results even in the case of the more flexible Cox processes. Note that the correlations between any two responses y_i and y_j may correspond to nonstationarity if their expectations under the density $f_{\boldsymbol{\theta}}(\cdot)$ are nonlinear in $\boldsymbol{\theta}$. However, there is no guarantee that the correlation tends to zero as $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \to \infty$.

On the other hand, our kernel convolution idea neatly solves this problem of attainment of nonstationarity via the first factor of our correlation structure given in (2.14). Indeed, the kernel $K(\boldsymbol{x}, \boldsymbol{\theta})$ can be chosen in the spirit of Higdon et al. (1999), for instance, such that $\mathrm{Corr}_{G_0}(K(\boldsymbol{x}_1, \boldsymbol{\theta}), K(\boldsymbol{x}_2, \boldsymbol{\theta}))$ does not depend upon $\boldsymbol{x}_1 - \boldsymbol{x}_2$ alone. In other words, by simply controlling the kernel we can ensure nonstationarity of our process $f(\cdot)$ even if the underlying ODDP is stationary or even isotropic. Of course, our process can be made stationary as well by choosing the kernel, say, in the spirit of Higdon (1998), and setting $U(\boldsymbol{x})$ to be of the forms specified by GS, when $\boldsymbol{x}$ consists of either only spatial co-ordinates or only temporal co-ordinate. When $\boldsymbol{x} = (\boldsymbol{s}', t)'$, then we set $U(\boldsymbol{x}) = D \times (-\infty, t]$, as already mentioned before.

We further note that our general space–time correlation structure given by (2.12) is nonseparable, that is, in general, $\mathrm{Corr}(f(\boldsymbol{s}_1, t_1), f(\boldsymbol{s}_2, t_2)) \neq \mathrm{Corr}_1(\boldsymbol{s}_1, \boldsymbol{s}_2) \times \mathrm{Corr}_2(t_1, t_2)$, where Corr_1 and Corr_2 are spatial and temporal structures, respectively. However, if desired, separability can be

easily induced by allowing the kernel to depend upon only the spatial location and by allowing the ordering $\boldsymbol{\pi}$ to depend only upon time, or the vice versa. Specifically, letting $K(\boldsymbol{x}, \boldsymbol{\theta}) = K(\boldsymbol{s}, \boldsymbol{\theta})$ and $\boldsymbol{\pi}(\boldsymbol{x}) = \boldsymbol{\pi}(t)$, we obtain

$$\text{Corr}(f(\boldsymbol{s}_1, t_1), f(\boldsymbol{s}_2, t_2)) = \text{Corr}_{G_0}(K(\boldsymbol{s}_1, \boldsymbol{\theta}), K(\boldsymbol{s}_2, \boldsymbol{\theta})) \times \text{Corr}(G_{t_1}, G_{t_2}), \tag{2.15}$$

and letting $K(\boldsymbol{x}, \boldsymbol{\theta}) = K(t, \boldsymbol{\theta})$ and $\boldsymbol{\pi}(\boldsymbol{x}) = \boldsymbol{\pi}(\boldsymbol{s})$, we obtain

$$\text{Corr}(f(\boldsymbol{s}_1, t_1), f(\boldsymbol{s}_2, t_2)) = \text{Corr}_{G_0}(K(t_1, \boldsymbol{\theta}), K(t_2, \boldsymbol{\theta})) \times \text{Corr}(G_{\boldsymbol{s}_1}, G_{\boldsymbol{s}_2}). \tag{2.16}$$

In contrast, under the ODDP approach of GS, it is clear from the correlation structure that $\text{Corr}(G_{\boldsymbol{x}_1}, G_{\boldsymbol{x}_2}) \neq \text{Corr}_1(\boldsymbol{s}_1, \boldsymbol{s}_2) \times \text{Corr}_2(t_1, t_2)$, showing that separability cannot be enforced if desired.

Thus, following our approach it is easy to construct nonparametric covariance structures that are either stationary or nonstationary, which, in turn, can be constructed as either separable or nonseparable, as desired. These illustrate the considerable flexibility inherent in our approach, while satisfying at the same time the desirable conditions that the correlation between $f(\boldsymbol{x}_1)$ and $f(\boldsymbol{x}_2)$ tends to 1 or zero accordingly as the distance between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ tends to zero or infinity.

2.3. Continuity and smoothness properties of our model

For stationary models, properties like continuity and smoothness can be quite generally characterized by the continuity and smoothness of the correlation function. In particular, continuity and smoothness of stationary processes typically depend upon the behavior of the correlation function at zero; see Yaglom (1987a) and Yaglom (1987b) for details. For nonstationary processes, however, such elegant theory is not available. Indeed, the structure of the correlation function itself may be difficult to get hold of, rendering it difficult to investigate the properties of the underlying nonstationary stochastic process. For our purpose, we utilize the notions of almost sure continuity, mean square continuity and mean square differentiability of stochastic processes (see, for example, Stein (1999), Banerjee and Gelfand (2003)) to study the properties of our nonstationary spatio-temporal process.

Definition 6. A process $\{X(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^d\}$ is L_2 continuous at $\boldsymbol{x}_0$ if $\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} E[X(\boldsymbol{x}) - X(\boldsymbol{x}_0)]^2 = 0$. Continuity in the L_2 sense is also referred to as mean square continuity and will be denoted by $X(\boldsymbol{x}) \xrightarrow{L_2} X(\boldsymbol{x}_0)$.

Definition 7. A process $\{X(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^d\}$ is almost surely continuous at $\boldsymbol{x}_0$ if $X(\boldsymbol{x}) \to X(\boldsymbol{x}_0)$ *a.s.* as $\boldsymbol{x} \to \boldsymbol{x}_0$. If the process is almost surely continuous for every $\boldsymbol{x}_0 \in \mathbb{R}^d$ then the process is said to have continuous realizations.

Theorem 8. *Assume the following conditions:*

(A1) For all $\boldsymbol{x}$ and $\boldsymbol{\theta}$, $|K(\boldsymbol{x}, \boldsymbol{\theta})| < M$ for some $M < \infty$.
(A2) Given any $\boldsymbol{\theta}$, $K(\boldsymbol{x}, \boldsymbol{\theta})$ is a continuous function of $\boldsymbol{x}$.

Then $f(\cdot)$ is both almost surely continuous and mean square continuous in the interior of $\cap_{k=1}^{\infty} A_{ki_k}$, where $A_{ki_k} = \{\boldsymbol{x} : \pi_k(\boldsymbol{x}) = i_k\}$, and for each $k = 1, 2, \ldots$, $i_k \in \{1, 2, \ldots\}$; $i_k \neq i_{k'}$ for any $k \neq k'$. On the other hand, $f(\cdot)$ is almost surely discontinuous at any point $\boldsymbol{x}_0 \in \cap_{k=1}^{\infty} A_{ki_k}$ lying on the boundary of A_{ki_k}, for any i_k.

See Section S-4 for a proof of this result. Now we examine mean square differentiability of our process.

Definition 9. A process $\{X(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^d\}$ is said to be mean square differentiable at $\boldsymbol{x}_0$ if for any direction $\boldsymbol{u}$, there exists a process $L_{\boldsymbol{x}_0}(\boldsymbol{u})$, linear in $\boldsymbol{u}$ such that

$$X(\boldsymbol{x}_0 + \boldsymbol{u}) = X(\boldsymbol{x}_0) + L_{\boldsymbol{x}_0}(\boldsymbol{u}) + R(\boldsymbol{x}_0, \boldsymbol{u}), \text{ where } \frac{R(\boldsymbol{x}_0, \boldsymbol{u})}{\|\boldsymbol{u}\|} \xrightarrow{L_2} 0.$$

Theorem 10. *Assume the following conditions:*

(B1) For all $\boldsymbol{x}$ *and* $\boldsymbol{\theta}$, $|K(\boldsymbol{x}, \boldsymbol{\theta})| < M$ *for some* $M < \infty$.
(B2) Given any $\boldsymbol{\theta}$, $K(\boldsymbol{x}, \boldsymbol{\theta})$ *is a continuously differentiable function of* $\boldsymbol{x}$.

Then $f(\cdot)$ *is mean square differentiable in the interior of* $\cap_{k=1}^{\infty} A_{ki_k}$.

See Section S-5 for a proof of this theorem.

In real life applications most of the spatio-temporal processes are expected to be irregular in nature. One of the desirable properties of a spatio-temporal model is that, it allows different degrees of smoothness across space and time. Our model has achieved this property regarding smoothness. For example, if we associate the ODDP prior only to the spatial locations, then the process becomes smoother across time than across space depending on the choice of the kernel.

2.4. Truncation of the infinite summand

Since our proposed model $f(\boldsymbol{x}) = \sum_{k=1}^{\infty} K(\boldsymbol{x}, \boldsymbol{\theta}_{\pi(\boldsymbol{x})}) p_i(\boldsymbol{x})$ is an infinite (random) series, for model-fitting purpose it is necessary to truncate the series to $f(\boldsymbol{x}) = \sum_{k=1}^{N} K(\boldsymbol{x}, \boldsymbol{\theta}_{\pi(\boldsymbol{x})}) p_i(\boldsymbol{x})$, where N is to be determined, or to implement variable-dimensional Markov chain methods. For the latter, N is to be considered a random variable so that the number of parameters associated with $f(\boldsymbol{x})$ is also a random variable.

Although we will describe and implement TTMCMC, we first prove a theorem with respect to truncation of the infinite random series. Note that in the context of traditional Dirichlet process characterized by Sethuraman's stick breaking construction (Sethuraman, 1994) which involves infinite random series, Ishwaran and James (2001) proposed a method of truncating the infinite series.

