Lagrangian Spatio-Temporal Covariance Functions for Multivariate Nonstationary Random Fields

Dissertation by
Mary Lai Salvaña

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

King Abdullah University of Science and Technology
Thuwal, Kingdom of Saudi Arabia

June, 2021
The dissertation of Mary Lai Salvaña is approved by the examination committee.

Committee Chairperson: Marc G. Genton
Committee Members: Hernando C. Ombao, Georgiy L. Stenchikov, Huiyan Sang
ABSTRACT

Lagrangian Spatio-Temporal Covariance Functions for Multivariate Nonstationary Random Fields
Mary Lai Salvaña

In geostatistical analysis, we are faced with the formidable challenge of specifying a valid spatio-temporal covariance function, either directly or through the construction of processes. This task is difficult as these functions should yield positive definite covariance matrices. In recent years, we have seen a flourishing of methods and theories on constructing spatio-temporal covariance functions satisfying the positive definiteness requirement. The current state-of-the-art when modeling environmental processes are those that embed the associated physical laws of the system. The class of Lagrangian spatio-temporal covariance functions fulfills this requirement. Moreover, this class possesses the allure that they turn already established purely spatial covariance functions into spatio-temporal covariance functions by a direct application of the concept of Lagrangian reference frame. In the three main chapters that comprise this dissertation, several developments are proposed and new features are provided to this special class. First, the application of the Lagrangian reference frame on transported purely spatial random fields with second-order nonstationarity is explored, an appropriate estimation methodology is proposed, and the consequences of model misspecification is tackled. Furthermore, the new Lagrangian models and the new estimation technique are used to analyze particulate matter concentrations over Saudi Arabia. Second, a multivariate version of the Lagrangian framework is established, catering to both second-order stationary and nonstationary spatio-temporal random fields. The capabilities of the Lagrangian spatio-temporal cross-covariance functions are demonstrated on a bivariate reanalysis climate model output dataset previously analyzed using purely spatial covariance
functions. Lastly, the class of Lagrangian spatio-temporal cross-covariance functions with multiple transport behaviors is presented, its properties are explored, and its use is demonstrated on a bivariate pollutant dataset of particulate matter in Saudi Arabia. Moreover, the importance of accounting for multiple transport behaviors is discussed and validated via numerical experiments. Together, these three extensions to the Lagrangian framework makes it a more viable geostatistical approach in modeling realistic transport scenarios.
AKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Professor Marc G. Genton, for the mentorship and support. Thank you for giving me a chance to work under your wing. I learned how to be a great scientist by working with you.

I would also like to thank Professor Hernando Ombao, Professor Georgiy Stenchikov, and Professor Huiyan Sang for accepting the invitation to serve in my dissertation committee. Thank you for your time and patience in reading my dissertation and for the valuable suggestions for this study.

I would also like to thank the KAUST Statistics faculty, Professor Ying Sun, Professor Håvard Rue, and Professor Raphäel Huser, who contributed directly or indirectly to the dissertation.

I would also like to thank my collaborators, Dr. Amanda Lenzi, Dr. Sameh Abdullah, Dr. Huang Huang, Dr. Hatem Ltaief, and Professor David Keyes. Thank you for generously sharing your expertise and for being available anytime to discuss our papers.

I would also like to thank the Ateneo Innovation Center, especially to our research director, Professor Gregory Tangonan. Thank you for introducing KAUST to me and for helping me chart a career back to academia.

I would also like to thank my friends, near and far, who made my whole Ph.D. journey 100x more bearable. Thank you Brent, Harold, Jeff, and Zhuo for the friendship, the laughter, and the happy distractions.

Finally, I would like to thank my family, especially my father, Armando, my mother, Leoba, my younger brother, Joseph Ali, and my uncle, Father Rolando, for the love and support. Thank you for always being there for me. You have been my source of wisdom and sympathetic ear. I love you dearly.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examination Committee Page</td>
<td>2</td>
</tr>
<tr>
<td>Copyright</td>
<td>3</td>
</tr>
<tr>
<td>Abstract</td>
<td>4</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>6</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>7</td>
</tr>
<tr>
<td>List of Symbols</td>
<td>8</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>10</td>
</tr>
<tr>
<td>1.1 Research Motivation</td>
<td>10</td>
</tr>
<tr>
<td>1.2 Univariate Geostatistics: Preliminaries</td>
<td>10</td>
</tr>
<tr>
<td>1.2.1 Univariate Spatio-Temporal Processes</td>
<td>11</td>
</tr>
<tr>
<td>1.2.2 Kriging</td>
<td>11</td>
</tr>
<tr>
<td>1.2.3 Properties of Spatio-Temporal Covariance Functions</td>
<td>12</td>
</tr>
<tr>
<td>1.3 Multivariate Geostatistics: Preliminaries</td>
<td>13</td>
</tr>
<tr>
<td>1.3.1 Multivariate Spatio-Temporal Processes</td>
<td>14</td>
</tr>
<tr>
<td>1.3.2 Cokriging</td>
<td>14</td>
</tr>
<tr>
<td>1.3.3 Properties of Spatio-Temporal Cross-Covariance Functions</td>
<td>15</td>
</tr>
<tr>
<td>1.4 Physically-Motivated Spatio-Temporal Models</td>
<td>16</td>
</tr>
<tr>
<td>1.5 The Lagrangian Framework in Geostatistics</td>
<td>18</td>
</tr>
<tr>
<td>1.6 Contributions and Outline</td>
<td>20</td>
</tr>
<tr>
<td><strong>2 Lagrangian Spatio-Temporal Nonstationary Covariance Functions</strong></td>
<td>21</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>21</td>
</tr>
<tr>
<td>2.2 Nonstationary Extension to the Lagrangian Framework</td>
<td>22</td>
</tr>
<tr>
<td>2.3 Non-Frozen Nonstationary Taylor’s Hypothesis</td>
<td>25</td>
</tr>
<tr>
<td>2.4 Estimation</td>
<td>29</td>
</tr>
<tr>
<td>2.4.1 Thin Plate Splines</td>
<td>29</td>
</tr>
</tbody>
</table>
2.4.2 Maximum Likelihood Estimation and Likelihood Approximations in the Temporal Domain ........................................... 32

2.4.3 Two-Step Maximum Likelihood Estimation ........................................... 33

2.5 Simulation Study: Lagrangian vs. Non-Lagrangian Spatio-Temporal Models ........................................................................ 34

2.5.1 Second-Order Stationarity ................................................................ 34

2.5.2 Second-Order Nonstationarity .............................................................. 40

2.6 Application to Particulate Matter Data ...................................................... 41

2.6.1 PM 2.5 Data ...................................................................................... 41

2.6.2 Models ............................................................................................ 42

2.7 Conclusion ............................................................................................ 44

3 Spatio-Temporal Nonstationary Cross-Covariance Functions under the Lagrangian Framework .............................................. 46

3.1 Introduction ......................................................................................... 46

3.2 The State-of-the-Art in Purely Spatial and Spatio-Temporal Cross-Covariance Function Models .................................................. 47

3.2.1 Purely Spatial Stationary Cross-Covariance Functions ....................... 47

3.2.2 Purely Spatial Nonstationary Cross-Covariance Functions ..................... 49

3.2.3 Spatio-Temporal Stationary Cross-Covariance Functions ....................... 51

3.2.4 Spatio-Temporal Nonstationary Cross-Covariance Functions ............... 53

3.3 Multivariate Nonstationary Extension to the Lagrangian Framework ................................................................. 54

3.3.1 Main Result ...................................................................................... 54

3.3.2 Examples ......................................................................................... 55

3.3.3 Estimation ....................................................................................... 58

3.3.4 Illustrations ...................................................................................... 58

3.4 Application to Regional Climate Model Output ........................................ 58

3.4.1 Spatio-Temporal Data Analysis ............................................................ 60

3.4.2 Model Performance ......................................................................... 63

3.5 Purely Spatial and Spatio-Temporal Nonstationary Cross-Covariance Functions on the Sphere ........................................ 66

3.6 Conclusion ............................................................................................ 67

4 Spatio-Temporal Cross-Covariance Functions under the Lagrangian Framework with Multiple Advections .............................................. 73

4.1 Introduction ......................................................................................... 73

4.2 Lagrangian Framework with Multiple Advections .................................... 74

4.3 Estimation ............................................................................................ 83
LIST OF SYMBOLS

cov  covariance
$\rho^S(h)$  univariate stationary correlation function
$C^S(s_1, s_2)$  purely spatial nonstationary covariance function
$C^S_{ij}(s_1, s_2)$  purely spatial nonstationary cross-covariance function
$C^S(h)$  purely spatial stationary covariance function
$C^S_{ij}(h)$  purely spatial stationary cross-covariance function
$C^T(u)$  purely temporal stationary covariance function
$C(s_1, s_2; t_1, t_2)$  spatio-temporal nonstationary covariance function
$C_{ij}(s_1, s_2; t_1, t_2)$  spatio-temporal nonstationary cross-covariance function
$C(h, u)$  spatio-temporal stationary covariance function
$C_{ij}(h, u)$  spatio-temporal stationary cross-covariance function
$\Sigma_V$  covariance matrix of advection
$\mu_V$  mean vector of advection
$\Sigma$  full spatio-temporal covariance matrix
$\in$  element of
$I_d$  $d \times d$ identity matrix
$K_\nu$  modified Bessel function of the second kind of order $\nu$
$M_\nu(; a)$  Matérn correlation function with smoothness parameter $\nu$ and scale parameter $a$
$(\cdot)_+$  $\max(\cdot, 0)$
$\mathcal{N}_d(\mu_V, \Sigma_V)$  $d$-dimensional Gaussian distribution with mean $\mu_V$ and covariance $\Sigma_V$
$|\cdot|$  $L_1$ norm
$\|\cdot\|$  $L_2$ norm
$d$  number of dimensions in space
$M$  number of independent replicates
$q$  number of parameters
$N$  number of spatial locations
$T$  number of temporal locations
$n$  number of spatio-temporal locations ($= NT$)
$p$  number of variables
$\Theta$  vector of parameters
\(\Theta\) vector of estimated parameters
\(O\) zero matrix
\(0\) zero vector
\(\mathbb{Z}^+\) set of all positive integers
\(\mathbb{R}\) set of real numbers
\(\mathbb{R}^d\) set of \(d\)-tuples of real numbers
\(\mathbb{R}^n\) set of \(n\)-tuples of real numbers
\(C_{NS}^{LGR}\) a Lagrangian spatio-temporal nonstationary model
\(C_{NS}^G\) a non-Lagrangian spatio-temporal nonstationary model
\(D_{NS}^{LGR}\) data generated from a Lagrangian spatio-temporal nonstationary model
\(D_{NS}^G\) data generated from a non-Lagrangian spatio-temporal nonstationary model
\(h\) spatial lag
\(s\) spatial location
\(s_i^*\) spatial location where the \(i\)-th basis function is centered
\(\eta\) spatial location in the latent dimension
\(s_0\) spatial location with no observation
\(u\) temporal lag
\(t\) temporal location
\(\xi\) temporal location in the latent dimension
\(t_0\) temporal location with no observation
\(\top\) transpose operator
\(\gamma^S(h)\) purely spatial stationary variogram
\(\mathcal{V}\) multiple advections vector
\(v\) constant advection velocity vector
\(\mathbf{V}\) random advection velocity vector
\(\tilde{Z}(s)\) zero-mean Gaussian purely spatial process
\(Z(s,t)\) zero-mean Gaussian spatio-temporal process
\(\tilde{Z}(s_0,t_0)\) simple kriging predictor of \(Z(s_0,t_0)\)
\(Y(s,t)\) Gaussian spatio-temporal process
\(\hat{Y}(s_0,t_0)\) ordinary kriging predictor of \(Y(s_0,t_0)\)
\(E\{Z(s_0,t_0) | Z\}\) conditional expectation of \(Z(s_0,t_0)\) given \(Z\)
Chapter 1

Introduction

1.1 Research Motivation

Traditionally, geostatistical models were used to produce high-quality mappings of heavy metal contamination, ore deposits, pollutants, sand distribution in shelf seas, temperature, and vegetation, to name a few. They were, and still are, used to model variables that are themselves required inputs in running physical models. Presently, the field is known to provide a suite of covariance function models that are capable of representing various physical systems. Theoretical advancements in the field have provided frameworks and tools for several important applications. The motivation of this research is to develop a class of spatio-temporal covariance functions for observations that are transported in space, over time. This transport behavior enables a purely spatial random field to change its location in space while retaining its spatial statistical (second-order) properties in time. By following a purely spatial random field as it gets transported in time, one operates under the Lagrangian framework and may derive its spatio-temporal statistical (second-order) properties.

1.2 Univariate Geostatistics: Preliminaries

Geostatistical modeling has emerged as an important complement to physics-based modeling of spatio-temporal phenomena because it is widely applicable and straightforward to use with sufficient data. It thrives in seeking correlations arising from causality and are adequate for many spatio-temporal phenomena. Its approach to the modeling problem is to consider the spatio-temporal process as a realization of a spatio-temporal random field. In this section, we
review the concept of spatio-temporal random fields and establish the role of spatio-temporal covariance functions.

1.2.1 Univariate Spatio-Temporal Processes

Consider a real-valued spatio-temporal process $Y(s, t), \ (s, t) \in \mathbb{R}^d \times \mathbb{R}$, where $(s, t)$ is the spatio-temporal location. For instance, $Y(s, t)$ may be the temperature, precipitation, or pollutant concentration at spatial location $s \in \mathbb{R}^d$, $d \geq 1$, and temporal location $t \in \mathbb{R}$. A common assumption to $Y(s, t)$ is that it is a Gaussian process comprised of a deterministic and a random component, i.e.,

$$Y(s, t) = \mu(s, t) + Z(s, t), \quad (1.1)$$

where $\mu(s, t)$ is a trend function and $Z(s, t)$ is a zero-mean Gaussian spatio-temporal process. Because of its Gaussianity, $Z(s, t)$ is completely characterized by its spatio-temporal covariance function, $C(s_1, s_2; t_1, t_2) = \text{cov} \{Z(s_1, t_1), Z(s_2, t_2)\}$. Owing to this convenient property of Gaussian processes, geostatisticians are constantly in pursuit of construction approaches that produce valid spatio-temporal covariance functions. A valid spatio-temporal covariance function ensures that the resulting spatio-temporal covariance matrix of the $n$-dimensional vector $Z = \{Z(s_1, t_1), \ldots, Z(s_n, t_n)\}^\top$ is positive definite, i.e., for any $n \in \mathbb{Z}^+$, for any finite set of points, $(s_1, t_1), \ldots, (s_n, t_n)$, and for any vector $\lambda \in \mathbb{R}^n$, we have $\lambda^\top \Sigma \lambda > 0$, where $\Sigma$ is an $n \times n$ matrix and $n$ is the number of spatio-temporal locations.

1.2.2 Kriging

The utility of spatio-temporal covariance functions extends to the prediction of missing values at spatio-temporal locations with no observations using the kriging technique. Mathematically, let $Z(s_0, t_0)$ be the unknown value at an unobserved spatio-temporal location $(s_0, t_0)$. Under the squared-error loss criterion, the simple kriging predictor of $Z(s_0, t_0)$ is the best
14

linear unbiased predictor \( \hat{Z}(s_0, t_0) = E \{ Z(s_0, t_0) | Z \} \) with closed form:

\[
\hat{Z}(s_0, t_0) = \Delta_0^\top \Sigma^{-1} Z,
\]

where \( \Delta_0 = \{ C(s_0, s_1; t_0, t_1), C(s_0, s_2; t_0, t_2), \ldots, C(s_0, s_n; t_0, t_n) \}^\top \).

1.2.3 Properties of Spatio-Temporal Covariance Functions

It is imperative to check that the chosen spatio-temporal covariance function matches the properties of the data being modeled. The literature is loaded with different classes of spatio-temporal covariance functions which can accommodate different dependence structures in space and time implied by any spatio-temporal process. Most of the early models operate under these three most common simplifying assumptions:

a) **(weak) stationarity and isotropy**: A spatio-temporal covariance function is stationary if it is a function of only the spatial and temporal separation vectors, i.e., \( C(s_1, s_2; t_1, t_2) \) simplifies to \( C(h, u) \), where \( h = s_1 - s_2 \) and \( u = t_1 - t_2 \). This means that the random field has homogenous spatial and temporal structures over its entire domain in space and time. If the covariance depends only on the magnitude of their separation, then \( C(s_1, s_2; t_1, t_2) \) further simplifies to \( C(||h||, |u|) \), where \( ||h|| = ||s_1 - s_2|| \) and \( |u| = |t_1 - t_2| \). A spatio-temporal random field with such spatio-temporal covariance structure is termed isotropic.

b) **space-time separability**: A spatio-temporal covariance function is space-time separable if it factors into its purely spatial and purely temporal components, i.e., \( C(h, u) = C^S(h)C^T(u) \), under stationarity, and \( C(s_1, s_2; t_1, t_2) = C^S(s_1, s_2)C^T(t_1, t_2) \), under non-stationarity, which implies independence of the space and time aspects. Here \( C^S(\cdot) \) and \( C^T(\cdot) \) are purely spatial and purely temporal covariance functions, respectively. Spatio-temporal covariance functions of this type model the dependence structure in space separately from the dependence structure in time. The resulting covariance ma-
trix is simply a Kronecker product of a purely spatial and purely temporal covariance matrix.

c) **full symmetry**: A spatio-temporal covariance function is fully symmetric if we have:

\[
C(h, u) = C(-h, u) = C(-h, -u) = C(h, -u),
\]

under stationarity, and

\[
C(s_1, s_2; t_1, t_2) = C(s_2, s_1; t_1, t_2) = C(s_2, s_1; t_2, t_1) = C(s_1, s_2; t_2, t_1),
\]

under nonstationarity. This means that the covariance does not change regardless of the direction in space or time.

The aforementioned properties are often invoked to simplify modeling and speed up computations. Several testing procedures were proposed to check whether the real data manifest these properties. When the results of the tests cannot support the usage of such simplified properties, other classes possessing more sophisticated properties should be utilized. Space-time symmetry, for instance, is often violated by environmental data influenced by a dominant flow. In that situation, asymmetric spatio-temporal models are preferred. Further discussions can be found in the reviews by Gneiting et al. (2007) and Chen et al. (2021). De Iaco et al. (2016) also provides a comprehensive roster of other useful properties.

### 1.3 Multivariate Geostatistics: Preliminaries

Usually there are multiple variables measured at each location, such as temperature, humidity, wind speed, and atmospheric pressure. These collocated variables may or may not depend on each other and on the variables at other locations. This multivariate modeling problem sets the stage for the use of multivariate Gaussian spatio-temporal random fields, the cornerstone of multivariate spatio-temporal geostatistics. Exposing relationships among
spatio-temporally referenced data is accomplished using multivariate spatio-temporal covariance functions, which are more commonly termed as spatio-temporal cross-covariance functions.

1.3.1 Multivariate Spatio-Temporal Processes

Consider a multivariate spatio-temporal process \( Y(s,t) = \{Y_1(s,t), \ldots, Y_p(s,t)\}^T \), \((s,t) \in \mathbb{R}^d \times \mathbb{R},\) such that at each spatial location \(s \in \mathbb{R}^d, d \geq 1,\) and at each time \(t \in \mathbb{R},\) there are \(p\) variables. Assume that \(Y(s,t)\) can be decomposed into a sum of a deterministic and a random component, i.e.,

\[
Y(s,t) = \mu(s,t) + Z(s,t),
\]

where \(\mu(s,t)\) is a trend function and \(Z(s,t)\) a zero-mean multivariate spatio-temporal Gaussian process. Under the Gaussian framework, \(Z(s,t)\) is completely specified by its spatio-temporal matrix-valued cross-covariance function \(C(s_1,s_2; t_1, t_2) = \{C_{ij}(s_1,s_2; t_1, t_2)\}_{i,j=1}^p,\)

where

\[
C_{ij}(s_1,s_2; t_1, t_2) = \text{cov}\{Z_i(s_1,t_1), Z_j(s_2,t_2)\},
\]

for \(i,j = 1, \ldots, p.\) Only functions that verify the well-known requirement of positive definiteness can be considered valid models for \(C_{ij}\). That is, for any \(n \in \mathbb{Z}^+,\) for any finite set of points \((s_1,t_1), \ldots, (s_n,t_n),\) and for any vector \(\lambda \in \mathbb{R}^{np},\) we have \(\lambda^T \Sigma \lambda > 0,\) where \(\Sigma\) is an \(np \times np\) matrix with \(n \times n\) block elements of \(p \times p\) matrices \(C(s_l,s_r; t_l,t_r), l,r = 1, \ldots, n.\)

1.3.2 Cokriging

A prediction location may have all or some variables that are missing. The first case may happen when there are locations with sensors that collect measurements for atmospheric variables like temperature, precipitation, and wind speed, for example, and one might be interested in predicting the values of these variables at locations with no sensors. The second case occurs when measurements of one variable are difficult or expensive to obtain while
measurements of another variable, correlated with the first one, are easy to collect. In this scenario, there will be more locations with data collection instruments for the cheaper variable, while observations will be sparse for the expensive one. The first case is more prevalent in environmental applications wherein sensors measuring different variables simultaneously were deployed at predetermined sites. Hence, throughout this work, we assume that all prediction locations are missing the measurements for all \( p \) variables.

The prediction of multiple variables using an optimal predictor can be done via cokriging which proceeds as follows. Suppose \((s_0, t_0)\) is a prediction location with an unknown vector of \( p \) variables, \( Z(s_0, t_0) \). Under the squared-error loss criterion, the best linear unbiased predictor of \( Z(s_0, t_0) \), given \( Z = \{Z(s_1, t_1)^\top, \ldots, Z(s_n, t_n)^\top\}^\top \in \mathbb{R}^{np} \), also called the cokriging predictor, is

\[
\hat{Z}(s_0, t_0) = c_0^\top \Sigma^{-1} Z, \tag{1.5}
\]

where \( c_0 \) is the \( pn \times p \) matrix formed by taking the cross-covariance between \( Z(s_0, t_0) \) and \( Z(s_r, t_r) \), \( r = 1, \ldots, n \), i.e.,

\[
c_0 = \{C(s_0, s_1; t_0, t_1), \ldots, C(s_0, s_n; t_0, t_n)\}^\top. \tag{1.6}
\]

### 1.3.3 Properties of Spatio-Temporal Cross-Covariance Functions

Similar to its univariate counterpart, a cross-covariance function can be classified based on the characteristics of its implied dependence in space and time. The three most common classes are:

a) **(weak) stationarity** and isotropy: A spatio-temporal cross-covariance function \( C_{ij}(s_1, s_2; t_1, t_2) \) is (weakly) stationary if it simplifies to \( C_{ij}(h, u) \) and is isotropic if the covariance depends only on the magnitude of their separation, i.e., \( C_{ij}(s_1, s_2; t_1, t_2) \) further simplifies to \( C_{ij}(\|h\|, |u|) \), for \( i, j = 1, \ldots, p \).

b) **space-time separability**: A spatio-temporal covariance function is space-time sepa-
rable if it factors into its purely spatial and purely temporal components, i.e., \( C_{ij}(h, u) = C_{ij}^S(h)C_{ij}^T(u) \), under stationarity, and \( C_{ij}(s_1, s_2; t_1, t_2) = C_{ij}^S(s_1, s_2)C_{ij}^T(t_1, t_2) \), under nonstationarity, for \( i, j = 1, \ldots, p \). Here \( C_{ij}^S(\cdot) \) and \( C_{ij}^T(\cdot) \) are purely spatial and purely temporal cross-covariance functions, respectively.

c) **full symmetry**: A spatio-temporal covariance function is fully symmetric if we have

\[
C_{ij}(h, u) = C_{ij}(-h, u) = C_{ij}(-h, -u) = C_{ij}(h, -u),
\]

under stationarity, and

\[
C_{ij}(s_1, s_2; t_1, t_2) = C_{ij}(s_2, s_1; t_1, t_2) = C_{ij}(s_2, s_1; t_2, t_1) = C_{ij}(s_1, s_2; t_2, t_1),
\]

under nonstationarity, for \( i, j = 1, \ldots, p \).

