TECHNOMETRICS
2021, VOL. 00, NO. 0, 1–14
https://doi.org/10.1080/00401706.2021.1883481

Taylor & Francis
Taylor & Francis Group

Estimation of Spatial Deformation for Nonstationary Processes via Variogram Alignment

Ghulam A. Qadir[a], Ying Sun[a], and Sebastian Kurtek[b]

[a]CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; [b]Department of Statistics, The Ohio State University, Columbus, OH

ABSTRACT

In modeling spatial processes, a second-order stationarity assumption is often made. However, for spatial data observed on a vast domain, the covariance function often varies over space, leading to a heterogeneous spatial dependence structure, therefore requiring nonstationary modeling. Spatial deformation is one of the main methods for modeling nonstationary processes, assuming the nonstationary process has a stationary counterpart in the deformed space. The estimation of the deformation function poses severe challenges. Here, we introduce a novel approach for nonstationary geostatistical modeling, using space deformation, when a single realization of the spatial process is observed. Our method is based on aligning regional variograms, where warping variability of the distance from each subregion explains the spatial nonstationarity. We propose to use multi-dimensional scaling to map the warped distances to spatial locations. We assess the performance of our new method using multiple simulation studies. Additionally, we illustrate our methodology on precipitation data to estimate the heterogeneous spatial dependence and to perform spatial predictions.

ARTICLE HISTORY
Received September 2019
Accepted January 2021

KEYWORDS
Distance warping; Functional data registration; Nonstationarity; Regional variograms

1. Introduction

Spatial statistics methods are widely used in various disciplines such as meteorology, hydrology and earth science, to model environmental processes for a better understanding of the latent dependence structure, and for making predictions at unobserved locations. Statistical analysis of spatial processes generally involves a second-order stationarity assumption stating that, for a random process $\{X(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d, d \geq 1\}$, the mean is a constant, that is, $E\big(X(\mathbf{s})\big) = c$ for some $c \in \mathbb{R}$, and the covariance between any two locations depends only on the lag vector between those two locations, that is, $\text{cov}\big(X(\mathbf{s}), X(\mathbf{s}+\mathbf{h})\big) = \text{C}(\mathbf{h})$. Isotropic and anisotropic processes are two special cases of a second-order stationary process. The former implies that the covariance function depends only on the $\mathbb{L}^2$ norm of a lag vector, that is, $||\mathbf{h}||$, whereas the latter is a small generalization that incorporates both length and direction into the covariance function through a linear transformation of the lag vector as $||\mathbf{A}^{-\frac{1}{2}}\mathbf{h}||$, where $\mathbf{A}$ is a $d \times d$ positive definite matrix known as the anisotropy matrix. Modeling spatial processes by assuming a translation-invariant spatial dependence is a convenient, but unrealistic approach, especially when the spatial domain is large and statistical features of the process vary in space; in this case, such an assumption is a misspecification of the process.

In recent decades, considerable research has been directed toward developing methods to model nonstationary processes. Risser (2016) and Fouedjio (2017) extensively reviewed the existing literature on this topic and published a comprehensive summary of nonstationary modeling approaches for univariate geostatistical data. Higdon (1998) proposed a process-convolution approach with a spatially varying convolution kernel to model the nonstationary dependence structure. Further adaptation of this approach in Higdon, Swall, and Kern (1999), Paciorek and Schervish (2006), and Calder (2008) resulted in a covariance function with spatially varying parameters. Subsequently, Fouedjio, Desassis, and Rivoirard (2016) generalized the idea of the process-convolution model by introducing a convolution with a spatially varying random weighting function. The process-convolution model with spatially varying parameters is also explored for efficient model estimation through local-polynomial fitting by Li and Sun (2019). Recent work by Nychka et al. (2018) introduces a computationally efficient method to model convolution type nonstationarity for large spatial datasets. Fuentes (2002) constructed a nonstationary process through convolution of locally stationary processes, which was later used by Reich et al. (2011) to introduce a novel spatio-temporal covariance function, addressing nonstationarity by using covariate information. Sampson and Guttorp (1992) published one of the first studies on nonstationary spatial modeling by pioneering the method of spatial deformation; their work served as the fundamental idea for further studies by Damian, Sampson, and Guttorp (2001), Schmidt and O'Hagan (2003), Iovleff and Perrin (2004), Anderes and Stein (2008), Anderes and Chatterjee (2009), Fouedjio, Desassis, and Romary (2015), and Kleiber (2016). Some other popular nonstationary spatial modeling approaches

CONTACT Ying Sun ying.sun@kaust.edu.sa CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/TECH.

include basis function methods (Nychka and Saltzman 1998; Nychka, Wikle, and Royle 2002; Stephenson et al. 2005), stochastic partial differential equations (SPDEs) approaches (Lindgren, Rue, and Lindström (2011); Fuglstad et al. (2015)), and moving window methods (Haas 1990a, 1990b; Lloyd and Atkinson 2000, 2002).

A prominent approach to model nonstationarity, using the method of spatial deformation introduced by Sampson and Guttorp (1992), involves mapping locations in a geographic space ($\mathcal{G}$) to transformed locations in a deformed space ($\mathcal{D}$), where the process is expected to be stationary and isotropic. This original concept provides an invaluable direction for modeling nonstationarity, but it fundamentally requires multiple independent realizations of the spatial process which, in practice, are often not observed. In addition, a major drawback of their method is the folding of space. This occurs if the estimated deformation function that maps geographical locations to the deformed space is not injective. In the presence of spatial data replicates, Damian, Sampson, and Guttorp (2001) and Schmidt and O'Hagan (2003) addressed the folding of space issue in a Bayesian framework, whereas Bornn, Shaddick, and Zidek (2012) addressed it in a frequentist framework by retaining the original locations of the geographic space and adding extra dimensions to embed a nonstationary field of lower dimensions to a higher dimensional stationary field. The problem of estimation of a spatial deformation by using only one realization of the spatial process was first addressed by Anderes and Stein (2008) and Anderes and Chatterjee (2009). However, their proposed quasi-conformal mappings-based methodology requires very dense spatial data and its application on a real dataset has not yet been illustrated. Fouedjio, Desassis, and Romary (2015) developed a method for estimating the deformation function, using a single realization of the spatial field that avoids the problem of folding of space, but their method relies heavily on many tuning parameters and subjective selection of anchor points.

Here, we propose a metric-based nonparametric method for estimating a spatial deformation by applying the functional data registration method, proposed by Srivastava et al. (2011), to spatial variograms. Our method avoids the strong assumption of replicates of spatial data and allows us to estimate the deformed space in higher dimensions, consequently avoiding the problem of folding of space. The key concept underlying the proposed method is based on aligning regional variograms belonging to different subregions of the entire spatial domain to estimate the warping variability in inter-point distances. The principal tools used in the proposed method are: (1) kernel smoothing, (2) classical (metric) multi-dimensional scaling (CMDS) (Torgerson 1958; Mardia, Kent, and Bibby 1979), and (3) a functional data registration algorithm (Srivastava et al. 2011); we use these tools to obtain a one-to-one mapping of locations in a geographic space ($\mathcal{G}$) to transformed locations in a deformed space ($\mathcal{D}$). Our method does not require the use of thin-plate splines (a key component in the methods of Sampson and Guttorp (1992), Bornn, Shaddick, and Zidek (2012), and Fouedjio, Desassis, and Romary (2015)) to estimate the deformed coordinates of unobserved locations. Both observed and unobserved locations can be mapped to their corresponding deformed coordinates in a single step, and hence can be used directly to obtain kriging estimates. Besides the estimation of a heterogeneous spatial dependence structure for spatial predictions, the proposed method also serves as a useful exploratory tool to visualize the degree of nonstationarity in spatial data. We illustrate the proposed method with a simulated example. We also apply it to precipitation data from the state of Colorado in the United States.

Section 2 describes the proposed spatial deformation estimation procedure, including a brief discussion of the functional data registration algorithm used in the proposed method. In Section 3, we illustrate our methodology on a simulated example. Section 4 discusses an application to the precipitation dataset, followed by a discussion in Section 5 highlighting the main contributions of this work and some directions for the future.

2. Methodology

According to Bornn, Shaddick, and Zidek (2012), "Environmental systems might exhibit behavior that looks locally stationary, yet when considered over large and heterogeneous domain they very often exhibit nonstationarity." Our method is motivated by such locally stationary behavior of environmental processes that can be well approximated by piecewise or regionwise stationary models. It involves a partitioning of the entire spatial domain into smaller subregions such that the process shows homogeneous spatial dependence within each subregion, but may exhibit heterogeneous spatial dependence across subregions. Such a partitioning approach commonly appears in the nonstationary spatial modeling literature (Fuentes and Smith 2001; Paciorek and Schervish 2006; Heaton, Christensen, and Terres 2017). One common way to quantify homogeneous spatial dependence is by using a stationary variogram that measures the variability in observations, depending on the distance between them. Therefore, heterogeneous spatial dependence across subregions implies that the regional variogram, as a function of distance, may vary across subregions. We treat these regional variograms as functional data. However, unlike the traditional functional data registration problem where functional data are directly observed, the regional variograms need to be estimated from spatial observations prior to alignment.