We now state our theorem on truncation, the proof of which is provided in Section S-6 of the supplement. But before stating the theorem it is necessary to define some required notation. Let

$$P_N(\boldsymbol{x}_i) = \sum_{i=1}^{N} K(\boldsymbol{x}_i, \theta_i) p_i, \quad \text{and} \quad P(\boldsymbol{x}_i) = \sum_{i=1}^{\infty} K(\boldsymbol{x}_i, \theta_i) p_i,$$

where N needs to be determined. Also let

$$P_N = (P_N(\boldsymbol{x}_1), \ldots, P_N(\boldsymbol{x}_n))' \quad \text{and} \quad P = (P(\boldsymbol{x}_1), \ldots, P(\boldsymbol{x}_n))',$$

and denote by Θ_N and Θ the sets of random quantities $(V_i, \boldsymbol{\theta}_i)$ associated with P_N and P respectively. We define the following marginal densities of the vector of observations $\boldsymbol{y} = \{y_1, \ldots, y_n\}$, where $[\cdot|\cdot]$ and $[\cdot]$ denote conditional and marginal densities, respectively:

$$\begin{aligned} m_N(\boldsymbol{y}) &= \int_{\Theta_N} [\boldsymbol{y}|P_N][P_N] d\Theta_N \\ &= \int_{\Theta} [\boldsymbol{y}|P_N][P] d\Theta, \end{aligned}$$

and

$$m_\infty(\boldsymbol{y}) = \int_{\Theta} [\boldsymbol{y}|P][P] d\Theta.$$

Theorem 11. *Under the assumption that* $\sup_{\boldsymbol{\theta}} K(x_i, \boldsymbol{\theta}) \le M$ *for* $i = 1, \ldots, n$, *where* $M > 0$ *is a finite constant, we have*

$$\int_{\mathbb{R}^n} |m_N(\boldsymbol{y}) - m_\infty(\boldsymbol{y})| \, d\boldsymbol{y} \le 4M^2 n \left(\frac{\alpha}{\alpha+2}\right)^N + 2\sqrt{\frac{2}{\pi}} Mn \left(\frac{\alpha}{\alpha+1}\right)^N.$$

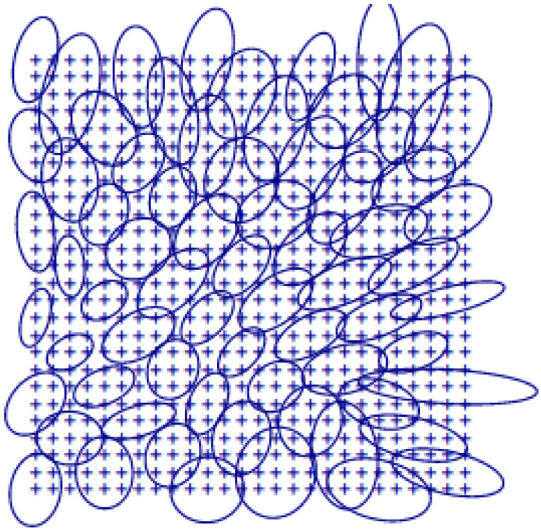


Fig. 2.4. Spatially Evolving Ellipse (Calder, 2014).

Since our proposed model involves a random infinite series, for model fitting one needs to either truncate the series or assume a random number of summands and adopt variable dimensional MCMC approaches. Although we adopt the latter framework for our applications, for the sake of completeness we also investigate the truncation approach. We consider the difference between the prior predictive models with and without truncation of the random infinite series, providing a bound that depends upon the truncation parameter. Thus, the truncation parameter can be chosen so that the bound falls below any desired threshold.

2.5. Choice of kernel, prior distributions and computational region

The choice of kernel $K(\cdot,\cdot)$ plays a crucial role in nonstationary spatio-temporal data analysis. For instance, if $K(\boldsymbol{x},\boldsymbol{\theta}) = K(\boldsymbol{x}-\boldsymbol{\theta})$, then the correlation between $Y(\boldsymbol{x}_1)$ and $Y(\boldsymbol{x}_2)$ turns out to be a function of $\boldsymbol{x}_1 - \boldsymbol{x}_2$, thus inducing stationarity. For the purpose of nonstationarity, it is necessary to make the parameters of the kernel depend upon space and time. In the spatial context such nonstationary kernels are considered in Higdon et al. (1999). In this paper, we consider a nonstationary space–time kernel; for the spatial part of the kernel we essentially adopt the dependence structure and the associated prior distributions proposed by Higdon et al. (1999) and for the temporal part we allow the relevant coefficient to be time varying, modeled by a stationary Gaussian process. Higdon et al. (1999) noted that there is a one-to-one mapping between a bivariate mean-zero Gaussian density and its one standard deviation ellipse. To achieve nonstationarity in space spatially-varying kernel matrices $\Sigma(\boldsymbol{s})$ are modeled using a spatially-varying family of ellipses, (Fig : 2.4). Specifically, Higdon et al. (1999) parameterizes the ellipses to be spatially- varying by allowing the coordinates of the focal points of $\Sigma(\boldsymbol{s})$ to be spatially-varying; these coordinates are assigned stationary Gaussian process prior distributions.

In particular, we consider the following kernel for our applications:

$$K(\boldsymbol{s},t,\boldsymbol{\theta},\tau) = \exp\left\{-\frac{1}{2}(\boldsymbol{s}-\boldsymbol{\theta})^T\Sigma(\boldsymbol{s})(\boldsymbol{s}-\boldsymbol{\theta}) - \delta(t)|t-\tau|\right\},$$

where $\Sigma(\boldsymbol{s})$ is a 2×2 positive definite dispersion matrix depending upon $\boldsymbol{s}$, and $\delta(t) > 0$ depends upon time t. We assume that $\log(\delta(t))$ is a zero mean Gaussian process with covariance $c_\delta(t_1,t_2) =$

$\sigma_\delta^2 \exp\left\{(t_1 - t_2)^2/a_\delta\right\}$. We set

$$\Sigma(\boldsymbol{s})^{\frac{1}{2}} = \varphi \begin{pmatrix} \left[\frac{\sqrt{4A^2+\|\psi(\boldsymbol{s})\|^4\pi^2}}{2\pi} + \frac{\|\psi(\boldsymbol{s})\|^2}{2}\right]^{\frac{1}{2}} & 0 \\ 0 & \left[\frac{\sqrt{4A^2+\|\psi(\boldsymbol{s})\|^4\pi^2}}{2\pi} - \frac{\|\psi(\boldsymbol{s})\|^2}{2}\right]^{\frac{1}{2}} \end{pmatrix} \times \begin{pmatrix} \cos\ \alpha(\boldsymbol{s}) & \sin\ \alpha(\boldsymbol{s}) \\ -\sin\ \alpha(\boldsymbol{s}) & \cos\ \alpha(\boldsymbol{s}) \end{pmatrix},$$

Where ψ_1, ψ_2 are the focal points of the ellipse, and are spatially varying. where $\|\psi(\boldsymbol{s})\|^2 = \psi_1^2(\boldsymbol{s}) + \psi_2^2(\boldsymbol{s})$ and $\alpha(\boldsymbol{s}) = \tan^{-1}\left(\frac{\psi_2(\boldsymbol{s})}{\psi_1(\boldsymbol{s})}\right)$. For the details regarding derivation of $\boldsymbol{\Sigma}(\boldsymbol{s})$, intuition and relevant diagrams for visual explanation, see Higdon et al. (1999).

We assume that $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are independent and identical zero mean Gaussian processes with stationary covariance $c_\psi(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sigma_\psi^2 \exp\left\{-\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|^2/b_\psi\right\}$. We put the $U(3, 200)$ prior on φ, a_δ, and b_ψ; we set $\sigma_\delta^2 = \sigma_\psi^2 = 1$. Also, we set $A = 3.5$. Since in our applications we center and scale the observed time points, for τ we specify the $N(0, 1)$ prior.

The use of Gaussian kernel functions, or a Gaussian correlation function, has the undesirable property of resulting in process realizations that are infinitely differentiable and, thus, too smooth for most spatio temporal applications, as detailed in Paciorek and Schervish (2006). To create a non-smooth covariance structure, we convolved this Gaussian Kernel with ODDP, which is discontinuous. However, different levels of smoothness can be attained by strategically utilizing ODDP, as was already mentioned in 2.3.

2.5.1. Choice of G_0

In our applications, we center and scale each of the two components $\{s_{1i};\ i = 1, \ldots, n\}$ and $\{s_{2i};\ i = 1, \ldots, n\}$ of the available spatial locations $\{\boldsymbol{s}_i = (s_{1i}, s_{2i});\ i = 1, \ldots, n\}$. Consequently, the choosing G_0 to be the bivariate normal distribution with both means zero, both variances equal to one, and correlation ρ appears to be reasonable. We estimate ρ by the empirical correlation between $\{s_{1i};\ i = 1, \ldots, n\}$ and $\{s_{2i};\ i = 1, \ldots, n\}$.

2.5.2. Prior selection for α

For the choice of prior distributions of the parameters associated with the ODDP we follow Griffin and Steel (2004) and GS. In particular, we put the inverted Beta distribution prior on α, given by

$$p(\alpha) = \frac{{n_0}^\eta \Gamma(2\eta)\alpha^{\eta-1}}{\Gamma(\eta)^2(\alpha + n_0)^{2\eta}},$$

where the hyperparameter n_0 is the prior median of α. Note that the prior variance exists if $\eta > 2$, and is a decreasing function of η. This prior implies that $\frac{\alpha}{\alpha+n_0}$ follows a $Beta(\eta, \eta)$ distribution.

2.5.3. Prior selection for λ

Note that, for small α, only the first few elements of stick breaking representation are important, so fewer number of points from the underlying Poisson process is needed to induce the second factor of the correlation structure (2.14), which roughly depends upon the ratio $\lambda/(\alpha + 1)$ for $U(\boldsymbol{x}) = D$ (spatial problem) and $U(x) = (-\infty, x]$ (temporal problem); see GS for the details. Thus a relatively small value of λ suffices in such cases. Similarly, when α is larger, larger λ is necessary to obtain the same correlation. Keeping these in mind, we select the log-normal prior for λ with mean $\log(\alpha)$ and variance b_λ, say. For our applications, we choose $b_\lambda = 20$, so that we obtain a reasonably vague prior.

2.5.4. Computational region

Following GS, we consider a truncated region for the point process Z which includes the range of the observed $\boldsymbol{x}$. This truncated region has been referred to as the computational region by GS. In particular, we choose a bounding box of the form $(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d)$ as the computational region, where $a_i = d_{a_i} - r$, $b_i = d_{b_i} - r$. Here d_{a_i} and d_{b_i} are the minimum and the maximum of $\boldsymbol{x}$ in dimension i, and $r = 2\left(\frac{\Gamma(d/2)d}{2\pi^{d/2}} \frac{\alpha+1}{\lambda} \log \frac{1}{\epsilon}\right)^{\frac{1}{d}}$, with $\epsilon = \exp\left\{-\frac{\lambda}{\alpha+1} \frac{2\pi^{d/2}}{\Gamma(d/2)d} \left(\frac{r}{2}\right)^d\right\}$. See GS for justification of these choices.