Further discussions can be found in the review by Genton and Kleiber (2015).

## 1.4 Physically-Motivated Spatio-Temporal Models

There is a demand for physically-meaningful spatio-temporal covariance function models that can surrogate physics-based models. This is because most environmental data obey some fundamental laws of nature. Thus, the underlying physical dynamics need to be acknowledged and included in the modeling. Unlike the dynamical spatio-temporal models such as partial differential equations (PDEs) (Wikle et al., 2001; Raissi et al., 2018) and integro-difference equations (IDEs) (Richardson et al., 2020), the covariance-based approach of geostatistics does not necessarily reflect fundamental scientific relationships (Wikle and Hooten, 2010). However, there are a growing number of exceptions, for instance, the Bayesian Hierarchical Models (BHM) which requires separate models for the data, process, and parameters. This approach involves both physical model output and observations (Liu et al., 2016). Covariate-dependent spatio-temporal covariance functions also attempt to incorpo-
rate scientific knowledge by including other explanatory variables in the model \cite{reich2011}. Covariance models preserving physical constraints, such as the divergence-free and curl-free requirements on vector fields processes, e.g. wind, ocean, electric, and magnetic fields, were also developed \cite{scheuerer2012, hewer2017}.

The transport phenomenon governed by balance equations (conservation laws) and rate equations (flux laws) was also formulated in the language of geostatistics. Gupta and Waymire \cite{gupta1987} defined a process

\[
Z(s, t) = \tilde{Z}(s - vt)
\]

with spatio-temporal stationary covariance function

\[
C(h, u) = C^S(h - vu)
\]

and called it the frozen field. Here, \(\tilde{Z}(s)\) is a purely spatial stationary random field and \(C^S\) its purely spatial stationary covariance function. Cox and Isham \cite{cox1988} offered more flexibility to the frozen field model by replacing the constant velocity, \(v \in \mathbb{R}^d\), with a random velocity, \(V \in \mathbb{R}^d\), resulting in a spatio-temporal covariance function model of the form

\[
C(h, u) = \mathbb{E}_V \left\{ C^S(h - Vu) \right\}.
\]

We call this the non-frozen random field model. The constant \(v\) in the frozen field model is usually treated as the mean of the random variable \(V\). The vectors \(V\) and \(v\) are commonly referred to as the random and constant transport or advection velocity vectors, respectively. Ma \cite{ma2003} established an umbrella theorem that formalizes the validity of the frozen and non-frozen models.

The covariance function in \eqref{1.9} is space-time asymmetric since the covariance between observations at site \(s_1\), at time \(t_1\), and those at site \(s_2\), at time \(t_2\), may not be equal to the
covariance between observations at site $s_1$, at time $t_2$, and at site $s_2$, at time $t_1$. This means that if in a certain geographic region the wind blows from West to East, the measurements at a station on the West coast, for example, would be highly correlated to measurements taken at stations on the East coast at a certain future time. This cannot be said about the correlation of measurements taken from a station on the East coast with future measurements taken on the West coast. This violates the assumption of full-symmetry. Such behavior is expected from variables occurring in the lowest part of the atmosphere which move due to turbulent flows caused by weather systems.

1.5 The Lagrangian Framework in Geostatistics

The stochastic representation (1.7) makes physical sense in modeling observations influenced by transport phenomena under the Lagrangian reference frame. This reference frame has its roots in physics and is a way of describing the development of a phenomenon in space and in time while moving or traveling with it. Covariance functions centered around modeling a process (1.7) are collectively termed “spatio-temporal covariance functions under the Lagrangian framework” (Gneiting, 2002a; Gneiting et al., 2007). These covariance functions use the Lagrangian reference frame to build spatio-temporal covariance functions from purely spatial covariance functions.

The frozen and non-frozen field models are stationary in their inception. From time to time, the frozen field model is used to model waves (Ailliot et al., 2011), wind (Gneiting et al., 2007; Ezzat et al., 2018), solar irradiance (Lonij et al., 2013; Inoue et al., 2012), and cloud cover data (Shinozaki et al., 2016). Gneiting et al. (2007) proposed to model an Irish wind dataset using a convex combination of a classical and a Lagrangian spatio-temporal covariance function. Because of prior knowledge of a prevailing westerly wind pattern, the Lagrangian spatio-temporal covariance function assumed the form $C(h,u) = (1 - \frac{1}{2v_1} |h_1 - uv_1|)_+$, where $h = (h_1, h_2)^T$, $v = (v_1, 0)^T$, and $(\cdot)_+ = \max(\cdot, 0)$. Christakos et al. (2017) used a different term for this random field and called it a traveling random field
that was then used to model the spread of diseases.

Although the Lagrangian framework easily extends purely spatial covariance functions to space-time, a model under this framework, the frozen field model, has the disadvantage that the spatio-temporal covariance functions it produces are not anisotropic, for any \( u \neq 0 \). One should not confuse anisotropy and asymmetry. Anisotropy is a property involving only the spatial arguments of the covariance function, whereas asymmetry involves both the spatial and temporal arguments. Porcu et al. (2006) proposed an anisotropic version of this model by partitioning the spatial lag and the advection velocity vector into smaller components:

\[
C(h, u) = E_{V_1, V_2}[\mathcal{L}\{\gamma_1(h_1 - V_1 u), \gamma_2(h_2 - V_2 u)\}], \quad h_1, V_1 \in \mathbb{R}^{d_1}, \quad h_2, V_2 \in \mathbb{R}^{d_2},
\]

where \( \gamma_1, \gamma_2 \) are purely spatial stationary variograms with \( \gamma_1(0) = \gamma_2(0) = 0 \) and \( d_1 + d_2 = d \). This is a valid asymmetric spatio-temporal component-wise anisotropic stationary covariance function in \( \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R} \). Here \( \mathcal{L} \) denotes a bivariate Laplace transform with representation

\[
\mathcal{L}(\theta_1, \theta_2) = \int_{[0, \infty)^2} \exp \left( -r_1 \theta_1 - r_2 \theta_2 \right) dF(r_1, r_2),
\]

where \( F \) is a bivariate probability measure and \( \mathcal{L}(0, 0) = 1 \).

Another major drawback of the frozen field model is that the model itself does not permit the dampening of the spatio-temporal covariance, meaning that the maximum covariances at different temporal lags are always equal. However, real data do not exhibit this property. Objects that are transported can be subjected to diffusion, i.e., they get transported to different directions at any time point (Hwang et al., 2018). This drawback was addressed by the non-frozen model (1.9) which allows a decreasing covariance as the temporal lag increased. The non-frozen model does not have an explicit form except for some special distributions of the random advection velocity vector \( V \) and purely spatial covariance \( C^S \). Schlather (2010) derived the explicit form when \( V \sim \mathcal{N}_d(\mu_V, \Sigma_V) \) and \( C^S \) is a covariance function of a normal-scale mixture, i.e.,

\[
C(h, u) = \frac{1}{\sqrt{I_d + \Sigma_V u^2}} C^S((h - \mu_V u)\top (I_d + \Sigma_V u^2)^{-1} (h - \mu_V u)). \quad (1.10)
\]
It is apparent from the model above that $C(h, u)$ decreases as $u$ increases and that the model can introduce anisotropy via $\Sigma_V$ at nonzero $u$.

The spatio-temporal covariance function in (1.9) is still ripe for further developments. A survey of existing literature suggests that there is no detailed Lagrangian formulation in the nonstationary, multivariate, and multiple advections arena. Hence, we took significant strides towards developing and unifying the modeling of transported data using specialized covariance functions under the Lagrangian framework in the next few chapters.

1.6 Contributions and Outline

The Lagrangian framework transforms what had been primarily a purely spatial covariance function into a spatio-temporal covariance function. Under this framework, the transport property is exploited, and one readily obtains substantial performance benefits of a spatio-temporal model using a primarily purely spatial one. The working premise is that the purely spatial random field retains its spatial properties while being transported and the models depend for effectiveness on the advection velocity.

The use of the Lagrangian framework to analyze spatio-temporal phenomena seems to be most widely used in engineering, and only few developments in theory and applications in geostatistics have been proposed. In Chapter 2 the modeling framework is developed to include second-order nonstationary behavior in space and time. In Chapter 3 the Lagrangian paradigm, originally formulated in the univariate setting, is extended to the multivariate nonstationary realm, along with a review of the recent advances in the multivariate spatio-temporal covariance functions modeling. Chapter 4 presents a new modification to the multivariate stationary formulation which involves different transport velocities affecting different variables. Chapter 5 summarizes the tangible contributions of this dissertation and closes with a discussion on some avenues for Lagrangian spatio-temporal research.
Chapter 2

Lagrangian Spatio-Temporal Nonstationary Covariance Functions

2.1 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 functions 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, 2003c), and the Gneiting class of spatio-temporal stationary covariance functions (Gneiting, 2002a). 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). 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 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. In this work, we 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 chapter is organized as follows: Section 2.2 formulates the nonstationary extension and introduces new families of Lagrangian spatio-temporal nonstationary covariance functions. Section 2.3 introduces a nonstationary version of a well-known property involving the purely spatial and purely temporal margins of Lagrangian spatio-temporal covariance functions. Section 2.4 discusses a viable estimation procedure. Section 2.5 presents a simulation study which illustrates the advantages of Lagrangian over non-Lagrangian spatio-temporal nonstationary covariance functions. Section 2.6 details the application of the new models to a spatio-temporal particulate matter dataset. Section 2.7 draws a conclusion.

2.2 Nonstationary Extension to the Lagrangian Framework

The model in (1.9) can be extended to allow for $C^S$ to be a purely spatial nonstationary covariance function.

Theorem 1. Let $V$ be a random vector on $\mathbb{R}^d$. If $C^S(s_1, s_2)$ is a valid purely spatial non-
stationary covariance function on $\mathbb{R}^d$, then,

$$C(s_1, s_2; t_1, t_2) = \mathbb{E}_V \left\{ C^S(s_1 - Vt_1, s_2 - Vt_2) \right\}, \quad (2.1)$$

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

The construction approach in Theorem 1 requires a purely spatial nonstationary covariance function, $C^S(s_1, s_2)$, 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 $V$. The transport behavior, dictated by $V$, influences the covariance through shifting the original spatial arguments of $C^S(s_1, s_2)$ by $Vt$. The derived Lagrangian spatio-temporal nonstationary covariance function $C(s_1, s_2; t_1, t_2)$ 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 $C^S(s_1, s_2)$ 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), Stephenson et al. (2005), and Risser (2015).

In the following figures, we show non-frozen Lagrangian spatio-temporal random fields for two models when $V \sim \mathcal{N}_2 \{(0.1, 0.1)^\top, 0.01 \times I_2\}$. Figure 2.1(a) plots the simulated $Z(s, t)$...
from the model:

\[
C(s_1, s_2; t_1, t_2) = \mathbb{E}_V \{ \sigma(s_1 - V t_1, s_2 - V t_2) \mathcal{M}_\nu \{ s_1 - s_2 - V(t_1 - t_2) \} \} \times D(s_1 - V t_1, s_2 - V t_2)^{-1} \{ s_1 - s_2 - V(t_1 - t_2) \}^{1/2},
\]

where \( \sigma(s_1 - V t_1, s_2 - V t_2) \) is the spatially varying variance parameter and the matrix \( D(s_1 - V t_1, s_2 - V t_2) \) serves as the spatially varying scale parameter (Kleiber and Nychka, 2012). Here \( \mathcal{M}_\nu(\cdot) \) is the univariate Matérn correlation with smoothness parameter \( \nu > 0 \), \( D(s_1, s_2) = \frac{1}{2} \{ D(s_1) + D(s_2) \} \), and \( \sigma(s_1, s_2) = |D(s_1)|^{1/4} |D(s_2)|^{1/4} |D(s_1, s_2)|^{-1/2} \). The matrix \( D(s) \) is parameterized through its spectral decomposition, i.e.,

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

Figure 2.1(b) illustrates the random field generated from the non-frozen Lagrangian deformation model:

\[
C(s_1, s_2; t_1, t_2) = \mathbb{E}_V \{ \sigma^2 \mathcal{M}_\nu \{ a \| f(s_1 - V t_1) - f(s_2 - V t_2) \| \} \},
\]

where \( f : \mathbb{R}^d \to \mathbb{R}^d \) is a deterministic non-linear smooth bijective deformation function and \( \sigma^2 \) and \( a \) are the variance and scale parameters, respectively. In the example in Figure 2.1(b), \( \sigma^2 = a = \nu = 1 \).

To illustrate the effect of the advection velocity \( V \sim \mathcal{N}_2 \{ (0.1, 0.1)^\top, 0.01 \times I_2 \} \) on the space-time dependence of the random fields in Figure 2.1 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 Figure 2.2(a), the first image in the first row gives the covariance between \( Z(s_{\text{Ref Loc 1}}, 1) \) and \( Z(s_l, 1) \), at every pixel location \( s_l \),
Figure 2.1: Simulated realizations in the unit square on a $50 \times 50$ grid from the non-frozen Lagrangian nonstationary covariance models in (2.2) and (2.3) with $\mathbf{V} \sim \mathcal{N}_2 \{(0.1, 0.1)^\top, 0.01 \times \mathbf{I}_2\}$, $\mathbf{I}_2$ is the $2 \times 2$ 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., $\mathbf{f}(\mathbf{s}) = \mathbf{b} + (\mathbf{s} - \mathbf{b})\{1 + 2\exp(-0.5\|\mathbf{s} - \mathbf{b}\|^2)\}$, $\mathbf{b} = (0.15, 0.15)^\top$. Reference locations 1 and 2 are marked with ‘×’.

$l = 1, \ldots, 2500$. The second image in the first row plots the covariance between $Z(\mathbf{s}_{\text{Ref Loc 1}}, 1)$ and $Z(\mathbf{s}_l, 2)$, at every pixel location $\mathbf{s}_l$, $l = 1, \ldots, 2500$. 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 $\mathbf{s}_l$, $l = 1, \ldots, 2500$. 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.

2.3 Non-Frozen Nonstationary Taylor’s Hypothesis

One special property of stationary covariance functions under the Lagrangian framework is that the advection velocity links the purely spatial and purely temporal margins. It is particularly useful when purely spatial second-order statistics are required but measurements in space are difficult to obtain while measurements in time are available. Taylor (1938) proposed that if observations in one dimension (temporal or spatial) are sparse, then the
observations in the dimension with much higher resolution can stand as surrogate. This idea is known as the Taylor’s hypothesis, named after its proponent. Suppose that a spatio-temporal random field has a spatio-temporal stationary covariance function $C(h, u) \in \mathbb{R}^d \times \mathbb{R}$, where $h \in \mathbb{R}^d$ is transported with velocity $v \in \mathbb{R}^d$. Then $C(h, u)$ satisfies Taylor’s hypothesis if

$$C(0, u) = C(vu, 0), \quad (2.4)$$

where $C(0, u)$ is the purely temporal stationary covariance function and $C(h, 0)$ is the purely spatial stationary covariance function. Ultimately, Taylor’s hypothesis implies a relationship between the purely spatial and purely temporal covariances, with $v$, also called the advection velocity, as the crucial linking parameter. The relationship in (2.4) is particularly useful when the desired marginal covariance is unavailable but the other one is readily obtainable.

Taylor’s hypothesis has played an important role in turbulence studies, where experiments rely on high resolution measurements in space (Kumar and Verma, 2018). For instance, the Reynolds number, an important parameter in computing the values of turbulent fluxes, which are themselves required inputs in running climate models (Kaimal et al., 1972; Cheng et al., 2017), cannot be computed properly without high resolution spatial data. How-
ever, by virtue of Taylor’s hypothesis, the Reynolds number can be computed using solely the temporal measurements of transported turbulent eddies. Taylor’s hypothesis also has a fundamental role in geophysics and hydrometeorology, specifically in modeling rainfall. Precipitation is a process with so much variability owing to its interactions with other environmental processes \cite{Schleiss2009}. Moreover, it influences other processes such as land surface hydrology, atmospheric circulation patterns, and land-atmosphere feedbacks \cite{Hurtado2009}. Modeling the evolution of rain in space and time is crucial for designing ground-based sensors for rainfall. Taylor’s hypothesis was also used to generate time series measurements from spatial measurements taken by a towed surface salinity profiler over the Pacific ocean using the ship’s speed over the ground as the conversion factor. The resulting time series of measurements was highly correlated to the simulated salinity measurements from the Generalized Ocean Turbulence Model \cite{Drushka2016}.

The breakdown in Taylor’s hypothesis has been reported several times and is often attributed to the heterogeneous features in the spatial domain. \cite{Lin1953} showed that the theory cannot be expected to hold in shear flows where spatial structures change. Other disagreements between the theory and experiments are listed in \cite{Shet2017}. The classical Taylor’s hypothesis was formulated in the second-order stationarity setting. This means that the variability of the process being transported at a rate $v$ should be the same everywhere. Once the stationarity of the second-order structure of the transported spatio-temporal random field cannot be guaranteed, it has not been established how the purely spatial nonstationary covariance relates with the purely temporal nonstationary covariance.

Taylor’s hypothesis was also shown to hold for only short periods of time. \cite{Zawadzki1973} showed that the empirical purely spatial and purely temporal covariances of rainfall rate from a particular storm occurrence, with the storm velocity as link, satisfy Taylor’s hypothesis for temporal lags less than 40 minutes. \cite{Crane1990} demonstrated using another rainfall rate dataset that the process follows Taylor’s hypothesis for temporal lags less than 30 minutes and spatial lags less than 20 kilometers. \cite{Poveda2005} performed similar
investigations on 12 storms and showed that Taylor’s hypothesis holds only in 3 storms for temporal lags less than 15 minutes. The authors conjectured that the advection velocity \( v \) changes with the passage of time, hence, the violation. The classical Taylor’s hypothesis, also called the frozen field hypothesis, only considers a constant transport behavior. Once the transport behavior changes, one can no longer derive any information about one marginal covariance given the other.

Improved versions of the classical Taylor’s hypothesis have been proposed but were mostly constructed for models of rainfall ([Lovejoy and Schertzer 1991]). Extensions of the theory were also suggested for models of velocities in turbulent boundaries. Creutin et al. (2015) introduced two velocities as the link between the two margins, namely, advection and Taylor’s velocities. Tennekes (1975) proposed to include an acceleration term in the transport. In this work, we propose a more general version of the classical Taylor’s hypothesis catering to spatio-temporal fields with spatio-temporal nonstationary covariance functions. This new Taylor’s hypothesis can accommodate processes with spatially varying second-order statistical properties. Furthermore, the new Taylor’s hypothesis replaces the constant linking parameter \( v \) with a random advection velocity \( V \in \mathbb{R}^d \). This relaxes the frozen field assumption and allows for changing transport behavior.

**Definition 1.** A spatio-temporal nonstationary covariance function \( C(s_1, s_2; t_1, t_2) \) on \( \mathbb{R}^d \times \mathbb{R} \) satisfies the non-frozen nonstationary Taylor’s hypothesis if there exist a random advection velocity \( V \in \mathbb{R}^d \) such that

\[
C(s, s; t_1, t_2) = \mathbb{E}_V[C\{s, s - V(t_2 - t_1); t_1, t_1\}].
\] (2.5)

Here \( C(s, s; t_1, t_2) \) is the purely temporal nonstationary covariance and \( \mathbb{E}_V[C\{s, s - V(t_2 - t_1); t_1, t_1\}] \) is the purely spatial nonstationary covariance. The classical Taylor’s hypothesis is a special case of (2.5), such that if \( C(s_1, s_2; t_1, t_2) \) is stationary, i.e., it can be written as a function of the spatial lag and temporal lag such that \( C(s_1 - s_2; t_1 - t_2) \), and the advection
velocity is constant, i.e., $V = v$, then (2.5) reduces to the classical Taylor’s hypothesis in (2.4).

The Lagrangian spatio-temporal nonstationary covariance functions in the previous section satisfies the non-frozen nonstationary Taylor’s hypothesis since its purely spatial margin can be written as $C(s, s; t_1, t_2) = E_V\{C^S(s - Vt_1, s - Vt_2)\}$ and its purely temporal margin is $E_V[C\{s, s - V(t_2 - t_1); t_1, t_1\}] = E_V\{C^S(s - Vt_1, s - Vt_2)\}$, such that the purely spatial and purely temporal margins are equal. The class of non-frozen Lagrangian spatio-temporal stationary covariance functions in (1.10) satisfies (2.5) when $V \sim \mathcal{N}(0, \sigma_V^2 I_d)$.

When the spatio-temporal random field follows Taylor’s hypothesis, it must be detected so one can take advantage of its implications. Most of the testing for the Taylor’s hypothesis have been done experimentally, not statistically. Li et al. (2009) devised a statistical procedure for detecting Taylor’s hypothesis. It was done through computing contrasts of covariance and obtaining test statistics based on these contrasts. The difficulty in this approach is that a set of spatio-temporal lags needs to be chosen and the test statistics are highly sensitive to this choice. Furthermore, the mechanism for choosing which spatio-temporal lags to include in the computations remains an open problem. Another way to test for the validity of Taylor’s hypothesis is using functional data analysis which would entail computing test functions. This is similar to the method in Huang and Sun (2019) and will be the subject of future work.

### 2.4 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. 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).

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

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

where $U(h) = h^2 \log h$, for $h > 0$, and zero otherwise, is a basis function, $A = (A_1, A_2, A_3)^\top \in$
\( \mathbb{R}^3 \) 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 (2.6), including multiple local maxima in the log-likelihood function and highly correlated parameters. Hence, following their recommendation, we adopt the form in (2.6) with \( w_i = \sum_{j=1}^{L-3} \beta_{j,i} g_{i,j} \), such that \( g_j = (g_{1,j}, \ldots, g_{L,j})^\top \in \mathbb{R}^L \), \( j = 1, \ldots, L - 3 \), also called the principal warps, are the last \( L - 3 \) eigenvectors of the bending energy matrix \( \mathcal{B} \) corresponding to its \( L - 3 \) nonzero eigenvalues.

The bending energy matrix \( \mathcal{B} \) is the upper left \( L \times L \) sub-matrix of \( \bar{\mathcal{B}} = [\mathbf{D} \mathbf{P}; \mathbf{P}^\top \mathbf{O}]^{-1} \in \mathbb{R}^{(L+d+1)\times(L+d+1)} \) with elements:

- \( \mathbf{D} \in \mathbb{R}^{L\times L} \) such that for \( l, r = 1, \ldots, L \), \( D_{lr} = d_{lr}^2 \log(d_{lr}) \), if \( l \neq r \), and \( D_{lr} = 0 \), otherwise, where \( d_{lr} = \|s^*_l - s^*_r\| \),

- \( \mathbf{P} \in \mathbb{R}^{L\times(d+1)} \), where the \( l \)-th row of \( \mathbf{P} \) is \( (1, \mathbf{s}_l^\top) \), \( \mathbf{s} \in \mathbb{R}^d \), and \( l = 1, \ldots, L \), 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, \( g_j \), are termed partial warps.