In this section, we give a brief introduction to the functional data registration algorithm (Section 2.1), followed by a detailed discussion of the proposed method for estimating spatial deformations. The estimation procedure can be broadly classified into two steps: (1) an "alignment step," and (2) a "construction step." The alignment step includes partitioning of the spatial domain into smaller subregions from which we estimate regional variograms and align them to estimate regional distance warping functions. The estimated regional distance warping functions are then used to warp the distance matrix of the entire spatial domain. The warped distance matrix is then supplied to CMDS in the construction step to estimate the deformed space. The details of Steps (1) and (2) are described in Section 2.2.1 and Section 2.2.2, respectively.

2.1. Functional Data Registration

We first introduce the functional data registration algorithm developed by Srivastava et al. (2011), Kurtek, Srivastava, and Wu (2011), and Srivastava and Klassen (2016) that we use in our work for variogram alignment. In those works, they defined the notion of "elastic functions," that is, functions with warping variability, and proposed a framework for separation of y-axis (magnitude) and x-axis (timing) variabilities in these elastic functions by warping the x-axis. They considered the following representation:

$$f_i = c_i(g \circ \phi_i) + e_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where "$\circ$" denotes function composition operator, f_i denote the observed (absolutely continuous) functions, $c_i \in \mathbb{R}^+$ are individual scalings, $e_i \in \mathbb{R}$ are vertical translations, and g is an underlying template. Each function f_i represents an observation of the template g under a random warping ϕ_i, and a random scaling and translation, c_i and e_i, respectively. For a given sample of functions $\{f_i\}$, the main task is to obtain a consistent estimator of the template g; this additionally results in estimates of the optimal warping functions ϕ_i^* (x-axis variation), and the set of optimally registered functions $f_i^* = f_i \circ \phi_i^{*-1}$ (y-axis variation). Standard solutions to the warping problem based on the $\mathbb{L}^2$ Hilbert space framework are known to have theoretical and practical issues, such as the lack of isometry of the $\mathbb{L}^2$ metric under the action of the warping group. This, in turn, results in degenerate warping solutions and the so-called pinching effect (Marron et al. 2015). To overcome these problems, Srivastava et al. (2011) proposed an approach based on the extended Fisher–Rao metric and the *square-root velocity function* (SRVF) representation of observed functional data. The SRVF allows for efficient computation of the optimal warping functions via dynamic programming (Robinson 2012). Their registration algorithm (available in the R-package `fdasrvf` (Tucker 2020)) has been extensively studied to demonstrate theoretical guarantees for the consistent estimation of the unknown template g (Kurtek, Srivastava, and Wu 2011; Lahiri, Robinson, and Klassen 2015; Srivastava and Klassen 2016). Furthermore, its practical efficiency has been explored in various applied contexts (Kurtek et al. 2013; Tucker, Wu, and Srivastava 2013; Samir et al. 2016; Meng et al. 2017). The registration method has also been adapted by Guan et al. (2019) to an applied spatial problem of model calibration; however, its application in spatial statistics is still nascent. For brevity, we skip the complete discussion of the registration algorithm, and instead refer the interested readers to Srivastava et al. (2011) and Srivastava and Klassen (2016) for details. In this work, we adapt this algorithm to the spatial setting for the registration of regional variograms.

2.2. Estimation of Spatial Deformation

Let $\{X(\mathbf{s}) : \mathbf{s} \in \mathcal{G} \subset \mathbb{R}^{d_\mathcal{G}}\}$ be a zero-mean nonstationary random field defined on the geographic space $\mathcal{G}$ of dimensionality $d_\mathcal{G}$, and $\{Y(\mathbf{u}) : \mathbf{u} \in \mathcal{D} \subset \mathbb{R}^{d_\mathcal{D}}\}$ be the corresponding zero-mean stationary random field defined on the deformed space $\mathcal{D}$ of dimensionality $d_\mathcal{D}$. Here, $d_\mathcal{D}$ is not necessarily equal to $d_\mathcal{G}$, and in fact, $\{d_\mathcal{D} = d_\mathcal{G} + \psi, \psi \in \{0, 1, 2, \ldots\}\}$, that is, the domain of the stationary process Y can have a higher dimensionality relative to the nonstationary process X. The primary objective is to estimate a deformation $\theta : \mathcal{G} \to \mathcal{D}$ such that $\{X(\mathbf{s}) = Y(\theta(\mathbf{s})), \mathbf{s} \in \mathcal{G}\}$ and $\{Y(\mathbf{u}) = X(\theta^{-1}(\mathbf{u})), \mathbf{u} \in \mathcal{D}\}$. This allows us to model the nonstationary covariance of X as

$$\text{cov}^{\text{NS}}(\mathbf{s}, \mathbf{s}') = \text{C}_\mathcal{D}(||\theta(\mathbf{s}) - \theta(\mathbf{s}')||), (\mathbf{s}, \mathbf{s}') \in \mathcal{G} \times \mathcal{G}, \tag{2}$$

where $\text{C}_\mathcal{D}(|| \cdot ||)$ represents any valid stationary and isotropic covariance function that depends only on the $\mathbb{L}^2$ distance between points in the deformed space. The corresponding nonstationary semivariogram (simply called variogram hereafter) of X is then given by $\gamma^{\text{NS}}(\mathbf{s}, \mathbf{s}') = \gamma_\mathcal{D}(||\theta(\mathbf{s}) - \theta(\mathbf{s}')||)$, where $\gamma_\mathcal{D}(||\cdot||)$ is a valid stationary and isotropic variogram model which is related to $\text{C}_\mathcal{D}(|| \cdot ||)$ via $\gamma_\mathcal{D}(||\mathbf{h}||) = \text{C}_\mathcal{D}(||\mathbf{0}||) - \text{C}_\mathcal{D}(||\mathbf{h}||)$.

Our method is based on a mild assumption of regional stationarity of the process $\{X(\mathbf{s}) : \mathbf{s} \in \mathcal{G} \subset \mathbb{R}^{d_\mathcal{G}}\}$, which implies that $\mathcal{G}$ can be partitioned into k mutually exclusive subregions $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_k$, $(\mathcal{G} = \cup_{i=1}^k \mathcal{G}_i)$, such that for each $i = 1, 2, \ldots, k$, $\{X(\mathbf{s}) : \mathbf{s} \in \mathcal{G}_i\}$ is a stationary process with spatial dependence described by the stationary and isotropic variogram model $\gamma_i(||\mathbf{h}||)$. The variogram models may differ from each other through various features such as smoothness, autocorrelation range, variance (sill) and nugget, making the process $X(\mathbf{s})$ nonstationary over the domain $\mathcal{G}$. For each subregion $\mathcal{G}_i$, we have a corresponding subregion $\mathcal{D}_i$ in the deformed space $\mathcal{D}$ such that $\mathcal{D} = \cup_{i=1}^k \mathcal{D}_i$.

Considering the regional variograms $\gamma_i(||\mathbf{h}||)$ as elastic functions results in the following representation (adaptation of Equation (1)):

$$\gamma_i(||\mathbf{h}||) = c_i(\gamma \circ \phi_i)(||\mathbf{h}||) + e_i, \quad i = 1, 2, \ldots, k. \tag{3}$$

In the spatial context, Equation (3) leads to the following interpretation: each regional variogram is an observation from the global stationary variogram model γ, under a regional distance warping function ϕ_i, with a scaling $c_i \in \mathbb{R}^+$ and a vertical translation $e_i \in R^+ \cup \{0\}$ (note that e_i is nonnegative because variogram values are always nonnegative). For instance, if we assume that the features of the global variogram model γ such as nugget and variance are 0 and 1, respectively, then c_i and e_i can be interpreted as the variance and nugget for the regional variogram γ_i. The application of functional data registration to Equation (3) sets γ to be the sample mean of γ_i, $i = 1, \ldots, k$, and allows us to estimate the k regional distance warping functions that are of paramount importance in estimating the deformation θ. Specifically, they inform us about the interpoint distances in different subregions $\mathcal{D}_i$, $i = 1, 2, \ldots, k$, in the deformed space. Consequently, θ can be defined locally for the ith subregion ($i = 1, 2, \ldots, k$) as $\theta : \mathcal{G}_i \to \mathcal{D}_i$, and the following condition drives its estimation.

Condition 1: For any two arbitrary locations $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{G}_i$, the distance between their corresponding locations in the deformed space is given by warping the distance between them in the geographic space with a warping function ϕ_i, that is, $||\theta(\mathbf{s}_1) - \theta(\mathbf{s}_2)|| = \phi_i(||\mathbf{s}_1 - \mathbf{s}_2||)$.

Following the interpretation of Equation (3) and imposing Condition 1 in the estimation of θ implies that the variogram models describing the spatial dependence for the processes $\{Y(\mathbf{u}) = X(\theta^{-1}(\mathbf{u})), \mathbf{u} \in \mathcal{D}_i, \ i = 1, 2, \ldots, k.\}$ share the same features, such as smoothness and autocorrelation range;

this indicates that the nonstationarity in smoothness and autocorrelation range can be addressed by variogram registration. However, the processes might have varying regional variances and nuggets. The functional data registration algorithm used in our method is invariant to scalings and vertical translations, and therefore cannot deal with the nonstationarity in those components. More specifically, the proposed method addresses the nonstationarity only in the correlation function to introduce nonstationarity in the covariance function.