2.6. Joint posterior and a briefing of TTMCMC for updating parameters in our variable dimensional modeling framework

Let k denote the random number of summands in

$$f_k(\boldsymbol{x}) = \sum_{i=1}^{k} K(\boldsymbol{x}, \boldsymbol{\theta}_{\pi_i(\boldsymbol{x})}) p_i(\boldsymbol{x}) \ , \ \forall \boldsymbol{x} \in D \subseteq \mathbb{R}^d. \tag{2.17}$$

Let $\boldsymbol{V} = (V_1, \ldots, V_k)$, $\boldsymbol{z} = (z_1, \ldots, z_k)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, with $\boldsymbol{\theta}_1 = (\theta_{11}, \ldots, \theta_{1k})$ and $\boldsymbol{\theta}_1 = (\theta_{21}, \ldots, \theta_{2k})$. Let also $\boldsymbol{\psi}_1 = (\psi_1(\boldsymbol{s}_1), \ldots, \psi_1(\boldsymbol{s}_n))$, $\boldsymbol{\psi}_2 = (\psi_2(\boldsymbol{s}_1), \ldots, \psi_2(\boldsymbol{s}_n))$ and $\boldsymbol{\delta} = (\delta(t_1), \ldots, \delta(t_n))$. The joint posterior is of the form

$$\begin{aligned} &\pi(k, \boldsymbol{V}, \boldsymbol{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\delta}, \tau, \sigma, \alpha, \lambda, b_\psi, a_\delta | \boldsymbol{Y}) \\ &\propto \pi(k)\pi(\boldsymbol{V}, \boldsymbol{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | k)\pi(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\delta})\pi(\tau, \varphi, b_\psi, a_\delta)\pi(\sigma)\pi(\alpha)\pi(\lambda|\alpha) \times L(\boldsymbol{V}, \boldsymbol{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma | k, \boldsymbol{Y}), \end{aligned} \tag{2.18}$$

where $L(\boldsymbol{V}, \boldsymbol{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma | k, \boldsymbol{Y})$ is the joint normal likelihood of $\boldsymbol{V}, \boldsymbol{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma$ under the model

$$y_i = f_k(\boldsymbol{x}_i) + \epsilon_i; \ \ \epsilon_i \overset{iid}{\sim} N(0, \sigma^2); \ \ i = 1, \ldots, n, \tag{2.19}$$

conditional on $f_k(\cdot)$.

For our applications, as the prior $\pi(k)$ on k, we assume the discrete uniform prior on $\{1, 2, \ldots, 30\}$; in our applications k never even reached 30. Under $\pi(\boldsymbol{V}, \boldsymbol{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | k)$, $V_i \overset{iid}{\sim} Beta(1, \alpha)$; $i = 1, \ldots, k$, $\boldsymbol{z}$ are realizations from the Poisson process with intensity λ, and for $i = 1, \ldots, k$, $(\theta_{1i}, \theta_{2i}) \overset{iid}{\sim} G_0$. Under $\pi(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\delta})$, ψ_1, ψ_2, δ are independent Gaussian processes, as detailed in Section 2.5. The prior distribution of $\tau, \varphi, b_\psi, a_\delta$, denoted by $\pi(\tau, \varphi, b_\psi, a_\delta)$, is already provided in Section 2.5. For the error standard deviation σ, the prior denoted by $\pi(\sigma)$ is the log-normal distribution with parameters 0 and 1, so that the mean and variance of σ are about 1.6 and 5, respectively. These quantities appear to be reasonable, and yielded adequate inference.

In order to obtain samples from the joint posterior (2.18) which involve the variable dimensional $f_k(\cdot)$, we implement the TTMCMC methodology. In a nutshell, TTMCMC updates all the parameters, both fixed and variable dimensional, as well as the number of parameters of the underlying posterior distribution in a single block using simple deterministic transformations of some low-dimensional random variable drawn from some fixed, but low-dimensional arbitrary distribution defined on some relevant support. The idea is an extension of Transformation based Markov Chain Monte Carlo (TMCMC) introduced by Dutta and Bhattacharya (2014) for updating high-dimensional parameters with known dimensionality in a single block using simple deterministic transformations of some low-dimensional (usually one-dimensional) random variable having arbitrary distribution on some relevant support. The strategy of updating high and variable dimensional parameters using very low-dimensional random variables clearly reduces dimensionality dramatically, thus greatly improving acceptance rate, mixing properties, and computational speed. In Section S-7 of the supplement we provide a detailed overview of TTMCMC, propose a general algorithm (Algorithm S-7.1) with certain advantages, and in Section S-8 of the supplement we specialize the algorithm to our spatio-temporal modeling set-up, providing full updating details (Algorithm S-8.1).

3. Results

3.1. Simulation study

To illustrate the performance of our model we first create a synthetic data generating process which is nonstationary and non-Gaussian. One popular method to create such process is the kernel convolution approach. However, since we have developed our spatio-temporal model itself using the kernel convolution approach, it is perhaps desirable to obtain the synthetic data from some nonstationary, non-Gaussian process created using some approach independent of the kernel convolution method. In Section 3.1.1 we detail such an approach. Then we fit our proposed model to the data pretending that the data-generating process is unknown.

3.1.1. A nonstationary non-Gaussian data generating process

Let $X(\cdot)$ denote a stationary Gaussian process with mean function $\mu(t, \boldsymbol{s}) = \beta_0 + \beta_1 t + \beta_2 s_1 + \beta_3 s_2$, with $\boldsymbol{s} = (s_1, s_2)$, and covariance function

$$A(i,j) = c((t_i, \boldsymbol{s}_i), (t_j, \boldsymbol{s}_j)) = \exp\left\{-0.5\left(\sqrt{(t_i - t_j)^2 + (s_{1i} - s_{1j})^2 + (s_{2i} - s_{2j})^2}\right)\right\},$$

for any $t_i, t_j, \boldsymbol{s}_i = (s_{1i}, s_{2i}), \boldsymbol{s}_j = (s_{1j}, s_{2j})$.

Let $\boldsymbol{X} = (X(t_1, \boldsymbol{s}_1), \ldots, X(t_n, \boldsymbol{s}_n))'$ denote observed data points from the Gaussian process X at the design points $\{(t_i, \boldsymbol{s}_i);\ i = 1, \ldots, n\}$. Let $\boldsymbol{t} = (t_1, \ldots, t_n)'$ and $\boldsymbol{S} = (\boldsymbol{s}_1', \ldots, \boldsymbol{s}_n')'$. Further, let us denote by $\boldsymbol{A} = (A(i,j); i = 1, \ldots, n;\ j = 1, \ldots, n)$ the covariance matrix and $\boldsymbol{\mu} = (\mu(t_1, \boldsymbol{s}_1), \ldots, \mu(t_n, \boldsymbol{s}_n))'$. Then the posterior process $[X(\cdot)|\boldsymbol{X}]$ is non-stationary Gaussian with mean function $\mu_X(t, \boldsymbol{s}) = \mu(t, \boldsymbol{s}) + \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$ and variance $\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$, where $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}$.

Let the posterior nonstationary Gaussian process $[X(\cdot)|\boldsymbol{X}]$ be denoted by $X^*(\cdot)$. Now, conditionally on the process $X^*(\cdot)$, consider another process $Y(\cdot)$ with mean function $\mu^*(t, \boldsymbol{s}) = X^*(t, \boldsymbol{s})$ and covariance function $c_Y((t_i, \boldsymbol{s}_i), (t_j, \boldsymbol{s}_j)) = \exp\left\{-0.5\left|X^*(t_i, \boldsymbol{s}_i) - X^*(t_j, \boldsymbol{s}_j)\right|\right\}$. Then marginally, $Y(\cdot)$ is a nonstationary non-Gaussian process.

For our illustration we will simulate the synthetic data set from the process $Y(\cdot)$. The algorithm for generation of this synthetic data is provided in supplementary material (Section S-9.1).

3.2. Results of fitting our model to the simulated data

Note that for this problem the number of parameters to be updated ranges between 300 to 400. Our TTMCMC based model implementation took 35 mins to yield 900000 realizations following a burn-in of 100000. Quite encouragingly, TTMCMC exhibited satisfactory acceptance rate and mixing properties. Traceplots are shown in Figure S-9.1 of supplement.

3.2.1. Leave-one-out cross-validation

We assess the predictive power of our model with the leave-one-out cross validation method. All the 95 cases were included in the 95% highest posterior densities of the corresponding leave-one-out posterior predictive densities, Fig. 3.1 displays the posterior predictive densities of six randomly selected space–time points, along with the true values, the latter denoted by the vertical lines. Thus, satisfactory performance of our proposed model is indicated by the results, particularly given the fact that our model does not assume knowledge of the true, data-generating, parametric model.

3.2.2. Correlation analysis

Though our simulation mechanism is completely different from our proposed model, the simulated data do exhibit the pattern that the correlations are close to zero for two widely separated locations and/or times. Indeed, from the structure of the covariance matrix $\boldsymbol{\Sigma}_{(95|5)}$, it is easily seen that the (i,j)-th element $(i \neq j)$ of $\boldsymbol{\Sigma}_{(95|5)}$ is close to zero whenever the distance between t_i and t_j and/or $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ is large.