A key ingredient in the TPS model is the set of landmarks, \( \{s^*_1, s^*_2, \ldots, s^*_L\} \). 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 grid, wire-
Figure 2.3: Marked in red are the observation locations on a regular $10 \times 10$ 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 $\mathbf{v} = (0.5, 0.5)\mathbb{T}$, and in green are the landmarks. The landmarks (green) may or may not coincide with the observation locations (red).

less 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 $Z(s, t)$ collected at observation location $s$ at time $t$ corresponds to the measurement $Z(s - \mathbf{v}t)$ at spatial location $s - \mathbf{v}t$ in the domain of the transported random field. Figure 2.3 shows a frozen Lagrangian spatio-temporal deformed random field traveling at a constant velocity of $\mathbf{v} = (0.5, 0.5)\mathbb{T}$. 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 nonstationarity 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.

2.4.2 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. Inference is performed through maximizing the log-likelihood

$$l(\Theta; Z) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma(\Theta)| - \frac{1}{2} Z^T \Sigma(\Theta)^{-1} Z$$

(2.7)

with respect to all the parameters collected in $\Theta \in \mathbb{R}^q$. Here $\Theta$ 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 $\Sigma(\Theta)$ is formed by a parametric spatio-temporal nonstationary covariance function. Penalties can be introduced to Equation (2.7) 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, $Z$ can be rewritten as $Z = (Z^T_1, \ldots, Z^T_T)^T \in \mathbb{R}^{NT}$ such that $Z_t = \{Z(s_1, t), \ldots, Z(s_N, t)\}^T \in \mathbb{R}^N$, for $t = 1, \ldots, T$. Here $N$ and $T$ specify the number of spatial and temporal locations, respectively, and $n = NT$. Furthermore, the log-likelihood function above can be approximated as follows:

$$l(\Theta; Z_1, \ldots, Z_T) \approx l(\Theta; Z_{1, t^*}) + \sum_{j=t^*+1}^{T} l(\Theta; Z_j|Z_{j-t^*,j-1})$$

(2.8)

where $Z_{a,b} = (Z_a^T, \ldots, Z_b^T)^T \in \mathbb{R}^{Nt^*}$, for $a < b$, and $t^*$ specifies the number of consecutive temporal locations included in the conditional distribution. Here $l(\Theta; Z_j|Z_{j-t^*,j-1})$ is the log-likelihood function based only on the vector of space-time measurements $Z_{j-t^*,j-1} = (Z^T_{j-t^*}, \ldots, Z^T_{j-1})^T$. 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, 2005).
2.4.3 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 (2.8). 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 (2.8). After the second step, it is likely that the optimization routine may still not reach the global maximum of (2.8). Hence, assuming the nonstationary model, iterating between the two steps several times is pursued until a stopping criterion is satisfied.

2.5 Simulation Study: Lagrangian vs. 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.

2.5.1 Second-Order Stationarity

For the Lagrangian spatio-temporal model, we hinge our simulation studies on a particular class of non-frozen models (1.10). When $\mu_V = 0$ and $\Sigma_V = \sigma_V^2 I_d$, the non-frozen Lagrangian
Figure 2.4: Values of the non-frozen Lagrangian spatio-temporal covariance model in (2.10) for \( \rho = 0, 0.1, 0.5, \) and 0.9, at temporal lags \( u = 1, 2, \) and 3, at every \( h = (h_x, h_y)^T \) such that \( \|h\| = 1. \) Note that the case \( \rho = 0 \) corresponds to the non-Lagrangian Gneiting model in (2.9).

The model reduces to

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

which is a spatio-temporal isotropic covariance function under the Gneiting class \citep{gneiting2002}. The Gneiting model in (2.9), therefore, corresponds to a particular Lagrangian model wherein the advection velocity vector has mean zero and has independent components with common variance. While \( \sigma_V^2 \) is interpreted as the marginal variance of each component of \( V \) in Lagrangian models, in non-Lagrangian models such as that in (2.9), \( \sigma_V^2 \) 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 (1.10) and compare it with (2.9). Suppose \( d = 2, \mu_V = 0, \) and \( \Sigma_V = \sigma_V^2 \left[ \begin{array}{c} 1 \\ \rho \\ 1 \end{array} \right] \) then (1.10) reduces to

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

(2.10)
Direct comparisons between (2.9) and (2.10) for different values of $\rho$ are not straightforward since the terms bearing $\rho$ 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 (2.10) for different $\rho$, $u$, 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 2.4 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 $u = 1, 2, 3$, 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 (2.9), 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 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 Figure 2.4 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 $\rho$ 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 $\rho$ 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 $D_{LGR} = \mathbf{Z} = (\mathbf{Z}_1^{\top}, \ldots, \mathbf{Z}_{10}^{\top})^{\top}$, such that
Figure 2.5: Estimates of $\sigma_Y^2$ in (2.9) when fitted to $D^{LGR}$ generated using (1.10) with (a) $V \sim N_2 \left\{ \mu_V = 0, \Sigma_V = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\}$ at different values of $\rho$, (b) $V \sim N_2 \left\{ \mu_V = 0, \Sigma_V = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \right\}$ at different values of $\sigma_y^2$, and (c) $V \sim N_2 \left\{ \mu_V = (\mu, \mu)^\top, \Sigma_V = I_2 \right\}$ at different values of $\mu$.

$Z_t = \{Z(s_1, t), \ldots, Z(s_{100}, t)\}^\top$, $(s, t) \in \mathbb{R}^2 \times \mathbb{R}$, are simulated from (1.10), with $a = 5$, on a $10 \times 10$ grid in the unit square, under the following distributions of $V$:

(a) $V \sim N_2 \left\{ \mu_V = 0, \Sigma_V = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\}$ at different values of $\rho$;

(b) $V \sim N_2 \left\{ \mu_V = 0, \Sigma_V = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \right\}$ at different values of $\sigma_y^2$;

(c) $V \sim N_2 \left\{ \mu_V = (\mu, \mu)^\top, \Sigma_V = I_2 \right\}$ at different values of $\mu$.

We reserve the values of $Z_{10}$ 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 (2.7). At this point, we question the effect of different values of $\rho$, $\sigma_y^2$, and $\mu$ on the estimates of $\sigma_Y^2$ in the non-Lagrangian model in (2.9). Figure 2.5 gives the boxplots of parameter estimates $\hat{\sigma}_Y^2$ for 100 rounds of fitting $C^G$ on $D^{LGR}$. The values of $\hat{\sigma}_Y^2$ reflect the changing degree of dependence in space-time as we change the values of the different parameters associated to the distribution $V$. In the first panel in Figure 2.5, 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 Figure 2.4. In the middle set of boxplots,
interestingly, the median of $\hat{\sigma}_V^2$ 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 $V$ gets farther from 0, the estimate for $\sigma_V^2$ has to compensate for a faster decorrelation in time which explains the increasing median of $\hat{\sigma}_V^2$ as $\mu$ increases. In the initial experiments concerning Figure 2.5(c), a number of experimental replicates obtained $\hat{\sigma}_V^2$ with values greater than 100 as $\mu$ increases. To obtain more compact boxplots, we re-ran the experiments and bounded the values that $\hat{\sigma}_V^2$ can take to 10. This does not alter the insights provided by the unconstrained version of the experiments for Figure 2.5(c) and the results presented in Figure 2.5(a) & 2.5(b). That is, as the non-frozen Lagrangian spatio-temporal model deviates from the non-Lagrangian scenario, i.e., $V \sim \mathcal{N}_2(\mu_V, \Sigma_V)$, where $\mu_V = 0$, and $\Sigma_V = \sigma_V^2 I_2$, the more disparate the models (1.10) and (2.9) become.

Next, we study the effect of $\rho$ 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):

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

where $\hat{Z}(s_l, 10)$ is the prediction for $Z(s_l, 10)$ at spatial location $s_l$, $l = 1, \ldots, 100$, and temporal location $t = 10$. Assuming the mean of the measurement vector that was used to estimate the parameters is 0, i.e., $E(Z_{1,9}) = 0$, where $Z_{1,9} = \{Z_1^T, \ldots, Z_9^T\}^T$, predictions are computed using the simple kriging predictor

$$\hat{Z}(s_l, 10) = c_l^T \Sigma(\Theta)^{-1} Z_{1,9}.$$

Here $c_l$ is the vector of $N \times (T - 1)$ covariance function values between $Z(s_l, 10)$ and $Z(s_r, t)$,
Figure 2.6: Values of the LOE and MOM at every spatial location when $C^G$ is fitted to $D^{LGR}$ simulated with $\rho = 0.1, \ldots, 0.9$. 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.

$r = 1, \ldots, N$ and $t = 1, \ldots, 9$, i.e.,

$$c_t = \{C(s_t, s_1; 10, 1), \ldots, C(s_t, s_N; 10, 1), C(s_t, s_1; 10, 2), \ldots, C(s_t, s_N; 10, 9)\}^\top.$$  \hspace{1cm} (2.11)

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) and Hong et al. (2019), namely, the Loss of Efficiency (LOE) and the Misspecification of the Mean Square Error (MOM). The LOE and MOM at space-time location $(s_t, t)$ are computed as follows:

$$\text{LOE}(s_t, t) = \frac{E_{tr,m}(s_t, t)}{E_{tr}(s_t, t)} - 1 \quad \text{and} \quad \text{MOM}(s_t, t) = \frac{E_{m}(s_t, t)}{E_{tr,m}(s_t, t)} - 1,$$  \hspace{1cm} (2.12)

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

$$E_j(s_t, t) = C(s_t, s_1; t, t) - c^j_{l}^\top \Sigma(\Theta^*)^{-1} c^j_{l}, \quad j = \{tr, m\},$$  \hspace{1cm} (2.13)

where $c^j_{l}$ and $\Sigma(\Theta^*)$ are computed using $\Theta^* = \Theta$, for model $tr$, and $\Theta^* = \hat{\Theta}^m$ for model $m$. Here $\Theta$ is the true parameter vector while $\hat{\Theta}^m$ is the estimated parameter vector under the
model $m$. On the other hand, $E_{tr,m}(s_t, t)$ 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

$$E_{tr,m}(s_t, t) = C(s_t, s_t; t, t) - 2c_t^{tr \top} \Sigma(\hat{\Theta}^m)^{-1}c_t^m + c_t^{m \top} \Sigma(\hat{\Theta}^m)^{-1} \Sigma(\Theta) \Sigma(\hat{\Theta}^m)^{-1} c_t^m. \tag{2.14}$$

Figure 2.6 plots the LOE and MOM values at every prediction location at $t = 10$. The LOE is closer to zero when $\rho$ 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 $\rho$ is different and is somehow dictated by the contours of the distribution of $V$. 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.

### 2.5.2 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 (2.9) proposed in Garg et al. (2012). It has the form

$$C(s_1, s_2; u) = \sigma(s_1, s_2) \left[ \frac{(s_1 - s_2)^\top D(s_1, s_2)^{-1}(s_1 - s_2)}{1 + a_t u^2} \right]^{1/2},$$

(2.15)

where $\sigma(s_1, s_2)$ and $D(s_1, s_2)$ are defined in Section 2.2 and the parameter $a_t > 0$ is the scale parameter in time. Data generated from (2.15) are labeled $D_{NS}^G$. On the other hand, $C_{NS}^{LGR}$ is the non-frozen Lagrangian spatio-temporal nonstationary model in (2.2) 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., 2019). Figure 2.7 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 $\rho$. It can be seen that at every $\rho$, 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 $\rho$. This should strongly advocate the use of Lagrangian models even when the random field does not appear to be transported.

### 2.6 Application to Particulate Matter Data

A spatio-temporal process that is known to be heavily influenced by the presence of a transport medium are 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
Figure 2.8: Snapshots of the log PM 2.5 residuals on January 1, 2017. The spatial images are 4 hours apart. Two reference locations are marked for ease of transport movement detection.

behavior to its spatio-temporal dependence structure is physically reasonable.

2.6.1 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. A 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, we regard as spatio-temporally independent. Since the measurements between any two years are at least 11 months apart, this independence assumption is reasonable. Figure 2.8 maps the log PM 2.5 residuals at 550 locations in Saudi Arabia at 4-hour 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.
2.6.2 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:
  
  \[
  C(h; u) = \sigma^2 E_V \{ \mathcal{M}_\nu (a\|s_1 - s_2 - V u\|) \};
  \]

- **M2**: Non-frozen Lagrangian spatio-temporal spatially varying parameters model in (2.2);

- **M3**: Non-frozen Lagrangian spatio-temporal deformation model in (2.3);

- **M4**: Non-Lagrangian spatio-temporal stationary covariance:
  
  \[
  C(h; u) = \frac{\sigma^2}{(a_t |u|^{2\alpha} + 1)^{3/2}} \mathcal{M}_\nu \left\{ \frac{a\|h\|}{(a_t |u|^{2\alpha} + 1)^{3/2}} \right\},
  \]
  
  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:
  
  \[
  C(s_1, s_2; u) = \frac{\sigma(s_1, s_2)}{(a_t |u|^{2\alpha} + 1)^{3/2}} \mathcal{M}_\nu \left\{ \frac{(s_1 - s_2)^T D(s_1, s_2)^{-1}(s_1 - s_2)}{(a_t |u|^{2\alpha} + 1)^{3/2}} \right\}^{1/2},
  \]
  
  a more flexible version of the model in (2.15); and

- **M6**: Non-Lagrangian spatio-temporal nonstationary covariance II:
  
  \[
  C(s_1, s_2; t_1, t_2) = \frac{\sigma(s_1, s_2)}{\{|(t_1 - t_2) D(t_1, t_2)|^{2\alpha} + 1\}^{\beta/2}} \mathcal{M}_\nu \left\{ \frac{(s_1 - s_2)^T D(s_1, s_2)^{-1}(s_1 - s_2)}{\{(t_1 - t_2) D(t_1, t_2)|^{2\alpha} + 1\}^{\beta/2}} \right\}^{1/2},
  \]
Table 2.1: 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.

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

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).

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

Each pixel in Figure 2.8 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 Section 2.4, 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 (2.8). 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 hours of January 31 and predict the measurements at all spatial locations. Table 2.1 reports the performance of the six models as measured by the MSE, Akaike (AIC*), and Bayesian information criteria (BIC*), where $\text{AIC}^* = -2l(\hat{\Theta}_1, \hat{\Theta}_2) + 2q$ and $\text{BIC}^* = -2l(\hat{\Theta}_1, \hat{\Theta}_2) + q \log(Mn)$. Here $l(\hat{\Theta}_1, \hat{\Theta}_2)$ is the value of the approximated log-likelihood function at the second esti-
mation step with parameter estimates $\hat{\Theta}_2$ while fixing the parameters $\hat{\Theta}_1$ 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 $V$ under M2 are $\hat{\mu}_V = (-0.0003, 0.0017)^T$ km/hr and $\hat{\Sigma}_V = (1.719 2.257 \ 2.257 3.301) \times 10^{-5}$ km$^2$/hr$^2$. This indicates that the estimated value of the correlation between the components of $V$ is $\hat{\rho} = 0.948$.

2.7 Conclusion

This chapter introduced the nonstationary version of the classical model of Cox and Isham (1988) and tackled the practicalities of using Lagrangian spatio-temporal covariance functions to model space-time data especially under second-order nonstationarity. The use of thin plate splines in modeling second-order nonstationarity parameters was demonstrated and the maximization of the approximated log-likelihood function when data are available at regular time intervals was proposed. Through several numerical studies the effect of fitting Lagrangian models to data generated from non-Lagrangian models, and vice versa, was shown. The predictions of non-Lagrangian models on Lagrangian data was found to be 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.

A 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. Furthermore, 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 may focus on these other classes.
Chapter 3

Spatio-Temporal Nonstationary Cross-Covariance Functions under the Lagrangian Framework

3.1 Introduction

The multivariate problem calls for spatio-temporal cross-covariance functions that can capture the spatio-temporal relationship of multiple variables. While every variable of interest can be modeled and predicted separately, it has been shown that more accurate predictions can be produced when modeling dependent variables jointly (Genton and Kleiber, 2015; Zhang and Cai, 2015; Salvaña et al., 2021a). Moreover, it provides a fertile ground for proposals of new models that can accommodate nonstationary behavior in the purely spatial or purely temporal dimension, or both. In recent years, we have seen a flourishing of methods and theories, both in the stationary and nonstationary realms. When proven to be a viable approach in the multivariate setting, the stationary and nonstationary Lagrangian spatio-temporal modeling can spawn a whole new class of spatio-temporal stationary and nonstationary cross-covariance functions. As a result, the available multivariate spatio-temporal models in the literature should increase by as much as the available multivariate purely spatial stationary and nonstationary models.

The rest of this chapter is organized as follows. An overview of the state-of-the-art advances in the field of multivariate spatial and spatio-temporal geostatistics is presented in Section 3.2 focusing mainly on results that appeared after the review of Genton and Kleiber (2015). In Section 3.3, the paper transitions from surveying recent works to the creation of a new class of multivariate spatio-temporal nonstationary models under the Lagrangian
framework. Section 3.4 illustrates the performance of the new models on a regional climate model output bivariate dataset. Section 3.5 outlines the theoretical challenges and practical considerations when the spatial locations are defined on a sphere and newly established models on the sphere are provided. Section 3.6 concludes with a discussion on new research avenues and remaining challenges.

3.2 The State-of-the-Art in Purely Spatial and Spatio-Temporal Cross-Covariance Function Models

We begin the review of the state-of-the-art methods and technical progress regarding model construction, starting with multivariate purely spatial models, and then proceed with the spatio-temporal ones.

3.2.1 Purely Spatial Stationary Cross-Covariance Functions

Genton and Kleiber (2015) reviewed three general methods of constructing stationary cross-covariance functions from existing univariate stationary covariance functions: linear model of coregionalization (LMC), convolution methods, and latent dimensions. The LMC considers the multivariate process as a linear combination of uncorrelated univariate spatial processes. One major shortcoming of this model is that it lacks flexibility as it bestows on all variables the smoothness of the roughest underlying univariate spatial process. The convolution methods, on the other hand, require convolving spatially varying kernel functions. The resulting cross-covariance function may or may not have a closed form. Lastly, the latent dimensions approach works by representing the components of $Z(s)$ as coordinates in a $k$-dimensional space, $1 \leq k \leq p$.

Another proposed model is the Matérn stationary cross-covariance function formulated by Gneiting et al. (2010). Additional work on the allowable parameter values for the stationary Matérn cross-covariance function was carried out by Apanasovich et al. (2012). Cressie et al. (2015) asserted the use of a multivariate spatial random effects model. Marcotte (2015)
developed a non-linear model of coregionalization (N-LMC), addressing the aforementioned critical drawback of the LMC. The N-LMC allows for different sets of uncorrelated univariate spatial processes in the marginals and the cross-covariances. Cressie and Zammit-Mangion (2016) introduced the conditional approach to model multivariate spatial dependence, with variable asymmetry as an additional feature. They modeled the cross-covariance structure using univariate conditional covariance functions based on the partitions of $Z(s)$. Ideally, the partitioning of $Z(s)$ should reflect the causal relationship between the variables, but this may not be easy to define with many variables. Gnann et al. (2018) proposed a bivariate correlation model that resembles the LMC, with the form $C_{12}(h) = \rho \sqrt{C_{11}(h)C_{22}(h)}$, where $-1 \leq \rho \leq 1$ is the usual colocated correlation coefficient. Although they did not provide a proof of its validity, the covariance matrices they obtained were positive definite. Marcotte (2019) issued some caution regarding the use of that model and provided four counterexamples, one for each exponential, squared exponential, Matérn, and spherical correlations, where the ordinary cokriging variance turned out to be negative. Finally, not unrelated, is the work of Bevilacqua et al. (2015), where two criteria, that compare the flexibility of two different cross-covariance functions, were defined.

New univariate models, such as the modified Matérn of Laga and Kleiber (2017), require mention. The modified Matérn has a spectral density

$$f(||\omega||) = \frac{(b^2 + ||\omega||)^{\xi}}{(a^2 + ||\omega||)^{\nu + d/2}}, \quad \omega \in \mathbb{R}^d,$$

where $a, \nu > 0$, $b \geq 0$, and $\xi < \nu$. The last condition is in place to make sure that the process has finite variance. When $\xi = 0$, one obtains the classical Matérn spectral density. The model presented above is more flexible than the classical one as the maximum spectrum can occur at a non-zero frequency. When $d = 2$, they derived its resulting covariance as follows:

$$C^S(h) = \frac{1}{2\pi} \frac{(-1)^{\nu}}{\nu!} \frac{\partial^\nu}{\partial(a^2)^\nu} \left\{ (b^2 - a^2)^{\xi} K_0(a||h||) \right\},$$
where $K_0$ is a modified Bessel function of the second kind of order zero. The resulting random field of this model can exhibit strong periodicities, which the random fields from the Matérn model of [Gneiting et al. (2010)] do not possess.

### 3.2.2 Purely Spatial Nonstationary Cross-Covariance Functions

The next important advancement is the development of cross-covariance functions that lead to multivariate purely spatial nonstationary behavior. These models made it possible to allow the purely spatial cross-covariance structure to depend on the spatial locations. The models in the previous section are only appropriate when applied to multivariate purely spatial stationary random fields. When stationarity in the second-order structure is untenable, one turns to multivariate purely spatial nonstationary models. Models that accommodate nonstationarity in the cross-covariance structure are more pertinent when studying large spatial domains, as the spatial second-order nonstationary behavior can be attributed to differing spatial features affecting the phenomena being investigated. The assumption of second-order stationarity is often reasonable when one studies a relatively small spatial domain or when one can substantiate that the spatial features in the whole domain are spatially invariable.

We now mention several studies that advanced the models available for studying multivariate purely spatial nonstationary behavior. [Genton and Kleiber (2015)] started the discussion on these models by highlighting the different nonstationary extensions of the LMC such as that of [Gelfand et al. (2004)] and [Fouedjio (2018)], and the nonstationary multivariate Matérn model of [Kleiber and Nychka (2012)]. These studies are based on similar approaches, i.e., they use multivariate purely spatial stationary models as the building blocks of complicated nonstationary models.

Another way of building multivariate purely spatial nonstationary models is to start with valid univariate purely spatial nonstationary models and extend them to the multivariate case. Hence, it is not surprising that many advances in univariate purely spatial
nonstationary models have been achieved. A survey of new approaches to building univariate nonstationary covariance functions was presented by Fouedjio (2017). In addition to the aforementioned survey are other papers on covariate-driven purely spatial nonstationary models (Risser, 2015; Risser and Calder, 2015), nonstationary convolution models (Fouedjio et al., 2016), space deformation approach (Sampson and Guttorp, 1992; Fouedjio et al., 2015; Kleiber, 2016), convolution models incorporating information regarding suspected potential sources, for instance pipes or reservoirs, with application on dosimetric data (Lajaunie et al., 2019), and a nonstationary Matérn model for data affected by boundaries, holes, or physical barriers (Bakka et al., 2019). Ton et al. (2018) merged principles from Fourier feature representations, Gaussian processes, and neural networks to create new nonstationary covariance functions. Multivariate extensions to these models are nontrivial and have yet to be proposed.

