

Lagrangian Spatio-Temporal Nonstationary Covariance Functions



Mary Lai O. Salvaña and Marc G. Genton

Abstract The Lagrangian reference frame has been used to model spatio-temporal dependence of purely spatial second-order stationary random fields that are being transported. This modeling paradigm involves transforming a purely spatial process to spatio-temporal by introducing a transformation in the spatial coordinates. Recently, it has been used to capture dependence in space and time of transported purely spatial random fields with second-order nonstationarity. However, under this modeling framework, the presence of mechanisms enforcing second-order nonstationary behavior introduces considerable challenges in parameter estimation. To address these, we propose a new estimation methodology which includes modeling the second-order nonstationarity parameters by means of thin plate splines and estimating all the parameters via two-step maximum likelihood estimation. In addition, through numerical experiments, we tackle the consequences of model misspecification. That is, we discuss the implications, both in the stationary and nonstationary cases, of fitting Lagrangian spatio-temporal covariance functions to data generated from non-Lagrangian models, and vice versa. Lastly, we apply the Lagrangian models and the new estimation technique to analyze particulate matter concentrations over Saudi Arabia.

1 Introduction

The need for models that explain spatio-temporal dependencies of environmental processes has been answered with a growing number of studies on spatio-temporal covariance functions. A number of the established spatio-temporal covariance func-

M. L. O. Salvaña · M. G. Genton (✉)
King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900,
Saudi Arabia
e-mail: marc.genton@kaust.edu.sa

M. L. O. Salvaña
e-mail: marylai.salvana@kaust.edu.sa

tions can only model spatio-temporal random fields that are second-order stationary in space and time. The list includes the spatio-temporal separable stationary covariance functions, spatio-temporal stationary mixture models (Ma 2003a), and the Gneiting class of spatio-temporal stationary covariance functions (Gneiting 2002). However, environmental processes are notorious for exhibiting second-order nonstationarity in space and/or time. The number of available spatio-temporal nonstationary covariance functions catering to this challenging second-order nonstationary behavior is slowly increasing but still lags behind its stationary counterpart. The construction approaches that define the current state-of-the-art for spatio-temporal nonstationary covariance functions modeling include the spatio-temporal dimension expansion (Shand and Li 2017), the spatio-temporal convolution (Garg et al. 2012), and the nonstationary Archimedean spectral densities (Porcu et al. 2009). Some spatio-temporal nonstationary models built from spatio-temporal stationary covariances and intrinsically stationary variograms were also proposed in Ma (2003b). Several other works on incorporating spatial nonstationarity focused on allowing the parameters in the covariance function to vary in space (Higdon et al. 1999; Neto et al. 2014; Paciorek and Schervish 2006; Stein 2005). These types of nonstationary covariance functions belong to a wider class of kernel convolution methods. Risser (2016), Sampson et al. (2001) feature comprehensive overviews of this wider class. Another flexible class of spatio-temporal nonstationary models termed the spatio-temporal random effects (STRE) models was put forward in Cressie et al. (2010). STRE combines the utilities of basis function approximations and Kalman filtering to achieve dimension reduction in space and fast and dynamic predictions in time. This class is highly useful in modeling large space-time nonstationary data.

A distinct class of spatio-temporal covariance functions has been championed for capturing a special behavior of a subset of spatio-temporal random fields. The class of Lagrangian spatio-temporal covariance functions was developed to model spatio-temporal dependence of transported purely spatial random fields through the use of the Lagrangian reference frame. Models springing from this technique obtain higher covariances along the direction of transport than the covariances lying in the other directions. However, much of the progress in this area was done in stationary variants such as Cox and Isham (1988), where this modeling technique was first proposed, and Salvaña et al. (2020), where the multivariate extension was explored. A recent treatment of this modeling scheme in the multivariate nonstationary setup was provided in Salvaña and Genton (2020). In this work, we formally establish the univariate nonstationary variant of the Lagrangian approach to spatio-temporal covariance construction. Moreover, we propose an efficient estimation methodology such that the novelty of the Lagrangian spatio-temporal nonstationary models translates to usability.

The rest of this paper is organized as follows. Section 2 reviews the developments in the Lagrangian spatio-temporal modeling and formulates the univariate nonstationary extension. Section 3 proposes a practical estimation procedure for nonstationary covariance models of the Lagrangian type. Section 4 presents some simulation studies designed to illustrate the advantages of Lagrangian spatio-temporal models over

other established spatio-temporal models. Section 5 details the application of the new models to a spatio-temporal particulate matter dataset. Section 6 draws a conclusion.

2 Lagrangian Spatio-Temporal Covariances

Under second-order stationarity of the purely spatial random field, Cox and Isham (1988) established that a new class of spatio-temporal stationary covariance functions can be constructed from purely spatial stationary covariance functions by utilizing the principles of Lagrangian reference frame. That is, define a spatio-temporal second-order stationary random field

$$Z(\mathbf{s}, t) = \tilde{Z}(\mathbf{s} - \mathbf{V}t), \quad (\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}, \mathbf{V} \in \mathbb{R}^d, d \geq 1,$$

such that $\tilde{Z}(\mathbf{s})$ is a purely spatial second-order stationary random field. Here \mathbf{V} is a random vector, independent from the purely spatial random field, that describes the velocity of the transport of $\tilde{Z}(\mathbf{s})$ and is often called the advection velocity vector. The resulting spatio-temporal stationary covariance function of $Z(\mathbf{s}, t)$ is

$$\text{cov}\{Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)\} = \text{cov}\{\tilde{Z}(\mathbf{s}_1 - \mathbf{V}t_1), \tilde{Z}(\mathbf{s}_2 - \mathbf{V}t_2)\} = E_{\mathbf{V}}\{C^S(\mathbf{h} - \mathbf{V}u)\}, \quad (1)$$

where $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$, $u = t_1 - t_2$, and $C^S(\cdot)$ is the purely spatial stationary covariance function of $\tilde{Z}(\mathbf{s})$ on \mathbb{R}^d . By introducing a transformation on the spatial arguments of $C^S(\cdot)$, the number of available spatio-temporal stationary covariance functions would greatly expand by as much as the number of valid purely spatial stationary covariance functions.

The model in (1) can be extended to accommodate multiple variables of interest as shown in Salvaña et al. (2020). That is, suppose at each spatio-temporal location (\mathbf{s}, t) there are $p > 1$ observations corresponding to p different features. This means that the purely spatial second-order stationary random field is now vector valued, i.e., $\tilde{\mathbf{Z}}(\mathbf{s}) = \{\tilde{Z}_1(\mathbf{s}), \dots, \tilde{Z}_p(\mathbf{s})\}^\top$. A multivariate spatio-temporal random field can be similarly defined as above, i.e., $\mathbf{Z}(\mathbf{s}, t) = \tilde{\mathbf{Z}}(\mathbf{s} - \mathbf{V}t) = \{\tilde{Z}_1(\mathbf{s} - \mathbf{V}t), \dots, \tilde{Z}_p(\mathbf{s} - \mathbf{V}t)\}^\top$, with matrix-valued spatio-temporal stationary cross-covariance function

$$\text{cov}\{\mathbf{Z}(\mathbf{s}_1, t_1), \mathbf{Z}(\mathbf{s}_2, t_2)\} = \text{cov}\{\tilde{\mathbf{Z}}(\mathbf{s}_1 - \mathbf{V}t_1), \tilde{\mathbf{Z}}(\mathbf{s}_2 - \mathbf{V}t_2)\} = E_{\mathbf{V}}\{\mathbf{C}^S(\mathbf{h} - \mathbf{V}u)\}, \quad (2)$$

where $\mathbf{C}^S(\cdot)$ is the $p \times p$ matrix-valued purely spatial stationary cross-covariance function of $\tilde{\mathbf{Z}}(\mathbf{s})$ on \mathbb{R}^d . This newly defined multivariate spatio-temporal random field is second-order stationary in space and time.

Using these two previous developments of spatio-temporal covariance functions, a recent review paper further developed the Lagrangian approach in the multivariate nonstationary arena. Salvaña and Genton (2020) established that the model in (2) can be tailored to accommodate an underlying cross-covariance function \mathbf{C}^S that is

nonstationary. This is particularly useful when the multivariate purely spatial random field being transported has nonnegligible second-order nonstationarity. Models arising from their proposal have the form

$$\text{cov}\{\mathbf{Z}(\mathbf{s}_1, t_1), \mathbf{Z}(\mathbf{s}_2, t_2)\} = \text{cov}\{\tilde{\mathbf{Z}}(\mathbf{s}_1 - \mathbf{V}t_1), \tilde{\mathbf{Z}}(\mathbf{s}_2 - \mathbf{V}t_2)\} = \text{Ev} \left\{ \mathbf{C}^S(\mathbf{s}_1 - \mathbf{V}t_1, \mathbf{s}_2 - \mathbf{V}t_2) \right\}, \quad (3)$$

where $\mathbf{C}^S(\cdot, \cdot)$ is a matrix-valued purely spatial nonstationary cross-covariance function of $\tilde{\mathbf{Z}}(\mathbf{s})$ on \mathbb{R}^d .