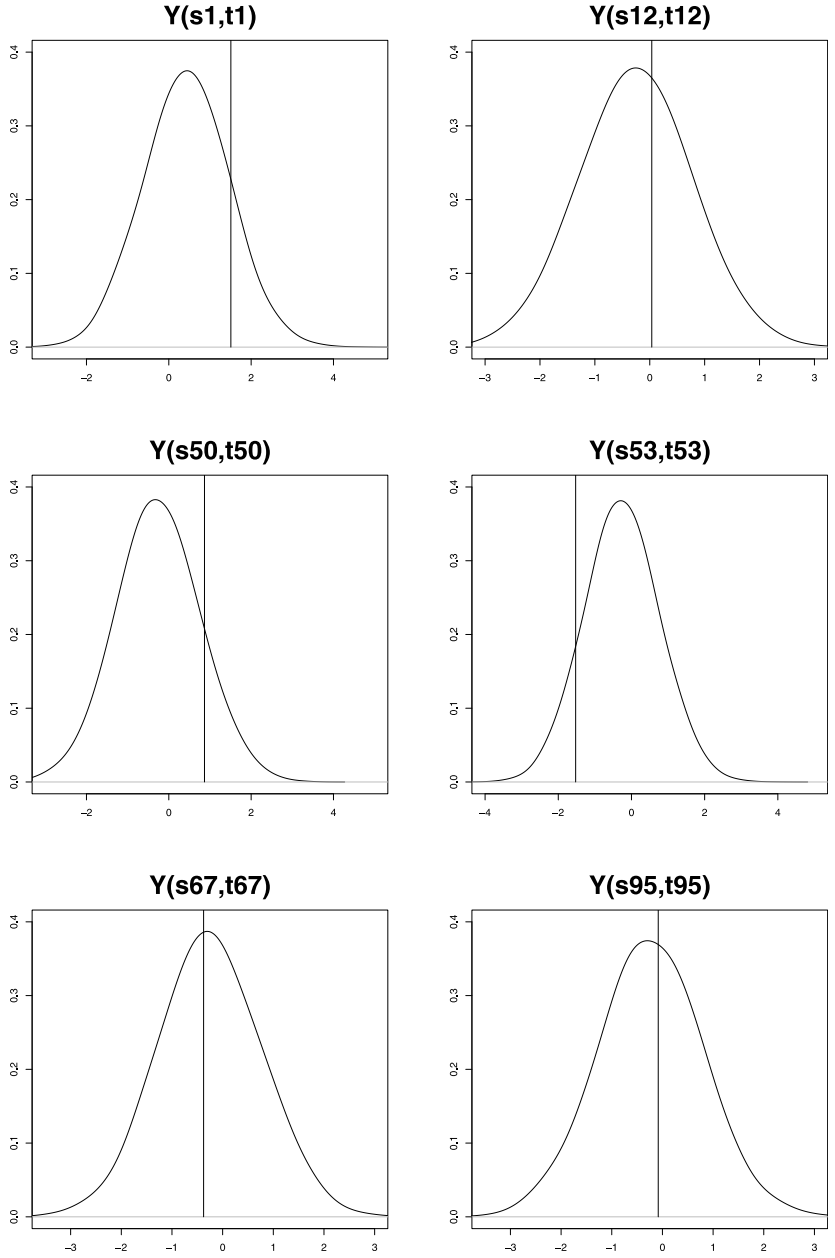


Fig. 3.1. Simulation study: Posterior predictive densities of $Y(\boldsymbol{s}, t)$ for the 6 different location-time pairs of our model — the corresponding true values are denoted by the vertical line.

We calculate the posterior densities of correlation for different pairs of space–time points. In formation of the pair, we select nearby locations, as well as locations which are widely separated, such that we obtain both high and low correlation values under the true, data-generating model. It

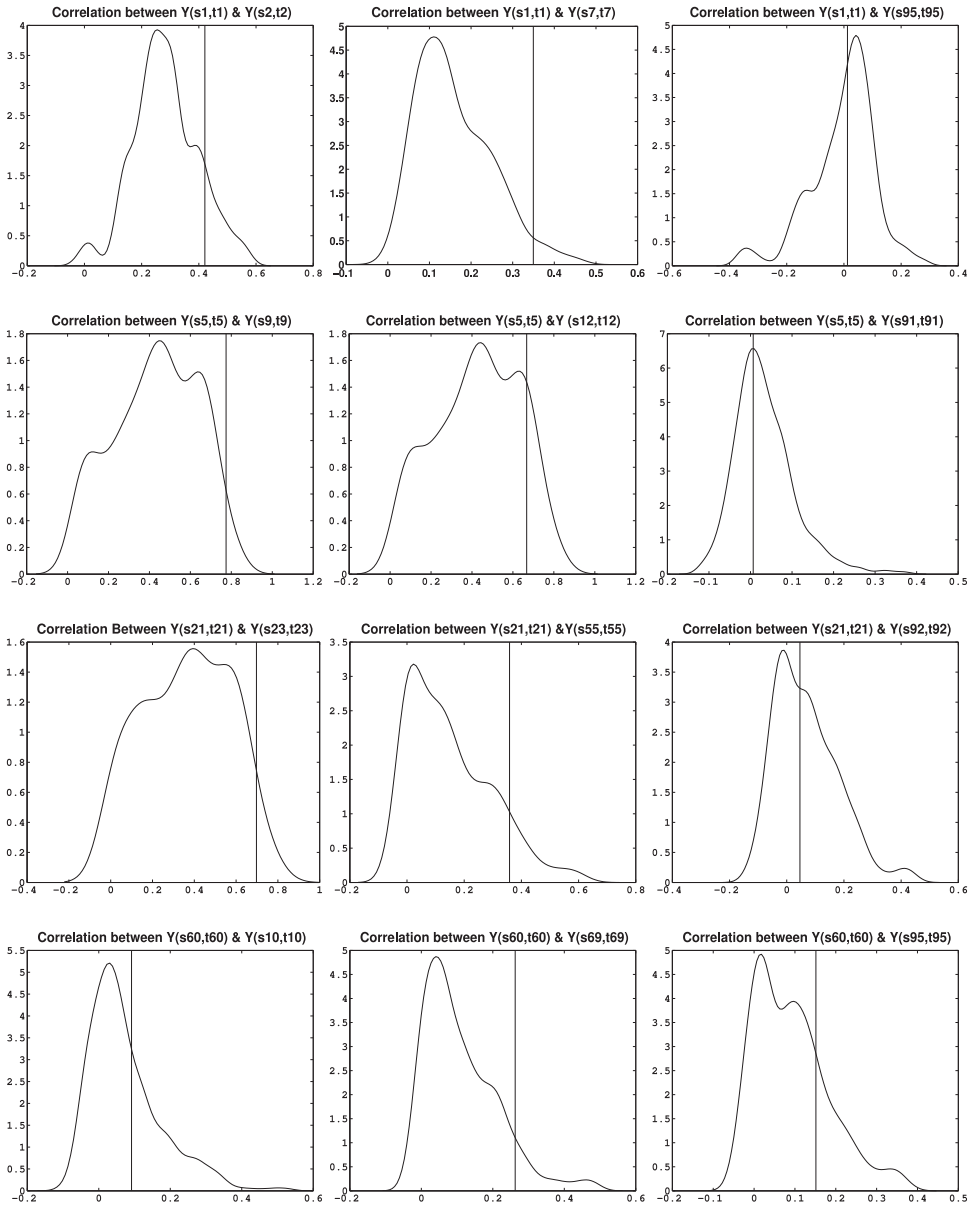


Fig. 3.2. Simulation study: Posterior densities of the correlations for the 12 different pairs of spatio-temporal points of our model; the vertical lines indicate the true correlations.

is evident from Fig. 3.2 that the true correlations, ranging from small to high values, lie well within their respective 95% credible intervals, vindicating reasonable performance of our model in terms of capturing the true correlation structure.

3.3. Comparative study with respect to FR's approach

We now compare the performance of our model with FR. For the purpose of comparison, we extended the exclusively spatial model of FR to space–time model, and apply the same to our simulated data.

We assess the predictive power of their model with the leave-one-out cross validation method. All the 95 cases were included in the 95% highest posterior densities of the corresponding leave-one-out posterior predictive densities. Fig. 3.3 displays the posterior predictive densities along with the true values obtained by employing FR's model for the same six locations that were investigated in our model. If we consider the CPO measure defined by $CPO_i = \pi(y_i^{obs}|y_{-i})$ (Conditional Predictive Ordinate) (Pettit, 1990; Geisser, 1993), except for a few locations where the CPO measure for our model is slightly smaller than the model proposed by FR, our model performance is significantly better for most of the locations. Moreover, variabilities of the leave-one-out posterior predictive densities associated with the model of FR are substantially larger for all the locations.

3.3.1. Correlation analysis

We calculate the posterior densities of the correlation for the same 12 pairs of space–time points that were investigated in our model. The main features of the correlation analysis are the following:

- The posterior densities, which are highly multimodal in nature, are in keeping with the trace plots of the correlations (not shown), which clearly indicate convergence to multimodal distributions.
- Analogous to the CPO measure described above, here we evaluate the correlation based performance of the models in terms of the densities of the true correlations under the corresponding posterior distributions. From Figs. 3.2 and 3.4, except for a few space–time pairs, our model significantly outperforms that of FR for all the remaining space–time pairs.
- Moreover, when the true correlations are close to zero, for all the space–time pairs, the densities of the true correlations under the corresponding posterior distributions are significantly higher than that of FR.
- The above facts strengthen our claim that, compared to other models, our correlation structure is sufficiently rich for capturing the actual correlations, specifically when the true correlation is close to zero for nonstationary models.

3.4. Real data analysis

3.4.1. Spatial data

According to the Clean Air Act certain air quality is to be maintained to protect the public health, and to maintain proper survival environment of animals and vegetations. As a measure of the quality of air, the Clean Air Act set standard limits for important air pollutants such as ozone. For our real data analysis we use the ozone metric called W126 metric. The impact of ozone exposure on trees, plants and ecosystems is often assessed using a seasonal index known as a "W126 index", which is the annual maximum of consecutive three month running total of weighted sum of hourly concentrations observed between 8AM and 8PM on each day during the high ozone season of April through October. A fundamental principle behind W126 metric is that higher hourly average ozone concentrations should be weighted more than middle and lower values when assessing human and environmental effects. The cumulative W126 exposure index uses a sigmoidally weighted function. The W126 index is a cumulative exposure index and not an "average" value. As indicated above, it is a biology based index, which is supported by research results (i.e., under both experimental and ambient conditions) that show that the higher hourly average ozone concentrations should be weighted greater than the mid- and lower-level values. The US EPA reviewed the National Ambient Air Quality Standards (NAAQS) for ozone in 2015, and determined that a 3-month W126 index level of 17 ppm-hrs is sufficient to protect the public welfare based on the latest science on effects of ozone on vegetation (US Federal Register, 2015). Also, we have information on Community Multiscale Air Quality indices (CMAQ), which is highly correlated with ozone level, so that we can use CMAQ as a covariate in our model.