Dimension expansions and covariance functions built as a linear combination of several local basis functions are two other widely popular univariate nonstationary models with multivariate nonstationary versions yet to be seen in the literature. Briefly, we propose straightforward extensions of these two approaches in the multivariate arena. Consider the multivariate Karhunen-Loève expansion (Theorem 5.2.2 of Wang (2008)) of the multivariate purely spatial random field

\[
Z(s) = \sum_{b=1}^{\infty} \{ \xi_{b,1} \lambda_{b,1} \phi_{b,1}(s) \ldots \xi_{b,p} \lambda_{b,p} \phi_{b,p}(s) \}^\top,
\]

such that \( \xi = (\xi_{b,1} \ldots \xi_{b,p})^\top \sim \mathcal{N}_p(0, I_{p \times p}) \), \( \lambda_{b,i} \in \mathbb{R} \), and \( \phi_{b,i}(s) \), \( i = 1, \ldots, p \), are the local basis functions. The nonstationary cross-covariance function of this random field is

\[
C_{ij}^S(s_1, s_2) = \sum_{b=1}^{\infty} \lambda_{b,i} \lambda_{b,j} \phi_{b,i}(s_1) \phi_{b,j}(s_2).
\]

In a landmark research in nonstationary spatial modeling, Bornn et al. (2012) showed how reducing the spatial dimensions can cause a nonstationary behavior in the covariance
structure. Following their construction approach, this time with \( p > 1 \), consider a multivariate purely spatial nonstationary random field \( \mathbf{Z}(\mathbf{s}) = \{Z_1(\mathbf{s}), \ldots, Z_p(\mathbf{s})\}^T \) such that \( \mathbf{Z}(\mathbf{s}, \cdot) = \{Z_1(\mathbf{s}, \eta_1), \ldots, Z_p(\mathbf{s}, \eta_p)\}^T \), where \( \mathbf{s} \in \mathbb{R}^d \) and \( \eta_i \in \mathbb{R}^{d'} \), \( d' > 0 \), is a multivariate purely spatial stationary random field, for \( i = 1, \ldots, p \). The components \( i, j \) of \( \mathbf{Z}(\mathbf{s}, \cdot) \), taken at spatial locations \((\mathbf{s}_1, \eta_1)\) and \((\mathbf{s}_2, \eta_2)\), after accounting for latent spatial dimensions, have a stationary cross-covariance structure \( C_{ij}^S \{(\mathbf{s}_1, \eta_1), (\mathbf{s}_2, \eta_2)\} \). The relationship between the stationary cross-covariance of the purely spatial stationary random field \( \mathbf{Z}(\mathbf{s}, \eta) \), and that of the purely spatial nonstationary random field \( \mathbf{Z}(\mathbf{s}) \) cannot be explicitly characterized, except when the cross-covariance function, \( C_{ij}^S \), is taken as the squared exponential stationary covariance function. In that case, \( C_{ij}^S(\mathbf{s}_1 - \mathbf{s}_2) = C_{ij}^S \{(\mathbf{s}_1, \eta_1) - (\mathbf{s}_2, \eta_2)\} / C_{ij}^S(\eta_1 - \eta_2) \). This dimension expansion approach allows one to use existing computationally more tractable multivariate stationary covariance functions on \( \mathbb{R}^{d + d'} \) for data exhibiting second-order nonstationary behavior on \( \mathbb{R}^d \).

Fuglstad et al. (2015) issued a caveat on using purely spatial nonstationary models hastily. They emphasized that a random field with a seemingly nonstationary second-order structure may possess that behavior due to an unaccounted nonstationarity in the mean structure. Spurious nonstationarity results in a misleading covariance structure. It is in the modeler’s interest to correctly identify the source of the nonstationarity because while nonstationarity in the mean structure is cheap, nonstationarity in the covariance structure is not.

### 3.2.3 Spatio-Temporal Stationary Cross-Covariance Functions

Combining spatial models with temporal information can tremendously improve modeling capabilities. The utility of spatio-temporal cross-covariance functions is predicated on the idea that closer objects tend to behave similarly than those that are distant to each other (Tobler 1970). Recently, this principle was recognized to hold when distance is taken with respect to the objects’ locations in time. This sparked enormous interest in building spatio-temporal models. Further, these models were constructed around the need to characterize the
behavior and interaction of multiple variables as they evolve in space and time. Nevertheless, it is essential to clarify that the focus of spatio-temporal geostatistics is typically not on the “how” of evolution; it is on describing the spatio-temporal mechanisms of an underlying process that may have generated the data.

Often, spatio-temporal datasets are modeled in the purely spatial context. When there are missing data, as temporal information is sometimes limited, one usually resorts to collapsing a spatio-temporal dataset to a spatial one by taking the spatial location-wise arithmetic mean. This is a perfectly legitimate approach as long as the scientific question to be answered is purely spatial in nature. But when the question is spatio-temporal, purely spatial models are insufficient.

The common genesis of many established spatio-temporal stationary cross-covariance functions is either a purely spatial stationary cross-covariance function or a univariate spatio-temporal stationary covariance function. A hybrid of these two approaches, the spatio-temporal (space-time) separable stationary cross-covariance function is arguably the easiest way to build multivariate spatio-temporal stationary models. Given a purely spatial stationary cross-covariance, $C_{ij}^S(h)$, and a univariate purely temporal stationary covariance, $C^T(u)$, then their product $C_{ij}(h,u) = C_{ij}^S(h)C^T(u)$ is a valid spatio-temporal (space-time) separable stationary cross-covariance function. However, multivariate spatio-temporal (space-time) separable models are always fully symmetric.

The different spatio-temporal extensions of the stationary LMC offer spatio-temporal separable stationary cross-covariance models with different types of separability. When the univariate covariances in the LMC are written as a product of two univariate covariance functions, one purely spatial and the other purely temporal, the LMC model is fully separable. Otherwise, when the univariate covariances are space-time nonseparable, the LMC model is a variable-separable model. A collection of different spatio-temporal stationary LMC is provided in De Iaco et al. (2019), and a list of different types of separability (full, space, time, and variable, to name a few) is discussed in Apanasovich and Genton (2010).
Other works, such as the several adaptations of the multivariate purely spatial Matérn to space-time were recently contributed in the literature such as the works of Bourotte et al. (2016) and Ip and Li (Ip and Li, 2016, 2017). In a Bayesian formalism, Zammit-Mangion et al. (2015) proposed a multivariate spatio-temporal model by merging stochastic partial differential equations with the spatio-temporal random field theory. Rodrigues and Diggle (2010) proposed a spatio-temporal extension of the stationary convolution models.


3.2.4 Spatio-Temporal Nonstationary Cross-Covariance Functions

Again, assuming second-order stationarity in space and in time is a convenient starting point, however, models that accommodate more realistic assumptions such as second-order nonstationarity in space and/or time are needed. These models offer more sophistication than those in the previous sections. Hence, a very sparse literature on spatio-temporal nonstationary cross-covariance functions is expected.

Ip and Li (2015) examined the possibility of changing the spatio-temporal covariance structure depending on the temporal location. The idea is simple and it addresses the problem of second-order nonstationarity in space and/or time. In their work, Ip and Li (2015) outlined several theorems to allow the $Np \times Np$ matrix, $C(\cdot; t_1, t_2)$, to assume different parametric forms for any $t_1$ and $t_2$.

A number of newly developed spatio-temporal nonstationary models were proposed only in the univariate setting such as the improved latent space approach (ILSA) of Xu and Gardoni (2018). A spatio-temporal Karhunen-Loève expansion developed in Choi (2014) may be used to construct covariance functions that are nonstationary in space and/or time.
Following the work of Bornn et al. (2012), Shand and Li (2017) formulated a spatio-temporal dimension expansion approach by performing a straightforward expansion of the temporal dimension. Again, a multivariate version of the approach has not yet been proposed and can be done as follows: Consider a multivariate spatio-temporal nonstationary random field $Z(s, t) = \{Z_1(s, t), \ldots, Z_p(s, t)\}^T$ such that

$$Z(s, \cdot; t, \cdot) = \{Z_1(s, \eta_1; t, \xi_1), \ldots, Z_p(s, \eta_p; t, \xi_p)\}^T,$$

where $s \in \mathbb{R}^d$, $\eta_i \in \mathbb{R}^{d'}$, and $\xi_i \in \mathbb{R}^{d''}$, $d' + d'' > 0$, is a multivariate spatio-temporal stationary random field, for $i = 1, \ldots, p$. The component $i, j$ of $Z(s, \cdot)$, taken at spatio-temporal locations $(s_1, \eta_i; t_1, \xi_i)$ and $(s_2, \eta_j; t_2, \xi_j)$, after accounting for the extra spatial and temporal dimensions, has a spatio-temporal stationary cross-covariance structure

$$C_{ij} \{(s_1, \eta_i), (s_2, \eta_j); (t_1, \xi_i), (t_2, \xi_j)\}.$$

### 3.3 Multivariate Nonstationary Extension to the Lagrangian Framework

Despite the many advantages of using covariance functions in the Lagrangian framework presented in the literature, the modeling framework has only been developed and applied in the univariate setting. One of the first mentions of the possible multivariate extension was by Christakos (2017) but no proof, applications, and/or theoretical aspects were discussed. Another earlier attempt was made by Apanasovich and Genton (2010) who did so using latent dimensions. The theorem below ensures the validity of the Lagrangian spatio-temporal nonstationary cross-covariance functions.
3.3.1 Main Result

**Theorem 2.** Let $V$ be a random vector on $\mathbb{R}^d$. If $C^S(s_1, s_2)$ is a valid purely spatial matrix-valued nonstationary covariance function on $\mathbb{R}^d$, i.e., $C^S(s_1, s_2) = \{C_{ij}^S(s_1, s_2)\}_{i,j=1}^p$, then

$$C(s_1, s_2; t_1, t_2) = \mathbb{E}_V\{C^S(s_1 - Vt_1, s_2 - Vt_2)\},$$

for $s_1, s_2 \in \mathbb{R}^d$, and $t_1, t_2 \in \mathbb{R}$, is a valid spatio-temporal matrix-valued nonstationary covariance function on $\mathbb{R}^d \times \mathbb{R}$ provided that the expectation exists.

Theorem 2 suggests that one way to build spatio-temporal nonstationary cross-covariance functions is to take any purely spatial nonstationary cross-covariance function and apply a Lagrangian transformation to the coordinates. The resulting covariance function is spatio-temporal, multivariate, nonstationary, and Lagrangian. A list of stylized examples applying Theorem 2 follows and an appropriate estimation procedure is discussed.

3.3.2 Examples

I. **Lagrangian Nonstationary Linear Model of Coregionalization**

There are several nonstationary versions of the classical LMC but we select the nonstationary LMC proposed by Fouedjio (2018), as an example, and extend it to space-time using Theorem 2. Let $V_r$, $r = 1, \ldots, R$, $1 \leq R \leq p$, be random vectors on $\mathbb{R}^d$ that characterize the different uncorrelated random advection velocities. If $C^S(s_1, s_2)$ is a valid purely spatial nonstationary LMC on $\mathbb{R}^d$, i.e.,

$$C^S(s_1, s_2) = \sum_{r=1}^R \rho^S_r \left[ \{(s_1 - s_2)^\top \mathbf{D}_r(s_1, s_2)^{-1}(s_1 - s_2)\}^{1/2} \right] \mathbf{A}_r(s_1)\mathbf{A}_r(s_2)^\top,$$
then
\[
C(s_1, s_2; t_1, t_2) = \sum_{r=1}^{R} \mathbb{E}_r \left( \rho_r^S \left[ \left( (s_1 - s_2 - V_r u) \right)^\top D_r (s_1 - V_r t_1, s_2 - V_r t_2)^{-1} \times (s_1 - s_2 - V_r u) \right]^{1/2} A_r (s_1 - V_r t_1) A_r (s_2 - V_r t_2)^\top \right),
\]

where \( \rho_r^S(h) \) is a valid univariate stationary correlation function of a normal scale-mixture type on \( \mathbb{R}^d \) and \( A_r \) is a \( p \times R \) matrix, is a valid spatio-temporal matrix-valued nonstationary covariance function on \( \mathbb{R}^d \times \mathbb{R} \), for any \( 1 \leq R \leq p \). The condition that \( V_r, r = 1, \ldots, R, \) are uncorrelated random vectors is set because the underlying univariate random fields are assumed to be uncorrelated. The case in which they may be dependent is on the works.

II. Lagrangian Spatially Varying Parameters Cross-Covariance Functions

Introducing spatially varying parameters in a cross-covariance function is a common approach of converting a stationary cross-covariance function to a nonstationary one. The purely spatial nonstationary LMC in (3.2) is in fact an example of the spatially varying parameters approach of Paciorek and Schervish (2006) for the normal scale-mixture type of covariance functions. Based on the univariate formulation of Paciorek and Schervish (2006), Kleiber and Nychka (2012) introduced the purely spatial Matérn nonstationary cross-covariance function. By applying Theorem 2, we obtain the Lagrangian spatio-temporal Matérn nonstationary cross-covariance function:

\[
C_{ij}(s_1, s_2; t_1, t_2) = \rho_{ij} \mathbb{E}_V \left[ \sigma_{ij} (s_1 - V t_1, s_2 - V t_2) \right. \\
\times \left\{ (s_1 - s_2 - V u) \right]^\top D_{ij} (s_1 - V t_1, s_2 - V t_2)^{-1} (s_1 - s_2 - V u) \}^{\nu_{ij}} \\
\times K_{\nu_{ij}} \left\{ (s_1 - s_2 - V u) \right]^\top D_{ij} (s_1 - V t_1, s_2 - V t_2)^{-1} (s_1 - s_2 - V u) \right\},
\]

for \( s_1, s_2 \in \mathbb{R}^d, t_1, t_2 \in \mathbb{R}, \) and \( i, j = 1, \ldots, p \). The purely spatial parameters are as follows: \( \nu_{ij} > 0 \) is the smoothness parameter, \( \rho_{ij} \in [-1, 1] \) is the colocated correlation parameter, and \( \sigma_{ij}(\cdot) \) is the spatially varying variance parameter, for \( i, j = 1, \ldots, p \). Note that \( \rho_{ij} \) may also
be allowed to vary as a function of its spatial location; see Kleiber and Nychka (2012). Here, $K_\nu$ is the modified Bessel function of the second kind of order $\nu$, $V$ is a random vector on $\mathbb{R}^d$, $\sigma_{ij}(s_1 - Vt_1, s_2 - Vt_2) = |D_i(s_1 - Vt_1)|^{1/4}|D_j(s_2 - Vt_2)|^{1/4}|D_{ij}(s_1 - Vt_1, s_2 - Vt_2)|^{-1/2}$, $D_{ij}(s_1, s_2) = \frac{1}{2} \{D_i(s_1) + D_j(s_2)\}$, and $D_i(s)$ is a $d \times d$ positive definite kernel matrix for variable $i, i = 1, \ldots, p$, at $s$ that controls the spatially varying local anisotropy and which can be defined via its spectral decomposition, i.e., for $d = 2$:

$$D_i(s) = \begin{bmatrix}
\cos \{\phi_i(s)\} & -\sin \{\phi_i(s)\} \\
\sin \{\phi_i(s)\} & \cos \{\phi_i(s)\}
\end{bmatrix}
\begin{bmatrix}
\lambda_1(s) & 0 \\
0 & \lambda_2(s)
\end{bmatrix}
\begin{bmatrix}
\cos \{\phi_i(s)\} & \sin \{\phi_i(s)\} \\
-\sin \{\phi_i(s)\} & \cos \{\phi_i(s)\}
\end{bmatrix},$$

where $\lambda_1(s), \lambda_2(s) > 0$ are the eigenvalues representing the spatial ranges and $\phi_i(s) \in (0, \pi/2)$ represents the angle of rotation.

### III. Lagrangian Multivariate Deformation Model

The univariate deformation model was first proposed by Sampson and Guttorp (1992). No multivariate extensions was proposed since. Here, we present a multivariate extension.

**Theorem 3.** If $\tilde{C}_{ij}^S(h)$ is a valid purely spatial stationary cross-covariance function on $\mathbb{R}^d$, then

$$C_{ij}^S(s_1, s_2) = \tilde{C}_{ij}^S \{\|f_i(s_1) - f_j(s_2)\|\},$$

where $f_i, i = 1, \ldots, p$, represent deterministic non-linear smooth bijective functions of the geographical space onto the deformed space, is a valid purely spatial nonstationary cross-covariance function on $\mathbb{R}^d$.

The nonstationary cross-covariance functions derived using Theorem 3 naturally yield models with variable asymmetry features. The simplest case is when $f_i = f$, for all $i = 1, \ldots, p$. When applying Theorem 2 to the cross-covariance functions in Theorem 3, we obtain the Lagrangian spatio-temporal multivariate deformation models as follows. Let $V$ be a random vector on $\mathbb{R}^d$. If $\tilde{C}_{ij}^S(h)$ is a valid purely spatial stationary cross-covariance function.
function on $\mathbb{R}^d$, then

$$C_{ij}(s_1, s_2; t_1, t_2) = \mathbb{E}_V \left[ \bar{C}_{ij} \left\{ \| f_i(s_1 - V t_1) - f_j(s_2 - V t_2) \| \right\} \right],$$  \hspace{1cm} (3.3)

where $f_i$, $i = 1, \ldots, p$, represent deterministic non-linear smooth bijective functions of the geographical space onto the deformed space, is a valid spatio-temporal nonstationary cross-covariance function on $\mathbb{R}^d \times \mathbb{R}$ provided that the expectation exists.

The deformation function can also vary in time, i.e., $f$ can also depend on the temporal location, $f^t$. As such, the model in (3.3) can be generalized in the following proposition.

**Proposition 1.** Let $V$ be a random vector on $\mathbb{R}^d$ and let $f^t_i$ be a time-varying deformation function, $i = 1, \ldots, p$. If $\bar{C}_{ij}^S(h)$ is a valid purely spatial stationary cross-covariance function on $\mathbb{R}^d$, then

$$C_{ij}(s_1, s_2; t_1, t_2) = \mathbb{E}_V \left[ \bar{C}_{ij}^S \left\{ \| f^{t_1}_i(s_1 - t_1 V) - f^{t_2}_j(s_2 - t_2 V) \| \right\} \right],$$  \hspace{1cm} (3.4)

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

### 3.3.3 Estimation

For stationary Lagrangian models, estimation can be done via least squares or maximum likelihood and may proceed in a multi-step fashion in both approaches, starting with retrieving the purely spatial parameters (marginals and crosses), followed by the advection vector parameters and other temporal parameters. The fact that, for stationary Lagrangian models, we can write out the spatial margins, free of any temporal parameters, is especially convenient for estimation. However, for nonstationary Lagrangian models, a joint estimation of purely spatial and advection vector parameter is necessary, as the spatial and temporal margins can no longer be split up; see Section 2 for a more in-depth discussion of the estimation of the nonstationary models.
### 3.3.4 Illustrations

In Figure 3.1 for $p = 2$, we illustrate the different proposed models, and their corresponding bivariate realizations are shown in Figure 3.2. In the Lagrangian nonstationary LMC example in Figure 3.1(a), we set $V_1 = v_1 = (0.1, 0.1)^\top$ and $V_2 = v_2 = (-0.1, -0.1)^\top$. In Figure 3.1(c), the first deformation is a point-source $f_1(s) = b + (s - b)\|s - b\|$, $b = (0.5, 0.5)^\top$, and the second deformation is of the form $f_2(s) = b + (s - b)\{1 + c_1 \exp(-c_2\|s - b\|^2)\}$, $b = (0.15, 0.15)^\top$, $c_1 = 6$, and $c_2 = 5$. Here $f_2$ is the same model used by Iovleff and Perrin (2004) in their simulation study.

![Figure 3.1: Heatmaps of the spatio-temporal marginals and cross-covariance functions for the proposed models: (a) Lagrangian nonstationary LMC, (b) Lagrangian spatially varying parameters model, and (c) Lagrangian deformation model. Reference locations 1 and 2 are marked in Figure 3.2.](image)

### 3.4 Application to Regional Climate Model Output

The aforementioned multivariate nonstationary spatial and spatio-temporal models are tested on a bivariate regional climate model output that includes both temperature and precipitation, the two being possibly influenced by transport, on a portion of the Midwest of the United States. The dataset is gridded and covers an area of approximately 1000 km $\times$ 1600
Figure 3.2: Simulated realizations for the proposed models: (a) Lagrangian nonstationary LMC, (b) Lagrangian spatially varying parameters model, and (c) Lagrangian deformation model. Two reference locations are represented by crosses, and are used in Figure 3.1.

km. The dataset is exactly the same as the one analyzed in Genton and Kleiber (2015). Contrary to the spatio-temporal vantage point we are proposing, after annual trend removal, they treated the yearly temporal replicates of temperature and precipitation measurements over the summer months (June, July, and August) for the years 1981-2004 as temporally independent and fitted eight purely spatial stationary and nonstationary cross-covariance function models to the dataset. This approach is limited because it misses the temporal structure due to the fact that the analysis is constrained to be purely spatial. In this work, using the same dataset used by Genton and Kleiber (2015), we view the repeated measurements in time as spatio-temporally dependent.

3.4.1 Spatio-Temporal Data Analysis

Figure 3.3 shows eight consecutive snapshots of the temperature and precipitation residual fields. From the figure, one can see that the year-on-year spatial profile of the average temperature and precipitation changes. Specifically, the lowest temperature occurs in different regions every year. Similarly, the region with the highest mean precipitation varies. Presence of atmospheric flows can cause this phenomenon. The prevailing advection direction can be detected visually by following the blue and red blobs for the temperature and precipitation
residual fields, respectively. Furthermore, the frozen field assumption can be outright dismissed as the transport direction and magnitude seem different for every two consecutive frames.

Figure 3.3 also shows that indeed, at any time point, temperature and precipitation may have equal correlation scales and that the two are negatively correlated. Moreover, the temperature residual field is smoother than the precipitation residual field and that their smoothness are consistent all throughout the temporal domain under study. Following Genton and Kleiber (2015), let \( T(s,t) \) and \( P(s,t) \) be the temperature and precipitation residual measurements at spatial location \( s \) and temporal location \( t \). Augmenting their purely spatial bivariate analysis, we seek to find the best bivariate spatio-temporal model for the phenomenon at hand. We fit one bivariate purely spatial and eight bivariate spatio-temporal models as follows:

- **M1**: Parsimonious bivariate purely spatial Matérn with spatially varying variances and colocated correlation coefficients modeled using thin plate splines.
- **M2**: Non-frozen Lagrangian stationary LMC with single advection velocity vector, i.e.,

\[
T(s,t) = A_{11}Z_1(s - V_1t) \quad \text{and} \quad P(s,t) = A_{21}Z_1(s - V_1t) + A_{22}Z_2(s - V_1t),
\]

where \( Z_1 \) and \( Z_2 \) are independent mean zero purely spatial processes generated from Matérn correlations, \( \mathcal{M}_{\nu_r}(h; a_r) \), and \( V_r \sim \mathcal{N}_2(\mu_{V_r}, \Sigma_{V_r}) \), \( r = 1, 2 \), are random advection
velocity vectors. Here $\mathcal{M}_\nu(h; a)$ is the univariate Matérn correlation with scale and smoothness parameters $a$ and $\nu$, respectively.

- **M3**: Non-frozen Lagrangian nonstationary LMC with single advection velocity vector and spatially varying coefficients modeled using thin plate splines, i.e., $T(s, t) = A_{11}(s - V_1t)Z_1(s - V_1t)$ and $P(s, t) = A_{21}(s - V_1t)Z_1(s - V_1t) + A_{22}(s - V_1t)Z_2(s - V_1t)$. $Z_1$ and $Z_2$ are the same as those in M2.