For a complete specification of the deformation θ, we need to define it globally, that is, $\theta : \mathcal{G} \to \mathcal{D}$, and therefore, a global distance warping function (GDWF) ϕ, which governs the inter-point distances in the deformed space $\mathcal{D}$, is required: $||\theta(\mathbf{s}) - \theta(\mathbf{s}')|| = \phi(\mathbf{s}, \mathbf{s}')$, for all $(\mathbf{s}, \mathbf{s}') \in \mathcal{G} \times \mathcal{G}$; this GDWF should also be consistent with Condition 1. We propose to define a GDWF $\phi : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^+ \cup \{0\}$ as a weighted linear combination of the regional distance warping functions as follows:

$$\phi(\mathbf{s}, \mathbf{s}') = \sum_{\mathcal{G}_i \in \mathcal{L}(\mathbf{s},\mathbf{s}')} \mathcal{W}_i(\mathbf{s}, \mathbf{s}')\phi_i(||\mathbf{s} - \mathbf{s}'||), \ (\mathbf{s}, \mathbf{s}') \in \mathcal{G} \times \mathcal{G}, \quad (4)$$

where $\mathcal{L}(\mathbf{s}, \mathbf{s}')$ is the set of subregions $\mathcal{G}_i$ such that the line segment joining the locations $\mathbf{s}$ and $\mathbf{s}'$ passes through all of the subregions in this set, and $\mathcal{W}_i(\mathbf{s}, \mathbf{s}')$ are the location-dependent weights for the ith regional distance warping function. We define the weights as $\mathcal{W}_i(\mathbf{s}, \mathbf{s}') = \frac{\mathcal{P}(i,\mathbf{s},\mathbf{s}')}{||\mathbf{s}-\mathbf{s}'||}$, where $\mathcal{P}(i, \mathbf{s}, \mathbf{s}')$ is the length of the line segment joining $\mathbf{s}$ and $\mathbf{s}'$ that lies in the subregion $\mathcal{G}_i$. The defined $\phi(\mathbf{s}, \mathbf{s}')$ is a semimetric, that is, $\phi(\mathbf{s}, \mathbf{s}') = 0 \Leftrightarrow \mathbf{s} = \mathbf{s}'$ and $\phi(\mathbf{s}, \mathbf{s}') = \phi(\mathbf{s}', \mathbf{s})$. This special choice of weights used in Equation (4) imparts robustness to our method under different subdivisions of the spatial domain (see Supplementary Material Section S4), and are specifically chosen to satisfy the following two properties that are crucial to our approach; the proofs are included in Supplementary Material Section S1.

Property 1. The GDWF $\phi : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^+ \cup \{0\}$ is consistent with Condition 1, that is, $\phi(\mathbf{s}, \mathbf{s'}) = \phi_i(||\mathbf{s} - \mathbf{s}'||)$ for all $(\mathbf{s}, \mathbf{s}') \in \mathcal{G}_i \times \mathcal{G}_i$, $i = 1, 2, \ldots, k$.

Property 2. If the process $\{X(\mathbf{s}) : \mathbf{s} \in \mathcal{G} \subset \mathbb{R}^{d_\mathcal{G}}\}$ is second-order stationary, then $\phi(\mathbf{s}, \mathbf{s}') = ||\mathbf{s} - \mathbf{s}'||$, implying that the geographic and deformed spaces are identical, up to a rotation and/or translation.

Due to the GDWF, the deformed space $\mathcal{D}$ is now known through the inter-point distances between different locations. Therefore, we propose to map these warped distances to deformed coordinates $(\theta(\mathbf{s}), \mathbf{s} \in \mathcal{G})$ using the CMDS algorithm (Torgerson 1958; Mardia, Kent, and Bibby 1979) (see Supplementary Material Section S5 for a sketch of the CMDS algorithm). For a given distance matrix, the application of CMDS seeks to find coordinates in a space of a specified dimensionality, such that the associated distance matrix is as close as possible to the given distance matrix (Ji and Zha 2004; Birchfield and Subramanya 2005). The distance matrix for the deformed space with n locations denoted by $\Delta_{(n\times n)} = \{\phi(\mathbf{s}_i, \mathbf{s}_j)\}_{i,j=1}^n$ is supplied to the CMDS algorithm for a given dimension $d_\mathcal{D} = d_\mathcal{G} + \psi$ to estimate the deformation $\hat{\theta}$. The estimation procedure is described in more detail in Sections 2.2.1 and 2.2.2.

2.2.1. Variogram Estimation and Registration

As already outlined in Section 2.2, our method is based on the regionwise stationary behavior of the spatial process on a vast domain, and hence requires the identification of homogeneous subregions. Informative covariates, or prior knowledge of the underlying physical process, can be instrumental in making this decision. In many real applications, fixed geographical factors such as terrain, topography and landform of the spatial domain of interest affect the spatial dependence of the process, and nonstationarity comes into play if there exists consequential heterogeneity in those factors. For example, if the spatial domain consists of both land and ocean, the temperature field exhibits nonstationarity with lower range of spatial dependence over the land than on the ocean (Genton and Kleiber 2015). In such cases, the division of the spatial domain into homogeneous subregions is straightforward, and is driven by the heterogeneity of the fixed geographical factors. Often, however, the nonstationary behavior of the process cannot be simply attributed to fixed geographical factors and there are other unknown factors causing the nonstationarity; in such cases, exploratory analysis along with expert information is needed for a proper division of the spatial domain into homogeneous subregions. To this end, one possible exploratory analysis is to locally fit a stationary variogram model to the entire spatial domain in a moving window fashion (Haas 1990a, 1990b; Lloyd and Atkinson 2000, 2002). Then, one can visualize the heat map of the locally estimated variogram parameters over the entire spatial domain, where the location of parameters corresponds to the center of their respective windows. These heat maps can guide the choice of the number and shape of the subregions on the basis of visual inspection of heterogeneity of parameter estimates on the spatial domain.

The variogram representation in Equation (3) requires the true regional variogram models $\{\gamma_i(||\mathbf{h}||), \ i = 1, 2, \ldots, k\}$ that are often unknown in practical situations. Therefore, we fit a valid stationary and isotropic variogram model for each subregion, and use the estimated variogram models $\{\hat{\gamma}_i(||\mathbf{h}||), \ i = 1, 2, \ldots, k\}$ to redefine Equation (3) as $\hat{\gamma}_i(||\mathbf{h}||) = c_i(\gamma \circ \phi_i)(||\mathbf{h}||) + e_i, \ i = 1, 2, \ldots, k$.

The choice of the number of subregions k is critical, as it controls the trade-off between the flexibility of the model and efficiency of our method. Higher values of k allow us to introduce a higher degree of nonstationarity, but potentially lead to inaccurate parameter estimates of the regional variogram models, due to a reduced number of observations per subregion. Similarly, lower values of k lead to better estimates of the regional variograms, but render a lower degree of nonstationarity in the model. Depending on the size of the data, an appropriate value of k should be chosen to maintain a balance between flexibility and estimation accuracy.

Once we have determined the appropriate value of k, the next step is to divide the geographic space $\mathcal{G}$ into k well-defined subregions $\mathcal{G}_1, \ldots, \mathcal{G}_k$, and to fit a valid stationary and isotropic variogram model for each subregion. We use the Matérn variogram model (Matérn 1986; Guttorp and Gneiting 2006) which is described by three parameters: variance ($\sigma^2 > 0$), spatial range ($a > 0$) and smoothness ($\nu > 0$). To represent the regional variograms as functions for registration, we evaluate the fitted variograms at a sequence of equally spaced points in

the interval $[0, ||\mathbf{h}_t||]$, where $||\mathbf{h}_t||$ is the distance at which all of the estimated variograms become numerically constant. We then apply the functional data registration algorithm to estimate the k regional distance warping functions $\hat{\phi}_i^{fdr}$, $i = 1, 2, \ldots, k$, which are defined over the domain $[0, ||\mathbf{h}_t||]$. To define the regional distance warping functions over the domain $[0, \infty)$, we augment identity warping to $\hat{\phi}_i^{fdr}$, $i = 1, 2, \ldots, k$, for any distance $||\mathbf{h}|| > ||\mathbf{h}_t||$ so that $\hat{\phi}_i$, $i = 1, 2, \ldots, k$, is now given by

$$\hat{\phi}_i(||\mathbf{h}||) = \begin{cases} \hat{\phi}_i^{fdr}(||\mathbf{h}||), & ||\mathbf{h}|| \leq ||\mathbf{h}_t|| \\ ||\mathbf{h}||, & ||\mathbf{h}|| > ||\mathbf{h}_t||. \end{cases} \tag{5}$$

Identity warping for large distances ensures that they remain unchanged, as beyond those distances all subregions exhibit spatial homogeneity (constant and identical variograms).

The regional distance warping functions serve as a valuable exploratory tool to visualize the degree of nonstationarity in the spatial data. Since the 45° line represents the identity warping, a larger deviation of regional distance warping functions from the identity warping indicates a higher degree of nonstationarity across subregions. Additionally, the warping functions tell us about the amount of stretching and compression required for different subregions to achieve stationarity in the deformed space. Note that the estimated regional distance warping functions represent the warped predetermined distances when evaluating the regional variograms. Thus, to be able to warp any distance in the interval $[0, ||\mathbf{h}_t||]$, we use kernel smoothing on the warped distances using a Gaussian kernel with a fairly low bandwidth. From the estimated regional distance warping functions in Equation (5), we estimate GDWF using Equation (4): $\hat{\phi}(\mathbf{s}, \mathbf{s}') = \sum_{\mathcal{G}_i \in \mathcal{L}(\mathbf{s}, \mathbf{s}')} \mathcal{W}_i(\mathbf{s}, \mathbf{s}') \hat{\phi}_i(||\mathbf{s} - \mathbf{s}'||)$. The GDWF is used to obtain the distance matrix for the deformed space $\mathcal{D}$ and to estimate θ.