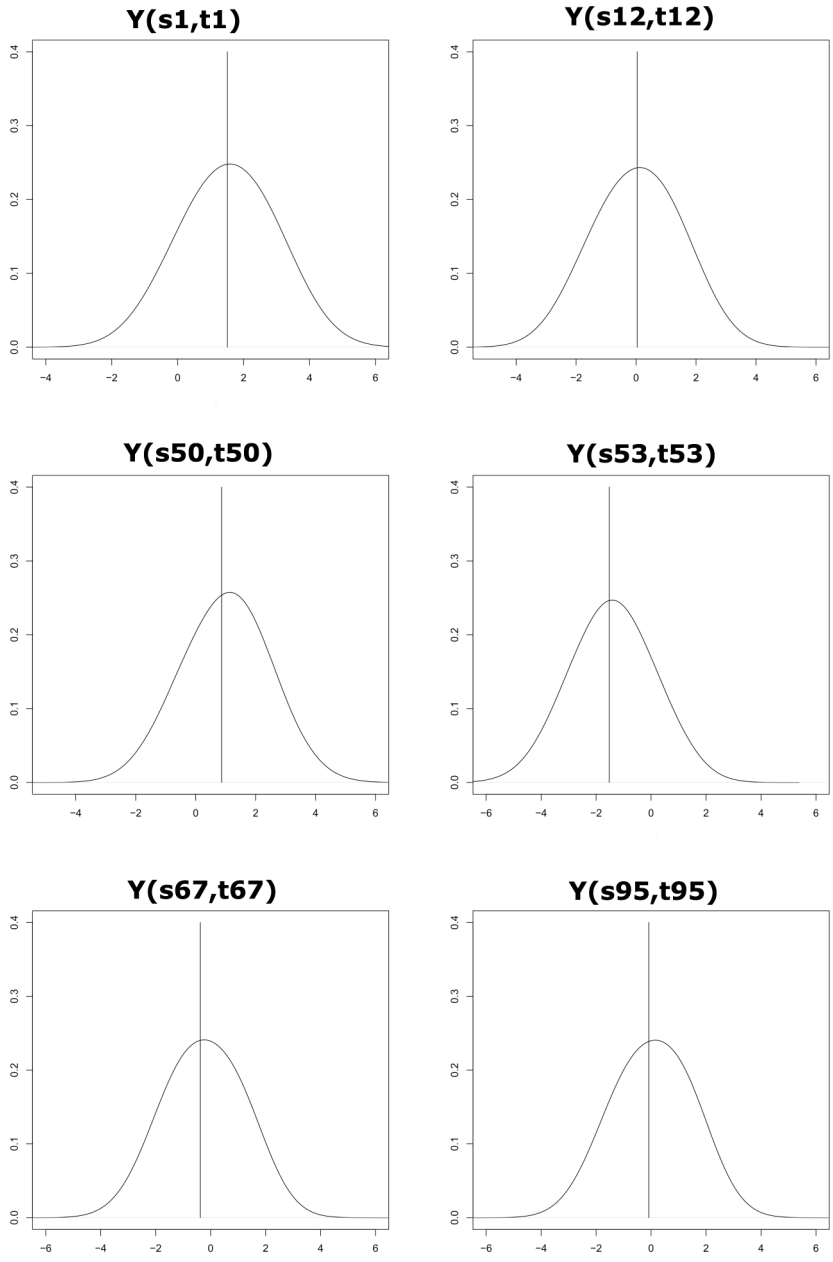


Fig. 3.3. Simulation study: Posterior predictive densities of $Y(\boldsymbol{s}, t)$ for the 6 different location-time pairs of FR — the corresponding true values are denoted by the vertical line.

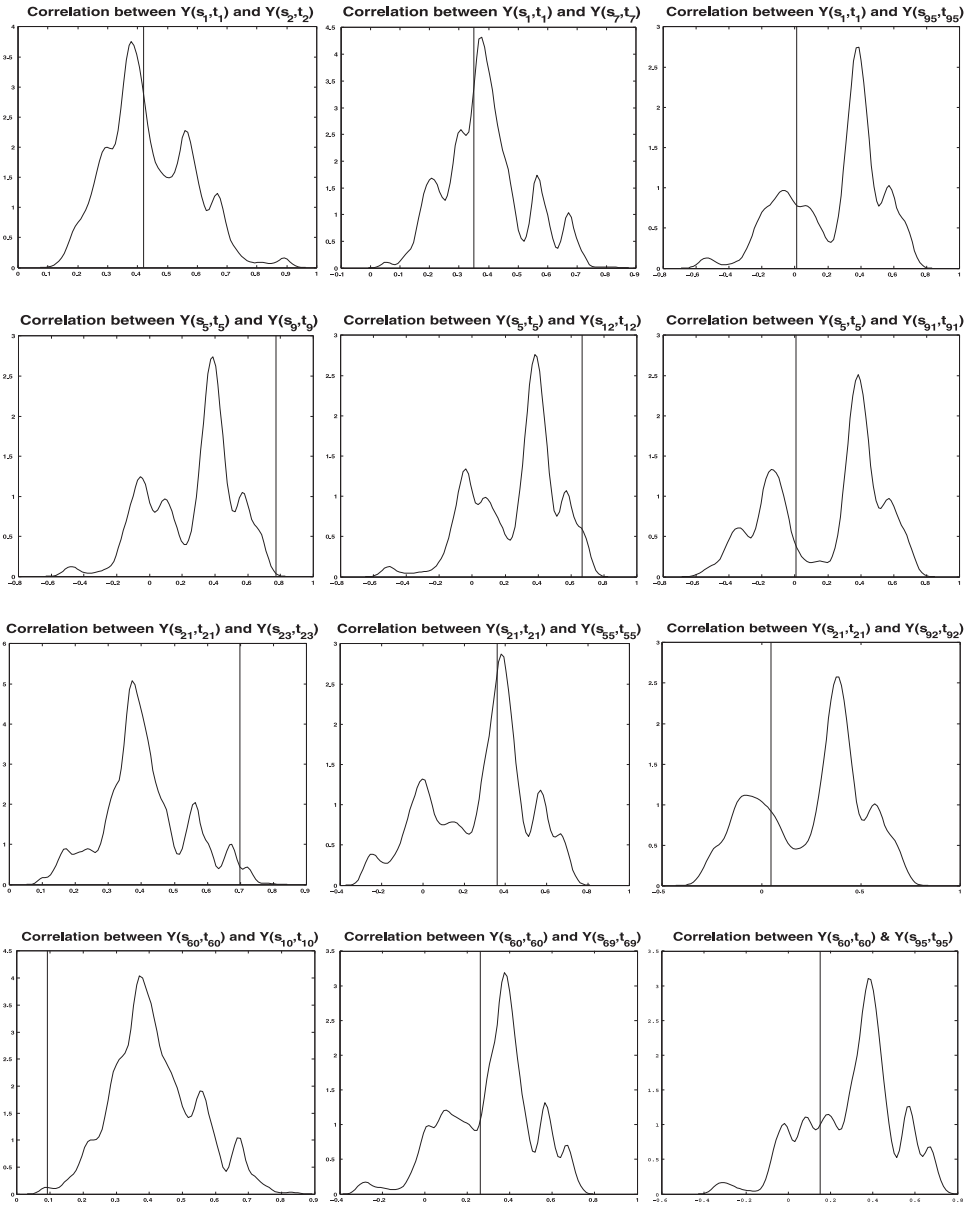


Fig. 3.4. Simulation study: Posterior densities of the correlations for the 12 different pairs of spatio-temporal points of FR; the vertical lines indicate the true correlations.

3.4.1.1. Calculating the W126 metric. Let $Q_l(\boldsymbol{s}, t)$ denote the observed ozone concentration level in parts per million (ppm) units at location $\boldsymbol{s}$ at hour l on day t, for $t = 1, \ldots, T$ and $l = 1, \ldots, 12$, where $T = 214$ days between April 1 and October 31 in a given year. The hours are the 12 day light hours between 8AM and 7PM. The W126 metric for site $\boldsymbol{s}$ is calculated as follows.

The weighted hourly metric is calculated using the transformation:

$$U_l(\boldsymbol{s}, t) = Q_l(\boldsymbol{s}, t) \times \left(\frac{1}{1 + 4403 \times \exp(-126 \times Q_l(\boldsymbol{s}, t))} \right).$$

This logistic transformation truncates the values smaller than 0.05 ppm to zero, but does not alter the magnitude of values larger than 0.10 ppm.

The daily index from the 12-hourly weighted values in each day is obtained as

$$Z(\boldsymbol{s}, t) = \sum_{l=1}^{12} U_l(\boldsymbol{s}, t).$$

The monthly index is calculated from the daily indices by summing and then adjusting for the number of days in the month as follows:

$$M_j(\boldsymbol{s}) = \sum_{t \in \text{month} j} Z(\boldsymbol{s}, t), j = 1, \ldots, 7,$$

where the summation is over all the days l that fall within the calender month j.

The three-month running totals are centered at the last month and are obtained as:

$$\bar{M}_j(\boldsymbol{s}) = \sum_{k=j-2}^{j} M_k(s), j = 3, \ldots, 7.$$

Finally, the annual W126 index value is calculated by:

$$Y(\boldsymbol{s}) = \max_{j=3}^{7} \bar{M}_j(\boldsymbol{s}).$$

The secondary ozone standard is met at a site $\boldsymbol{s}$ at a given year when the true value of $Y(\boldsymbol{s})$ is less than 21 ppm-hours.

Corresponding to each observed ozone concentration $Q_l(\boldsymbol{s}, t)$ we have a CMAQ model output $v_l(A, t)$, where the site $\boldsymbol{s}$ is contained in the unique grid cell A. Using the output $v_l(A, t)$ and the above details daily and annual indices of CMAQ values namely $X(A, t)$ and $X(A)$ are constructed.

We have data on annual indices of ozone values ($W126$) $Y(\boldsymbol{s})$, and corresponding CMAQ $X(A)$ values for 76 locations in the US. Now we fit our model to this real data set. Here we model the data on the log scale; we also use the log transformation of the CMAQ values. In other words, we consider

$$\log(Y(\boldsymbol{s})) = \alpha_0 + \alpha_1 \log(X(A_i)) + f(\boldsymbol{s}_i) + \epsilon_i, \ \ i = 1, \ldots, 76,$$

where α_0 and α_1 are regression coefficients, $f(\boldsymbol{s}_i)$ is an annual level spatial random effect at location $\boldsymbol{s}_i$ and ϵ_i is an independent nugget effect with variance σ^2. Here $f(\boldsymbol{s}_i)$ is our proposed spatial model based on kernel convolution with ODPP.

It is worth mentioning that we had initially considered a stationary kernel for convolution, but obtained a poor fit. This possibly suggested nonstationary process as an appropriate model, but until recently, we were not aware of any formal method for checking stationarity and nonstationarity in a completely nonparametric setup. Indeed, Roy and Bhattacharya (2020) proposed a novel recursive Bayesian methodology for characterizing stationarity and nonstationarity for general stochastic processes, among various other characterizations, and illustrated their ideas with ample examples in fields as varied as time series, MCMC convergence diagnosis, spatial and spatio-temporal setups, point processes, as well as (multiple) frequency determination of oscillating time series. With their ideas, they also analyze this ozone data to check stationarity. The details of their analyses and the results, presented in Section 13.7.1 of their paper, indicate that the ozone data is indeed nonstationary. Further, a simple quantile–quantile plot shows non-normality of the data.

The above arguments justify our nonparametric model choice and nonstationary kernel used for convolution with ODDP. All the prior distributions are the same as mentioned before. For the additional parameters α_0 and α_1, we use the vague prior distribution $N(0, 10^4)$, and for σ we use

the log-normal prior with mean zero and variance 10^4. The TTMCMC trace plots shown in Figure S-10.1 of the supplement bear out adequate performance of our model and methodologies.