- **M4**: Non-frozen Lagrangian stationary LMC with multiple advection velocity vectors, i.e., $T(s, t) = A_{11}Z_1(s - V_1t)$ and $P(s, t) = A_{21}Z_1(s - V_1t) + A_{22}Z_2(s - V_2t)$. $Z_1$ and $Z_2$ are the same as those in M2.

- **M5**: Non-frozen Lagrangian nonstationary LMC with multiple advection velocity vectors and spatially varying coefficients modeled using thin plate splines, i.e., $T(s, t) = A_{11}(s - V_1t)Z_1(s - V_1t)$ and $P(s, t) = A_{21}(s - V_1t)Z_1(s - V_1t) + A_{22}(s - V_2t)Z_2(s - V_2t)$. $Z_1$ and $Z_2$ are the same as those in M2.

- **M6**: Non-frozen Lagrangian parsimonious bivariate stationary Matérn:

$$C_{ij}(h, u) = \rho_{ij}\sigma_i\sigma_jE_V\{\mathcal{M}_\nu(h - Vu; a)\}.$$ 

- **M7**: Non-frozen Lagrangian parsimonious bivariate Matérn with spatially varying variances and colocated correlation coefficients modeled using thin plate splines:

$$C_{ij}(s_1, s_2; t_1, t_2) = E_V\{\rho_{ij}(s_1 - Vt_1, s_2 - Vt_2)\sigma_{ij}(s_1 - Vt_1, s_2 - Vt_2)\mathcal{M}_{\nu ij}(h - Vu; a)\}.$$ 

- **M8**: Bivariate spatio-temporal Gneiting-Matérn of Bourote et al. (2016) with a frozen Lagrangian parsimonious bivariate stationary Matérn. This model is a linear combination of a bivariate spatio-temporal fully symmetric stationary covariance function and a bivariate spatio-temporal asymmetric stationary covariance function of the form:

$$C_{ij}(h, u) = \rho_{ij}\sigma_i\sigma_j\left\{\frac{1}{\alpha|u|^2 + 1}\mathcal{M}_{\nu ij}(h; a) + \Lambda\mathcal{M}_{\nu ij}(h - Vu; a)\right\},$$
where $\alpha > 0$ and $\xi \in (0, 1]$ describe the temporal range and smoothness, respectively. Here $\Lambda \in [0, 1]$ is the temporal asymmetry parameter which represents the degree of lack of symmetry in time. This temporal asymmetry parameter is key to detect possible transport effect. When $\Lambda \neq 0$ and $\mathbf{v} \neq \mathbf{0}$, the variables are most likely influenced by an advection velocity and are being transported. The model above is very flexible since a wide range of multivariate spatio-temporal random fields can be modeled, from static to moving.

- M9: Bivariate spatio-temporal Gneiting-Matérn of Bourotte et al. (2016) with a frozen Lagrangian parsimonious bivariate Matérn (similar to M8) with spatially varying variances and colocated correlation coefficients modeled using thin plate splines:

$$C_{ij}(s_1, s_2; t_1, t_2) = \rho_{ij}(s_1 - vt_1, s_2 - vt_2)\sigma_{ij}(s_1 - vt_1, s_2 - vt_2) \times \{(1 - \Lambda) \frac{1}{\alpha|u|^2\xi + 1} \mathcal{M}_{\nu_{ij}}(h; a) + \Lambda \mathcal{M}_{\nu_{ij}}(h - \mathbf{v}u; a)\}.$$  

A few remarks regarding the chosen models above are in order. Because frozen models generally do not perform well when fitted to random fields that are not frozen, as it does not allow diffusion or dissipation of covariances and cross-covariances at nonzero temporal lags, we do not fit frozen versions of models M2 to M7. We still included, however, a variant of the frozen field models such as that in models M8 and M9 as the non-Lagrangian portion of the models takes care of the dissipation of covariances and cross-covariances at nonzero temporal lags. For brevity, we limit the nonstationary models to only capture spatially varying variances and cross-correlation coefficients since that was the approach undertaken by Genton and Kleiber (2015), to which we aim to make a comparison regarding purely spatial vs. spatio-temporal fits. Moreover, prior knowledge of the topography of the region under study signify that a spatially varying variance and colocated correlation coefficients model is sufficient as every site in the region is subjected to almost similar, mainly agricultural, topographical features. The formulation of the LMC models is tailored after the technique of
Genton and Kleiber (2015) to bestowed on the temperature variable a smoother spatial random field. Finally, unlike Genton and Kleiber (2015), we do not have the luxury of independent spatio-temporal replicates to produce empirical estimates of the spatially varying variance and colocated correlation coefficients. Hence, the spatially varying parameters are assumed to vary smoothly over space and are modeled via thin plate splines; refer to Salvaña and Genton (2021) for a justification of this estimation approach.

### 3.4.2 Model Performance

The model parameters are estimated via maximum likelihood. The negative log-likelihoods are minimized using the `optim` function with quasi-Newton method “BFGS” in R (R Core Team 2019). It took approximately 3 hours to fit the purely spatial models while fitting spatio-temporal models took 20 hours using a 32-core Intel Xeon Gold 6148 with 2.6GHz clock speed. The interpolation performance of the models are evaluated by the Akaike and Bayesian information criteria. Table 1 collects the maximum likelihood parameter estimates of the best performing frozen and non-frozen models in terms of the BIC (see Table 2). Indeed, the two fields are negatively correlated with an average correlation coefficient of $-0.573$ under stationary models M6 and M8. The estimated value for the spatio-temporal asymmetry parameter $\Delta$ is nonzero, i.e., $\hat{\Delta} = 0.442$. This means that there is a transport behavior that will ultimately be missed if one results to using only spatio-temporal non-Lagrangian models. The advection velocity vector estimates from all the Lagrangian models imply approximately the same Northwest mean direction of transport.

We want to find out if spatio-temporal models have additional benefits over purely spatial

<table>
<thead>
<tr>
<th>Model</th>
<th>$\nu_1$</th>
<th>$\nu_2$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$\alpha$</th>
<th>$\Delta$</th>
<th>$\nu_1$</th>
<th>$\Sigma_{\nu_1}$</th>
<th>$\nu_2$</th>
<th>$\Sigma_{\nu_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M5</td>
<td>0.261</td>
<td>0.307</td>
<td>355</td>
<td>1026</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$(−1.096,1.441)^\top$</td>
<td>0.011</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$(−1.139,2.643)^\top$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$(0.001,0.002)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M9</td>
<td>1.275</td>
<td>0.602</td>
<td>323</td>
<td>-</td>
<td>360</td>
<td>0.442</td>
<td>$(−0.313,0.205)^\top$</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Maximum likelihood parameters estimates of the best performing frozen and non-frozen models in terms of the BIC. The advection velocity parameters are in degrees while the scale parameters $a_1, i = 1, 2$, are in kilometers. The spatially varying variance and colocated correlation coefficients are no longer shown.
models in interpolation and prediction. Comparisons of their performance can be done directly simply by using the likelihood. Also a viable approach in performance comparison between the two modeling paradigms, purely spatial and spatio-temporal, is introducing a magnitude adjustment to the likelihood function; see Ribatet et al. (2012). This was the approach taken by Sharkey and Winter (2019) to quantify loss of information when fitting purely spatial models given spatio-temporally dependent data. Another approach, which we follow in this chapter, is to conduct a pseudo cross-validation study and measure cokriging performance. In particular, we introduce different degrees of data screening. In the first round, we screen 5% of the 620 available spatial locations, at each $t$, $t = 1, \ldots, 24$. Then, we increase the number of values screened at an increment of 5%. When a spatial location is chosen to be screened, all the variables observed on that location are screened. At each round, we compute the average root mean square error (RMSE). The average RMSE is defined as:

$$\text{RMSE}_{\text{ave}} = \sqrt{\frac{1}{|S|T} \sum_{t=1}^{T} \sum_{r \in S} \|Z(s_r, t) - \hat{Z}(s_r, t)\|^2}, \quad (3.5)$$

where $T = 24$ and $S$, with cardinality $|S|$, is the set of screened spatial locations indices and $S$ does not change across $t$. Here $\hat{Z}(s_r, t)$ is the predicted values of variables at time $t$ at the unobserved location $s_r$, $r \in S$, and $\hat{Z}(s_r, t)$ is computed using the cokriging formula.

Table 3.2: A summary of the models and their in-sample (log likelihood, AIC, and BIC) and out-of-sample prediction scores (RMSE$_{\text{ave}}$). The in-sample scores were computed using the full data. The lower the AIC, BIC, and RMSE values, the better. The reverse is true for the log likelihood. The best scores are in bold. For concise comparison, we include the fit of three models in Genton and Kleiber (2015) and their corresponding out-of-sample prediction scores.
Each round is repeated ten times with different sets of randomly chosen screened spatial locations. We expect that spatio-temporal models will have lower RMSE_{ave} than the purely spatial ones since they can borrow more information from neighboring temporal sites to more accurately predict screened data. Table 3.2 summarizes the spatio-temporal cokriging performances of the different models. The log likelihood values of all spatial and spatio-temporal models are at par with each other. The nonstationary models generally perform better than their stationary counterparts. Furthermore, the spatio-temporal non-frozen Lagrangian models and the frozen Lagrangian model M7 have some of the best interpolation performance, with the non-frozen Lagrangian nonstationary LMC with multiple advection velocities as the preferred model in all metrics. This was expected since the model offers more flexibility by allowing different magnitudes and directions of advection. While the in-sample metrics (log likelihood, AIC, and BIC) provide limited evidence that spatio-temporal modeling should be pursued on this dataset, the out-of-sample metrics in Table 3.2 say otherwise. The cokriging RMSE is less when using spatio-temporal models on this bivariate dataset. Moreover, the discrepancies between the prediction performance of purely spatial and spatio-temporal models are more pronounced as more spatial locations are screened. Hence, we conclude that spatio-temporal models provide a large improvement over the purely spatial models.

3.5 Purely Spatial and Spatio-Temporal Nonstationary Cross-Covariance Functions on the Sphere

All the methods previously mentioned produce valid cross-covariance functions on the sphere when evaluated using the chordal distance. The chordal distance is the length of the shortest straight line between two locations on the sphere. However, the concept of a straight line does not make sense on a sphere. On curved surfaces such as the sphere, the amount of departure that the surfaces make from being a plane should be accounted for (Jacobson and Jacobson 2005). As a consequence, the shortest path between two locations on the sphere is rightfully
represented by a curve or a geodesic. The length of the curve separating two locations on the sphere is called the great circle distance. The great circle distance, however, renders positive definiteness of a covariance function model a serious concern. Furthermore, replacing the great circle distance with chordal distance, to which the latter has an extensive array of positive definite functions in Euclidean space to choose from, may lead to certain problems in interpolation and prediction, as the chordal distance underestimates the great circle distance (Porcu et al., 2016). These complications provided the impetus for extending existing models on the Euclidean space to work with the great circle distance and for developing new methods that work on data obtained on the sphere. Much recent progress has been made including the results of Jeong and Jun (2015), Guinness and Fuentes (2016), Jeong et al. (2017), Porcu et al. (2018), and White and Porcu (2019b). The papers from Arafat Hassan Mohammed (2017) and Guella et al. (2018) provided rigorous characterization of strictly positive definite covariance and cross-covariance functions on the sphere, respectively. An analogue of the univariate purely spatial stationary Matérn that works on the sphere was introduced by Alegria et al. (2018). Variable asymmetry on spherical processes was also studied in that paper. Lastly, White and Porcu (2019a) modeled air pollution using valid models on the sphere.

Other studies detailing the construction and characterization of nonstationary covariance functions on the sphere were published by Jun and Stein (2007), Jun and Stein (2008), Hitzchenko and Stein (2012), and Jun (2014). A study by Jun (2011) involved deriving models from scalar potentials using differential operators. A physically-motivated construction was used in the models of Fan et al. (2018), specifically for divergence-free and curl-free random vector fields. A paper by Li and Zhu (2016) extended the kernel convolution approach of Paciorek and Schervish (2006) to introduce nonstationary models on the sphere.

Alegria and Porcu (2017) and Porcu et al. (2018) were the first to discuss the validity of the covariance functions under the Lagrangian framework on the sphere, i.e., the covariance functions were evaluated without the use of the Euclidean distance but the great circle
distance instead. More specifically, the transport was modeled through a random rotation matrix $\mathcal{R} \in \mathbb{R}^{(d+1) \times (d+1)}$, and not through the random advection velocity vector $V \in \mathbb{R}^{d+1}$. Moreover, $\mathcal{R}$ was chosen such that it is an orthogonal matrix with a determinant equal to 1. Consider the sphere $S^2$ with unit radius, i.e., $S^2 = \{s \in \mathbb{R}^3, \|s\| = 1\}$, where $\|\cdot\|$ is the usual Euclidean distance, as our spatial domain. The spatial location $s \in \mathbb{R}^3$ has a spherical coordinate representation $s = (\phi, \theta)^\top$, where $\phi = L\pi/180$ and $\theta = l\pi/180$ are the polar and azimuthal angles, and $(L, l) \in [-90^\circ, 90^\circ] \times [-180^\circ, 180^\circ]$ is the spatial location given in latitude, $L$, and longitude, $l$. The Lagrangian spatio-temporal covariance function $C(s_1, s_2; t_1, t_2) = \mathbb{E}_{\mathcal{R}} \{C^S(s_1^\top \mathcal{R}^u s_2)\}, s_1, s_2 \in S^2, \mathcal{R} \in \mathbb{R}^{3 \times 3}, u = t_2 - t_1$, where $C^S$ is a purely spatial covariance function evaluating its arguments using the great circle distance, $d_{GC}(s_1, s_2) = \arccos(\langle s_1, s_2 \rangle) = \arccos\{\sin \phi_1 \sin \phi_2 + \cos \phi_1 \cos \phi_2 \cos(\theta_1 - \theta_2)\}$, and $\mathcal{R}' = Q [\text{diag}\{\exp(i\kappa_k t)\}]_{k=1}^3 Q^{-1}$, such that the eigenvalues $\lambda_k$ of $\mathcal{R}$ can be uniquely written as $\lambda_k = \exp(i\kappa_k)$, for $k = 1, 2, 3$, is a valid spatio-temporal covariance function on the sphere provided that the expectation exists. Extending this model to accommodate $p > 1$ variables is straightforward. Let $\mathcal{R}$ be a random rotation matrix on $\mathbb{R}^{3 \times 3}$. Suppose $C_{ij}^S(s_1, s_2)$ is a valid purely spatial stationary cross-covariance function on $S^2$, then $C_{ij}(s_1, s_2; t_1, t_2) = \mathbb{E}_{\mathcal{R}} \{C_{ij}^S(\mathcal{R}'_1 s_1, \mathcal{R}'_2 s_2)\}$, is a valid spatio-temporal cross-covariance function on $S^2 \times \mathbb{R}$, provided that the expectation exists. The proof can be found in the Appendix.

3.6 Conclusion

Important research progress has been made since the review paper of Genton and Kleiber (2015). Many modeling techniques and approaches were developed, and many research avenues were explored. In this chapter, we have reviewed recent advances in the field of multivariate spatio-temporal geostatistics, and presented a variety of models that can adequately describe different behaviors of multivariate spatio-temporal datasets. We devoted a significant part of the chapter to introduce and formulate new spatio-temporal covariance models under the Lagrangian framework. The Lagrangian framework provides a recipe for
extending purely spatial models to space-time and the models derived from this formulation are generally space-time asymmetric. Although here we attribute the space-time asymmetry to transport caused by an advection velocity, the modeling approach can still be used as long as the space-time asymmetry behavior is observed. The only limitation of the models under the Lagrangian framework is that they are more appropriately applied when the random field is transported.

The Lagrangian formulation of known spatial copula models is an interesting research problem. Significant work has been done in the copula space for multivariate nonstationary random fields. Krupskii and Genton (2019) proposed a new copula model that can capture more complex dependence behavior such as strong joint tail dependence and variable asymmetry. The concept of spatial asymmetry is a feature that is also recently studied in the copula space. Báróssy and Hörning (2017) offered a procedure in detecting spatial asymmetry by using the concept of reversibility in time series to purely spatial random fields. The novelty of their work lies in the ability to detect directional dependence from a single purely spatial snapshot of a spatio-temporal random field. These two papers may provide a key starting point for constructing methodologies to detect and model space-time asymmetries from a purely spatial dataset.

An additional aim of research should be to identify distributions that lead to explicit forms of (1.9). A major challenge in this area is to hasten the evaluation of the non-frozen Lagrangian covariance model. The usage of the non-frozen model is inherently difficult because of the presence of the expectation. Since only a few known specialized cases result in explicit forms, sophisticated models have to be evaluated numerically. Development of techniques in performing this numerical evaluation rapidly is also an open problem. Approximations to the non-frozen model may be attempted.

The Lagrangian framework can be used to extend multivariate purely spatial variograms to space-time. Given the relationship between cross-covariance functions and cross-variograms under joint second-order stationarity, i.e., $\gamma(h) = C(0) - \frac{1}{2}\{C(h) + C(-h)\}$, one obtains a
spatio-temporal cross-variogram: \( \gamma(h, u) = C(-uv) - \frac{1}{2}\{C(h - vu) + C(vu - h)\} \). Hence, multivariate models such as those proposed in [Chen and Genton (2019)] can readily be extended to space-time.

For the sake of conciseness, we only provided sufficient discussions on concepts and models, highlighted only their distinctive features, and restricted the discussions on models with clear avenues for future research. However, three essential topics that were omitted require mention. First, computational issues when fitting massive multivariate spatial and spatio-temporal datasets is a practical consideration that should be addressed. Furthermore, fitting complex models consumes a lot of computing power. This is largely due to heavy parameterization of more complex models. Parameter estimation and prediction becomes excruciatingly slow as \( n \) and \( p \) increase. Cost in computation should not far exceed the gain in prediction. Otherwise, there is substantial disincentive in fitting more advanced models. Nevertheless, this challenge presents an opportunity to support usage of sophisticated models on large datasets. [Ton et al. (2018)] highlighted three viable strategies to overcome scalability issues, including low rank approximations, sparse approximation methods, and spectral methods. Low rank approximations involve approximating the full covariance matrix with a matrix of smaller rank. Often, basis functions at pre-specified knots are utilized for this purpose. A recent work of [Kleiber et al. (2019)] utilized basis function representations, with coefficients taken from a multivariate lattice process, and gave alternatives to commonly used multivariate purely spatial models. Dimension reduction may also be achieved by clustering via Dirichlet processes. A complete treatment of this model is found in [Shirotu et al. (2019)]. [Baugh and Stein (2018)] proposed an approximation to the full likelihood for purely spatial nonstationary Gaussian processes using recursive skeletonization factorizations. The full recursive skeletonization factorization procedure is laid out in [Minden et al. (2017)]. [Litvinenko et al. (2019)] introduced the hierarchical matrix or \( \mathcal{H} \)-matrix approximation of a dense log-likelihood. A known technique in linear algebra, the \( \mathcal{H} \)-matrix approximation involves partitioning the full covariance matrix into sub-blocks, followed by low-rank approximation
of the majority of the sub-blocks.

The second approach, the sparse approximation methods, introduces sparsity in the dense full covariance matrix via compactly supported covariance functions. Hence, for this purpose, a great deal of attention is being given to flexible compactly supported covariance function models and covariance tapering; see Genton and Kleiber (2015) and references therein for a full discussion on this second approach. Porcu et al. (2020) provided spatio-temporal compactly supported models. The compact supports in their models are dynamical in the sense that the compact supports depend on the spatial and temporal lags. Bevilacqua et al. (2016) studied the implications of fitting multivariate covariance tapered models on two fronts: statistical efficiency and computational complexity. They concluded that their proposed models lead to some loss in computational efficiency but kept the estimation equations unbiased. Another alternative to compactly supported covariance functions are the nearest neighbor Gaussian process (NNGP) models (Datta et al., 2016). These models induce sparsity on the full precision matrix and they work under the graphical models framework. Recently, Taylor-Rodriguez et al. (2019) combined this approach with spatial factor models (SFM) to come up with the SF-NNGP model for LIDAR and ground measurements of forest variables, with large $p$ and large $n$. A specialized treatment is demanded for SFM with large $p$ and large $n$, but not all variables are observed on the spatial locations under study. This problem was tackled by Ren and Banerjee (2013) using an adaptive Bayesian factor model. Hybrid approaches involving low rank approximations and sparse approximation methods are also done in practice and were thoroughly reviewed in Zhang et al. (2019).

Lastly, spectral approaches exploit the spectral representation of the full covariance matrix. Mosammam (2016) proposed the half spectral composite likelihood targeted for large $n$ problems. His approach involves rewriting the full likelihood as a function of the periodogram and the spectral density function evaluated at $(h, \tau)$, where $h$ is the spatial lag and $\tau$ is the temporal frequency. This avoids the expensive inversion and determinant computation of the large full covariance matrix. Other spectral approaches are listed in Ton et al. (2018).
When one includes spatial (and temporal) nonstationarity into the mix of complex features present in the data, the models above cannot be appropriately applied as they are defined only in the stationary case. New models addressing large multivariate spatio-temporal nonstationary phenomenon, similar to the work of Kleiber and Porcu (2015) in the purely spatial stationary case, are demanded.

The scalability issues mentioned in the previous paragraphs may be overcome using high performance computations such as the ExaGeoStat software developed mainly for large \( n \) problems with dense full covariance matrices (Abdulah et al., 2018a). ExaGeoStat employs the most advanced parallel architectures, combined with cutting edge dense linear algebra libraries. ExaGeoStat was also fine-tuned to work on the Tile Low-Rank representation of the dense full covariance matrix (Abdulah et al., 2018b).

The second important topic regarding multivariate spatio-temporal modeling which was not yet mentioned in this work are the efficient estimation techniques for large \( n \) and \( p \) problems. A good estimation technique is necessary to provide good prediction performance. Already numerous estimation techniques have been developed: least squares, maximum likelihood, restricted maximum likelihood, composite likelihood, and other nonparametric approaches. However, theoretical developments in estimation techniques in the multivariate nonstationary context lag behind and should be attempted. Tajbakhsh et al. (2019) formulated the generalized sparse precision matrix selection (GSPS) algorithm for fitting variable separable purely spatial cross-covariance models and guaranteed theoretical convergence of the estimators. The GSPS method is predicated on a linear algebra result which states that “if the elements of a matrix show a decay property, then the elements of its inverse also show a similar behavior” (Jaffard, 1990; Benzi, 2016). The GSPS is a two-stage approach. The first stage involves approximating the precision matrix of the full data by an unparameterized sparse matrix using Gaussian Markov random field (GMRF) approximation via maximum likelihood. The second stage entails inversion of the fitted precision matrix and fitting a parametrized cross-covariance matrix via least squares. The convexity of the preci-
sion matrix in the first stage makes computation less demanding. Castrillon-Candás et al. (2016) formulated a new set of contrasts for their proposed multi-level restricted maximum likelihood. Horrell and Stein (2015) highlighted the complications brought by the composite likelihood to datasets with magnanimous spatial and temporal separation lags. According to them, the composite likelihood has no clear criteria in choosing the subsets of the data and their corresponding conditioning sets. In practice, observations with small spatial and temporal lags are grouped together. However, this is not the case with their polar-orbiting satellite dataset. Hence, they developed the Interpolation likelihood or I-likelihood which eradicates all these issues.