2.2.2. Estimation of the Coordinates in Deformed Space

The GDWF is defined for any arbitrary pair of locations in the geographic space $\mathcal{G}$, that is, for any observed or unobserved pair of locations. Therefore, we can compute their corresponding pairwise distance in the deformed space. Let $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n \in \mathcal{G}$ be n observed locations, and $\mathbf{s}_{n+1}, \mathbf{s}_{n+2}, \ldots, \mathbf{s}_{n+m} \in \mathcal{G}$ be m unobserved locations. The aim is to estimate θ such that the approximation $||\hat{\theta}(\mathbf{s}_i) - \hat{\theta}(\mathbf{s}_j)|| \approx \hat{\phi}(\mathbf{s}_i, \mathbf{s}_j)$ is true for all $i, j = 1, 2, \ldots, n + m$.

To achieve this goal, we compute the transformed distance matrix $\Delta_{(n+m)\times(n+m)} = \{\hat{\phi}(\mathbf{s}_i, \mathbf{s}_j)\}_{i,j=1}^{n+m}$ and apply CMDS to $\Delta_{(n+m)\times(n+m)}$ to estimate $\hat{\theta}$ for a space of dimensionality $d_{\mathcal{D}}$. As we increase $d_{\mathcal{D}}$, the approximation improves, but an exact distance match is not guaranteed, even for a large $d_{\mathcal{D}}$. The optimal choice of $d_{\mathcal{D}} = d_{\mathcal{G}} + \psi$ is based on the accuracy of the CMDS approximation. Therefore, we estimate the deformed space for various values of $\psi \in \{0, 1, \ldots, \psi_U\}$, $\psi_U \in \mathbb{Z}_+$, and choose the one for which the difference between the transformed distance matrix Δ and the distance matrix of the corresponding estimated deformed space is minimized based on some measure of difference, for example, root mean squared error (RMSE) or normalized mean squared error (NMSE). While one can consider any measure of difference, in our implementation we have used NMSE for which values close to one indicate small differences. The value of ψ can be increased to $\psi + 1$ in the case of co-located deformed coordinates to tackle the space-folding problem.

The following is an algorithmic description that summarizes the proposed approach:

Step 1: Divide the spatial domain $\mathcal{G}$ into k homogeneous subregions and estimate k regional variogram models $\gamma_i(||\mathbf{h}||)$, $i = 1, \ldots, k$, corresponding to the k subregions (Section 2.2.1).

Step 2: Evaluate the estimated regional variograms $\hat{\gamma}_i(||\mathbf{h}||)$, $i = 1, \ldots, k$, at a sequence of equally spaced points in the interval $[0, ||\mathbf{h}_t||]$ and register them to obtain regional distance warping functions $\hat{\phi}_i(||\mathbf{h}||)$, $i = 1, \ldots, k$, as defined in (5).

Step 3: Apply Gaussian kernel smoothing with a fairly low bandwidth to each $\hat{\phi}_i(||\mathbf{h}||)$, such that $\hat{\phi}_i(||\mathbf{h}||)$ can be evaluated at any arbitrary $\mathbf{h}$.

Step 4: Using the smoothed $\hat{\phi}_i(||\mathbf{h}||)$, $i = 1, \ldots, k$, obtain the GDWF $\phi(\mathbf{s}, \mathbf{s})$ as defined in (4) (Section 2.2.1).

Step 5: Use the GDWF to compute the transformed distance matrix Δ. Apply CMDS to Δ for some $\psi \in \mathbb{Z}_+ \cup \{0\}$ to estimate the deformation θ (Section 2.2.2).

The execution of the proposed approach requires the user to tune the following components: (1) k (Step 1), (2) partitions $\{\mathcal{G}_1, \ldots, \mathcal{G}_k\}$ (Step 1), (3) equally spaced interval $[0, ||\mathbf{h}_t||]$ (Step 2), (4) kernel bandwidth (Step 3), and (5) ψ (Step 5). Among all of these components, k and the partitions $\{\mathcal{G}_1, \ldots, \mathcal{G}_k\}$ are the most critical and some guidelines for their specification are given in Section 2.2.1. The choice of the equally spaced interval $[0, ||\mathbf{h}_t||]$ and kernel bandwidth are less critical, because the estimated deformation function is much less sensitive to their specification. The value of ψ is also important as it determines the dimensionality of the estimated deformed space; a discussion on choosing an optimal value of ψ is provided in Section 2.2.2.

3. Simulation Study

To explore the performance of the proposed method, we conduct a simulation study in which we apply our method to a two dimensional Gaussian process, which has a locally stationary spatial dependence. Specifically, we consider a zero-mean Gaussian process X over a spatial domain $\mathcal{G} = [0, 2]^2$, with a spatial dependence described by the following nonstationary Matérn covariance function (Paciorek and Schervish 2006):

$$C^{NS}(\mathbf{s}_i, \mathbf{s}_j : \tilde{\eta}) = \sigma(\mathbf{s}_i)\sigma(\mathbf{s}_j) \frac{|\Sigma(\mathbf{s}_i)|^{1/4} |\Sigma(\mathbf{s}_j)|^{1/4}}{2^{\nu-1}\Gamma(\nu)} \left| \frac{\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j)}{2} \right|^{-1/2} \left(2\sqrt{\nu Q_{ij}}\right)^{\nu} K_\nu \left(2\sqrt{\nu Q_{ij}}\right), \tag{6}$$

where $\tilde{\eta}$ represents the vector of parameters, $\sigma(\mathbf{s})$ is a location-dependent standard deviation, ν is the smoothness parameter, $Q_{ij} = (\mathbf{s}_i - \mathbf{s}_j)^T [\{\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j)\}/2]^{-1} (\mathbf{s}_i - \mathbf{s}_j)$ is the Mahalanobis distance between a pair of locations $\mathbf{s}_i = (x_i, y_i)^T$ and $\mathbf{s}_j = (x_j, y_j)^T$, K_ν is a modified Bessel function of second order, and $\Sigma(\mathbf{s})$ is a spatially varying kernel matrix that supervises the range and direction of spatial dependence.

We simulate 100 realizations of X at a 70×70 regularly spaced grid points on $\mathcal{G}$, with constant smoothness $\nu = 0.8$ and standard deviation $\{\sigma(\mathbf{s}) = \sqrt{5}, \mathbf{s} \in \mathcal{G}\}$. However,

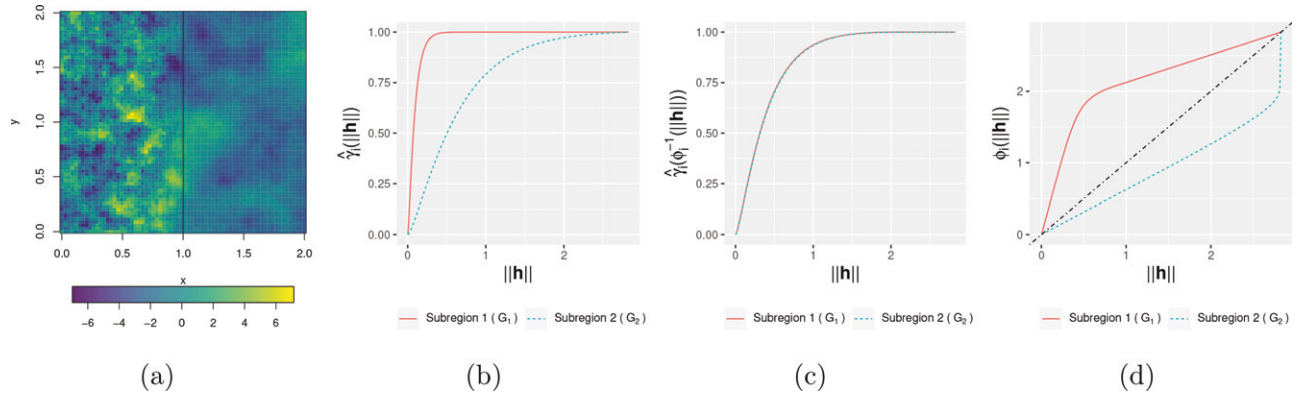

Figure 1. (a) A realization of the zero-mean locally stationary Gaussian process, with a solid black line indicating the partitioning of the geographic space. (b) Estimated standardized regional variograms. (c) Registered variograms. (d) Regional distance warping functions.

to introduce locally stationary spatial dependence, we specify the kernel matrix to be a spatially varying diagonal matrix $\Sigma(\mathbf{s}) = a(\mathbf{s})^2\mathbb{I}_{(2\times 2)}$, $a(\mathbf{s})^2 = \sum_{i=1}^{4} a_i^2 \exp(-\frac{||\mathbf{s}-\mathbf{s}_i^*||^2}{\lambda}) / \sum_{i=1}^{4} \exp(-\frac{||\mathbf{s}-\mathbf{s}_i^*||^2}{\lambda})$, where $a_1 = a_3 = 0.1$, $a_2 = a_4 = 0.9$, $\mathbf{s}_1^* = (0.5, 1.5)^{\mathrm{T}}, \mathbf{s}_2^* = (1.5, 1.5)^{\mathrm{T}}, \mathbf{s}_3^* = (0.5, 0.5)^{\mathrm{T}}, \mathbf{s}_4^* = (1.5, 0.5)^{\mathrm{T}}$, and $\lambda = 0.1$. Here, $a(\mathbf{s})$ represents the spatial range parameter at location $\mathbf{s}$ and λ controls the degree of smoothness with which $a(\mathbf{s})$ varies away from locations $\mathbf{s}_1^*, \ldots, \mathbf{s}_4^*$. The parameter λ leads to the departure from a locally stationary behavior to continuously varying nonstationarity as we increase its value within the interval (0,1]. Therefore, the specified $a(\mathbf{s})$ with a small value of $\lambda = 0.1$, as shown in Figure 4(a), allows us to simulate a locally stationary process with nonstationarity only in the spatial range.