The models in Eqs. (1)–(3) suggest how the Lagrangian framework can be used to create spatio-temporal covariance functions when one has at one's disposal purely spatial covariance functions that are either univariate stationary, multivariate stationary, or multivariate nonstationary. The univariate nonstationary formulation of the Lagrangian construction can be readily established from (3) when $p = 1$. For completeness, we state this as a theorem below.

Theorem 1 *Let \mathbf{V} be a random vector on \mathbb{R}^d . If $C^S(\mathbf{s}_1, \mathbf{s}_2)$ is a valid purely spatial nonstationary covariance function on \mathbb{R}^d , then,*

$$C(\mathbf{s}_1, \mathbf{s}_2; t_1, t_2) = \text{Ev} \left\{ C^S(\mathbf{s}_1 - \mathbf{V}t_1, \mathbf{s}_2 - \mathbf{V}t_2) \right\}, \quad \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^d, t_1, t_2 \in \mathbb{R}, \quad (4)$$

is a valid spatio-temporal nonstationary covariance function on $\mathbb{R}^d \times \mathbb{R}$ provided that the expectation exists.

The validity of this theorem follows because it is a special case ($p = 1$) of a theorem proved for general p in Salvaña and Genton (2020). The construction approach in Theorem 1 requires a purely spatial nonstationary covariance function, $C^S(\cdot, \cdot)$, and returns a spatio-temporal covariance function that is nonstationary in both space and time. Theorem 1 implies a purely spatial random field with second-order nonstationarity that is transported to new locations at a velocity \mathbf{V} . The transport behavior, dictated by the velocity \mathbf{V} , influences the covariance through shifting the original spatial arguments of $C^S(\cdot, \cdot)$ by $\mathbf{V}t$. The derived Lagrangian spatio-temporal nonstationary covariance function $C(\mathbf{s}_1, \mathbf{s}_2; t_1, t_2)$ is nonstationary in space, as its fundamental building block is a purely spatial nonstationary covariance function, and is also nonstationary in time, as the transformation from purely spatial to spatio-temporal depends on time t .

There is a rich literature on valid purely spatial nonstationary covariance functions from which we can choose $C^S(\cdot, \cdot)$ including the dimension expansion (Bornn et al. 2012), deformation approach (Sampson and Guttorp 1992), kernel-based methods (Higdon et al. 1999), convolution-based methods (Heaton et al. 2014; Higdon 1998, 2002), spectral methods (Fuentes 2002), orthogonal expansions (Nychka and Saltzman 1998), spatially varying parameters (Neto et al. 2014; Paciorek and Schervish 2006; Gelfand et al. 2004), piece-wise Gaussian process (Kim et al. 2005), covariate-driven approaches (Schmidt et al. 2011), and basis function models (Nychka et al. 2002; Wikle 2010; Chang et al. 2010). Other purely spatial nonstationary models to which Theorem 1 can be applied are discussed in Sampson et al. (2001), Risser (2015), and Stephenson et al. (2004).

Lagrangian spatio-temporal random fields can be classified into two general categories, namely, frozen and non-frozen random fields. The former characterizes Lagrangian spatio-temporal random fields with a constant advection velocity, that is, $\mathbf{V} = \mathbf{v}$. Meanwhile, Lagrangian spatio-temporal random fields that are termed non-frozen are those transported with a random advection velocity \mathbf{V} . Salvaña and Genton (2020) showed realizations of frozen Lagrangian spatio-temporal random fields simulated from (3) when $p = 2$ using prominent classes of purely spatial nonstationary cross-covariance functions, such as the multivariate spatially varying parameters and the multivariate deformation models. Realizations of frozen Lagrangian spatio-temporal nonstationary random fields from the model in (4) can be obtained similarly by assuming that Z_1 and Z_2 in Fig. 2 of Salvaña and Genton (2020) are independent. In the following figures, we show non-frozen Lagrangian spatio-temporal random fields for two models when $\mathbf{V} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Figure 1a plots the simulated $Z(\mathbf{s}, t)$ from the model

$$C(\mathbf{s}_1, \mathbf{s}_2; t_1, t_2) = \text{E}_{\mathbf{V}} \left(\sigma(\mathbf{s}_1 - \mathbf{V}t_1, \mathbf{s}_2 - \mathbf{V}t_2) \mathcal{M}_{\nu} \left[\{\mathbf{s}_1 - \mathbf{s}_2 - \mathbf{V}(t_1 - t_2)\}^{\top} \right. \right. \\ \left. \left. \times \mathbf{D}(\mathbf{s}_1 - \mathbf{V}t_1, \mathbf{s}_2 - \mathbf{V}t_2)^{-1} \{\mathbf{s}_1 - \mathbf{s}_2 - \mathbf{V}(t_1 - t_2)\} \right]^{1/2} \right), \quad (5)$$

where $\sigma(\mathbf{s}_1 - \mathbf{V}t_1, \mathbf{s}_2 - \mathbf{V}t_2)$ is the spatially varying variance parameter and the matrix $\mathbf{D}(\mathbf{s}_1 - \mathbf{V}t_1, \mathbf{s}_2 - \mathbf{V}t_2)$ serves as the spatially varying scale parameter (Kleiber and Nychka 2012). Here $\mathcal{M}_{\nu}(\cdot)$ is the univariate Matérn correlation with smoothness parameter $\nu > 0$, $\mathbf{D}(\mathbf{s}_1, \mathbf{s}_2) = \frac{1}{2} \{\mathbf{D}(\mathbf{s}_1) + \mathbf{D}(\mathbf{s}_2)\}$, and $\sigma(\mathbf{s}_1, \mathbf{s}_2) = |\mathbf{D}(\mathbf{s}_1)|^{1/4} |\mathbf{D}(\mathbf{s}_2)|^{1/4} |\mathbf{D}(\mathbf{s}_1, \mathbf{s}_2)|^{-1/2}$. The matrix $\mathbf{D}(\mathbf{s})$ is parameterized through its spectral decomposition, i.e.

$$\mathbf{D}(\mathbf{s}) = \begin{bmatrix} \cos \{\phi(\mathbf{s})\} & -\sin \{\phi(\mathbf{s})\} \\ \sin \{\phi(\mathbf{s})\} & \cos \{\phi(\mathbf{s})\} \end{bmatrix} \begin{bmatrix} \lambda_1(\mathbf{s}) & 0 \\ 0 & \lambda_2(\mathbf{s}) \end{bmatrix} \begin{bmatrix} \cos \{\phi(\mathbf{s})\} & \sin \{\phi(\mathbf{s})\} \\ -\sin \{\phi(\mathbf{s})\} & \cos \{\phi(\mathbf{s})\} \end{bmatrix}.$$

Figure 1b illustrates the random field generated from the non-frozen Lagrangian deformation

$$C(\mathbf{s}_1, \mathbf{s}_2; t_1, t_2) = \text{E}_{\mathbf{V}} \left[\sigma^2 \mathcal{M}_{\nu} \{a \|\mathbf{f}(\mathbf{s}_1 - \mathbf{V}t_1) - \mathbf{f}(\mathbf{s}_2 - \mathbf{V}t_2)\|\} \right], \quad (6)$$

where $\mathbf{f} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a deterministic nonlinear smooth bijective deformation function and σ^2 and a are the variance and scale parameters, respectively. In the example in Fig. 1b, $\sigma^2 = a = \nu = 1$.

To illustrate the effect of the advection velocity $\mathbf{V} \sim \mathcal{N}_2 \{(0.1, 0.1)^{\top}, 0.01 \times \mathbf{I}_2\}$ on the space-time dependence of the random fields in Fig. 1, we examine two locations, marked with ‘×’, which we call “reference locations”. We plot as heatmaps the covariance between the observations at each reference location and the observations at all locations, including the reference locations themselves. For example, in Fig. 2a, the first image in the first row gives the covariance between $Z(\mathbf{s}_{\text{Ref Loc } 1}, 1)$ and $Z(\mathbf{s}_l, 1)$, at every pixel location $\mathbf{s}_l, l = 1, \dots, 2500$. The second image in the first

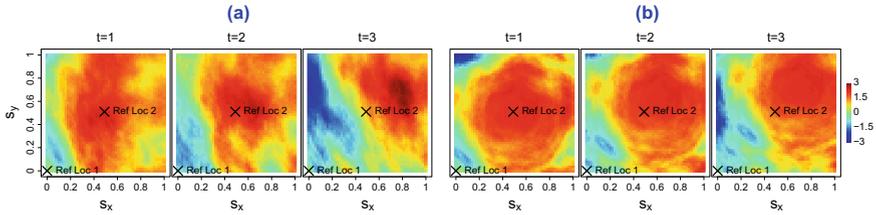


Fig. 1 Simulated realizations in the unit square on a 50×50 grid from the non-frozen Lagrangian nonstationary covariance models in (5) and (6) with $\mathbf{V} \sim \mathcal{N}_2 \{ (0.1, 0.1)^\top, 0.01 \times \mathbf{I}_2 \}$, \mathbf{I}_2 is the 2×2 identity matrix. (a) The spatially varying parameters have the following representations: for $\mathbf{s} = (s_x, s_y)^\top$, $\phi(\mathbf{s}) = (s_x - 0.5) + 2(s_y - 0.5) + (s_y - 0.5)^2$, $\lambda_1(\mathbf{s}) = -3 - 6(s_x - 0.5)^2 - 7(s_y - 0.5)^2$, and $\lambda_2(\mathbf{s}) = -5 + (s_x - 0.5)^2 - 4(s_y - 0.5)^2$. (b) The deformation function assumed is the point-source deformation, i.e., $\mathbf{f}(\mathbf{s}) = \mathbf{b} + (\mathbf{s} - \mathbf{b})\{1 + \exp(-0.5\|\mathbf{s} - \mathbf{b}\|^2)\}$, $\mathbf{b} = (0.15, 0.15)^\top$. Reference locations 1 and 2 are marked with ‘x’