Lastly, new constructing principles that are capable of modeling environmental phenomenon more realistically, without sacrificing critical features to much simpler assumptions, should be explored. The pervasiveness of large spatio-temporal data has given us the ability to extract even the most hidden features of a dataset. These features should be represented in the spatio-temporal cross-covariance functions. Active areas of work such as Bayesian models and stochastic partial differential equations (SPDE) were not discussed here explicitly, but these offer different perspectives and strategies in modeling.
Chapter 4

Spatio-Temporal Cross-Covariance Functions under the Lagrangian Framework with Multiple Advections

4.1 Introduction

When analyzing the spatio-temporal dependence in most environmental and earth sciences variables such as pollutant concentrations at different levels of the atmosphere, a special property is observed: the covariances and cross-covariances are stronger in certain directions. This property is attributed to the presence of natural forces, such as wind, which cause the transport and dispersion of these variables. This spatio-temporal dynamics prompts the integration of the Lagrangian reference frame to any Gaussian spatio-temporal geostatistical model. Under this modeling framework, a whole new class was birthed and is known as the class of spatio-temporal covariance functions under the Lagrangian framework, with several developments already established in the univariate setting, in both stationary and nonstationary formulations, but less so in the multivariate case. Despite the many advances in this modeling approach, efforts have yet to be directed to probing the case for the use of multiple advections, especially when several variables are involved. Accounting for multiple advections makes the Lagrangian framework a more viable approach in modeling realistic multivariate transport scenarios.

In the previous chapter, the multivariate extension such that the model in (1.9) remains valid when the underlying purely spatial covariance function is a matrix-valued nonstationary cross-covariance function $C_S(s_1, s_2)$ on $\mathbb{R}^d$, $s_1, s_2 \in \mathbb{R}^d$ was presented. The establishing theorem relies on a single $V$ which implies that every component of $Z(s, t)$ is transported by
the same advection velocity. However, different variables may experience different transport patterns which render the model in (3.1) inadequate. The Lagrangian spatio-temporal cross-covariance function that is a linear combination of uncorrelated univariate Lagrangian spatio-temporal covariance functions, each depending on different advections, is a good first step to addressing this multiple advections problem. When the marginal and cross-advections are introduced, i.e., $V_{ij} \in \mathbb{R}^d, i, j = 1, \ldots, p$, several questions arise regarding the validity of the extended model, including which values of $V_{ij}, i \neq j$, will preserve the positive definiteness of the cross-covariance matrix resulting from (3.1). In this work, we aim at answering such a fundamental question and providing a comprehensive treatment to the Lagrangian spatio-temporal cross-covariance functions with multiple advections, with a main focus on underlying purely spatial cross-covariance functions that are stationary.

The remainder of the chapter is organized as follows. Section 4.2 presents the proposed extension of (3.1) with multiple advections and introduces some examples. Section 4.3 details the estimation procedure. Section 4.4 investigates the consequences of neglecting multiple advections in multivariate Lagrangian spatio-temporal modeling. Section 4.5 compares the performance of the proposed models with other benchmark models in the literature using a bivariate pollutant dataset of particulate matter in Saudi Arabia. Section 4.6 concludes.

## 4.2 Lagrangian Framework with Multiple Advections

The validity of Lagrangian spatio-temporal cross-covariance functions with different advection for every variable can be established by considering a zero-mean multivariate spatio-temporal random field

$$Z(s, t) = \{\tilde{Z}_1(s - V_{11}t), \ldots, \tilde{Z}_p(s - V_{pp}t)\}^\top,$$  

(4.1)

such that $\tilde{Z}(s) = \{\tilde{Z}_1(s), \ldots, \tilde{Z}_p(s)\}^\top$ is a zero-mean multivariate purely spatial random field and every component of $\tilde{Z}$ is transported by different random advections $V_{ii} \in \mathbb{R}^d, i =$
1, \ldots, p$. The resulting matrix-valued spatio-temporal cross-covariance function of the process in (4.1) is given in the following theorem.

**Theorem 4.** Let $V_{11}, V_{22}, \ldots, V_{pp}$ be random vectors on $\mathbb{R}^d$. If $C^S(h)$ is a valid purely spatial matrix-valued stationary cross-covariance function on $\mathbb{R}^d$ then

$$C(h; t_1, t_2) = E[V^T V] C^S(h - V_{i1} t_1 + V_{j1} t_2)]_{i,j=1}^p, \quad (4.2)$$

where the expectation is taken with respect to the joint distribution of $V = (V_{11}^T, V_{22}^T, \ldots, V_{pp}^T)^T$, is a valid matrix-valued spatio-temporal cross-covariance function on $\mathbb{R}^d \times \mathbb{R}$ provided that the expectation exists.

When $V_{ii} = V$, for all $i$, the above model reduces to the single advection case. Moreover, the model in (4.2) can be rewritten to resemble the form in (1.9) such that the temporal lag $u$ appears, i.e.,

$$C(h; t_1, t_2) = E[V^T V] C^S(h - \nabla_{ij} u + (V_{jj} - V_{ii})m)]_{i,j=1}^p, \quad (4.3)$$

where $\nabla_{ij} = \frac{V_{ii} + V_{jj}}{2}$, $i, j = 1, \ldots, p$, and $m = \frac{t_1 + t_2}{2}$. It can be seen that for a stationary $C^S(h)$ and for $i \neq j$, nonstationarity in time is introduced by the Lagrangian shift in the cross-covariances. However, when $i = j$, the term that depends on the midpoint between $t_1$ and $t_2$ disappears. Hence, the marginal covariances remain stationary in time.

An explicit form of (4.2) can also be derived similar to (1.10) and is given in the following theorem.

**Theorem 5.** For $p > 2$, let $V = (V_{11}^T, V_{22}^T, \ldots, V_{pp}^T)^T \sim N_{pd}(\mu_V, \Sigma_V)$. If $C^S(h)$ is a matrix-valued normal scale-mixture cross-covariance function, then

$$C_{ii}(h, u) = \frac{C^S_i(h - e^T_{(di-1):(di)} \mu_V u)^T (I_d + e^T_{(di-1):(di)} \Sigma_V u^2)^{-1} (h - e^T_{(di-1):(di)} \mu_V u)}{|I_d + e^T_{(di-1):(di)} \Sigma_V u^2|^{1/2}}, \quad (4.4)$$
where \( e_{(di-1):(di)} \) is the sub-matrix of \( I_{pd} \), comprised of its \((di-1)\)-th and \((di)\)-th rows, for \( i = 1, \ldots, p \), and

\[
C_{ij}(h; t_1, t_2) = \frac{C_{ij}^S((h - \bar{e}^\top \mu_V)^\top[I_d + T\{T^\top T + (\bar{e}^\top \Sigma_V)^{-1}\}^{-1}T^\top](h - \bar{e}^\top \mu_V))}{|I_{2d} + (\bar{e}^\top \Sigma_V)T^\top T|^{1/2}} \tag{4.5}
\]

where \( T = (t_1 I_d - t_2 I_d) \), \( \bar{e} = e_{((di-1):(di)\backslash(j-1):(dj))} \), such that \( e_{((di-1):(di)\backslash(j-1):(dj))} \) is the sub-matrix of \( I_{pd} \) comprised of its \((di-1)\)-th, \((di)\)-th, \((dj-1)\)-th, and \((dj)\)-th rows, for \( i, j = 1, \ldots, p, i \neq j \).

Several properties of non-frozen Lagrangian spatio-temporal cross-covariance functions with multiple advections can be identified based on the forms given in Theorem 5. First, the spatial lag at which maximum value occurs is at \( e^\top((di-1):(di)\backslash(u)) \mu_V \cdot u \), for the marginals, and at \( T \bar{e}^\top \mu_V \), for the cross-covariances. While the spatial lag \( e^\top((di-1):(di)\backslash(u)) \mu_V \cdot u \) is the same for any \( t_1 \) and \( t_2 \) such that \( u = t_1 - t_2 \), the spatial lag \( T \bar{e}^\top \mu_V \) is different for different values of \( t_1 \) and \( t_2 \). A significant implication of this is that for any variable, regardless of its exact spatial and temporal location, it is highly dependent with itself that is situated at a location \( e^\top((di-1):(di)\backslash(u)) \mu_V \cdot u \) away and \( u \) time steps away. Similar dynamics cannot be observed regarding the dependence between two variables with different advections. Second, there is purely spatial variable asymmetry whenever \( t_1 = t_2 \) (or \( u = 0 \)) but \( t_1, t_2 \neq 0 \). Purely spatial variable asymmetry is a property of the purely spatial marginal cross-covariance function wherein \( C_{ij}(h; t, t) \neq C_{ji}(h; t, t) \), for any \( t \) such that \( t = t_1 = t_2 \) \cite{LiZhang2011,Huang2020}. A consequence of this property is that the maximum purely spatial cross-covariance does not occur at spatial lag \( 0 \). As aforementioned, the maximum value of \( (4.5) \) occurs at \( T \bar{e}^\top \mu_V \), which is a nonzero vector unless \( t_1 = t_2 = 0 \). Furthermore, the spatial lag \( T \bar{e}^\top \mu_V \) is time-varying, i.e., it changes as the values of \( t_1 \) and \( t_2 \) change.

Figure 4.1 shows simulated Lagrangian spatio-temporal bivariate random fields from \( (4.4) \) and \( (4.5) \) on a \( 50 \times 50 \) regular grid in the unit square \([0, 1]^2 \) for \( p = 2 \) and \( d = 2 \), with the parsimonious Matérn cross-covariance function as \( C^S(h) \) \cite{Gneiting2010}. Panels (a)-
Figure 4.1: Simulated realizations of the Lagrangian spatio-temporal parsimonious Matérn cross-covariance function for $p = 2$ and $d = 2$, on a $50 \times 50$ regular grid in the unit square $[0, 1]^2$ with purely spatial parameters, namely, $\nu_{11} = 0.5$, $\nu_{22} = 1.5$, $a = 0.23$, $\rho = 0.5$, $\sigma_{11}^2 = \sigma_{22}^2 = 1$. The plots on the left hand side show the bivariate spatio-temporal random fields simulated from (4.4) and (4.5), where a common $Z_1(s, t)$ is simulated for every configuration shown in Panels (a)-(c). The different realizations of $Z_2(s, t)$ under varying degrees of dependence between $V_{11}$ and $V_{22}$, namely, (a) $V_{11} = 0.9V_{22}$, (b) $V_{11}$ and $V_{22}$ are independent, and (c) $V_{11} = -0.9V_{22}$ are displayed in Panels (a)-(c). The plots on the right hand side show the bivariate spatio-temporal random fields simulated from (4.8) with $d' = 1$, $s'_{11} = 0.2$, $s'_{22} = 0.23$. Similarly, a common $Z_1(s, t)$ is simulated for different realizations of $Z_2(s, t)$ in Panels (d)-(f), where (d) $\text{cov}(V'_{11}, V'_{22}) = 0.9$, (e) $\text{cov}(V'_{11}, V'_{22}) = 0$, and (f) $\text{cov}(V'_{11}, V'_{22}) = -0.9$.

(c) correspond to three different joint distributions of $V_{11}$ and $V_{22}$, namely, (a) $V_{11} = 0.9V_{22}$, (b) $V_{11}$ and $V_{22}$ are independent, and (c) $V_{11} = -0.9V_{22}$. The purely spatial parameters were chosen such that the practical spatial range of the variable with a less smooth field is equal to 0.7, i.e., $C_{11}(h, 0)/C_{11}(0, 0) \approx 0.05$ when $\|h\| = 0.7$. The marginal mean advection parameters are as follows: $E(V_{11}) = \mu_{V_{11}} = (0.1, 0.1)^\top$, $E(V_{22}) = \mu_{V_{22}} = (-0.1, 0.1)^\top$, $\text{var}(V_{22}) = 0.1 \times I_2$ in (a)-(c), and $\text{var}(V_{11}) = 0.1 \times I_2$ in (b), where $I_d$ is the $d \times d$ identity matrix.
The three scenarios described above imply different strengths of spatio-temporal dependence. For ease of comparison, we simulate the same $Z_1(s, t)$ for every configuration and contrast different simulated $Z_2(s, t)$. It can be seen in Panels (a)-(c) that while the direction of transport of $Z_2(s, t)$ is to the North West in all three cases, the spatio-temporal random fields are substantially different, with the different scenarios in the joint distribution of $\mathbf{V}_{11}$ and $\mathbf{V}_{22}$ having visible consequences in the values of $Z_2$ as time progresses.

The comparisons can also be done by directly examining the values of $C(h; t_1, t_2)$ at...
some spatial lag $h$ and temporal locations $t_1$ and $t_2$. Since $\text{cov}(V_{11}, V_{22})$ does not affect the marginals, the corresponding values of $C_{11}$ and $C_{22}$ are the same in all three scenarios. The difference lies in the values of $C_{12}$ and $C_{21}$. Panels (a)-(c) in Figure 4.2 visualize the corresponding spatio-temporal cross-covariance structure of the simulations in Panels (a)-(c) in Figure 4.1 at different $(h; t_1, t_2)$ combinations via heatmaps. The plots in Figure 4.2 narrate how strong the dependence is between $Z_1$ and $Z_2$ taken at two spatial locations that are $h$ units apart and $Z_1$ is either behind, ahead, or at the same time as $Z_2$. Figure 4.2(a) can be read as follows: the first plot in the first row of the $3 \times 3$ panel indicates that when $Z_1$ and $Z_2$ are both taken at the same temporal location, i.e., $t_1 = t_2 = 1$, then $Z_1$ has the highest degree of dependence with $Z_2$ taken at a spatial location that is $\mu_{V_{11}} - \mu_{V_{22}}$ units away from $Z_1$. Next, the second plot in the first row shows that if $Z_2$ is taken at $t_2 = 1$ and $Z_1$ is taken at a future time $t_1 = t_2 + 1$, then $Z_1$ is most highly dependent with the $Z_2$ at a site that is $2\mu_{V_{11}} - \mu_{V_{22}}$ units away. The other plots can also be interpreted in the same fashion.

The difference between scenarios (a)-(c) is more pronounced in Figure 4.2 than in Figure 4.1. Although the direction of advection is the same in all three, the strength of dependence is remarkably different. Strong cross-covariance is sustained when $V_{11}$ and $V_{22}$ are highly positively correlated, as in (a). On the other hand, the cross-covariance weakens rapidly when $V_{11}$ and $V_{22}$ are highly negatively correlated, as in (c). The nonstationarity in time is revealed when looking at the heatmaps along the diagonal with values of the maximum cross-covariance decreasing as time goes farther away from zero. Furthermore, the location at which the maximum cross-covariance occurs changes and moves away from $h = 0$. The single advection model is highly restrictive and does not allow for this nonstationarity in time of the cross-covariance. This implies that failing to acknowledge multiple advections can lead to overestimation or underestimation of the cross-covariances.

In the models so far, the cross-advections $V_{ij}$, $i \neq j$, cannot be prescribed independently of the temporal locations, i.e., the transport behavior experienced in the cross-covariance
remains a function of the marginal advections $V_{i1}$ and the temporal locations. An interesting problem is to find a form for $V_{ij}$ such that

$$C(h, u) = E_V[C_{ij}^S(h - V_{ij}u)]_{i,j=1}^P,$$

(4.6)

is valid. Here $V$ is the vector of marginal and cross-advections. Suppose from (4.6) that we build the covariance matrix $\Sigma$ as follows:

$$\Sigma = \begin{bmatrix} C_{11,V_{11}} & C_{12,V_{12}} & \cdots & C_{1P,V_{1P}} \\ C_{21,V_{21}} & C_{22,V_{22}} & \cdots & C_{2P,V_{2P}} \\ \vdots & \vdots & \ddots & \vdots \\ C_{P1,V_{p1}} & C_{P2,V_{p2}} & \cdots & C_{PP,V_{pp}} \end{bmatrix} \in \mathbb{R}^{np \times np},$$

(4.7)

where $C_{ij,V_{ij}} = (E_V[C_{ij}^S(s_{il} - s_{ir} - V_{ij}(t_l - t_r))])_{i,j=1}^n \in \mathbb{R}^{n \times n}$, for $i, j = 1, \ldots, p$. For the model in (4.6) to be valid, $\Sigma$ has to be positive definite. By Theorem 1 in [Ip and Li (2015)], $\Sigma$ is positive definite if and only if the $(np \times np)$ matrix $K$ with entries $K = (C_{ii,V_{i1}}C_{ij,V_{ij}}C_{jj,V_{j2}})^{1/2}_{i,j=1} \in \mathbb{R}^{np \times np}$ is positive definite. Here $A^{1/2}$ is the square root of a square matrix $A$ such that $A^{1/2}A^{1/2} = A$. This result gives us more control regarding the transport or advection behavior in the cross-covariances. However, it remains a challenge to more precisely characterize $V_{ij}$ such that $K$ is indeed positive definite.

Manipulating the transport behavior in the cross-covariances can also be done by defining new dimensions in space or time and allowing variable specific advections in those extra dimensions. The following theorem establishes a Lagrangian spatio-temporal cross-covariance model that augments the spatial dimensions.

**Theorem 6.** Let $V_{11}, V_{22}, \ldots, V_{pp}$ be random vectors on $\mathbb{R}^d$ and $V'_{11}, V'_{22}, \ldots, V'_{pp}$ be random vectors on $\mathbb{R}^{d'}$. If $C_{ij}^S(h, h')$ is a valid purely spatial stationary cross-covariance function
on \( \mathbb{R}^{d+d'} \), then

\[
C_{ij}\{(h, h'_ij); t_1, t_2\} = E \tilde{\Phi}\{C_{ij}^{S}(h - V_{ii}t_1 + V_{jj}t_2, h'_ij - V'_{ii}t_1 + V'_{jj}t_2)\}, \tag{4.8}
\]

where \( h'_ij = s'_{ii} - s'_{jj} \), for \( s'_{ii}, \ldots, s'_{pp} \in \mathbb{R}^{d'} \), and the expectation is taken with respect to the joint distribution of \( \tilde{\Phi} = \{(V_{11}^T, V_{11}')^T, (V_{22}^T, V_{22}')^T, \ldots, (V_{pp}^T, V_{pp}')^T\}^T \), is a valid spatio-temporal cross-covariance function on \( \mathbb{R}^{d+d'} \times \mathbb{R} \) provided that the expectation exists.

When \( d' = 1 \), \( s'_{ii} \in \mathbb{R} \) can possibly be the altitude or the location in the z-axis at which variable \( i \) was taken and \( V_{ii} \in \mathbb{R} \) is the component of the advection velocity along that axis, \( i = 1, \ldots, p \). Augmenting the temporal dimensions can also be done similarly. However, introducing a vector of temporal locations brings an unnatural physical interpretation to the Lagrangian transport phenomenon. Even in the classical univariate model of [Cox and Isham (1988)] in (1.9), the form of the Lagrangian shift when the temporal argument becomes a vector is nontrivial and has not yet been explored anywhere. A Lagrangian shift of the form \( h - \tilde{V}u \), where \( h \in \mathbb{R}^d \), \( u \in \mathbb{R}^{d''} \), and \( \tilde{V} \in \mathbb{R}^{d \times d''} \) is the advection velocity matrix, can be pursued when faced with multiple dimensions in time. The columns of \( \tilde{V} \) indicate the component of the transport velocity in every dimension of time. However, due to a lack of useful physical interpretation of Lagrangian models with an advection velocity matrix, we discuss only (4.8) and pursue the idea of advection velocity matrix in another work.

We simulate from (4.8) using the closed forms in (4.4) and (4.5) such that \( Z_1 \) is taken at \( (s^T, 0.2)^T \) and \( Z_2 \) at \( (s^T, 0)^T \), i.e., \( Z_1 \) and \( Z_2 \) have the same locations in the \( xy \)-axis but are separated 0.2 units away in the \( z \)-axis. Moreover, we set the advection of \( Z_1 \) and \( Z_2 \) in the \( xy \)-axis to be the same, i.e., \( \mu_{V_{11}} = \mu_{V_{22}} = (0.1, 0.1)^T \), but we augment them with different vertical components. In particular, we set \( V'_{11} \) and \( V'_{22} \) such that \( Z_1 \) gets transported downwards, while \( Z_2 \) gets transported upwards, i.e., \( E(V'_{11}) = -0.05 \) and \( E(V'_{22}) = 0.05 \). Furthermore, we consider three different strengths of dependence between the vertical components, namely (a) \( \text{cov}(V'_{11}, V'_{22}) = 0.9 \), (b) \( \text{cov}(V'_{11}, V'_{22}) = 0 \), and (c) \( \text{cov}(V'_{11}, V'_{22}) = -0.9 \). The
realizations and the corresponding heatmaps of the resulting bivariate Lagrangian spatio-
temporal random field are shown in Panels (d)-(f) in Figures 4.1 and 4.2 respectively. From
the figures, it can be seen that more sophisticated spatio-temporal dependencies spring from
adding dimensions in space. For instance, in Figure 4.2(d), the maximum cross-covariance
occurs at \( t_1 = t_2 = 2 \) and not at \( t_1 = t_2 = 1 \), which is the case in Figure 4.2(a). This
is expected because while \( Z_1 \) and \( Z_2 \) are colocated in \( \mathbb{R}^2 \), they are actually 0.2 units apart
in \( \mathbb{R}^3 \). As \( Z_1 \) travels downwards and \( Z_2 \) travels upwards, they decrease the spatial separa-
tion between them, which increases their spatial dependence. However, this behavior is not
manifested in Figure 4.2(e)-(f). This is because independent or negatively highly correlated
advections bring about greater reduction in the maximum purely spatial cross-covariance.
That is, \( Z_1 \) and \( Z_2 \) may be colocated in \( \mathbb{R}^3 \) after some time but variability from the mean
advections distorts the transported purely spatial random fields, thereby reducing the max-
imum purely spatial cross-covariance attainable.

Another way to introduce multiple advections in the Lagrangian framework while remain-
ing stationary in time is by using latent uncorrelated univariate transported purely spatial
random fields, each influenced by different advection velocities, suggested in Chapter 3, with
latent transported purely spatial random fields that are second-order nonstationary. That
model of course remains valid when the latent transported purely spatial random fields are
second-order stationary. We formalize such models in the following theorem.

**Theorem 7.** Let \( V_r, r = 1, \ldots, R, \) be random vectors on \( \mathbb{R}^d \). If \( \rho_r(h) \) is a valid univariate
stationary correlation function on \( \mathbb{R}^d \), then

\[
C(h, u) = \sum_{r=1}^{R} E_{V_r} \{ \rho_r(h - V_r u) \} T_r
\]

(4.9)
is a valid spatio-temporal matrix-valued stationary cross-covariance function on \( \mathbb{R}^d \times \mathbb{R} \), for
any \( 1 \leq R \leq p \) and \( T_r, r = 1, \ldots, R, \) are positive semi-definite matrices.

The model in (4.9) is the resulting Lagrangian spatio-temporal cross-covariance function
of the following multivariate spatio-temporal process:

\[ Z(s,t) = AW(s,t) = A[W_1(s - V_1t), W_2(s - V_2t), \ldots, W_R(s - V_Rt)]^\top, \quad (4.10) \]

where \( A \) is a \( p \times R \) matrix and the components of \( W(s,t) \) in \( \mathbb{R}^R \) are independent but not identically distributed. Each component \( W_r \) has a univariate Lagrangian spatio-temporal stationary correlation function \( \rho_r(h - V_ru) \), \( r = 1, \ldots, R \). Here, \( T_r = a_ra_r^\top \), where \( a_r \) is the \( r \)th column of \( A \). Moreover, when \( V_1 = V_2 = \cdots = V_R = V \), we return to the single advection velocity vector case and retrieve the Lagrangian spatio-temporal version of the linear model of coregionalization (LMC); see Gelfand et al. (2002) and Wackernagel (2003) for the discussion of such class of purely spatial cross-covariance functions.