As before, we assess the predictive power of the model using leave-one-out cross validation. For all the locations, the true value of ozone concentration lies within the 95% credible interval of the respective cross-validation posterior. This is summarized in the top panel of Fig. 3.5, where the middle surface represents the observed data; the lower and the upper surfaces represent the lower and the upper 95% credible regions associated with the respective leave-one-out posterior predictive densities. The surface in the middle of the bottom panel are the posterior medians, while the lower and the upper surfaces denote the 95% credible intervals as before. For the convenience of visually comparing the observed data and the posterior medians, we include Fig. 3.6, which also contains the 95% credible intervals. The plots clearly show that our proposed model is quite adequate for the ozone data.

Posterior densities of correlations, for 6 pairs of sites, are shown in Figure S-10.2 of supplement. All of them seem to give high posterior probability to the approximate range (0.1, 0.3).

3.5. Spatio-temporal data analysis

'Particulate matter' (PM) is the general term used for a mixture of solid particles and liquid droplets found in the air. Airborne PM comes from many different sources. "Primary" particles are released directly into the atmosphere from sources such as cars, trucks, heavy equipment, forest fires, and other burning activities. An extensive body of scientific evidence shows that there are adverse effects of this PM particles on health, including cardiovascular problems, premature death and many more. Ambient air monitoring stations generally measure air concentrations of different ranges of particles, but most monitoring station is for two size ranges: $PM_{2.5}$ and PM_{10}.

3.5.1. Data

Our data is a part of a big data set analyzed by Paciorek et al. (2009) (Data Source: http://www.stat.berkeley.edu/~paciorek/data/pm/). They specify stationary spatial structures through the use of penalized thin plate splines. Assumption of stationarity leads to an important simplification in their model. The assumption of stationarity is particularly appropriate for $PM_{2.5}$ values, but there is evidence of nonstationarity for PM_{10} values. Indeed, Roy and Bhattacharya (2020) infer with their novel Bayesian recursive methodology that the PM_{10} data is strictly, as well as weakly nonstationary (Section 13.7.2 of their paper) and that the $PM_{2.5}$ data is strictly stationary (Section 13.7.3 of their paper).

For illustration purpose, we fit a nonstationary spatio temporal model for a smaller section of the full data set. We analyze monthly average values of PM_{10} for the year 1988–2002 (180 time points) at 50 locations. There are few locations with fewer sample points. Our model will be appropriate for this kind of data, since we are using the spatial locations and time points as arguments of our proposed mean functional. Our data consists of total 3934 observations for monthly PM_{10} values. To increase the predictive performance of the model, we have used available covariate information for different spatial locations and time points. It is expected that inclusion of covariates may better explain the spatio-temporal heterogeneity. The details of the covariate selection are discussed in Yanosky et al. (2008a,b). The non-time-varying covariates are as follows: distances to the nearest road within four road size classes; particulate point source emissions within 1 and 10 km buffers; the proportion of urban land use of within 1 km; elevation; and block group, tract, and county population density from the 1990 US Census. The time varying covariates are wind speed, precipitation and barometric pressure, with hourly values averaged to the month at each station.

We also analyze the properties of the empirical correlations for increasing spatio-temporal lags with respect to the complete data set consisting of 70572 observations. Fig. 3.7, obtained from the raw correlations after taking moving averages of length 50 for better visualization, shows that the correlations tend to zero with increasing lags, as realistically expected, in spite of the data being nonstationary. Moreover, a simple quantile–quantile plot (not shown for brevity) shows that the data is far from normality. These very much support our modeling idea.

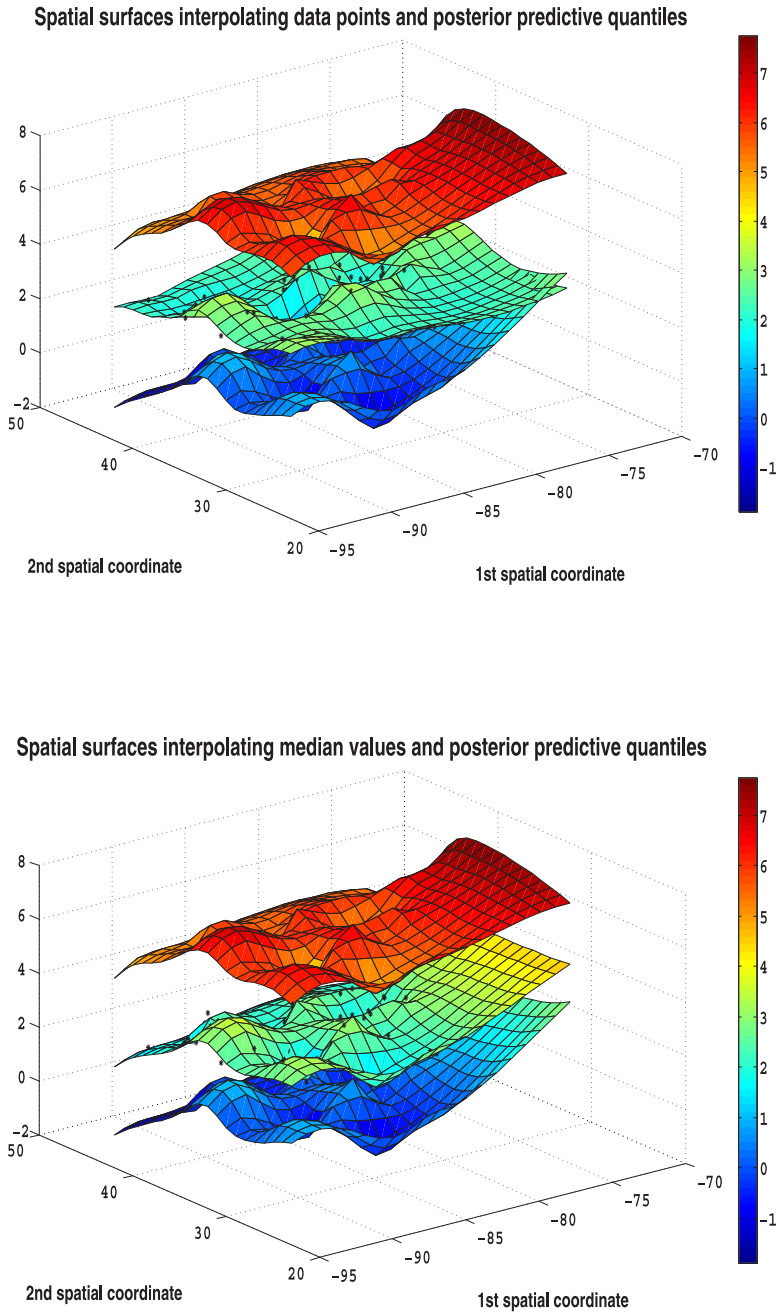


Fig. 3.5. Real spatial data analysis: The top panel shows the surface plot of ozone concentrations (middle), the lower and the upper 95% credible intervals associated with the leave-one-out posterior predictive densities, denoted by the lower and the upper surfaces, respectively. The bottom panel shows the surface plot of the posterior medians (middle) along with the lower and the upper 95% credible intervals associated with the leave-one-out posterior predictive densities (lower and the upper surfaces, respectively). The observed data points are indicated by '*'.

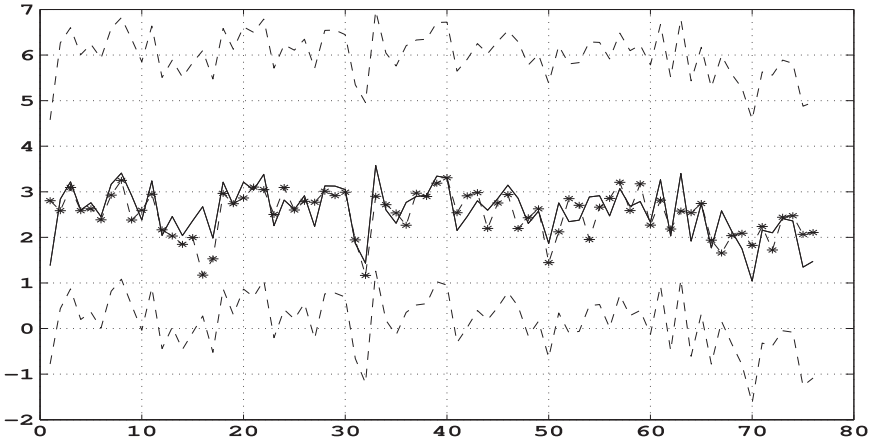


Fig. 3.6. Real spatial data analysis: Posterior predictive distributions summarized by the median (middle line) and the 95% credible intervals as a function of $\boldsymbol{s}$. The observed data points are denoted by '*'.

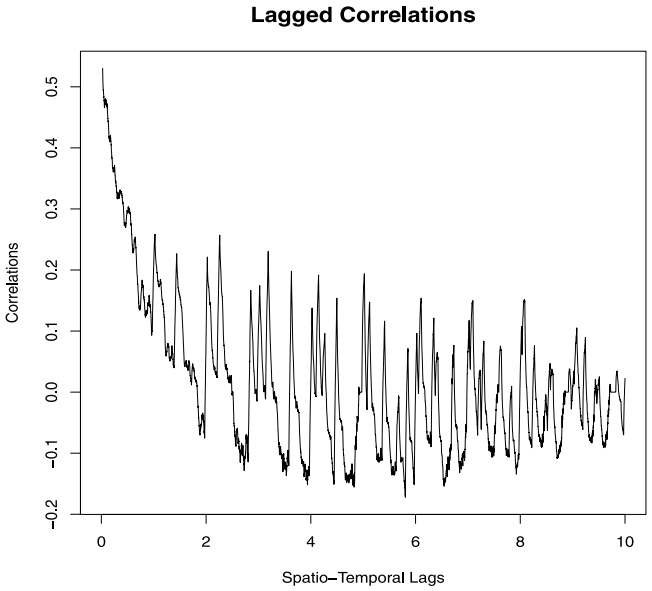


Fig. 3.7. Empirical correlations for increasing spatio-temporal lags. The raw correlations are smoothed by taking moving averages of size 50 for better visualization.