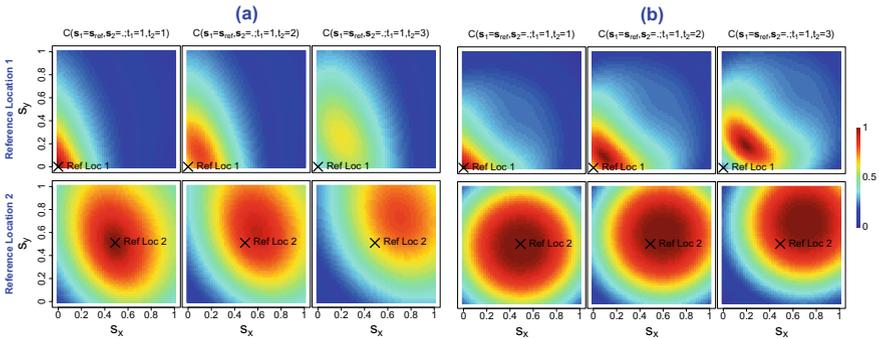


Fig. 2 Heatmaps of the non-frozen Lagrangian nonstationary covariance models in (5) and (6) observed at two reference locations marked with ‘x’. (a) shows the strengths of dependence between any two locations in space and time under the spatially varying parameters model and (b) under the deformation model. See Fig. 1 for the corresponding random field realizations

row plots the covariance between $Z(\mathbf{s}_{\text{Ref Loc } 1}, 1)$ and $Z(\mathbf{s}_l, 2)$, at every pixel location \mathbf{s}_l , $l = 1, \dots, 2500$. Lastly, the third image in the first row plots the covariance between $Z(\mathbf{s}_{\text{Ref Loc } 1}, 1)$ and $Z(\mathbf{s}_l, 3)$, at every pixel location \mathbf{s}_l , $l = 1, \dots, 2500$. All the other plots are organized in the same manner. Notice that among the covariances taken at the same temporal locations, i.e., $t_1 = t_2$, the maximum covariance occurs at the reference location. However, among the covariances taken between any two space-time locations that are one time step apart, the maximum covariance no longer occurs at the reference location. Instead, it can be observed at a spatial location $(0.1, 0.1)^\top$ away from the reference location. A similar observation can be made when taking covariances between any two space-time locations that are two time steps apart.

3 Estimation

The parameters for any spatio-temporal nonstationary covariance functions spawned by the Lagrangian approach include both purely spatial and advection velocity parameters. The estimation methods to recover the former depend on the form of C^S and are already fully developed in their respective references; see Sect. 2. Here we propose a way to extend those estimation methods to space-time in order to recover both the purely spatial and the additional advection velocity parameters. We focus on an estimation strategy that operates on the spatio-temporal nonstationary covariance matrix built using all the spatio-temporal locations. This allows inferences regarding the second-order nonstationarity structure of the transported purely spatial random field possible. However, alternative estimation strategies which involve fitting local spatio-temporal stationary models can also be considered (Kuusela and Stein 2018).

3.1 Thin Plate Splines

Throughout the remainder of this work, we narrow our attention to Lagrangian spatio-temporal nonstationary models whose C^S are the deformation and spatially varying parameters models. We focus on these two classes because their second-order nonstationarity parameters can be considered a surface and we aim to leverage a technique used to model surfaces, namely, thin plate splines (TPS). The TPS is a basis function and is used to interpolate surfaces using a predetermined set of landmarks or the locations where the basis functions are centered (Bookstein 1989; Wahba 1990; Donato and Belongie 2002; Chen and Geman 2014). TPS is a central topic in morphometrics and has found a wide range of applications including biomedical, computer vision, data mining, and engineering (Whitbeck and Guo 2006; Hegland et al. 1997; Tenakoon et al. 2013; Chen et al. 2017; Bazen and Gerez 2003). This section describes how TPS can be appropriately applied to model the second-order nonstationarity parameters of the Lagrangian spatio-temporal nonstationary models.

Suppose $\psi(\mathbf{s})$ is an unknown second-order nonstationarity parameter of interest at spatial location \mathbf{s} . This parameter might be the x – or y –coordinate in the new spatial domain for the deformation model or the spatially varying parameters $\lambda_1(\mathbf{s})$, $\lambda_2(\mathbf{s})$, or $\phi(\mathbf{s})$. The TPS model for $\psi(\mathbf{s})$ is

$$\psi(\mathbf{s}) = A_1 + A_2s_x + A_3s_y + \sum_{i=1}^L w_i U(\|\mathbf{s}_i^* - \mathbf{s}\|^2), \quad (7)$$

where $U(h) = h^2 \log h$, for $h > 0$, and zero otherwise, is a basis function, $\mathbf{A} = (A_1, A_2, A_3)^\top \in \mathbb{R}^3$ and $\mathbf{w} \in \mathbb{R}^L$ are the parameters responsible for the affine and nonlinear components of the transformation, respectively, and L is the number of landmarks. Sampson (2015) pointed out several problems springing from the formulation in (7), including multiple local maxima in the log-likelihood function and

highly correlated parameters. Hence, following their recommendation, we adopt the form in (7) with $w_i = \sum_{j=1}^{L-3} \beta_j g_{i,j}$, such that $\mathbf{g}_j = (g_{1,j}, \dots, g_{L,j})^\top \in \mathbb{R}^L$, $j = 1, \dots, L-3$, also called the principal warps, are the last $L-3$ eigenvectors of the bending energy matrix \mathbf{B} corresponding to its $L-3$ nonzero eigenvalues. The bending energy matrix \mathbf{B} is the upper left $L \times L$ sub-matrix of $\mathbf{B} = [\mathbf{D} \mathbf{P}; \mathbf{P}^\top \mathbf{O}]^{-1} \in \mathbb{R}^{(L+d+1) \times (L+d+1)}$ with elements:

- $\mathbf{D} \in \mathbb{R}^{L \times L}$ such that for $l, r = 1, \dots, L$, $D_{lr} = d_{lr}^2 \log(d_{lr})$, if $l \neq r$, and $D_{lr} = 0$, otherwise, where $d_{lr} = \|\mathbf{s}_l^* - \mathbf{s}_r^*\|$,
- $\mathbf{P} \in \mathbb{R}^{L \times (d+1)}$, where the l -th row of \mathbf{P} is $(1, \mathbf{s}_l^\top)$, $\mathbf{s} \in \mathbb{R}^d$, and $l = 1, \dots, L$, and
- \mathbf{O} is a zero matrix in $\mathbb{R}^{(d+1) \times (d+1)}$.

Together, the linear combinations of the coefficients, $\beta_{i,j}$, and the principal warps, \mathbf{g}_j , are termed partial warps.

A key ingredient in the TPS model is the set of landmarks, $\{\mathbf{s}_1^*, \mathbf{s}_2^*, \dots, \mathbf{s}_L^*\}$. The TPS model interpolates at these landmark points while preserving maximal smoothness (Bazen and Gerez 2003). The placement of these landmarks dictates the quality of the parameter estimates (Lewis et al. 2004). The landmarks and the number of landmarks are fixed prior to modeling and the choice is left to the discretion of the modeler. In the morphometrics literature, the landmarks are often positioned where important features can be observed (Gunz and Mitteroecker 2013). In the spatial statistics literature, the observation locations are commonly designated as landmarks (Kleiber et al. 2014).

In studying Lagrangian spatio-temporal random fields, there is a need to distinguish between the observation locations and the domain of the transported random field. The former refers to the predefined locations where measurements are obtained, e.g., regular latitude/longitude grid, wireless sensor networks, wind turbine sites, meteorological towers, and many others. The latter has its own coordinate system. The measurements contained in the transported random field get picked up by the data collection tools at the observation locations as the random field travels past them. In frozen Lagrangian spatio-temporal random fields, the measurement $Z(\mathbf{s}, t)$ collected at observation location \mathbf{s} at time t corresponds to the measurement $Z(\mathbf{s} - \mathbf{v}t)$ at spatial location $\mathbf{s} - \mathbf{v}t$ in the domain of the transported random field. Figure 3 shows a frozen Lagrangian spatio-temporal deformed random field traveling at a constant velocity of $\mathbf{v} = (0.5, 0.5)^\top$. While the observation locations are fixed at any time, the corresponding locations in the Lagrangian random field are not. Choosing the observation locations as landmarks, therefore, will not suffice in capturing the non-stationarity of the entire Lagrangian spatio-temporal random field as every region in the domain should be represented by these landmarks. Assuming that the domain of the Lagrangian spatio-temporal random field is larger than the domain of observation locations, we advocate to situate the landmarks on a regular grid that covers the entire Lagrangian spatio-temporal random field. In practice, unfortunately, the appropriate size and resolution of this regular grid of landmarks cannot be identified prior to modeling. However, cross-validation studies can be performed to determine the suitable positioning and number of landmarks.