For each of the 100 simulated realizations, we randomly split the simulated data into training and validation sets of 1200 and 3700 observations, respectively, and divide $\mathcal{G}$ into two subregions $\mathcal{G}_1 = [0,1] \times [0,2]$ and $\mathcal{G}_2 = (1,2] \times [0,2]$. We then proceed to fit the isotropic Matérn variogram model using the training set for both subregions via maximum likelihood estimation (MLE), and register the two estimated regional variograms. Results of the registration step for one realization are shown in Figure 1. Figure 1(a) shows the simulated realization in the geographic space, with a solid black line depicting the chosen partitioning. The estimated regional variograms, standardized using their respective regional variances, shown in Figure 1(b), exhibit varying spatial range for the two subregions, and their registration eliminates this variability (Figure 1(c)). The estimated regional distance warping functions are shown in Figure 1(d). The extent of nonstationarity in the simulated data can be assessed visually by looking at this figure, where the large deviation of both warping functions from the identity corresponds to a high degree of nonstationarity. It also suggests that stretching in $\mathcal{G}_1$ and compression in $\mathcal{G}_2$ are required to achieve stationarity.

We then estimate the deformation θ for each of the 100 realizations with their corresponding optimal choice of ψ as discussed in Section 2.2.2. Figure 2 shows a comparison of the geographic space and the estimated deformed space (in the first three dimensions of maximum variation) for the same realization as in Figure 1. The highly correlated observations corresponding to $\mathcal{G}_1$ are placed at higher inter-point distances in the deformed space, leading to a higher spatial range relative to $\mathcal{G}_1$. On the other hand, the spatial range is lowered in the deformed space for the observations corresponding to $\mathcal{G}_2$, due to compression. The compression and stretching bring the spatial range of both subregions to nearly the same level, allowing the spatial dependence structure to be adequately modeled with a stationary variogram model in the deformed space. To evaluate the performance of our method in capturing the true spatial dependence, we use MLE to fit an isotropic Matérn covariance function in the deformed space, that is, the nonstationary covariance model in Equation (2) with $C_{\mathcal{D}}(\cdot)$ specified to be the isotropic Matérn covariance function, using the training sets of each of the 100 simulation runs. For one simulation run, the deformed space of which is shown in Figure 2, Figure 3 shows the true nonstationary correlations at three locations and their comparison to the estimated nonstationary correlations at the corresponding locations. Specifically, we show a heat map of the estimated and true nonstationary correlation between $\mathbf{s}_f$ and $\mathbf{s} \in \mathcal{G}$, where $\mathbf{s}_f$ is fixed to $(0.55, 1.78)^{\mathrm{T}}, (1.19, 0.72)^{\mathrm{T}}$, and $(1.42, 0.72)^{\mathrm{T}}$. The similarities between the true correlations and the estimated nonstationary correlations demonstrate the effectiveness of our approach in capturing the nonstationary spatial dependence as the deformation-based model satisfactorily recovers the varying spatial range for the two subregions.

To evaluate the gains in prediction performance using the proposed deformation-based nonstationary method over the stationary method, we perform kriging (Cressie 1993) on 3700 validation locations for each of the 100 simulation runs. In addition to the estimated nonstationary deformation-based Matérn covariance model, we also fit the stationary Matérn covariance model in the geographic space via MLE on the training set, and use the two models as candidate covariance models for kriging. To quantify the prediction performance of both models, we compute multiple prediction quality assessment metrics on the validation set, which gauge the accuracy of the prediction and its uncertainty. In particular, we compute: (1) RMSE, (2) NMSE, (3) mean absolute error (MAE), (4) mean logarithmic score (mLogS), and (5) mean continuous ranked probability score (mCRPS) (Gneiting and Raftery 2007). While RMSE, NMSE, and MAE only measure the accuracy of the predicted value, mLogS and mCRPS take into account both

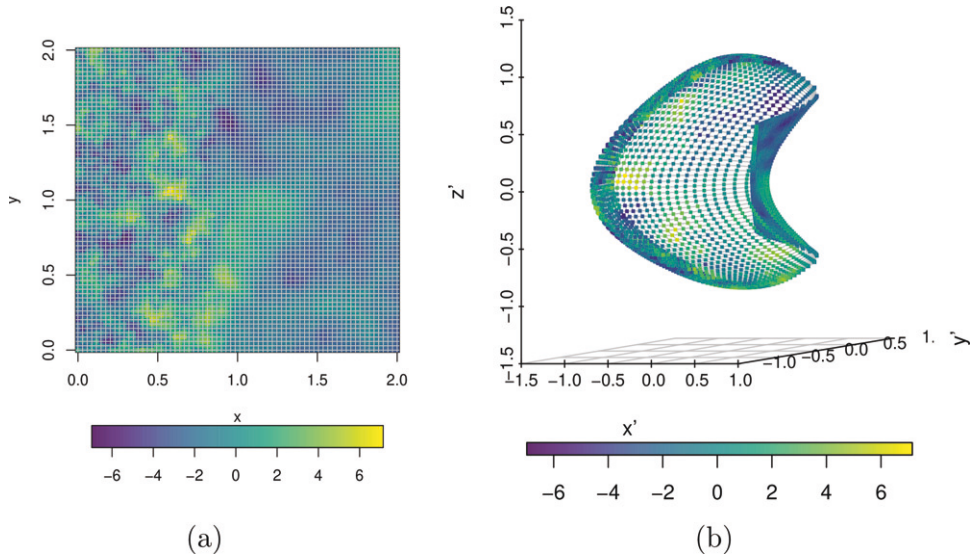

Figure 2. A realization in the (a) geographic space, and (b) estimated deformed space (first three dimensions of maximum variation). Gray lines in (a) show the grid in the geographic space, and their corresponding mapping in the estimated deformed space is in (b).

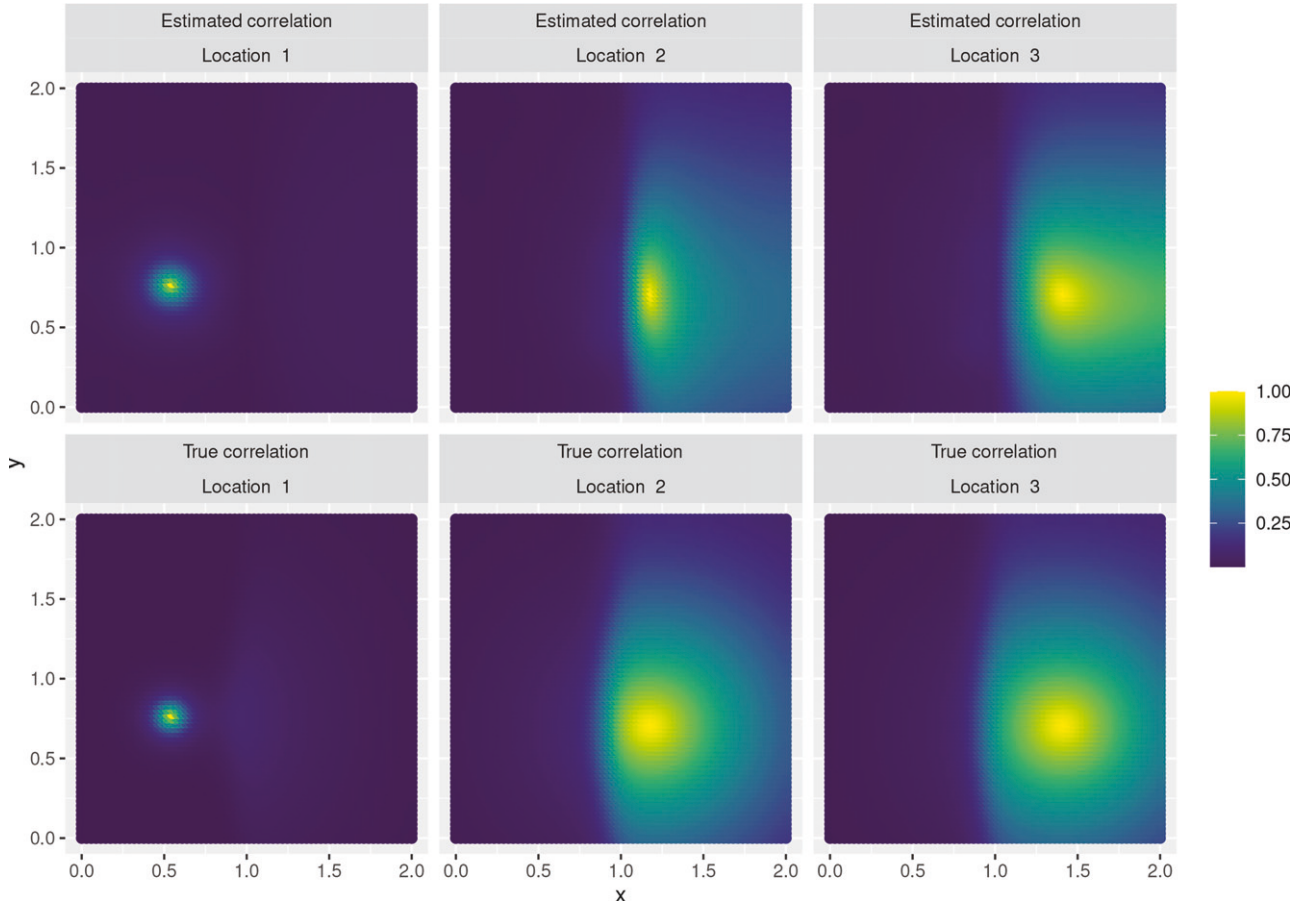

Figure 3. Estimated nonstationary correlations (top row) and true nonstationary correlations (bottom row) for three locations on the simulated grid, $\mathbf{s} \in \mathcal{G}$. First, second and third columns show the correlations between location 1 $= (0.55, 1.78)^{\mathsf{T}}$, location 2 $= (1.19, 0.72)^{\mathsf{T}}$, and location 3 $= (1.42, 0.72)^{\mathsf{T}}$, respectively, and every other location $\mathbf{s} \in \mathcal{G}$. The similarity of the top and bottom rows demonstrates effective recovery of the true nonstationary correlations using the proposed deformation-based model.