3.5.2. Model

We propose the following model for the real data:

$$\log y_{it} = \alpha_0 + f(\boldsymbol{s}_i, t) + g_1(\tilde{\boldsymbol{z}}_i) + g_2(\boldsymbol{z}_{it}) + \epsilon_{it},\ i = 1, \ldots, 50, t = 1, \ldots, 180,$$

where α_0 is intercept term, $f(\boldsymbol{s}_i, t)$ is our proposed spatio-temporal model based on kernel convolution with ODDP. In the above, g_1 and g_2 are function is of non-time varying covariates $\tilde{\boldsymbol{z}}_i$ and time varying covariates $\boldsymbol{z}_{it}$, respectively. We assume a Gaussian process prior on g_1 such that $\mu_1(\boldsymbol{z}) = E\left[g_1(\boldsymbol{z})\right] = \boldsymbol{\beta}'\boldsymbol{z}$ and $Cov\left(g_1(\boldsymbol{z}_i), g_1(\boldsymbol{z}_j)\right) = \exp\left(-\frac{1}{2}\|\boldsymbol{z}_i - \boldsymbol{z}_j\|\right)$.

We set g_2 as a linear function of time varying covariates: $g_2(\boldsymbol{z}) = \boldsymbol{\gamma}'\boldsymbol{z}$. The assumption of linearity in g_2 will simplify our computation to a great extent. Also there is evidence from the previous analysis that using linear terms in places of the unknown function led to only negligible decrease in predictive ability. In our model, ϵ_{it} are independent nugget effects with variance σ^2.

All the prior distributions are the same as mentioned before. For the additional parameters $\alpha_0, \boldsymbol{\beta}, \boldsymbol{\gamma}$, we use the vague prior distribution $N(0, 10^4)$, and for σ we use the log-normal prior with mean zero and variance 10^4.

3.5.3. Implementation

Note that here we have total 3934 number of observations. We have to update the number of parameters ranging between 300 to 400. Our TTMCMC based algorithm took 25 minutes to generate 5000 observations following a burn in of 20000. As in the other cases, TTMCMC exhibited satisfactory acceptance rate and mixing properties, as evident from the trace plots displayed in n Section S-10.3 of supplement.

3.5.4. Leave-one-out cross validation

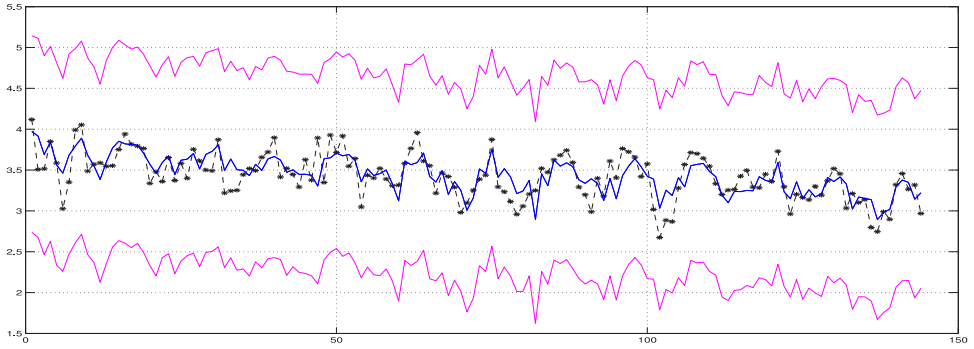
As before, we assess the predictive power of the model using leave-one-out cross validation. For all the spatio-temporal points, the true value of PM_{10} lies within the 95% credible interval of the respective cross-validation posterior. Also we calculate the mean square prediction error ($MSPE$), given by $\frac{\sum(y_{it}-\hat{y}_{it})^2}{n}$, where $\hat{y}_{it}$ is the median of the posterior predictive density at the spatial location $(\boldsymbol{s}_i, t)$. In this case, we obtain $MSPE = 0.101$. Fig. 3.8 displays the observed data and posterior medians for at three spatial locations having data for more than 10 years, which also contains 95% credible intervals. We have also reported $MSPE$ for these three locations. The values are significantly lower than overall $MSPE$. It reveals the fact that our model have captured more precise information for the spatial locations, having larger number of time points. We also provide a visual representation of the model performance at 50 locations summarized over time points. In Fig. 3.9, the surface represents the posterior median values, averaged over all month-specific predictions for 50 spatial locations. From the plots, it is clear that our model performs quite satisfactorily for the data. Posterior densities of correlations, for 6 pairs of sites, are shown in Figure S-10.4 of the supplement.

4. Summary and conclusion

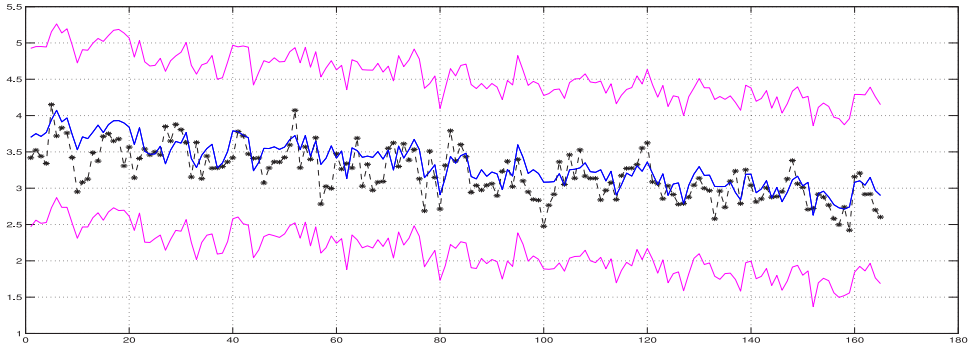
In this article, we have developed a nonstationary, non-Gaussian spatio-temporal model based on kernel convolution of ODDP. Dependence is induced in the weights through similarities in the ordering of the atoms. Using this property, we could ensure that our model-based correlation between two random data points, which are widely separated, will be close to zero. We incorporated non-stationarity via appropriate kernels, which would be convolved with ODDP. Although our proposed model is nonstationary and non-separable, it includes stationarity and separability as special cases. Moreover, since our model is based on kernel convolution, replication is unnecessary for inference. If one wishes to achieve different degrees of smoothness across space and across time, then that is also allowed by our model framework. For example, if we associate the ODDP prior only to the spatial locations, then the process will become smoother across time than across space, depending on the choice of the kernel.

From the computational point of view, we have developed a fast and efficient TTMCMC-based algorithm for implementing our variable dimensional spatio-temporal model. Indeed, our model consists of a large number of variables, where the number of variables associated with the summands is random. Using TTMCMC, we could update all the parameters and the number of the parameters simultaneously, using simple deterministic transformations of some one-dimensional random variable.

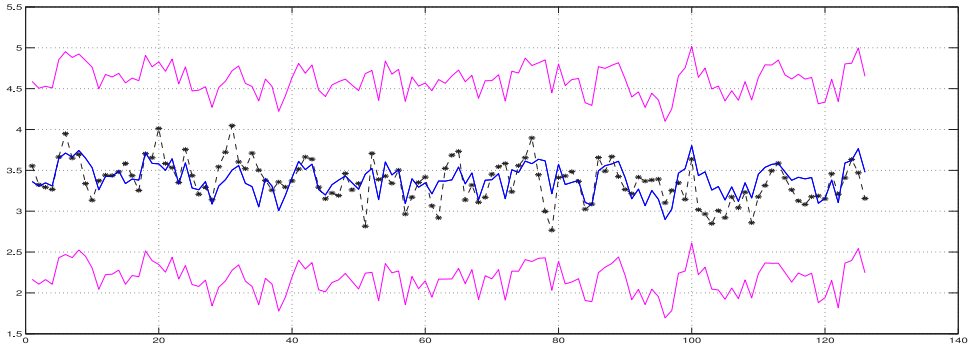
We illustrated the performance of our model with a simulation study and compared our model's performance with the FR's model. The comparative study supports our claim that our model can capture the zero correlations between two widely separated data points (in space and/or time) more precisely. We have also applied our model and methods to two real data examples of spatial and



(a) spatial location:(38.270833 -85.740278), MSPE:0.0344



(b) spatial location:(41.600278 -87.334722), MSPE:0.051



(c) spatial location:(39.766111,-86.129167), MSPE:0.048

Fig. 3.8. Real spatio-temporal data analysis: Posterior predictive distributions summarized by the median (middle line) and the 95% credible intervals as a function of t for three randomly chosen spatial locations. The observed data points are denoted by '*'.

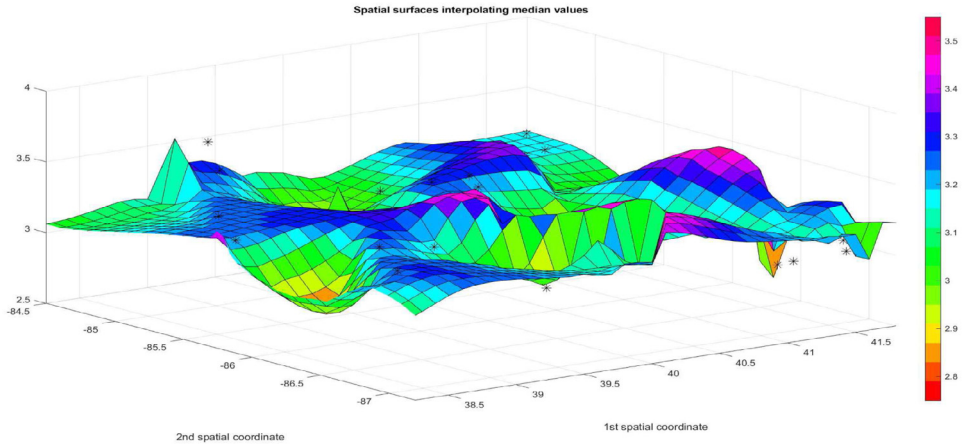


Fig. 3.9. Real spatio-temporal data analysis: the surface plot of posterior median values at 50 locations, averaged over all month specific predictions from 1988–2002. The observed data points are indicated by '*'.

spatio-temporal dependence. As illustrated in detail, our model exhibited excellent performance in both cases.

Although for the current paper we restricted ourselves to spatio-temporal applications only, our model is readily applicable in the functional data context. In fact, in the context of nonparametric function estimation, a new class of prior distributions can be introduced through our proposed model. Note that unknown functions can be modeled as a limit of a weighted sum of kernels or generator functions indexed by continuous parameters. In our model the weights will be the p_i's of ODDP, and kernels are indexed by θ_i, where for $i = 1, 2, \ldots, \theta_i \overset{iid}{\sim} G_0$. We have already obtained some sufficient conditions ensuring that our model converges in L_p norm and Besov semi-norm. These results make our proposed model a promising candidate for function estimation.