### 4.3 Estimation

In this section, we outline a viable estimation procedure involving Lagrangian spatio-temporal cross-covariance functions with multiple advections. Let \( Y = \{Y(s_1,t_1)^\top, \ldots, Y(s_n,t_n)^\top\}^\top \in \mathbb{R}^{np} \) be an \( np \)-vector of multivariate spatio-temporal observations such that \( n \) is the total number of spatio-temporal locations and \( p \) is the number of variables. Assume that the mean function in (1.3) can be characterized as a linear combination of some covariates \( X_1, X_2, \ldots, X_M \). Denote by \( X = \{I_p \otimes X(s_1,t_1)^\top, I_p \otimes X(s_2,t_2)^\top, \ldots, I_p \otimes X(s_n,t_n)^\top\}^\top \in \mathbb{R}^{np \times Mp} \), where \( X(s,t) = \{X_1(s,t), \ldots, X_M(s,t)\} \in \mathbb{R}^M \) and by \( \beta = (\beta_1^\top, \ldots, \beta_p^\top)^\top \in \mathbb{R}^{Mp} \) the vector of mean parameters, where \( \beta_i = (\beta_{1,i}, \ldots, \beta_{M,i})^\top \in \mathbb{R}^M \), for \( i = 1, \ldots, p \). The model in (1.3) becomes

\[ Y(s,t) = \{I_p \otimes X(s,t)^\top\} \beta + Z(s,t). \]

Furthermore, denote by \( \Sigma(\Theta) \) the \( np \times np \) covariance matrix, parameterized by \( \Theta \in \mathbb{R}^q \) such that \( \Sigma(\Theta) = \{C_{ij}(s_l - s_r; t_l, t_r|\Theta)\}_{i,j=1}^p \}_{l,r=1}^n \). The mean parameters, \( \beta \), and the cross-covariance parameters, \( \Theta \), are estimated via restricted maximum likelihood estimation (REML) which proceeds by maximizing, through an iterative procedure (Cressie and
The iteration procedure begins with an initialization of $\beta$ which we set to the ordinary least squares estimate (OLS), $\hat{\beta}_{\text{OLS}} = (X^TX)^{-1}X^TY$, and which we plug-in to the likelihood equations above wherever $\beta$ appears. Next, we estimate $\Theta$ in a multi-step fashion. Splitting the estimation problem into several parts has been routinely employed when groups of parameters in the cross-covariance function can be estimated sequentially (Apanasovich and Genton, 2010; Bourotte et al., 2016; Qadir et al., 2021). Furthermore, it has been established that under some fairly general conditions, the multi-step MLE yields consistent estimators of the parameters in the last step (Murphy and Topel, 2002; Zhelonkin et al., 2012; Greene, 2014). Based on the properties of the Lagrangian spatio-temporal cross-covariance functions with multiple advections, maximizing (4.11) with respect to $\Theta$ can be done in two steps. Suppose $\Theta = \{(\theta^S_M)^\top, (\theta^S_C)^\top, (\theta^{ST})^\top\}^\top$, where $\theta^S_M$ is the vector of marginal purely spatial parameters, $\theta^S_C$ is the vector of purely spatial cross-covariance parameters, and $\theta^{ST}$ is the vector of advection velocity parameters. The dimensions of each vector depend on the underlying purely spatial cross-covariance function and the assumed distribution of the advection velocity vectors. The elements of $\Theta$ are estimated sequentially as follows:

1. Since the purely spatial marginal covariance functions can be derived, independently of the cross-covariance and advection parameters, the first step involves finding the vector $\hat{\theta}^S_M$ that maximizes the marginal purely spatial version of (4.11).

2. Since the cross-covariance functions are nonstationary in time, the cross-covariance and advection parameters need to be jointly estimated. That is, embedding the previously-found MLEs, $\hat{\theta}^S_M$, find $\hat{\theta}^S_C$ and $\hat{\theta}^{ST}$ that maximize (4.11). Note that to ensure $\Sigma_Y$ remains positive definite, its entries are parameterized via its Cholesky decomposition.
Once \( \Theta \) is obtained, we solve for \( \hat{\beta}_{\text{GLS}} \), where \( \hat{\beta}_{\text{GLS}} \) is the vector of estimates of the regression coefficients via generalized least squares (GLS) of the form

\[
\hat{\beta}_{\text{GLS}} = \{X^\top \Sigma(\Theta)^{-1}X\}^{-1}X^\top \Sigma(\Theta)^{-1}Y
\]

and loop again through the above multi-step estimation of \( \Theta \). The procedure is terminated when a stopping criterion is reached.

When \( T \) is large and there is negligible dependence between observations that are very distant in the temporal sense, according to Stein (2005), (4.11) can be approximated as:

\[
l_{\text{REML}}(\Theta, \beta; Y) \approx l_{\text{REML}}(\Theta, \beta; Y_{1,t^*}) + \sum_{j=t^*+1}^{T} l_{\text{REML}}(\Theta, \beta; Y_j | Y_{j-t^*,j-1}),
\]

(4.12)

where \( Y_{a,b} = (Y_a^\top, \ldots, Y_b^\top) \in \mathbb{R}^{Np} \), \( Y_t = \{Y(s_1, t)^\top, \ldots, Y(s_N, t)^\top\}^\top \in \mathbb{R}^{Np} \), for \( a < b \), and \( t^* \) specifies the number of consecutive temporal locations included in the conditional distribution. Here \( l_{\text{REML}}(\Theta, \beta; Y_j | Y_{j-t^*,j-1}) \) is the log-likelihood function based only on the vector of spatio-temporal measurements \( Y_{j-t^*,j-1} = (Y_{j-t^*}^\top, \ldots, Y_{j-1}^\top)^\top \). In the subsequent sections, we do not perform any approximations since the \( T \) in our real data application is small and thus, full REML computations are feasible.

### 4.4 Simulation Study

Under the single advection setting and for \( p = 1 \), Chapter 3 showed that when the components of \( V \sim \mathcal{N}_d(\mu_V, \Sigma_V) \) have zero mean, the same variance, and are uncorrelated, i.e., \( \mu_V = 0 \) and \( \Sigma_V = \sigma_V^2 I_d \), for any \( \sigma_V^2 > 0 \), the univariate Lagrangian spatio-temporal models with normal scale-mixture \( C^S \) reduce to univariate non-Lagrangian spatio-temporal isotropic covariance functions belonging to the Gneiting class (Gneiting, 2002b). A breakdown on any of the above-mentioned restrictions dichotomizes univariate Lagrangian spatio-temporal models from their non-Lagrangian counterparts. Their simulation studies can be adopted
for $p > 1$ and similar conclusions can be drawn.

When faced with a multivariate Lagrangian spatio-temporal random field with multiple advections, one can either fit multivariate Lagrangian spatio-temporal models with multiple advections, as in (4.3), or marginally fit univariate Lagrangian spatio-temporal models, as in (1.9), each with different advections. While it has been shown in the literature that multivariate modeling generally yields lower prediction errors as the presence of the other variables essentially increases the sample size of one variable (Genton and Kleiber 2015; Zhang and Cai 2015; Salvaña et al. 2021a), it remains to be explored how the dependence between any two advection velocities affects the accuracy of predictions. To answer this inquiry, we perform experiments to identify scenarios where multivariate Lagrangian spatio-temporal models with multiple advections are favorable over multiple univariate Lagrangian spatio-temporal ones.

Another objective of this section is to show the consequences of using a bivariate Lagrangian spatio-temporal covariance function with single advection to model a bivariate Lagrangian spatio-temporal random field with advections $V_{11} \sim \mathcal{N}_d(\mu_{V_{11}}, \Sigma_{V_{11}})$ and $V_{22} \sim \mathcal{N}_d(\mu_{V_{22}}, \Sigma_{V_{22}})$ such that $\mu_{V_{11}} = \mu_{V_{22}}$ and $\Sigma_{V_{11}} = \Sigma_{V_{22}}$ but $V_{11} \neq V_{22}$. Such bivariate random fields appear to be driven by single advection when in fact they are not. Again, the dependence between $V_{11}$ and $V_{22}$ introduces some interesting dynamics that may or may not be useful in modeling or prediction. In this section, we aim to expose such consequences.

4.4.1 Design

All the simulation studies are framed under the assumption that $d = 2$ and $p = 2$. Consider the following Lagrangian spatio-temporal models:

- **M1**: univariate Lagrangian spatio-temporal model in (1.10), where $C^S$ is the Matérn covariance function, with purely spatial parameters $\sigma, a, \nu$;
• M2: bivariate Lagrangian spatio-temporal model with single advection, i.e.,

\[ C_{ij}(h, u) = \frac{\rho \sigma_{ij}}{\sqrt{I_d + \Sigma V u^2}} \mathcal{M}\{(h - \mu V u)^\top (I_d + \Sigma V u^2)^{-1} (h - \mu V u); a, \nu_{ij}\}, \quad i, j = 1, 2, \]

where \( \mathcal{M}(h; a, \nu) \) is the univariate Matérn correlation with spatial scale and smoothness parameters \( a \) and \( \nu \), respectively; and

• M3: bivariate Lagrangian spatio-temporal model with multiple advections in (4.4) and (4.5), where \( C_{ij}^S \) is the parsimonious Matérn cross-covariance function, with purely spatial parameters \( \rho, \sigma_{ij}, a, \nu_{ij}, i, j = 1, 2 \).

We simulate 100 sample zero-mean bivariate spatio-temporal Gaussian random fields, \( Z_1(s, t) \) and \( Z_2(s, t) \), with purely spatial parameters as in Figure 4.1 containing \( N = 529 \) spatial observations, on a \( 23 \times 23 \) grid in the unit square, at time \( t = 0, 1, \ldots, 5 \). This number of spatio-temporal locations from which the data is generated is chosen to mimic the setup in the real spatio-temporal data to be analyzed in Section 4.5. In the following experiments, we remove the observations at \( t = 5 \) and use the remaining observations to fit the models. Upon obtaining the parameter estimates, we predict the previously removed values using simple kriging for the univariate model and simple cokriging for the bivariate models and report the error of the predictions measured by the Mean Square Error (MSE), MSE = \( \frac{1}{pn} \sum_{i=1}^{p} \sum_{l=1}^{N} \{\hat{Z}_l(s_i, 5) - Z_l(s_i, 5)\}^2 \).

In order to understand the effect of the dependence between the two advection velocities, \( V_{11} \) and \( V_{22} \), on the accuracy of predictions, we simulate from M3 under different assumptions on the joint distribution of \( V_{11} \) and \( V_{22} \), namely, \( \mu_{V_{11}} = (0.1, 0.1)^\top \), \( \mu_{V_{22}} = (-0.1, 0.1)^\top \) and (a) \( V_{11} = 0.9 V_{22} \), (b) \( V_{11} \) and \( V_{22} \) are independent, and (c) \( V_{11} = -0.9 V_{22} \), for several values of the spatial cross-correlation parameter, i.e., \( \rho = \pm 0.3, \pm 0.6, \pm 0.9 \). It is worthwhile to note that configuration (a) is the closest to the single advection model, in which case \( V_{11} = V_{22} \), while (c) is the farthest. On the simulated values, we fit the true model, M3, and a simpler alternative model, M1.
To reveal the consequences of using a single advection model instead of a multiple advections model, we simulate from M3 with $\mu V_{11} = \mu V_{22} = (0.1, 0.1)^\top$ with different $\text{cov}(V_{11}, V_{22})$, namely, (d) $\text{cov}(V_{11}, V_{22}) = k \left( \begin{smallmatrix} 1 & 0.9 \\ 0.9 & 1 \end{smallmatrix} \right) \otimes I_2$, (e) $\text{cov}(V_{11}, V_{22}) = kI_4$, (f) $\text{cov}(V_{11}, V_{22}) = k \left( \begin{smallmatrix} 1 & -0.9 \\ -0.9 & 1 \end{smallmatrix} \right) \otimes I_2$, for $k = 0.001, 0.1, 1$, and $\rho = \pm 0.3, \pm 0.6, \pm 0.9$. Scenarios (d) and (f) represent the highly positive and negative dependence between the corresponding components of $V_{11}$ and $V_{22}$, respectively, while (e) establishes that $V_{11}$ and $V_{22}$ are independent. The parameter $k$ acts as the scale parameter in time, with higher values of $k$ implying faster drop in correlation in the temporal domain. Both the true model, M3, and a simpler model, M2, are fitted to the aforementioned simulated random fields.

### 4.4.2 Results and Analysis

In the first set of experiments, we compare the accuracy of the predictions between the two modeling paradigms, namely, marginally fitting multiple univariate Lagrangian spatio-temporal models with different advections (M1) and fitting a bivariate Lagrangian spatio-temporal model with multiple advections (M3), given that the true model is the latter.
Figure 4.3 summarizes the boxplots of the MSEs under different strengths of dependence between $V_{11}$ and $V_{22}$ and different values of the colocated correlation $\rho$. It can be seen that the prediction performance of M1 is fairly similar regardless of the true value of $\rho$ and the dependence between the different advections. Using a bivariate model (M3), on the other hand, improves predictions for nonzero $\rho$, with higher gains in accuracy as $\rho$ gets farther away from 0. Furthermore, M3 yields more accurate predictions when the true model consists of a nonzero $\rho$ with advection vectors that are highly dependent. When utilizing M1 over M3, one disregards a possible purely spatial variable dependence and possible dependence between the advections, which can help improve predictions.

In the second set of experiments, we study the effect of fitting a single advection model to a bivariate random field that appears to be simulated from a single advection model on two fronts: 1) estimates of the common parameters; and 2) prediction performances of models M2 and M3. Figure 4.4 gives the summary of the centered and scaled MLEs of M2, denoted by $\hat{\Theta} = (\hat{\sigma}_{11}^2, \hat{\sigma}_{22}^2, \hat{\mu}_{11}, \hat{\mu}_{22}, \hat{\nu}_{11}, \hat{\nu}_{22}, \hat{\rho}, \hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}_{111}, \hat{\Sigma}_{222}, \hat{\Sigma}_{112}, \hat{\Sigma}_{122})^\top$, where $\hat{\mu}_V = (\hat{\mu}_1, \hat{\mu}_2)^\top$ and $\hat{\Sigma}_V = \begin{pmatrix} \hat{\Sigma}_{111} & \hat{\Sigma}_{112} \\ \hat{\Sigma}_{121} & \hat{\Sigma}_{222} \end{pmatrix}$. Clearly, the most impacted parameter is $\rho$. As the dependence between $V_{11}$ and $V_{22}$ goes from highly positive to highly negative, $\hat{\rho} \to 0$ for any value of $\rho$ that we
considered. This is expected for the following reason. The cross-covariance function in M2 is stationary in space and time, which implies that the purely spatial colocated correlation is constant. M3, on the other hand, has a cross-covariance function that is nonstationary in time with a time-varying and decreasing purely spatial colocated correlation as \( t_1 \) and \( t_2 \) move away from 0. Furthermore, the more negatively dependent \( V_{11} \) and \( V_{22} \) are, the faster the decline in the purely spatial colocated correlation; see Section 2 and Figure 4.2. Thus, when a model that can only handle a constant purely spatial colocated correlation parameter is fitted to a bivariate random field that possesses decreasing purely spatial colocated correlation, the optimization routine is required to find the best compromise for \( \hat{\rho} \).

Additionally, we scrutinize the errors, summarized via boxplots in Figure 4.5, when fitting M2 and M3 to data generated from M3. It can be seen that M2 does not predict as well as M3 in all cases but the difference in accuracy is the highest under scenario (f). This is because in scenario (f), M2 severely overvalues the cross-covariances compared to the ones obtained by M3, resulting in a grave misspecification of spatio-temporal dependence between \( Z_1 \) and \( Z_2 \). All in all, in both experiments, the results demonstrate that neglecting the multiple advection phenomenon leads to substantial losses in prediction accuracy.
4.5 Application to Particulate Matter Data

Wind is recognized as a major driver of pollutant transport in the atmosphere. The Modern-Era Retrospective Analysis for Research and Applications v.2 (MERRA-2) simulated particulate matter (PM) concentrations over Saudi Arabia plotted in Figure 4.6 demonstrate such transport behavior (Buchard et al., 2017). Any physical model on PM factors in mechanisms of transport such as the wind fields at different levels of the troposphere (Kallos et al., 2007), jet streams, pyroconvective events, and boundary layer turbulence (National Research Council, 2010). The transport behavior enables the propagation of suspended PM to locations far from the original source and this often results to higher PM concentrations on its path of transport.

The modeling of PM concentrations is an active field of research in computational fluid dynamics (CFD). In that area of study, emphasis is placed on airflow motion and turbulence in determining PM concentrations at any location in space and time (Zhang and Chen, 2007; Katra et al., 2016). CFD involves tracking a large number of particles and employs highly specialized models, e.g., particle concentration equations coupled with momentum and turbulence equations, which need to be fed with various model input parameters (Knox, 1974). Lack of domain expert knowledge impedes adoption of these physically consistent models by non-experts in CFD.

The proliferation of PM measurements, which can be obtained from several sources such as ground stations, online databases (e.g., NASA Earthdata website), satellite remote sensing, lidar networks, and other outdoor monitoring systems, has launched a new wave of statistical modeling methodologies focused on modeling and predicting particulate matter concentrations (Shao et al., 2011). The list includes Bayesian models (Sahu et al., 2006; Calder, 2008), generalized additive models (Paciorek et al., 2009; Munir et al., 2013), geographically weighted regression (Van Donkelaar et al., 2016), stochastic partial differential equations (Cameletti et al., 2013), time series models (Goyal et al., 2006), and machine learning models (Mehdipour et al., 2018). These modeling approaches typically require only
Figure 4.6: MERRA-2 simulated particulate matter concentrations on January 18, 2019, 9:00 to 21:00, at two pressure levels in log scale of $\mu g/m^3$. Two sites are marked with “×” to aid detection of transport behavior.

Historical measurements of PM and other associated variables. In this work, we tackle the problem on hand by combining the strengths of spatio-temporal Gaussian geostatistical modeling and the concept of transport in CFD through the class of Lagrangian spatio-temporal models.

4.5.1 Bivariate PM 2.5 Dataset

The dataset used in the present study was obtained from [NASA Earthdata](https://earthdata.nasa.gov) and it contains measurements of PM (black carbon) concentrations, measured every 3 hours at two different pressure levels, 880 hPa ($\sim 1.2$ km above sea level) and 985 hPa ($\sim 0.7$ km above sea level), on a regular grid with pixel size $0.5^\circ \times 0.625^\circ$, with no missing observations. A logarithmic transformation is applied to obtain close to normally distributed measurements; see [Paciorek et al. 2009](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2711563), [Sahu 2012](https://link.springer.com/article/10.1007%2Fs10659-011-0359-7), and [Cameletti et al. 2013](https://link.springer.com/article/10.1007%2Fs10659-013-9325-7) for similar treatments. Additionally, the spatial coordinates originally in degree latitude and longitude are converted to their corresponding Cartesian coordinates. This means that the distance between any two locations is measured in kilometers (km), with an overall minimum distance between any two stations equal to 16.9 km.
In this study, we consider only the five consecutive 3-hourly measurements starting from 9:00 to 21:00 on January 18, at 550 locations over Saudi Arabia, over the period 1980-2019. In this work, we assume each year as an independent replicate coming from the same underlying spatio-temporal Gaussian process. Figure 4.6 maps the log PM concentrations over the spatial and temporal domains under consideration for the year 2019. Using the reference locations as aid, the transport behavior can be detected, with a seemingly South East direction of movement at both pressure levels.

The spatio-temporal modeling is carried out following the linear model in (1.3), where $Y_1(s,t)$ and $Y_2(s,t)$ are the log PM measurements at 880 hPa and 925 hPa, respectively, at spatial location $s \in D$, $D$ being the region inside the borders of Saudi Arabia, and temporal location $t \in T$, $T = \{0, \ldots, 4\}$. Moreover, we make the following assumptions on $\mu(s,t)$ and $Z(s,t)$. First, the mean function $\mu(s,t)$ can be modeled by a set of covariates comprising of the relative humidity and temperature at spatio-temporal location $(s,t)$. Both covariates are accessible in the same database where the log PM measurements were obtained. Second, the resulting residuals $Z(s,t)$ are assumed to be Lagrangian or transported.

NASA Earthdata also provides simulated wind vectors at the same spatial and temporal resolutions as the log PM concentrations at each pressure level. These simulated wind vectors were used as inputs by MERRA-2 to simulate the raw PM concentrations (Randles et al., 2017; Ukhov et al., 2020). In Figure 4.7, we plot the pairwise empirical bivariate distributions of the components of the wind vectors in m/s at the two pressure levels. A plausible estimate of $\mu_V$ is the empirical mean vector. When inspecting the second plot in the first row and the last plot in the third row, one may hypothesize that the bivariate spatio-temporal random field is from a single advection model since the bivariate distributions of the wind vectors of the two variables appear to be identical. However, as we have shown in Section 4.4, the appearance of a single advection phenomenon does not guarantee that the true model is a single advection model. Another observation is that the corresponding components of the advection vectors of the two variables are not equal but are highly positively dependent as
Figure 4.7: Empirical bivariate distributions of the MERRA-2 simulated wind vectors (in m/s) over the spatio-temporal domain under consideration. Here $v_x$ ($v_y$) denotes the first (second) component or the component along the x-axis (y-axis) of the wind vector. Superimposed are $\mu_v$ (in m/s) in M1, M2 (in red and green), $\mu_{v_y}$ (in m/s) in M3 (blue), and $\mu_{v_y}$ (in m/s) in M4 (purple), respectively. The lines in the bivariate plots enclose the 95% probability region. Only M3 and M4 are applicable in the 2 x 2 panels from the top right corner. Here M3 and M4 almost coincide, with M3 being the best performing model according to the BIC*.

shown in the third plot in the first row and the last plot in the second row, a scenario that we also considered and studied in Section 4.4 Thus, a multiple advections model may be the most suitable for this dataset.
4.5.2 Models

We consider six different spatio-temporal cross-covariance functions with bivariate parsimonious Matérn purely spatial margins, three of which were utilized in Chapter 4.4, namely, M1, M2, and M3, and the other three are as follows:

- M4: bivariate Lagrangian spatio-temporal model with variable specific advection in (4.8), with $d' = 1$;
- M5: bivariate Lagrangian LMC in (4.9); and
- M6: bivariate non-Lagrangian fully symmetric Gneiting-Matérn, i.e.,

$$C_{ij}(h, u) = \frac{\rho \sigma_i \sigma_{jj}}{\alpha |u|^{2\xi} + 1} \mathcal{M} \left\{ \frac{h}{(\alpha |u|^{2\xi} + 1)^{b/2}}; a, \nu_{ij} \right\}, \quad i, j = 1, 2,$$

where $\alpha > 0$ and $\xi \in (0, 1]$ describe the temporal range and smoothness, respectively (Bourotte et al., 2016). The parameter $b \in [0, 1]$, also called the “nonseparability parameter”, represents the strength of the spatio-temporal interaction.

Preliminary data analyses suggest that there is an anisotropic behavior in the marginal and cross-covariance structures. Hence, instead of evaluating the models at $h$, we use $Rh$, where $R = \begin{pmatrix} R_1 \cos R_3 & R_1 \sin R_3 \\ -R_2 \sin R_3 & R_2 \cos R_3 \end{pmatrix}$ is the anisotropy matrix parameterized by the anisotropy scale parameters, $R_1$ and $R_2$, and the anisotropy angle parameter, $R_3$; see Paciorek and Schervish (2006) and Hewer et al. (2017) for discussions on this modeling approach and more general parameterizations of $R$.