the predicted value and prediction uncertainty. As a decision rule, lower values of RMSE, MAE, mLogS, and mCRPS indicate better predictions, whereas NMSE closer to one indicates better predictions.

Kriging under the Gaussianity assumption provides us with a predictive distribution at unobserved locations, which is a conditional Gaussian distribution with the predicted (kriged) value and the prediction (kriging) variance as its mean and variance,

respectively. Let $\{X(\mathbf{s}_j)\}_{j=1}^{3700}$ denote the validation set. Then, the estimated conditional distribution for $X(\mathbf{s}_j)$ is $\mathcal{N}(\hat{X}(\mathbf{s}_j), \sigma_j^2)$, where $\hat{X}(\mathbf{s}_j)$ and σ_j^2 are the kriged value and the kriging variance for $X(\mathbf{s}_j)$, respectively. The estimated predictive distribution enables us to construct a p-prediction interval (p-PI) for $X(\mathbf{s}_j)$ as $\left(\hat{Q}_{\frac{(1-p)}{2}}(j), \hat{Q}_{\frac{(1+p)}{2}}(j)\right)$, where $\hat{Q}_p(j)$ represents the pth quantile of $\mathcal{N}(\hat{X}(\mathbf{s}_j), \sigma_j^2)$. By construction, the true value of $X(\mathbf{s}_j)$ falls in this interval with probability $0 < p < 1$. We also explore the accuracy of the p-PI provided by the stationary and the nonstationary methods, through goodness statistic (G) (Deutsch 1997; Goovaerts 2001; Moyeed and Papritz 2002), the accuracy plot ($\tilde{\kappa}(p)$ vs. p) and the average width plot ($W(p)$ vs. p) (Fouedjio and Klump 2019). The value of $\tilde{\kappa}(p)$ is defined $\tilde{\kappa}(p) = \frac{1}{m}\sum_{j=1}^{m} \kappa_j(p)$, $\kappa_j(p) = \mathbb{1}_{X(\mathbf{s}_j)\in p-\text{PI}}$, where $\mathbb{1}$ is an indicator function; the p-PI is accurate if $\tilde{\kappa}(p) > p$ for all $p \in [0,1]$. The closeness of $\tilde{\kappa}(p)$ and p is quantified as $G = 1 - \int_0^1 [3a(p) - 2][\tilde{\kappa}(p) - p]dp$, $a(p) = \mathbb{1}_{\tilde{\kappa}(p)>p}$. The value of G lies in the interval $[0,1]$; for maximum goodness corresponding to the case $\tilde{\kappa}(p) = p$, for all $p \in [0,1]$, $G = 1$, whereas for the opposite case $\tilde{\kappa}(p) = 0$, for all $p \in [0,1], G = 0$. In general, G and the accuracy plot evaluate the coverage accuracy of the p-PI, where higher values of G and points closer to the 45° line in the accuracy plot are preferred. In addition to accurate coverage of the p-PI, the width of the p-PI is desired to be consistently narrow for $p \in [0,1]$, and this width is quantified using $W(p) = \frac{1}{m\tilde{\kappa}(p)}\sum_{j=1}^{m} \kappa_j(p)[\hat{Q}_{\frac{(1+p)}{2}}(j) - \hat{Q}_{\frac{(1-p)}{2}}(j)]$. In principle, if two methods provide qualitatively equivalent accuracy in terms of G and the accuracy plot, then the one which has lower $W(p)$ in the average width plot is preferred.

Figures 4(b)–(i) report the prediction quality assessment metrics for the nonstationary and stationary methods, across 100 simulation runs. Figures 4(b)–(g) show the boxplots of the RMSE, NMSE, MAE, mCRPS, mLogS, and G, respectively. Additionally, they also report the mean and the standard deviation of those scores in their respective plots. Figures 4(h) and (i) show the accuracy plot and the average width plot, respectively, averaged over the same 100 simulation runs. In terms of RMSE, NMSE, and MAE, which are based only on the predicted values, the stationary method provides slightly lower scores than the nonstationary method. However, the differences are very small, and thus both approaches result in predicted values of qualitatively equivalent accuracy. On the other hand, on the basis of mCRPS and mLogS, which take into account the predicted value and prediction uncertainty, the nonstationary method significantly outperforms the stationary method; the scores are much lower for the nonstationary method. These results are similar to the findings based on the case study of Fuglstad et al. (2015) who stated that the use of nonstationary models over stationary ones largely affects the prediction variance and not the predicted values. Hence, the improvements in our study are mainly reflected in mCRPS and mLogS. In terms of G, there is a notable improvement in the accuracy of the p-PI computed using the nonstationary method as compared to the stationary method; the former produces significantly higher values of G. This improvement is also reflected in the accuracy plot where the points corresponding to the nonstationary method are much closer to the 45° line than those corresponding to the stationary method, especially at higher values of p. In addition, the nonstationary method also provides much narrower prediction intervals than those provided by the stationary method as seen in the average width plot.

Overall, the simulation results suggest that there is a significant improvement in prediction performance by the proposed nonstationary method over the stationary method. The improvement is mostly attributed to more accurate prediction uncertainty and prediction intervals provided by the nonstationary method; both methods provide predicted values with similar accuracy. This improvement is expected as the underlying true covariance function of the simulated process is nonstationary in a locally stationary sense, which is well approximated by the proposed nonstationary approach based on the assumption of regional stationarity. The stationary method, on the other hand, is theoretically incapable of modeling such locally varying spatial dependence, and hence results in inferior prediction performance. In Supplementary Material Section S2, we also explore the performance of the proposed method when the underlying true nonstationary covariance function shifts from the locally stationary behavior to continuously varying nonstationarity (i.e., for increasing values of λ). The aim is to measure the extent to which the proposed model can capture the continuously varying nonstationary process. Additionally, in Supplementary Material Section S3, we conduct a simulation study to explore how well the proposed method can recover the true known deformation function.

4. Application to Precipitation Data

In this section, we illustrate the application of our method to the total annual precipitation data for the state of Colorado in the United States. The data came from Colorado's climate record provided by the Geophysical Statistics Project (GSP) at the National Center for Atmospheric Research (NCAR) (*http://www.image.ucar.edu/GSP/Data/US.monthly.met/CO.html*). It contains monthly total precipitation (in mm) recorded from a network of weather stations located across the state of Colorado over the period of 1895–1997. The spatial domain of interest has a varied topography with noticeable distinction between the mountainous region in the West and the flat plains in the East, as shown in Figure 5. Furthermore, the topographical variability in Western Colorado is much higher than in Eastern Colorado. The diverse topography induces landform driven nonstationarity in the precipitation data, which has been studied previously by Paciorek and Schervish (2006). For our analysis, we consider the log-transformed total annual precipitation data for the year 1992 (shown in Figure 5), since the number of weather stations (254) having nonmissing recordings for total annual precipitation is highest for this year. The distribution of log-transformed precipitation data is approximately Gaussian, which makes it suitable for modeling as a Gaussian process. We apply the proposed deformation-based approach to model the nonstationary spatial dependence in the data by assuming regional stationarity within the mountainous subregion in the West and flat plains in the East. Based on the estimated nonstationary model, we interpolate the sparsely observed data to a fine grid of resolution