Acknowledgments

We are grateful to Prof. Sujit Sahu for kindly providing us with the ozone data set and to Dr Suman Guha and Dr. Maria Xose Rodriguez for helpful discussions. The main work of this paper was carried out during PhD work of Moumita Das under the supervision of Sourabh Bhattacharya. This research was supported by PhD Fellowship provided to Moumita Das from the Indian Statistical Institute and Basque Center of Applied Mathematics (BCAM) Severo Ochoa accreditation (SEV-2017-0 718); and project MTM2017-82 379-R (Agencia Estatal de Investigación/Fondo Europeo de Desarrollo Regional, Unión Europea; principal investigator Dr. Maria Xose Rodriguez, BCAM).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spasta.2023.100751.

References

Banerjee, S., 2017. High-dimensional Bayesian geostatistics. Bayesian Anal. 12 (2), 583–614.
Banerjee, S., Gelfand, A.E., 2003. On smoothness properties of spatial processes. J. Multivariate Anal. 84, 85–100.
Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. J. R. Stat. Soc. Ser. B Stat. Methodol. 70 (4), 825–848.
Bhattacharya, S., 2021. Bayesian Lévy-dynamic spatio-tepmoral process: Towards big data analysis. Available at "arXiv:2105.08451".

Calder, J., 2014. An overview of nonstationary spatial modeling. https://cpb-us-w2.wpmucdn.com/u.osu.edu/dist/5/5306/files/2015/12/nonstationary-overview-rau7u9.pdf.
Chang, Y.-M., Hsu, N.-J., Huang, H.-C., 2011. Semiparametric estimation and selection for nonstationary spatial covariance functions. J. Comput. Graph. Statist. 19, 117–139.
Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. J. R. Stat. Soc. Ser. B Stat. Methodol. 70 (1), 209–226.
Damian, D., Sampson, P.D., Guttorp, P., 2001. Bayesian estimation of semi-parametric non-stationary spatial covariance structures. Environmetrics 12, 161–178.
Das, M., Bhattacharya, S., 2019a. Supplement to "Nonstationary, Nonparametric, Nonseparable Bayesian Spatio-Temporal Modeling Using Kernel Convolution of Order Based Dependent Dirichlet Process". arXiv preprint.
Das, M., Bhattacharya, S., 2019b. Transdimensional transformation based Markov chain Monte Carlo. Braz. J. Probab. Stat. 33, 87–138.
Duan, J.A., Gelfand, A.E., Sirmans, C.F., 2009. Modeling space-time data using stochastic differential equations. Bayesian Anal. 4, 733–758.
Duan, J.A., Guindani, M., Gelfand, A.E., 2007. Generalized spatial Dirichlet process models. Biometrika 94, 809–825.
Dutta, S., Bhattacharya, S., 2014. Markov chain Monte Carlo based on deterministic transformations. Stat. Methodol. 16, 100–116, Also available at http://arxiv.org/abs/1106.5850. Supplement available at http://arxiv.org/abs/1306.6684.
Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. Ann. Statist. 1, 209–230.
Ferguson, T.S., 1974. Prior distributions on spaces of probability measures. Ann. Statist. 2, 615–629.
Fuentes, M., 2002. Spectral methods for nonstationary spatial processes. Biometrika 89, 197–210.
Fuentes, M., Reich, B., 2013. Multivariate spatial nonparametric modelling via kernel process mixing. Statist. Sinica 23, 75–97.
Fuentes, M., Smith, R.L., 2001. A new class of nonstationary spatial models. Technical Report, Department of Statistics, North Carolina State University.
Geisser, S., 1993. Predictive inference : An for introduction. Chapman & Hall, London.
Gelfand, A.E., Kottas, A., MacEachern, S.N., 2005. Bayesian nonparametric spatial modeling with Dirichlet process mixing. J. Amer. Statist. Assoc. 100, 1021–1035.
Gilani, O., Berrocal, V.J., Batterman, S.A., 2016. Non-stationary spatio-temporal modeling of traffic-related pollutants in near-road environments. Spatial and Spatio-Temporal Epidemiol. 18, 24–37.
Griffin, J.E., Steel, M.F.J., 2004. Semiparametric Bayesian inference for stochastic frontier models. J. Econometrics 123, 121–152.
Griffin, J.E., Steel, M.F.J., 2006. Order-based dependent Dirichlet processes. J. Amer. Statist. Assoc. 101, 179–194.
Guttorp, P., Sampson, P.D., 1994. Methods for estimating heterogeneous spatial covariance functions with environmental applications. In: Patil, G.P., Rao, C.R. (Eds.), Handbook of Statistics XII: Environmental Statistics. Elsevier/North Holland, New York, pp. 663–690.
Guttorp, P., Schmidt, A.M., Bartlett, M., Besag, J., 2013. Covariance structure of spatial and spatiotemporal processes. WIREs Comput. Stat. 279–287.
Haas, T.C., 1995. Local prediction of a spatio-temporal process with an application to wet sulfate deposition. J. Amer. Statist. Assoc. 90, 1189–1199.
Higdon, D., 1998. A process-convolution approach to modeling temperatures in the North Atlantic Ocean. Environ. Ecol. Stat. 5, 173–190.
Higdon, D., 2001. Space and space-time modeling using process convolutions. In: Barnett, C.W.A.V., Chatwin, P.C., El-Sharaawi, A.H. (Eds.), Quantitative Methods for Current Environmental Issues. Springer-Verlag, London, pp. 37–56.
Higdon, D., Swall, J., Kern, J., 1999. Non-stationary spatial modeling. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 6. Oxford University Press, Oxford, pp. 761–768.
Ingebrigtsen, R., Lindgren, F., Steinsland, I., 2014. Spatial models with explanatory variables in the dependence structure. Spatial Stat. 8, 20–38.
Ishwaran, H., James, L.F., 2001. Gibbs sampling methods for stick-breaking prior. J. Amer. Statist. Assoc. 96, 161–173.
Katzfuss, M., 2013. Bayesian nonstationary spatial modeling for very large datasets. Environmetrics 24 (3), 189–200.
Kim, H.-m., Mallick, B.K., Holmes, C.C., 2005. Analyzing nonstationary spatial data using piecewise Gaussian processes. J. Amer. Statist. Assoc. 653–668.
Kottas, A., Duan, J.A., Gelfand, A.E., 2007. Modeling disease incidence data with spatial and spatio-temporal Dirichlet process mixtures. Biom. J. 49, 1–14.
Mazumder, S., Banerjee, S., Bhattacharya, S., 2022. A New Spatio-Temporal Model Exploiting Hamiltonian Equations. ArXiv Preprint.
Neto, J.H.V., Schmidt, A.M., Guttorp, P., 2014. Accounting for spatially varying directional effects in spatial covariance structures. J. R. Stat. Soc. Ser. C. Appl. Stat. 63, 103–122.
Nott, D.J., Dunsmuir, W.T.M., 2002. Estimation of nonstationary spatial covariance structure. Biometrika 89, 819–829.
Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., Sain, S., 2015. A multiresolution Gaussian process model for the analysis of large spatial datasets. J. Comput. Graph. Statist. 24 (2), 579–599.
Paciorek, C.J., 2003. Nonstationaty Gaussian Process for Regression and Spatial Modeling (Doctoral thesis). Carnegie Mellon University.
Paciorek, C., Schervish, M., 2006. Spatial modelling using a new class of nonstationary covariance functions. Environmetrics 17 (5), 483–506.
Paciorek, C.J., Yanosky, J.D., Puett, R.C., 2009. Practical large-scale spatio-temporal modeling of particulate matter concentrations. Ann. Appl. Stat. 3, 370–397.

Petrone, S., Guindani, M., Gelfand, A.E., 2009. Hybrid Dirichlet mixture models for functional data. J. R. Stat. Soc. Ser. B Stat. Methodol. 71, 755–782.
Pettit, L., 1990. The conditional of predictive-ordinate for the normal distribution. J. R. Stat. Soc. Ser. B Stat. Methodol. 52, 175–184.
Reich, B.J., Fuentes, M., Dunson, D.B., 2011. Bayesian spatial quantile regression. J. Amer. Statist. Assoc. 106, 6–20.
Ren, Q., Banerjee, S., 2013. Hierarchical factor models for large spatially misaligned data: A low-rank predictive process approach. Biometrics 69 (1), 19–30.
Risser, M.D., Calder, C.A., 2015. Regression-based covariance functions for nonstationary spatial modeling. Environmetrics 26, 284–297.
Risser, M.D., Calder, C.A., Berrocal, V.J., Berrett, C., 2019. Nonstationary spatial prediction of soil organic carbon: Implications for stock assessment decision making. Ann. Appl. Stat. 13, 165–188.
Roy, S., Bhattacharya, S., 2020. Bayesian Characterizations of Properties of Stochastic Processes with Applications. ArXiv Preprint.
Sampson, P.D., Guttorp, P., 1992. Nonparametric estimation of nonstationary spatial covariance structure. J. Amer. Statist. Assoc. 87, 108–119.
Schmidt, A.M., Guttorp, P., O'Hagan, A., 2011. Considering covariates in the covariance structure of spatial processes. Enironmetrics 22, 487–500.
Schmidt, A.M., O'Hagan, A., 2003. Bayesian inference for nonstationary spatial covariance structure via spatial deformations. J. R. Stat. Soc. Ser. B Stat. Methodol. 65, 743–758.
Sethuraman, J., 1994. A constructive definition of Dirichlet priors. Statist. Sinica 4, 639–650.
Shand, L., Li, B., 2017. Modeling nonstationarity in space and time. Biometrics 73 (3), 759–768.
Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer-Verlag, New York.
Stein, M.L., 2014. Limitations on low rank approximations for covariance matrices of spatial data. Spatial Stat. 8, 1–19.
Wolpert, R.L., Clyde, M.A., Tu, C., 2011. Stochastic expansions using continuous dictionaries: Lévy adaptive regression kernels. Ann. Statist. 39, 1916–1962.
Yaglom, A.M., 1987a. Correlation Theory of Stationary and Related Random Functions–Volume-I: Basic Results. Springer-Verlag, New York.
Yaglom, A.M., 1987b. Correlation Theory of Stationary and Related Random Functions–Volume-II: Supplemtary Notes and References. Springer-Verlag, New York.
Yanosky, J.D., Paciorek, C.J., Schwartz, J., Laden, F., Puett, R.C., Suh, H.H., 2008a. Spatio-temporal modeling of chronic PM10 exposures of nurses health study. Atmos. Environ. 47, 4047–4062.
Yanosky, J.D., Paciorek, C.J., Suh, H.H., 2008b. Predicting chronic fine particulate exposures using spatio-temporal models for the northeastern and midwestern U.S.. Environ. Health Perspect. 117, 522–529.