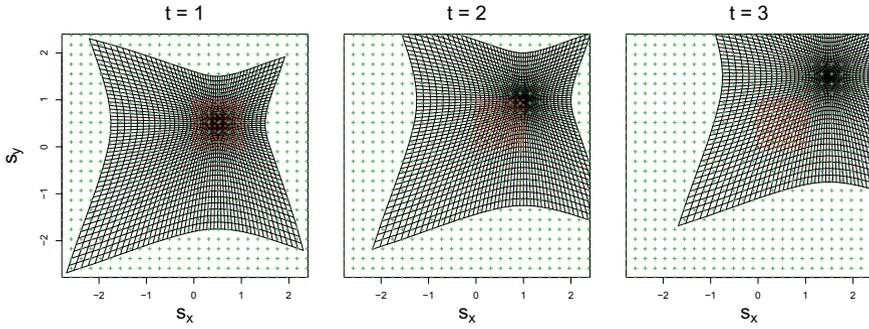


Fig. 3 Marked in red are the observation locations on a regular 10×10 grid. Superimposed in black are the spatial locations on the domain of the frozen Lagrangian spatio-temporal deformed random field which travels past the observation locations with an advection velocity $\mathbf{v} = (0.5, 0.5)^\top$, and in green are the landmarks. The landmarks (green) may or may not coincide with the observation locations (red)

3.2 Maximum Likelihood Estimation and Likelihood Approximations in the Temporal Domain

Having established the representation of the unknown nonstationarity parameters, we introduce the estimation procedure carried out in this work. Suppose $\mathbf{Z} = \{Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2), \dots, Z(\mathbf{s}_n, t_n)\}^\top$ is a zero mean measurement vector where $n \in \mathbb{Z}^+$ is the total number of space-time locations. Inference is performed through maximizing the log-likelihood

$$l(\boldsymbol{\Theta}; \mathbf{Z}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}(\boldsymbol{\Theta})| - \frac{1}{2} \mathbf{Z}^\top \boldsymbol{\Sigma}(\boldsymbol{\Theta})^{-1} \mathbf{Z} \tag{8}$$

with respect to all the parameters collected in $\boldsymbol{\Theta} \in \mathbb{R}^q$. Here $\boldsymbol{\Theta}$ includes all the purely spatial, advection velocity, and the TPS parameters, and q is the total number of parameters. The $n \times n$ covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\Theta})$ is formed by a parametric spatio-temporal nonstationary covariance function. Penalties can be introduced to Equation (8) such as the L_1 penalty for the deformation models in order to avoid folding of the surface (Sampson 2015).

For spatio-temporal measurements that are regularly spaced in time, \mathbf{Z} can be rewritten as $\mathbf{Z} = (\mathbf{Z}_1^\top, \dots, \mathbf{Z}_T^\top)^\top \in \mathbb{R}^{N \cdot T}$ such that $\mathbf{Z}_t = \{Z(\mathbf{s}_1, t), \dots, Z(\mathbf{s}_N, t)\}^\top \in \mathbb{R}^N$, for $t = 1, \dots, T$. Here N and T specify the number of spatial and temporal locations, respectively, and $n = N \cdot T$. Furthermore, the log-likelihood function above can be approximated as follows:

$$l(\boldsymbol{\Theta}; \mathbf{Z}_1, \dots, \mathbf{Z}_T) \approx l(\boldsymbol{\Theta}; \mathbf{Z}_{1,t^*}) + \sum_{j=t^*+1}^T l(\boldsymbol{\Theta}; \mathbf{Z}_j | \mathbf{Z}_{j-t^*, j-1}), \tag{9}$$

where $\mathbf{Z}_{a,b} = (\mathbf{Z}_a^\top, \dots, \mathbf{Z}_b^\top)^\top \in \mathbb{R}^{Mt^*}$, for $a < b$, and t^* specifies the number of consecutive temporal locations included in the conditional distribution. Here $l(\boldsymbol{\Theta}; \mathbf{Z}_j | \mathbf{Z}_{j-t^*,j-1})$ is the log-likelihood function based only on the vector of space-time measurements $\mathbf{Z}_{j-t^*,j-1} = (\mathbf{Z}_{j-t^*}^\top, \dots, \mathbf{Z}_{j-1}^\top)^\top$. This kind of approximation is usually preferred when T is large and the dependence in time relies heavily only on the more recent measurements (Stein 2005c).

3.3 Two-Step Maximum Likelihood Estimation

The inclusion of the nonstationarity parameters in the model increases the dimension of the estimation problem. This kind of setup is known to run into numerical difficulties and complications (Kathuria et al. 2019; Zhu and Wu 2010; Li and Sun 2018). Therefore, as a practical alternative to joint estimation of all the parameters, in this work, the estimation problem is split into two parts. First, a Lagrangian spatio-temporal stationary model is assumed and all the associated purely spatial and advection parameters are estimated by maximizing the approximated log-likelihood in (9). Second, fixing the estimates found in the first step, the nonstationary version of the model is assumed and the parameters involved in the TPS are estimated also by maximizing (9). After the second step, it is likely that the optimization routine may still not reach the global maximum of (9). Hence, assuming the nonstationary model, iterating between the two steps several times is pursued until a stopping criterion is satisfied.

4 Simulation Study: Lagrangian Versus Non-Lagrangian Spatio-Temporal Models

The Lagrangian spatio-temporal covariance functions are primarily used to model transported space-time data. There are other classes of spatio-temporal covariance functions that model space-time data that are not necessarily transported. In this section, we investigate the outcome of fitting a non-Lagrangian model to transported space-time data and the outcome of fitting a Lagrangian model to space-time data that are not transported. We conduct the study under both second-order stationarity and nonstationarity assumptions.

4.1 Second-Order Stationarity

For the Lagrangian spatio-temporal model, we hinge our simulation studies on a particular class of non-frozen models whose explicit forms were derived in Schlather

(2010). When $\mathbf{V} \sim \mathcal{N}_d(\boldsymbol{\mu}_V, \boldsymbol{\Sigma}_V)$ and C^S is the stationary squared exponential covariance function, the model in (1) takes the form

$$C(\mathbf{h}, u) = \frac{1}{\sqrt{|\mathbf{I}_d + \boldsymbol{\Sigma}_V u^2|}} \exp \left\{ -a (\mathbf{h} - \boldsymbol{\mu}_V u)^\top (\mathbf{I}_d + \boldsymbol{\Sigma}_V u^2)^{-1} (\mathbf{h} - \boldsymbol{\mu}_V u) \right\}, \quad (10)$$

where $a > 0$ is a scale parameter in space inherited from C^S , and $\boldsymbol{\mu}_V$ and $\boldsymbol{\Sigma}_V$ are the Lagrangian parameters. When $\boldsymbol{\mu}_V = \mathbf{0}$ and $\boldsymbol{\Sigma}_V = \sigma_V^2 \mathbf{I}_d$, the Lagrangian model above reduces to

$$C(\mathbf{h}, u) = \frac{1}{(1 + \sigma_V^2 u^2)^{d/2}} \exp \left(-\frac{a \|\mathbf{h}\|^2}{1 + \sigma_V^2 u^2} \right), \quad (11)$$

which is a spatio-temporal isotropic covariance function under the Gneiting class (Gneiting 2002). The Gneiting model in (11), therefore, corresponds to a particular Lagrangian model wherein the advection velocity vector has mean zero and has independent components with common variance. While σ_V^2 is interpreted as the marginal variance of each component of \mathbf{V} in Lagrangian models, in non-Lagrangian models such as that in (11), σ_V^2 serves as a scale parameter in time, whose inverse controls the range of dependence in time.