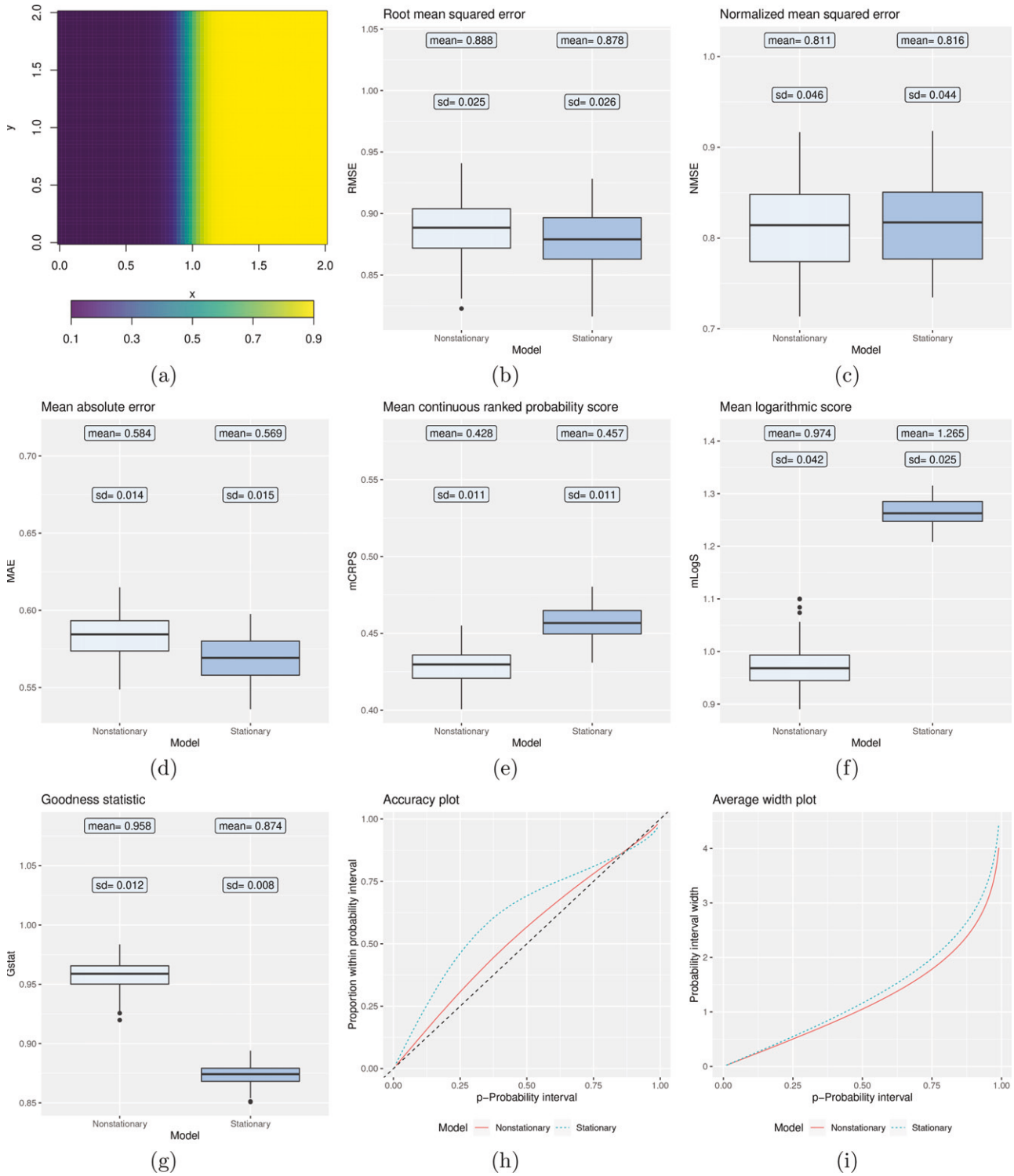

Figure 4. (a) True spatially varying range parameters for $\lambda = 0.1$. Boxplots of (b) root mean squared error, (c) normalized mean squared error, (d) mean absolute error, (e) mean continuous ranked probability score, (f) mean logarithmic score, and (g) goodness statistic. (h) Accuracy plot. (i) Average width plot. The results are based on 100 simulation runs. Above the boxplots, we report the mean and standard deviations.

0.17° longitude × 0.10° latitude by kriging. Additionally, we also interpolate the data to the fine grid by using the stationary method. However, prior to the interpolation on the fine grid, we want to evaluate the prediction performance of the two methods on this dataset, which we perform via cross-validation.

We begin by standardizing the data and splitting the entire region into Western and Eastern subregions demarcated by the longitude 104.873° W. The chosen partitioning, which was also considered by Paciorek and Schervish (2006) in their analysis, is motivated by the fact that the resulting two subregions significantly differ in their topographic features. Additionally, due to the Eastern and Western subregions being mostly flat plains and mountainous, respectively, the process can be reasonably assumed to be regionally stationary. For our cross-validation

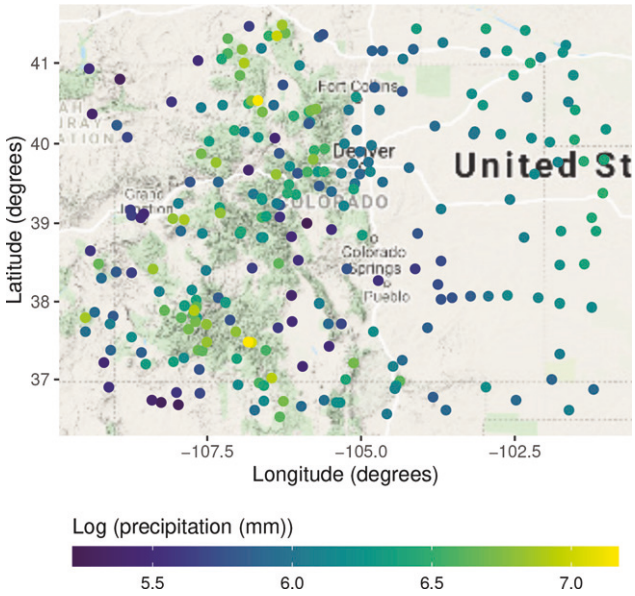

Figure 5. Observed log-transformed total annual precipitation for the year 1992, overlaid on the Google map (Kahle and Wickham 2013) of Colorado, with topography showing a mountainous Western region and Eastern flat plains.

study, we randomly divide the 254 observed data points into a training set of 224 points and a validation set of 30 points; we repeat this division 100 times to produce 100 training and validation sets for 100 cross-validation runs. Similar to the simulation study in Section 3, we consider a stationary Matérn model for the regional variograms, which we estimate via MLE using the training sets. For each run, we estimate the deformation θ with the optimal choice of ψ as discussed in Section 2.2.2. Next, we use MLE to fit the nonstationary covariance model in Equation (2) with $C_{\mathcal{D}}(\cdot)$ specified to be the isotropic Matérn covariance function. For the stationary method, we fit the stationary Matérn model in the geographic space via MLE using the training data for each of the 100 cross-validation runs. Based on the estimated nonstationary and stationary models, we perform kriging on the corresponding validation sets, and compute the prediction quality assessment metrics RMSE, NMSE, MAE, mCRPS, mLogS, G, as well as the accuracy and average width plots. The prediction performance metrics and the two plots are shown in Figure 6. The joint assessment of the prediction performance metrics from this cross-validation study indicates that there is an appreciable improvement in performance by the nonstationary method over the stationary one. Overall, the nonstationary method produces superior RMSE, NMSE, MAE, mCRPS, and mLogS (see Figures 6(a)–(e)); the improvement is largest in mLogS. This again suggests that when both the predicted value and prediction variance are considered, the proposed nonstationary method significantly outperforms the stationary method. Both methods provide comparable coverage accuracy of the p-PI in terms of G and the accuracy plot (see Figures 6(f) and (g)). However, the nonstationary method provides much narrower prediction intervals as revealed by the average width plot (see Figure 6(h)). Therefore, the proposed method is clearly preferred over the stationary method for interpolation of data on a fine grid in this application.

For the purpose of interpolation on a fine grid, we follow the specifications of the cross-validation study, but instead of using the training data, we use the entire dataset of 254 points to estimate the nonstationary and stationary models. The alignment results based on the entire data and the corresponding estimated deformed space are explained in the Supplementary Material Section S7. The deformation-based nonstationary correlation function estimated using the entire dataset is shown in Figure 7, where the heat maps visualize the correlation of ten randomly selected locations with all other locations on the fine grid chosen for interpolation. Our model perfectly captures the regionally varying spatial dependence structure, with strong spatial correlations in the Eastern subregion and relatively weaker spatial correlations in the Western subregion.

Using the estimated nonstationary and stationary models, we perform kriging on the fine grid locations chosen for interpoltion. Figures 8(a) and (c) show the kriged values for the stationary and nonstationary models, respectively. As the proposed method assumes the process to be regionally stationary, it takes into account the local features of each subregion, whereas the stationary model overlooks these local features. This causes the kriged values associated with the two models to look slightly different. Kriged values in the Eastern subregion associated with the nonstationary model exhibit wider patches of highly correlated values as compared to those based on the stationary model. There is a remarkable difference in the kriging standard deviations; those estimated using the stationary model are shown in Figure 8(b), while those estimated using the nonstationary model are shown in Figure 8(d). Kriging standard deviations from the stationary model are nearly homogeneous throughout the entire domain, with more certain predictions in the Western subregion due to the availability of more dense observations. The stationary approach does not take into account the higher prediction uncertainty that arises due to larger topographical variability in the Western subregion as compared to the Eastern subregion. On the other hand, kriging standard deviations based on the nonstationary model are more realistic as they exhibit lower prediction uncertainty in the Eastern subregion and higher in the Western subregion; this can be attributed to the strong and weak spatial dependencies in these two subregions, respectively. Stronger spatial dependence provides more information for prediction, which leads to more certain predictions while weaker spatial dependence leads to more uncertainty in prediction. In the context of deformation, stronger spatial dependence is equivalent to a compressed subregion, where more observations are available in the neighborhood of a prediction location, leading to a reduction in prediction uncertainty.