4.5.3 Results and Discussions

Through the procedure described in Section 4.3, we fit the models using all 40 spatio-temporal random fields and predict the concentrations at all spatial locations at $t = 6$ for each year. The anisotropy parameters included in $R$ are estimated in the first step alongside the marginal purely spatial parameters. Figure 4.8 shows the resulting map of predicted log con-
Figure 4.8: Map of predicted log residuals particulate matter concentrations on January 19, 2019, 0:00, under the six different models.

Table 4.1: In-sample (log-likelihood, AIC∗, and BIC∗) and out-of-sample (MSE) scores. The lower the values, the better. The best scores are given in bold.

<table>
<thead>
<tr>
<th>Model</th>
<th>log-likelihood</th>
<th>In-Sample</th>
<th>Out-of-Sample</th>
<th>No. of parameters</th>
<th>Computation Time (secs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AIC∗</td>
<td>BIC∗</td>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>461,825</td>
<td>−923,614</td>
<td>−923,439</td>
<td>0.0521</td>
<td>18</td>
</tr>
<tr>
<td>M2</td>
<td>479,159</td>
<td>−958,290</td>
<td>−958,156</td>
<td>0.0546</td>
<td>14</td>
</tr>
<tr>
<td>M3</td>
<td>484,070</td>
<td>−968,094</td>
<td><strong>967,873</strong></td>
<td>0.0516</td>
<td>23</td>
</tr>
<tr>
<td>M4</td>
<td><strong>484,150</strong></td>
<td>−968,210</td>
<td>−967,777</td>
<td><strong>0.0514</strong></td>
<td>45</td>
</tr>
<tr>
<td>M5</td>
<td>470,852</td>
<td>−941,658</td>
<td>−941,437</td>
<td>0.1602</td>
<td>23</td>
</tr>
<tr>
<td>M6</td>
<td>477,480</td>
<td>−954,936</td>
<td>−954,821</td>
<td>0.0601</td>
<td>12</td>
</tr>
</tbody>
</table>

The results show that models with multiple advections, i.e., M1, M3, and M4, obtain superior prediction performance compared to those with single (M2) or no advection at all (M6). Furthermore, all the Lagrangian spatio-temporal models, M1-M4, outperform the non-Lagrangian spatio-temporal benchmark model, M6, in prediction. This is due to the Lagrangian models being able to capture the transport behavior that is seen to have
occurred when connecting Figure 4.6 and the first column of Figure 4.8. The Lagrangian
LMC model, M5, appears to be the worst performing model in terms of prediction. Even
though M5 accommodates multiple advections, it does so in an unnatural way. Furthermore,
the drawbacks possessed by the purely spatial LMC (Gneiting et al., 2010) is inherited by its
Lagrangian spatio-temporal extension. M4 has the best performance among all the models in
3 out of 4 metrics, but at the cost of more parameters to be estimated and more computation
time. The BIC* supports the usage of M3 as its performance is very close to M4 but with
less parameters to estimate. Based on these results and the computational efficiency of the
models, M3 emerges as the best model overall.

To visually validate the fit of the best performing model in terms of BIC*, we plot the
empirical spatio-temporal marginal and cross-covariances of the resulting log residuals under
M3 in Figure 4.9. The empirical plots of the resulting residuals reveal some anisotropy
in both the marginal and cross-covariances. Furthermore the spatial lags at which the
maximum values of the marginal and cross-covariances occur, i.e., locations marked with “×”,
hover around $h = 0$. These observations provide further support on including anisotropy
parameters in the models and they also suggest that there may be no purely spatial variable
asymmetry at any time. The fitted marginal and cross-covariances are juxtaposed beside
their empirical counterparts. Clearly, M3 closely mirrors the spatio-temporal dependence
profile of the real bivariate spatio-temporal dataset.

To further validate the in-sample and out-of-sample performance of the models, we boot-
strap through the 40 available spatio-temporal replicates, which we assume are independent
and identically distributed, by performing 100 rounds of estimation and prediction using
a subset of randomly sampled 30 spatio-temporal random fields. Figure 4.10 shows the
boxplots of the log-likelihood and MSE values. Figure 4.11 displays the boxplots and, con-
sequently, the uncertainties of the estimated parameters of M3, such that highlighted by a
red line are the parameters in the model fitting performed to obtain the results in Table 4.1.
It can be seen that the estimated parameters are substantiated by the empirical marginal
Figure 4.9: Empirical ($C_{11}$ and $C_{22}$) and fitted ($\hat{C}_{11}$ and $\hat{C}_{22}$) marginals and empirical ($C_{12}$) and fitted ($\hat{C}_{12}$) cross-covariance values at every spatial and temporal lag are plotted as heatmaps. The spatial resolution is in $\times 10^2$ kilometers while the temporal resolution is 3 hours. The spatial lags at which the maximum value of the marginal and cross-covariance occur are marked with “$\times$”.

Additionally, the estimates associated with the transport phenomenon are validated against the real transport behavior observed in the spatio-temporal domain under study. In Figure 4.7, the estimated distribution of the advection velocity vectors in M1-M4 are superimposed on the real wind vectors. Visually, it can be seen that the advection parameters obtained by models M3 and M4 adequately captures the empirical distribution. Furthermore, M1 and M2 obtains higher values for the marginal variance parameters of the components
Figure 4.10: The log-likelihood and MSE values of the different models under the bootstrap study.

Figure 4.11: Boxplots of estimated parameters of M3. In red, the estimates obtained using all the 40 spatio-temporal random fields are plotted. Here \( \mathbf{V}_{11} = \{ V_x(880 \text{ hPa}), V_y(880 \text{ hPa}) \}^\top \), \( \mathbf{V}_{22} = \{ V_x(925 \text{ hPa}), V_y(925 \text{ hPa}) \}^\top \), \( \hat{\mathbf{\mu}}_\mathbf{V} = \{ E(\mathbf{V}_{11})^\top, E(\mathbf{V}_{22})^\top \}^\top = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4)^\top \) (in m/s), and \( \hat{\mathbf{\Sigma}}_\mathbf{V} = \begin{bmatrix} \text{cov}(\mathbf{V}_{11}, \mathbf{V}_{11}) & \text{cov}(\mathbf{V}_{11}, \mathbf{V}_{22}) \\ \text{cov}(\mathbf{V}_{22}, \mathbf{V}_{11}) & \text{cov}(\mathbf{V}_{22}, \mathbf{V}_{22}) \end{bmatrix} \) (in \( m^2/s^2 \)) such that \( \text{cov}(\mathbf{V}_{11}, \mathbf{V}_{11}) = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix} \), \( \text{cov}(\mathbf{V}_{11}, \mathbf{V}_{22}) = \begin{pmatrix} \Sigma_{1,3} & \Sigma_{1,4} \\ \Sigma_{2,3} & \Sigma_{2,4} \end{pmatrix} \), and \( \text{cov}(\mathbf{V}_{22}, \mathbf{V}_{22}) = \begin{pmatrix} \Sigma_{3,3} & \Sigma_{3,4} \\ \Sigma_{4,3} & \Sigma_{4,4} \end{pmatrix} \).

of \( \mathbf{V}_{11} \) and \( \mathbf{V}_{22} \) compared to those in M3 and M4. This may be due to the unaccounted dependence between \( \mathbf{V}_{11} \) and \( \mathbf{V}_{22} \) when using models M1 and M2. Although the parameter estimates in Figure 4.11 show that \( \hat{\mathbf{V}}_{11} \) and \( \hat{\mathbf{V}}_{22} \) are almost identical, it is still paramount
to consider and use a multiple advections model because from the real wind data, the wind vectors in 880 hPa and 925 hPa are not identical.

4.6 Conclusion

We successfully pursued the development of Lagrangian spatio-temporal cross-covariance functions with multiple advections. The proposed framework offers a suite of data-driven models which is more flexible and realistic. We also outlined an estimation procedure to get sensible estimates of all parameters. We showed through numerical experiments involving bivariate spatio-temporal random fields how failing to account for multiple advections produces poor predictions and how utilizing simpler univariate Lagrangian spatio-temporal models cannot capitalize on the purely spatial variable dependence. With wind recognized as the main driver of the transport behavior, the proposed models, along with other benchmark spatio-temporal models, were tested on PM concentrations over Saudi Arabia. Indeed, a model under our proposed class emerged as the best model and its estimate of the multivariate distribution of the advection velocities checks out with the prevailing behavior of wind in the spatio-temporal domain under consideration.
Chapter 5

Concluding Remarks and Other Avenues for Research

The field of geostatistics has an impressive array of multivariate nonstationary spatio-temporal covariance functions. Despite a saturation of spatio-temporal models in the field, there is only a handful of physically-meaningful models. This dissertation sought to expand the roster of models grounded on physical laws by furthering our grasp on geostatistical models specializing on transport phenomenon. Even though transport phenomenon is ubiquitous and the modeling paradigm enables purely spatial models to operate in a spatio-temporal context, Lagrangian spatio-temporal modeling has not been thoroughly investigated in the nonstationary and multivariate settings. Theories that support its usage on the two important areas aforementioned were lacking.

Building on the seminal work of Cox and Isham (1988), who laid down the first stochastic model embedding a parameter responsible for transport, we proposed three main theorems that provide more flexibility and features designed to cover a wider range of transport scenarios. Our results provide justifications for the inclusion of purely spatial covariance functions that are nonstationary and/or multivariate to the pool of models upon which we can apply the Lagrangian reference frame. Our work also sheds light on the limitations of some variants of Lagrangian spatio-temporal covariance functions, such as the negative effects of forcing a single advection model on multivariate data when every variable may have different transport behaviors and the lack of known closed forms of non-frozen models for distributions of the advection vector other than the Gaussian distribution.

Extensions to our work may revolve around Lagrangian spatio-temporal modeling for very large $n$ and $p$ via high performance implementations or stochastic partial differential
equations (SPDEs). In the following, we set the significance of these two areas of research.

5.1 High Performance Lagrangian Spatio-Temporal Modeling

Gaussian geostatistical modeling relies heavily on the operations done on the full covariance matrix. In the early stages of modeling, the covariance matrix needs to be created by evaluating the spatio-temporal cross-covariance function at \( n \) spatio-temporal locations for each of the \( p \) variables. Parameters then have to be estimated, with the cross-covariance function evaluated every time new sets of parameters are assumed. Further, the Gaussian likelihood requires the inverse and the determinant of the \( np \times np \) full covariance matrix. Lastly, the cokriging equations involve the inverse of the \( np \times np \) full covariance matrix and the covariance matrix formed by the observed and unobserved spatio-temporal locations. The computing power demanded for these operations grows as \( n \) and \( p \) grow. Furthermore, a one day forecast might take several days or weeks for large \( n \) and \( p \). Hence, a highly accurate model does not necessarily translate into a usable model when the dataset being analyzed is massive. This problem created the need for ultra fast or high performance routines for geostatistical computations which would allow geostatistical models to be utilized for large \( n \) and \( p \) datasets in reasonable execution time.

The open source software \( R \) is one of the preferred analytic environment and language for statistical computing owing to its extensive collection of packages for statistical analysis (\cite{RCoreTeam2019}). However, \( R \) was not built as a high performance language. Doing operations on very large matrices using standard functions in \( R \) is futile, but this problem is not without any solutions. In the previous decades, high-end computing resources such as multicore CPUs, accelerators, and coprocessors have been widely available and are changing the game in scientific computations. Utilizing these upgrades in computer performance can dramatically improve software implementations. Nevertheless, software implementations have not matched up to the advances in hardware until recently (\cite{Dongarraetal1998}). High performance computing was born out of the need to tap into these high-end computing resources
to boost software performance. Programming languages such as C, Fortran, and R can now utilize the underlying high performance systems using certain programming syntaxes.

Currently, there are several existing fast routines in R which speed up computations regularly used in geostatistics. For calculating the inverse of a covariance matrix, one can use the `spdinv` function in the library `Rfast` (Papadakis et al., 2019) or solve the inverse via the Cholesky decomposition of the covariance matrix, i.e., `chol` then `chol2inv`. These two approaches are a lot faster than the custom R function for matrix inverse, `solve`. Other fast routines rely on the covariance matrix being sparse by exploiting the properties of sparse covariance matrices. The list includes the `spam` library (Furrer, 2019) which provides Fortran routines for sparse matrices, Cholesky factorization, and other sparse matrix algebra operations and `SparseM` (Koenker and Ng, 2019) which features a compressed sparse row format as the primary storage mode. For parameter estimation, the function `optimParallel` (Gerber, 2019) in the library `optimParallel` is available and is the parallelized version of `optim`, the generic minimization function in R. Other parallel implementations in R are reviewed in Arora (2016).

ExaGeoStat is a software that utilizes state-of-the-art parallel architectures and is specialized for geostatistical computations (Abdulah et al., 2018a). ExaGeoStat unifies Chameleon (Agullo et al., 2019) and HiCMA (Akbudak et al., 2019), two libraries with high-performance solvers for dense and sparse linear algebra, StarPU, a task scheduler on multi-core architectures, and parallel systems such as Intel KNL, NVIDIA GPU, and distributed-memory systems. ExaGeoStat was tested on soil moisture data from the Mississippi River basin region, USA (Abdulah et al., 2018a) and sea surface temperature data from Agulhas and surrounding areas off the shore of South Africa (Abdulah et al., 2019). An R implementation of ExaGeoStat is also available in the package `ExaGeoStatR`. Through `ExaGeoStatR`, R users can perform geostatistical computations without needing a level of expertise in high performance computing.

Salvaña et al. (2021a) added large-scale multivariate purely spatial modeling capabilities
to ExaGeoStat. Their large-scale implementation illustrates how multivariate purely spatial models can be deployed on many current hardware environments such as shared-memory systems and distributed-memory systems. [Salvaña et al. (2021b)] presented the large-scale univariate spatio-temporal implementation in ExaGeoStat. Moreover, to target both the log-likelihood computations and optimization problems, they proposed a two-level parallelization technique. At the inner level, complex matrix operations required in MLE and kriging was relied on state-of-the-art dense linear algebra libraries and parallel runtime systems. At the outer level, a particle swarm optimization (PSO) algorithm was used to parallelize the optimization process. Combining these new capabilities of ExaGeoStat, multivariate Lagrangian spatio-temporal nonstationary modeling is possible. ExaGeoStat will play a major role in accelerating the speed in computing (1.9), (2.1), (3.1), and (4.2), analytically or numerically, and in deploying it for applications with very large $n$ and $p$.

5.2 Lagrangian Spatio-Temporal SPDE

Another alternative to purely spatial Gaussian random fields that relieves the computational bottlenecks of Cholesky factorization under large-scale modeling are the purely spatial Gaussian Markov random fields. In fact, a purely spatial Gaussian random field can be characterized using SPDEs with a purely spatial Gaussian Markov random field as its approximate solution (Lindgren et al. 2011). By virtue of its Markovian property, a purely spatial Gaussian Markov random field is easier to deal with, computationally, because it obtains a sparse precision matrix, the inverse of the covariance matrix. For instance, the purely spatial Gaussian random field with covariance function being the purely spatial isotropic Matérn, i.e.,

$$C(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (a\|h\|)^\nu K_\nu (a\|h\|),$$

(5.1)
where $K_\nu$ is the modified Bessel function of the second kind of order $\nu$, and $\sigma > 0$, $\nu > 0$, and $a > 0$ are the variance, smoothness, and scale parameters, is a solution to the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} \{ \tau Z(s) \} = \mathcal{W}(s),$$

where $\mathcal{W}(s)$ is a Gaussian purely spatial white noise process. Here $\Delta$ is the Laplacian defined as $\sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ and the parameters of (5.1) and (5.2) are related such that $a = \kappa$, $\nu = \alpha - d/2$, and $\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)} \frac{1}{4\pi^{d/2} \kappa^{d/2}}$. Using the Finite Element Method (FEM), an approximate solution to the SPDE may be derived. FEM involves dividing the spatial domain into a set of non-intersecting triangles leading to a triangulated mesh with $n'$ nodes and $n'$ basis functions. The basis function $\phi_k(r)$ are defined as piecewise linear functions on each triangle that is equal to 1 at vertex $k$, and equal to 0 at other vertices. Thus, the purely spatial random field $Z(s)$ can be represented via finite basis functions defined on the triangulated mesh, i.e.,

$$Z(s) = \sum_{k=1}^{n'} \phi_k(s) \lambda_k,$$

where $n'$ is the number of vertices of the triangulation and $\lambda_1, \ldots, \lambda_{n'}$ are zero-mean Gaussian-distributed weights. It should be interesting to explore the corresponding SPDE and Gaussian spatio-temporal Markov random field for the Lagrangian spatio-temporal Matérn stationary covariance function:

$$C(h, u) = \mathbb{E}_V \left\{ \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)} (a\|h - Vu\|)^\nu K_\nu (a\|h - Vu\|) \right\},$$

and its multivariate and multiple advections variants.
REFERENCES


Porcu, E., Bevilacqua, M., and Genton, M. G. (2016). Spatio-temporal covariance and cross-
covariance functions of the great circle distance on a sphere. *Journal of the American


covariance functions. *Stochastic Environmental Research and Risk Assessment*, 21(2):113–
122.

Porcu, E., Gregori, P., and Mateu, J. (2009). Archimedean spectral densities for nonstation-

1239.


R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Founda-
tion for Statistical Computing, Vienna, Austria.


Randles, C., Da Silva, A., Buchard, V., Colarco, P., Darmenov, A., Govindaraju, R.,
Smirnov, A., Holben, B., Ferrare, R., Hair, J., et al. (2017). The MERRA-2 aerosol reanalysis,

5(4):2265.

Ren, Q. and Banerjee, S. (2013). Hierarchical factor models for large spatially misaligned


Appendix

Proof of Theorem 1 Let $\lambda_l \in \mathbb{R}$. Then:

$$\sum_{l=1}^{n} \sum_{r=1}^{n} \lambda_l C(s_l, s_r; t_l, t_r) \lambda_r = \sum_{l=1}^{n} \sum_{r=1}^{n} \lambda_l \mathbb{E} \left\{ C^S(s_l - V t_l, s_r - V t_r) \right\} \lambda_r$$

$$= \mathbb{E} \left\{ \sum_{l=1}^{n} \sum_{r=1}^{n} \lambda_l C^S(\bar{s}_l, \bar{s}_r) \lambda_r \right\} \geq 0$$

for all $n \in \mathbb{Z}^+$. The last inequality follows from the fact that $C^S$ is a valid purely spatial nonstationary covariance function on $\mathbb{R}^d$. \hfill \Box

Proof of Theorem 2 Let $\lambda_l \in \mathbb{R}^p$. Then,

$$\sum_{l=1}^{n} \sum_{r=1}^{n} \lambda_l^T C(s_l, s_r; t_l, t_r) \lambda_r = \sum_{l=1}^{n} \sum_{r=1}^{n} \lambda_l^T \mathbb{E} \left\{ C^S(s_l - V t_l, s_r - V t_r) \right\} \lambda_r$$

$$= \mathbb{E} \left\{ \sum_{l=1}^{n} \sum_{r=1}^{n} \lambda_l^T C^S(\bar{s}_l, \bar{s}_r) \lambda_r \right\} \geq 0$$

for all $n \in \mathbb{Z}^+$. The last inequality follows from the assumption that $C^S$ is a valid purely spatial matrix-valued nonstationary covariance function on $\mathbb{R}^d$. \hfill \Box

Proof of Theorem 3 The validity is established by considering a purely spatial random field $Z(s) = [Z_1 \{ f_1(s) \}, \ldots, Z_p \{ f_p(s) \}]^\top$. \hfill \Box

Proof of Proposition 1 Given a multivariate purely spatial random field $\tilde{Z}(s)$, with second-order nonstationarity, define a multivariate deformed spatio-temporal random field $Z(s, t) = \left[ \tilde{Z}_1 \{ f_1^t(s - V t) \}, \ldots, \tilde{Z}_p \{ f_p^t(s - V t) \} \right]^\top$, where $f_i^t$ is a temporally varying spatial deformation, $i = 1, \ldots, p$. The covariance between variable $i$ taken at spatio-temporal
location \((s_1, t_1)\) and variable \(j\) taken at spatio-temporal location \((s_2, t_2)\) is

\[
\text{cov} \{ Z_i(s_1, t_1), Z_j(s_2, t_2) \} = \mathbb{E}_V \left( \text{cov} \left[ \tilde{Z}_i \{ f_i^{t_1} (s_1 - V t_1) \}, \tilde{Z}_j \{ f_j^{t_2} (s_2 - V t_2) \} \right] \right)
\]

\[
= \mathbb{E}_V \left[ \tilde{C}_{ij} \{ f_i^{t_1} (s_1 - V t_1) - f_j^{t_2} (s_2 - V t_2) \} \right].
\]

Proof of Theorem 4 Let \(\lambda_l \in \mathbb{R}^p, l = 1, \ldots, n\). Then:

\[
\sum_{l=1}^{n} \sum_{r=1}^{n} \lambda_l^\top C(s_l, s_r; t_l, t_r) \lambda = \sum_{l=1}^{n} \sum_{r=1}^{n} \lambda_l^\top \mathbb{E}_V \left( \left[ C_{ij}^S \{ (s_l - V_{ii} t_l) - (s_r - V_{jj} t_r) \} \right]_{i,j=1}^p \right) \lambda_r
\]

\[
= \mathbb{E}_V \left[ \sum_{l=1}^{n} \sum_{r=1}^{n} \lambda_l^\top \left( C_{ij}^S (s_l - s_r - V_{ii} t_l + V_{jj} t_r) \right)_{i,j=1}^p \lambda_r \right] \geq 0
\]

for all \(n \in \mathbb{Z}^+\) and \(\{(s_1, t_1), \ldots, (s_n, t_n)\} \in \mathbb{R}^d \times \mathbb{R}\). The last inequality follows from the assumption that \(C^S\) is a valid purely spatial matrix-valued stationary cross-covariance function with variable asymmetry on \(\mathbb{R}^d\); see Li and Zhang (2011), Genton and Kleiber (2015), and Qadir et al. (2021) for discussions on this class of cross-covariance functions.

Proof of Theorem 5 See derivations found in this website:

https://github.com/ladybug-hash/multiple-advections

Proof of Theorem 6 The validity is guaranteed by the dimension expansion approach in Bornn et al. (2012).

Proof of Theorem 7 The validity of (4.9) is established as it is the resulting spatio-temporal cross-covariance function of the process in (4.10).

Derivation of Equation (4.3)

\[
C_{ij} \left( h - V_{ii} t_l + V_{jj} t_r \right) = C_{ij} \left\{ h - V_{ii} \left( m + \frac{u}{2} \right) + V_{jj} \left( m - \frac{u}{2} \right) \right\}
\]

\[
= C_{ij} \left\{ h - V_{ij} u + (V_{jj} - V_{ii}) m \right\},
\]

where \(u = t_l - t_r, m = \frac{t_l + t_r}{2}, \) and \(\nabla_{ij} = \frac{V_{ii} + V_{jj}}{2}\).
Papers Published and Submitted


Brief Biography

Mary Lai Salvàna is a PhD student in the Spatio-Temporal Statistics & Data Science group at King Abdullah University of Science and Technology (KAUST). She received her BS and MS degrees in Applied Mathematics in 2015 and 2016 from Ateneo de Manila University, Philippines. Her research interest includes multivariate spatio-temporal statistics and high performance computing for large spatial and spatio-temporal datasets.