A question of scientific interest is how the two models differ when the components of the advection velocity are no longer uncorrelated or when they do not share a common variance or when the advection velocity vector has a nonzero mean. To answer the first inquiry, we can scrutinize the form in (10) and compare it with (11). Suppose $d = 2$, $\boldsymbol{\mu}_V = \mathbf{0}$, and $\boldsymbol{\Sigma}_V = \sigma_V^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ then (10) reduces to

$$C(\mathbf{h}, u) = \frac{1}{\sqrt{(1 + \sigma_V^2 u^2)^2 - (\rho \sigma_V^2 u^2)^2}} \exp \left[-a \left\{ \frac{(h_x^2 + h_y^2)(1 + \sigma_V^2 u^2) - 2h_x h_y \rho \sigma_V^2 u^2}{(1 + \sigma_V^2 u^2)^2 - (\rho \sigma_V^2 u^2)^2} \right\} \right]. \quad (12)$$

Direct comparisons between (11) and (12) for different values of ρ are not straightforward since the terms bearing ρ involve the temporal lag u and the components of the spatial lag $\mathbf{h} = (h_x, h_y)^\top$. However, we can plot the values of (12) for different ρ , u , and \mathbf{h} , in order to visualize how the non-frozen Lagrangian spatio-temporal model deviates from the non-Lagrangian spatio-temporal model when the components of \mathbf{V} are correlated. Figure 4 provides such illustrations. It juxtaposes the covariance function values of the non-frozen Lagrangian spatio-temporal model, C^{LGR} for notational convenience, at different combinations of spatial lags with Euclidean norm equal to 1, at $u = 1, 2$, and 3, and at different strengths of dependence between the components of the advection velocity. In the plots, the values of the covariance function are plotted as the distance from the origin $(0, 0)$ to (h_x, h_y) . Note that the case $\rho = 0$ corresponds to the spatio-temporal Gneiting model in (11), denoted as C^G . The isotropy of C^G , at any u , manifests by the constant value of C^G when evaluated at any (h_x, h_y) . Another standout observation is that the value of C^{LGR} depends on

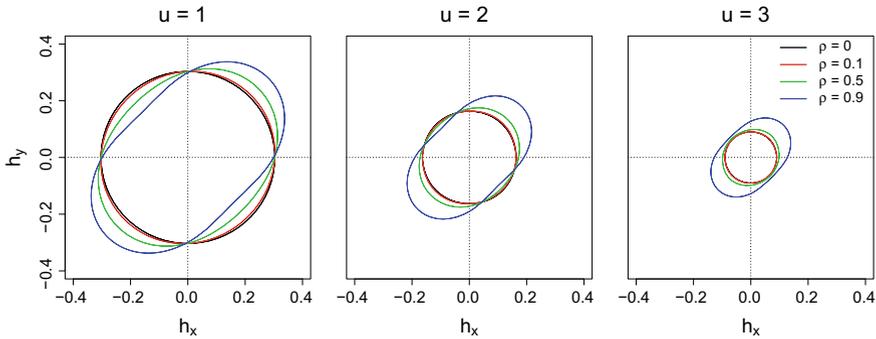


Fig. 4 Values of the non-frozen Lagrangian spatio-temporal covariance model in (12) for $\rho = 0, 0.1, 0.5,$ and $0.9,$ at temporal lags $u = 1, 2,$ and $3,$ at every $\mathbf{h} = (h_x, h_y)^\top$ such that $\|\mathbf{h}\|_2 = 1.$ Note that the case $\rho = 0$ corresponds to the non-Lagrangian Gneiting model in (11)

the signs of the components of the spatial lag and the magnitude of the correlation parameter $\rho.$

It can also be seen in the example in Fig. 4 that at $u = 1,$ when h_x and h_y have the same signs, C^G is less than $C^{LGR}.$ However, when h_x and h_y have different signs, C^G is greater than $C^{LGR}.$ This relationship between C^G and C^{LGR} at $u = 1$ does not persist as the temporal lag increases as other scenarios are observed. At $u = 3,$ for example, C^G and C^{LGR} are almost identical when ρ is near zero. However, when $\rho = 0.5$ or $\rho = 0.9,$ C^G is less than C^{LGR} in any direction. The difference, therefore, between C^G and C^{LGR} under the presence of a nonzero dependence parameter between the components of \mathbf{V} is not clear-cut but can be explored under some scenarios. Nevertheless, the deviation of C^{LGR} from C^G gets more pronounced as ρ increases.

We turn to some numerical experiments to answer the other unexplored questions, including what happens when C^G is fitted to data simulated from $C^{LGR},$ denoted $D^{LGR},$ such that the components of \mathbf{V} have different marginal variances or \mathbf{V} has a nonzero mean. Suppose $T = 10, N = 100, d = 2.$ The values $D^{LGR} = \mathbf{Z} = (\mathbf{Z}_1^\top, \dots, \mathbf{Z}_{10}^\top)^\top,$ such that $\mathbf{Z}_t = \{Z(\mathbf{s}_1, t), \dots, Z(\mathbf{s}_{100}, t)\}^\top, (\mathbf{s}, t) \in \mathbb{R}^2 \times \mathbb{R},$ are simulated from (10), with $a = 5,$ on a 10×10 grid in the unit square, under the following distributions of $\mathbf{V}:$

- (a) $\mathbf{V} \sim \mathcal{N}_2 \left\{ \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\}$ at different values of $\rho;$
- (b) $\mathbf{V} \sim \mathcal{N}_2 \left\{ \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \right\}$ at different values of $\sigma_y^2;$
- (c) $\mathbf{V} \sim \mathcal{N}_2 \left\{ \boldsymbol{\mu} = (\mu, \mu)^\top, \boldsymbol{\Sigma} = \mathbf{I}_2 \right\}$ at different values of $\mu.$

We reserve the values of \mathbf{Z}_{10} for prediction purposes and use the remaining 900 spatio-temporal realizations for estimation. Given the small problem size, full maximum likelihood estimation is performed; see (8). At this point, we question the effect of different values of $\rho, \sigma_y^2,$ and μ on the estimates of σ_y^2 in the non-Lagrangian model

in (11). Figure 5 gives the boxplots of parameter estimates $\hat{\sigma}_{\mathbf{V}}^2$ for 100 rounds of fitting C^G on D^{LGR} . The values of $\hat{\sigma}_{\mathbf{V}}^2$ reflect the changing degree of dependence in space-time as we change the values of the different parameters associated to the distribution of \mathbf{V} . In the first panel in Fig. 5, for example, when $\rho = 0.9$, the median of the estimates is 0.887 which translates to a stronger dependence in time, a fact also established in Fig. 4. In the middle set of boxplots, interestingly, the median of $\hat{\sigma}_{\mathbf{V}}^2$ is approximately equal to $(1 + \sigma_y^2)/2$. This result cannot be easily explained mathematically. Numerically, however, this is expected as the optimization routine finds the isotropic model parameters that maximize the log-likelihood given data simulated from a model with elliptical contours that are stretched in the x-axis. Lastly, as the mean of \mathbf{V} gets farther from $\mathbf{0}$, the estimate for $\sigma_{\mathbf{V}}^2$ has to compensate for a faster decorrelation in time which explains the increasing median of $\hat{\sigma}_{\mathbf{V}}^2$ as μ increases. In the initial experiments concerning Fig. 5c, a number of experimental replicates obtained $\hat{\sigma}_{\mathbf{V}}^2$ with values greater than 100 as μ increases. To obtain more compact boxplots, we re-ran the experiments and bounded the values that $\hat{\sigma}_{\mathbf{V}}^2$ can take to 10. This does not alter the insights provided by the unconstrained version of the experiments for Fig. 5c and the results presented in Fig. 5a and b. That is, as the non-frozen Lagrangian spatio-temporal model deviates from the non-Lagrangian scenario, i.e., $\mathbf{V} \sim \mathcal{N}_2(\boldsymbol{\mu}_{\mathbf{V}}, \boldsymbol{\Sigma}_{\mathbf{V}})$, where $\boldsymbol{\mu}_{\mathbf{V}} = \mathbf{0}$, and $\boldsymbol{\Sigma}_{\mathbf{V}} = \sigma_{\mathbf{V}}^2 \mathbf{I}_2$, the more disparate the models (10) and (11) become.

Next, we study the effect of ρ on the predictions and the quality of those predictions. Often, the assessment of the quality of the predictions is done by computing the Mean Square Error (MSE)

$$\text{MSE} = \frac{1}{100} \sum_{l=1}^{100} \left\{ \hat{Z}(\mathbf{s}_l, 10) - Z(\mathbf{s}_l, 10) \right\}^2,$$

where $\hat{Z}(\mathbf{s}_l, 10)$ is the prediction for $Z(\mathbf{s}_l, 10)$ at spatial location $\mathbf{s}_l, l = 1, \dots, 100$, and temporal location $t = 10$. Assuming the mean of the measurement vector that was used to estimate the parameters is $\mathbf{0}$, i.e., $E(\mathbf{Z}_{1,9}) = \mathbf{0}$, where $\mathbf{Z}_{1,9} = (\mathbf{Z}_1^\top, \dots, \mathbf{Z}_9^\top)^\top$, predictions are computed using the simple kriging predictor

$$\hat{Z}(\mathbf{s}_l, 10) = \mathbf{c}_l^\top \boldsymbol{\Sigma}(\boldsymbol{\Theta})^{-1} \mathbf{Z}_{1,9}.$$

Here \mathbf{c}_l is the vector of $N \times (T - 1)$ covariance function values between $Z(\mathbf{s}_l, 10)$ and $\mathbf{Z}(\mathbf{s}_r, t), r = 1, \dots, N$ and $t = 1, \dots, 9$, i.e.

$$\mathbf{c}_l = \{C(\mathbf{s}_l, \mathbf{s}_1; 10, 1), \dots, C(\mathbf{s}_l, \mathbf{s}_N; 10, 1), C(\mathbf{s}_l, \mathbf{s}_1; 10, 2), \dots, C(\mathbf{s}_l, \mathbf{s}_N; 10, 9)\}^\top. \tag{13}$$

Nevertheless, the MSE is unable to give an appropriate measure of the loss of statistical efficiency in cases when a different model is used instead of the true model. In this regard, we turn to the proposed criteria of Stein (1999), namely, the Loss of Efficiency (LOE) and the Misspecification of the Mean Square Error (MOM). The

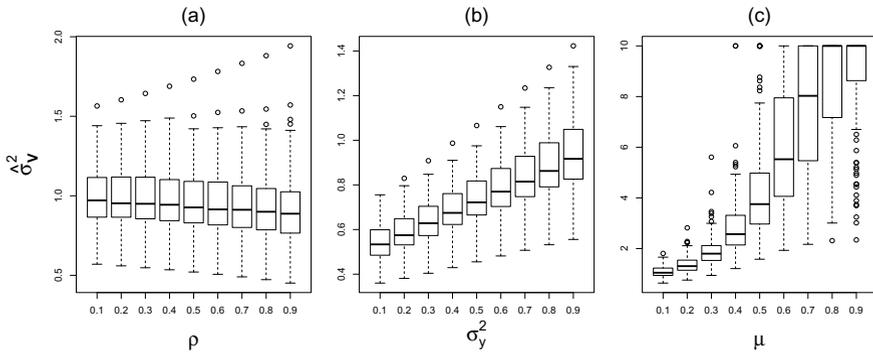


Fig. 5 Estimates of σ_y^2 in (11) when fitted to D^{LGR} generated using (10) with (a) $\mathbf{V} \sim \mathcal{N}_2 \left\{ \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\}$ at different values of ρ , (b) $\mathbf{V} \sim \mathcal{N}_2 \left\{ \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \right\}$ at different values of σ_y^2 , and (c) $\mathbf{V} \sim \mathcal{N}_2 \left\{ \boldsymbol{\mu} = (\mu, \mu)^\top, \boldsymbol{\Sigma} = \mathbf{I}_2 \right\}$ at different values of μ

LOE and MOM at space-time location (\mathbf{s}_l, t) are computed as follows:

$$\text{LOE}(\mathbf{s}_l, t) = \frac{E_{tr,m}(\mathbf{s}_l, t)}{E_{tr}(\mathbf{s}_l, t)} - 1 \quad \text{and} \quad \text{MOM}(\mathbf{s}_l, t) = \frac{E_m(\mathbf{s}_l, t)}{E_{tr,m}(\mathbf{s}_l, t)} - 1, \quad (14)$$

where $E_{tr}(\mathbf{s}_l, t)$ and $E_m(\mathbf{s}_l, t)$ are the mean square errors of the predictors under the true, tr , and misspecified, m , models, respectively, and are calculated as follows:

$$E_j(\mathbf{s}_l, t) = C(\mathbf{s}_l, \mathbf{s}_l; t, t) - \mathbf{c}_l^{j\top} \boldsymbol{\Sigma}(\boldsymbol{\Theta}^*)^{-1} \mathbf{c}_l^j, \quad j = \{tr, m\}, \quad (15)$$

where \mathbf{c}_l^j and $\boldsymbol{\Sigma}(\boldsymbol{\Theta}^*)$ are computed using $\boldsymbol{\Theta}^* = \boldsymbol{\Theta}$, for model tr , and $\boldsymbol{\Theta}^* = \hat{\boldsymbol{\Theta}}^m$ for model m . Here $\boldsymbol{\Theta}$ is the true parameter vector while $\hat{\boldsymbol{\Theta}}^m$ is the estimated parameter vector under the model m . On the other hand, $E_{tr,m}(\mathbf{s}_l, t)$ is the mean square error, with respect to the true model, of the predictor that is derived from the misspecified model, and is given as

$$E_{tr,m}(\mathbf{s}_l, t) = C(\mathbf{s}_l, \mathbf{s}_l; t, t) - 2\mathbf{c}_l^{tr\top} \boldsymbol{\Sigma}(\hat{\boldsymbol{\Theta}}^m)^{-1} \mathbf{c}_l^m + \mathbf{c}_l^{m\top} \boldsymbol{\Sigma}(\hat{\boldsymbol{\Theta}}^m)^{-1} \boldsymbol{\Sigma}(\boldsymbol{\Theta}) \boldsymbol{\Sigma}(\hat{\boldsymbol{\Theta}}^m)^{-1} \mathbf{c}_l^m. \quad (16)$$

Figure 6 plots the LOE and MOM values at every prediction location at $t = 10$. The LOE is closer to zero when ρ is near zero compared to the LOE when $\rho = 0.9$. An LOE near zero indicates that the misspecified model is similar to the true model. Furthermore, the change in the LOE at each prediction location as we increase ρ is different and is somehow dictated by the contours of the distribution of \mathbf{V} . This means that the quality of predictions is not equal everywhere and the worst misspecification occurs in the direction where the highest correlation under C^{LGR} occurs. The plots for the MOM convey the same story.

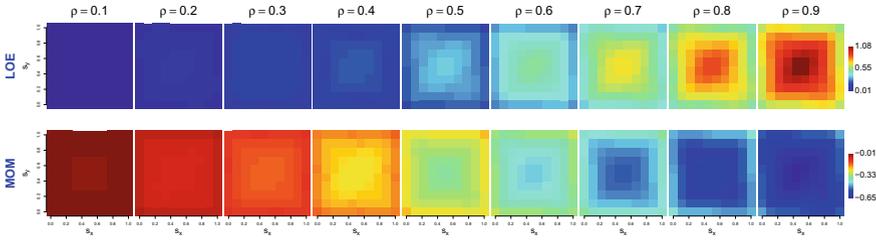


Fig. 6 Values of the LOE and MOM at every spatial location when C^G is fitted to D^{LGR} simulated with $\rho = 0.1, \dots, 0.9$. The closer the LOE values are to zero or the bluer the plots are, the better. Similarly, the closer the values of the MOM are to zero or the redder the plots are, the better

4.2 Second-Order Nonstationarity

Similar analyses cannot be easily adapted to the nonstationary counterparts of the models in the previous section since the covariances may depend on arbitrary nonstationarity parameters at each spatio-temporal location. However, we can draw insights on the consequences of fitting C^G to data generated from C^{LGR} and vice versa, under second-order nonstationarity, by again looking at the quality of predictions.

The non-Lagrangian nonstationary covariance model used in the succeeding numerical experiments, C_{NS}^G , is the nonstationary version of (11) proposed in Garg et al. (2012). It has the form

$$C(\mathbf{s}_1, \mathbf{s}_2; u) = \frac{\sigma(\mathbf{s}_1, \mathbf{s}_2)}{(1 + a_t u^2)^{d/2}} \mathcal{M}_v \left[\frac{\{(\mathbf{s}_1 - \mathbf{s}_2)^\top \mathbf{D}(\mathbf{s}_1, \mathbf{s}_2)^{-1} (\mathbf{s}_1 - \mathbf{s}_2)\}^{1/2}}{(1 + a_t u^2)^{1/2}} \right], \quad (17)$$

where $\sigma(\mathbf{s}_1, \mathbf{s}_2)$ and $\mathbf{D}(\mathbf{s}_1, \mathbf{s}_2)$ are defined in Sect. 2 and the parameter $a_t > 0$ is the scale parameter in time. Data generated from (17) are labeled D_{NS}^G . On the other hand, C_{NS}^{LGR} is the non-frozen Lagrangian spatio-temporal nonstationary model in (5) and data from this model are tagged as D_{NS}^{LGR} . We assess the quality of the predictions by comparing the mean LOEs (MLOE) and mean MOMs (MMOM) when C_{NS}^G is fitted to D_{NS}^{LGR} and when C_{NS}^{LGR} is fitted to D_{NS}^G (Hong et al. 2021). Figure 7 plots the medians of the computed MLOE and MMOM for both scenarios after 100 rounds of parameter estimation via maximization of the full log-likelihood at different values of ρ . It can be seen that at every ρ , the median MLOE is greater when C_{NS}^G is fitted to D_{NS}^{LGR} compared to the median MLOE when C_{NS}^{LGR} is fitted to D_{NS}^G . Moreover, both scenarios of model misspecification yield median MMOMs that are far from zero. However, the median MMOMs are more favorable in cases when C_{NS}^{LGR} is fitted to D_{NS}^G at larger values of ρ . This should advocate the use of Lagrangian models even when the random field does not appear to be transported.

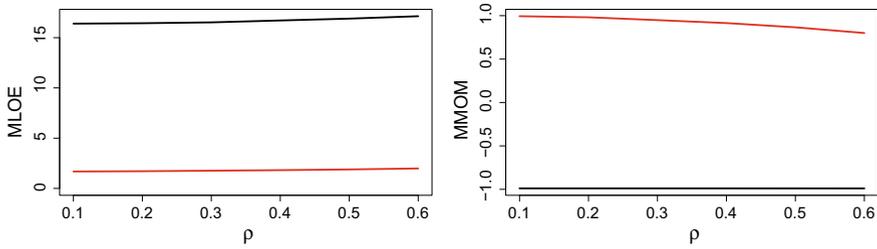


Fig. 7 Medians of the MLOE and MMOM when C_{NS}^G is fitted to D_{NS}^{LGR} (black) and when C_{NS}^{LGR} is fitted to D_{NS}^G (red) at different values of ρ

5 Application to Particulate Matter Data

A spatio-temporal process that is known to be heavily influenced by the presence of a transport medium is pollutant measurements. Pollutants in the atmosphere are transported by the wind to neighboring sites over time (National Research Council 2010). This behavior causes the pollutant measurements at one site to be strongly correlated to the pollutant measurements at a site along the path of transport. Thus, a model incorporating this transport behavior to its spatio-temporal dependence structure is physically reasonable.

5.1 PM 2.5 Data

We study the spatio-temporal dependence of log particulate matter (log PM 2.5) residuals. We retrieve the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) reanalyses dataset of hourly PM 2.5 measurements from NASA Earthdata. Preliminary processing of the raw PM 2.5 data was done to ensure that the resulting spatio-temporal residuals fulfill the modeling assumptions of Gaussianity and zero mean. We consider the first 744 hourly measurements for each year from 1980–2019, at 550 spatial locations, as spatio-temporally dependent, while measurements across years are regarded as spatio-temporally independent. Since the measurements between any two years are at least 11 months apart, this independence assumption is reasonable. Figure 8 maps the log PM 2.5 residuals at 550 locations in Saudi Arabia, at 4 h intervals, starting from 0:00 of January 1, 2017. The transport behavior is evident and can be identified when following the red, blue, and yellow blobs. The direction of transport at every spatial and temporal location appears to be different as the displacements of the red blob indicate transport to the South or South East direction while a North or North West movement can be detected from the yellow blob.

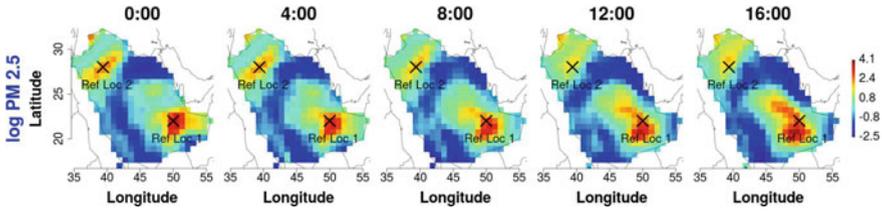


Fig. 8 Snapshots of the log PM 2.5 residuals on January 1, 2017. The spatial images are 4 h apart. Two reference locations are marked for ease of transport movement detection

5.2 Models

We fit six different spatio-temporal covariance functions with Matérn spatial margins. The models under consideration are the following:

- M1: Non-frozen Lagrangian spatio-temporal stationary covariance:

$$C(\mathbf{h}; u) = \sigma^2 E_{\mathbf{V}} \{ \mathcal{M}_v (a \| \mathbf{s}_1 - \mathbf{s}_2 - \mathbf{V}u \|) \};$$

- M2: Non-frozen Lagrangian spatio-temporal spatially varying parameters model in (5);
- M3: Non-frozen Lagrangian spatio-temporal deformation model in (6);
- M4: Non-Lagrangian spatio-temporal stationary covariance:

$$C(\mathbf{h}; u) = \frac{\sigma^2}{(a_t |u|^{2\alpha} + 1)^{\beta d/2}} \mathcal{M}_v \left\{ \frac{a \| \mathbf{h} \|}{(a_t |u|^{2\alpha} + 1)^{\beta/2}} \right\},$$

where $\alpha \in (0, 1]$ is the smoothness parameter in time and $\beta \in [0, 1]$ is the space-time interaction parameter;

- M5: Non-Lagrangian spatio-temporal nonstationary model:

$$C(\mathbf{s}_1, \mathbf{s}_2; u) = \frac{\sigma(\mathbf{s}_1, \mathbf{s}_2)}{(a_t |u|^{2\alpha} + 1)^{\beta d/2}} \mathcal{M}_v \left[\frac{\{(\mathbf{s}_1 - \mathbf{s}_2)^\top \mathbf{D}(\mathbf{s}_1, \mathbf{s}_2)^{-1} (\mathbf{s}_1 - \mathbf{s}_2)\}^{1/2}}{(a_t |u|^{2\alpha} + 1)^{\beta/2}} \right],$$

a more flexible version of the model in (17); and

- M6: Non-Lagrangian spatio-temporal nonstationary covariance II:

$$C(\mathbf{s}_1, \mathbf{s}_2; t_1, t_2) = \frac{\sigma(\mathbf{s}_1, \mathbf{s}_2)}{\{|(t_1 - t_2)D(t_1, t_2)|^{2\alpha} + 1\}^\beta} \mathcal{M}_v \left[\frac{\{(\mathbf{s}_1 - \mathbf{s}_2)^\top \mathbf{D}(\mathbf{s}_1, \mathbf{s}_2)^{-1} (\mathbf{s}_1 - \mathbf{s}_2)\}^{1/2}}{\{|(t_1 - t_2)D(t_1, t_2)|^{2\alpha} + 1\}^{\beta/2}} \right],$$

where $D(t_1, t_2) = \frac{1}{2} \{D(t_1) + D(t_2)\}$ and $D(t)$ controls the temporally varying parameters. This is a more general nonstationary version of model M5; see Garg et al. (2012).

Table 1 A summary of the models fitted to the log PM 2.5 residuals and their corresponding AIC*, BIC*, and MSE. The lower the values, the better. The best scores are in bold. The number of parameters, q , are also reported

Model	q	AIC*	BIC*	MSE
M1 (S)	8	-13, 823, 238	-13, 823, 121	0.0050
M2 (NS)	37	-15, 051, 228	-15, 050, 688	0.0018
M3 (NS)	28	-14, 859, 980	-14, 859, 571	0.0023
M4 (S)	6	-13, 408, 544	-13, 408, 456	0.0171
M5 (NS)	35	-13, 808, 486	-13, 807, 975	0.0081
M6 (NS)	44	-14, 315, 594	-14, 314, 951	0.0035

The expectations in models M1, M2, and M3 are evaluated numerically with respect to $\mathbf{V} \sim \mathcal{N}_2(\boldsymbol{\mu}_V, \boldsymbol{\Sigma}_V)$. Furthermore, the covariance matrix $\boldsymbol{\Sigma}_V$ is parametrized using its Cholesky decomposition to guarantee positive definiteness.

Each pixel in Fig. 8 is $0.5^\circ \times 0.625^\circ$. The spatial coordinates are transformed to their appropriate projections in kilometers (km). This means that the unit of the advection velocity is in km/hr. The minimum distance between any two stations is 16.9 km. Following the techniques presented in Sect. 3, we order the measurements based on their locations in time and group them into blocks of 6 consecutive purely spatial random fields and maximize the approximated log-likelihood in (9). Moreover, we perform a two-step estimation where we retrieve first the estimates of the space and time parameters of the stationary versions and plug in those estimates to the nonstationary models in the next round of maximizing the approximated log-likelihood with respect to the nonstationarity parameters.

To validate the models, we leave out the spatio-temporal observations in the last 5 h of January 31 and predict the measurements at all spatial locations. Table 1 reports the performance of the six models as measured by the MSE, Akaike (AIC*), and Bayesian information criteria (BIC*), where $\text{AIC}^* = -2l(\hat{\boldsymbol{\Theta}}_1, \hat{\boldsymbol{\Theta}}_2) + 2q$ and $\text{BIC}^* = -2l(\hat{\boldsymbol{\Theta}}_1, \hat{\boldsymbol{\Theta}}_2) + q \log(Mn)$. Here $l(\hat{\boldsymbol{\Theta}}_1, \hat{\boldsymbol{\Theta}}_2)$ is the value of the approximated log-likelihood function at the second estimation step with parameter estimates $\hat{\boldsymbol{\Theta}}_2$ while fixing the parameters $\hat{\boldsymbol{\Theta}}_1$ obtained at the first estimation step and M is the number of independent replicates of the spatio-temporal random field. The nonstationary models show more favorable AIC* and BIC* values compared to their stationary counterparts. The additional nonstationarity parameters provided the nonstationary models more flexibility to model the space-time data. In terms of prediction, the Lagrangian models report lower MSEs than the non-Lagrangian models. Overall, the non-frozen Lagrangian spatially varying parameters model M2 is the best performing model across all criteria. The estimated mean and covariance matrix of \mathbf{V} under M2 are $\hat{\boldsymbol{\mu}} = (-0.0003, 0.0017)^\top$ km/hr and $\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 1.719 & 2.257 \\ 2.257 & 3.301 \end{pmatrix} \times 10^{-5}$ km²/hr². This indicates that the estimated value of the correlation between the components of \mathbf{V} is $\hat{\rho} = 0.948$.

6 Conclusion

The theme of this work focused on the practicalities of using Lagrangian spatio-temporal covariance functions to model space-time data, especially under second-order nonstationarity assumptions. The work undertaken in this article aims to illustrate the usability and utility of Lagrangian spatio-temporal models.

We demonstrated the use of thin plate splines in modeling second-order nonstationarity parameters. We also advocated the maximization of the approximated log-likelihood function when data are available at regular time intervals. We showed through several numerical studies the effect of fitting Lagrangian models to data generated from non-Lagrangian models, and vice versa. We found that the predictions of non-Lagrangian models on Lagrangian data are of inferior quality compared to the quality of predictions of Lagrangian models on non-Lagrangian data. This should provide support to using Lagrangian models even when the spatio-temporal random field is not transported.

Further work would be to validate the estimated distribution of the advection velocity vector against the real wind data used as inputs to a partial differential equation which generated the PM 2.5 measurements under study. The equivalence between Lagrangian spatio-temporal models and physical models such as the advection-dispersion equations in Physics is not straightforward and is worth exploring.

The models used in this work as underlying purely spatial nonstationary covariance functions were limited to only two classes. There are other classes in the literature whose Lagrangian formulations deserve attention in terms of model interpretation and parameter estimation, such as the dimension expansion and basis functions models. Future work will focus on these other classes.

References

- Bazen, A. M., & Gerez, S. H. (2003). Fingerprint matching by thin-plate spline modelling of elastic deformations. *Pattern Recognition*, *36*(8), 1859–1867.
- Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(6), 567–585.
- Bornn, L., Shaddick, G., & Zidek, J. (2012). Modeling nonstationary processes through dimension expansion. *Journal of the American Statistical Association*, *107*(497), 281–289.
- Chang, Y.-M., Hsu, N.-J., & Huang, H.-C. (2010). Semiparametric estimation and selection for nonstationary spatial covariance functions. *Journal of Computational and Graphical Statistics*, *19*(1), 117–139.
- Chen, Y., Zhao, J., Deng, Q., & Duan, F. (2017). 3D craniofacial registration using thin-plate spline transform and cylindrical surface projection. *PLoS One*, *12*(10), e0185567. Public Library of Science, San Francisco, CA, USA.
- Chen, T.-L., & Geman, S. (2014). Image warping using radial basis functions. *Journal of Applied Statistics*, *41*(2), 242–258.
- Cox, D. R., & Isham, V. (1988). A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, *415*, 317–328.

- Cressie, N., Shi, T., & Kang, E. (2010). Fixed rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics*, 19(3), 724–745.
- Donato, G., & Belongie, S. (2002). Approximate thin plate spline mappings. In *European conference on computer vision* (pp. 21–31). Springer.
- Fouedjio, F., Desassis, N., & Romary, T. (2015). Estimation of space deformation model for non-stationary random functions. *Spatial Statistics*, 13(1), 45–61.
- Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, 89(1), 197–210.
- Garg, S., Singh, A., & Ramos, F. (2012). Learning non-stationary space-time models for environmental monitoring. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.
- Gelfand, A. E., Schmidt, A. M., Banerjee, S., & Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *TEST*, 13(2), 263–312.
- Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97(458), 590–600.
- Gunz, P., & Mitteroecker, P. (2013). Semilandmarks: A method for quantifying curves and surfaces. *Hystrix, the Italian Journal of Mammalogy*, 24(1), 103–109.
- Heaton, M., Katzfuss, M., Berrett, C., & Nychka, D. (2014). Constructing valid spatial processes on the sphere using kernel convolutions. *Environmetrics*, 25(1), 2–15.
- Hegland, M., Roberts, S., & Altas, I. (1997). *Finite element thin plate splines for data mining applications* Citeseer.
- Higdon, D. (2002). Space and space-time modeling using process convolutions. In *Quantitative methods for current environmental issues* (pp. 37–56). Springer.
- Higdon, D. (1998). A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*, 5(2), 173–190.
- Higdon, D., Swall, J., & Kern, J. (1999). Non-stationary spatial modeling. *Bayesian Statistics*, 6(1), 761–768.
- Hong, Y., Abdulah, S., Genton, M. G., & Sun, Y. (2021). Efficiency assessment of approximated spatial predictions for large datasets. *Spatial Statistics*, 43, 100517.
- Kathuria, D., Mohanty, B. P., & Katzfuss, M. (2019). A nonstationary geostatistical framework for soil moisture prediction in the presence of surface heterogeneity. *Medical Imaging 2004: Image Processing*, 55(1), 729–753.
- Kaufman, C., Schervish, M., & Nychka, D. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484), 1545–1555.
- Kim, H.-M., Mallick, B. K., & Holmes, Co. (2005). Analyzing nonstationary spatial data using piecewise Gaussian processes. *Journal of the American Statistical Association*, 100(470), 653–668.
- Kleiber, W., & Nychka, D. (2012). Nonstationary modeling for multivariate spatial processes. *Journal of Multivariate Analysis*, 112, 76–91.
- Kleiber, W., Sain, S., & Wiltberger, M. (2014). Model calibration via deformation. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1), 545–563.
- Kuusela, M., & Stein, M. L. (2018). Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proceedings of the Royal Society A*, 474(2220), 20180400.
- Lewis, J. P., Hwang, H.-J., Neumann, U., & Enciso, R. (2004). Smart point landmark distribution for thin-plate splines. *Medical Imaging 2004: Image Processing*, 5370, 1236–1243.
- Li, Y., & Sun, Y. (2018). Efficient estimation of non-stationary spatial covariance functions with application to high-resolution climate model emulation. *Statistica Sinica*, 29, 1209–1231.
- Ma, C. (2003a). Spatio-temporal stationary covariance models. *Journal of Multivariate Analysis*, 86(1), 97–107.
- Ma, C. (2003b). Nonstationary covariance functions that model space-time interactions. *Statistics and Probability Letters*, 61(4), 411–419.
- National Research Council. (2010). *Global sources of local pollution: An assessment of long-range transport of key air pollutants to and from the United States*, National Academies Press.

- Neto, J. H. V., Schmidt, A. M., & Guttorp, P. (2014). Accounting for spatially varying directional effects in spatial covariance structures. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63(1), 103–122.
- Nychka, D., & Saltzman, N. (1998). Design of air-quality monitoring networks. In *Case Studies in Environmental Statistics* (pp. 51–76). Springer.
- Nychka, D., Wikle, C., & Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, 2(4), 315–331.
- Paciorek, C. J., & Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5), 483–506.
- Porcu, E., Gregori, P., & Mateu, J. (2009). Archimedean spectral densities for nonstationary space-time geostatistics. *Statistica Sinica*, 19(1), 273–286.
- Risser, M. D. (2015). *Spatially-varying covariance functions for nonstationary spatial process modeling*. PhD thesis, The Ohio State University.
- Risser, M. D. (2016). Nonstationary spatial modeling, with emphasis on process convolution and covariate-driven approaches. *arXiv preprint arXiv:1610.02447*.
- Salvaña, M. L., & Genton, M. G. (2020). Nonstationary cross-covariance functions for multivariate spatio-temporal random fields. *Spatial Statistics*, 37, 100411.
- Salvaña, M. L., Lenzi, A., & Genton, M. G. (2020). Spatio-temporal cross-covariance functions under the Lagrangian framework with multiple advections. Unpublished Manuscript.
- Sampson, P. (2015). A partial warp parameterization for the spatial deformation model for nonstationary covariance. *Joint Statistical Meeting*.
- Sampson, P., Damian, D., & Guttorp, P. (2001). Advances in modeling and inference for environmental processes with nonstationary spatial covariance. In *GeoENV III - Geostatistics for environmental applications* (pp. 17–32). Springer.
- Sampson, P. D., & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417), 108–119.
- Schlather, M. (2010). Some covariance models based on normal scale mixtures. *Bernoulli*, 16(3), 780–797.
- Schmidt, A. M., Guttorp, P., & O'Hagan, A. (2011). Considering covariates in the covariance structure of spatial processes. *Environmetrics*, 22(4), 487–500.
- Shand, L., & Li, B. (2017). Modeling nonstationarity in space and time. *Biometrics*, 73(3), 759–768.
- Stein, M. L. (2005a). Nonstationary spatial covariance functions. *Unpublished technical report*.
- Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. New York: Springer.
- Stein, M. L. (2005c). Statistical methods for regular monitoring data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5), 667–687.
- Stephenson, J., Holmes, C., Gallagher, K., & Pintore, A. (2005). A statistical technique for modelling non-stationary spatial processes. In *Geostatistics Banff 2004* (pp. 125–134). Springer.
- Tennakoon, R. B., Bab-Hadiashar, A., Suter, D., & Cao, Z. (2013). Robust data modelling using thin plate splines. *2013 international conference on digital image computing: Techniques and applications (DICTA)* (pp. 1–8).
- Wahba, G. (1990). *Spline models for observational data* (Vol. 59). SIAM.
- Whitbeck, M., & Guo, H. (2006). Multiple landmark warping using thin-plate splines. *IPCV*, 6, 256–263.
- Wikle, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of Spatial Statistics* (pp. 114–125). CRC Press.
- Zhu, Z., & Wu, Y. (2010). Estimation and prediction of a class of convolution-based spatial nonstationary models for large spatial data. *Journal of Computational and Graphical Statistics*, 19(1), 74–95.