5. Discussion

We introduced a method to estimate a spatial deformation for modeling nonstationary spatial processes using functional data alignment of estimated regional variograms. The proposed approach allows for efficient estimation of the underlying nonstationary dependence structure by taking into account the stationary dependence structure in various subregions. It allows for estimation of the deformation with a single realization of a spatial process; it also avoids the problem of space folding, by

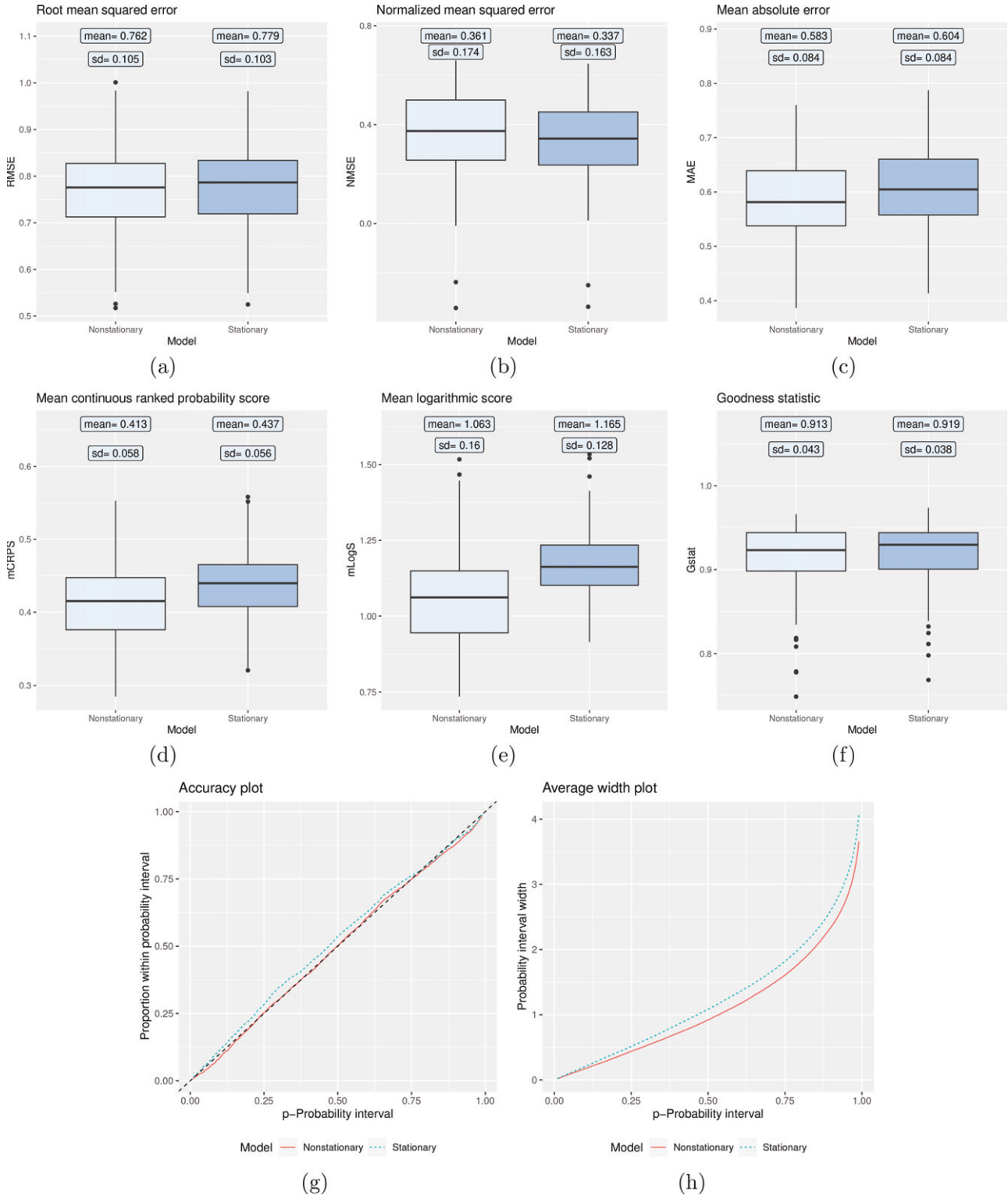

Figure 6. Boxplots of (a) root mean squared error, (b) normalized mean squared error, (c) mean absolute error, (d) mean continuous ranked probability score, (e) mean logarithmic score, and (f) goodness statistic. (g) Accuracy plot. (h) Average width plot. The results are based on 100 cross-validation runs. Above the boxplots, we report the mean and standard deviations.

allowing the deformation to be estimated in higher dimensions. Moreover, the estimated regional distance warping functions provide a neat exploratory tool that visualizes the degree of nonstationarity.

As pointed out in Section 2.2, variogram alignment cannot account for the nonstationarity in sill and nugget. However, such nonstationarity can be included trivially by allowing the nugget and the square root of the sill (marginal standard deviation) to vary spatially in the deformed space during covariance estimation.

While in theory the proposed method exhibits some robustness to different partitionings of the spatial domain (see

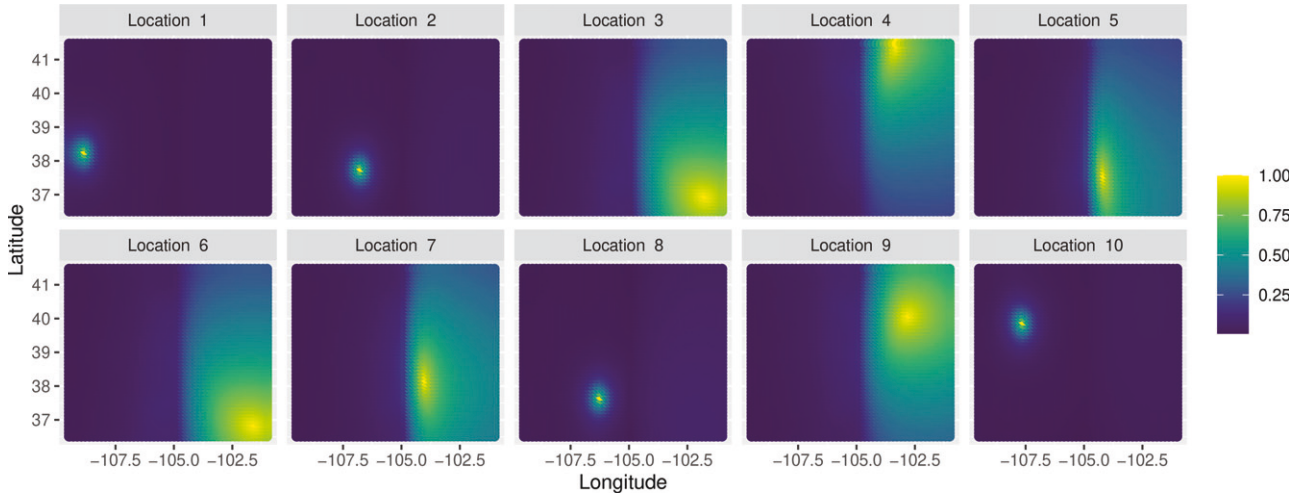

Figure 7. Heat maps for the nonstationary correlation function estimated using the entire dataset. Each heat map shows the estimated nonstationary correlation of a reference location with every other location on a fine grid. The ten locations were selected randomly.

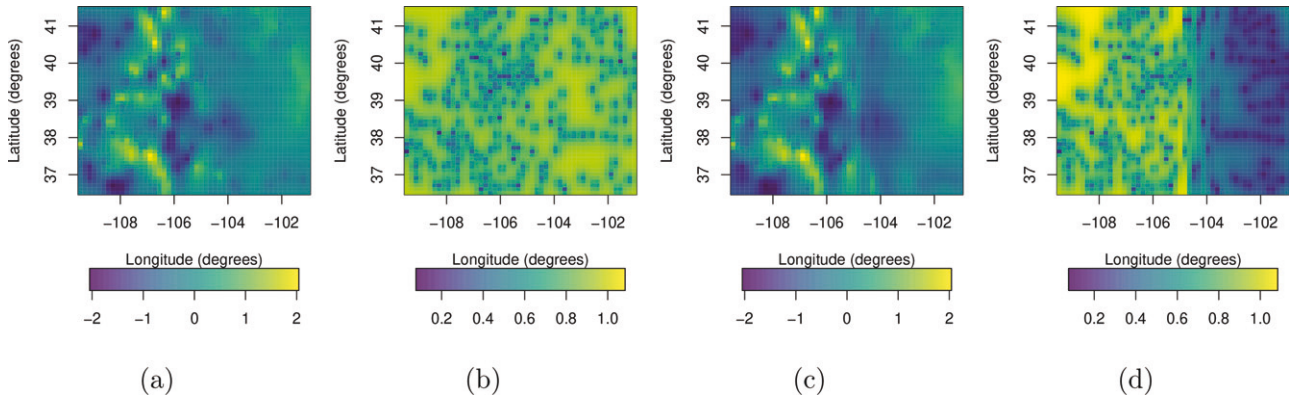

Figure 8. Kriged values and kriging standard deviations for standardized log-transformed total annual precipitation data using the stationary approach (a, b) and the nonstationary approach (c, d).

Supplementary Material Section S4), in practice, the partitioning plays a critical role and the estimated deformation can look very different for different choices of partitioning. Indeed, the proposed method might not provide any advantage over stationary modeling if the partitions are not selected appropriately. Thus, developing a method for objective selection or an adaptive scheme for selection of subregions is one potential direction for future research. Additionally, as shown in Supplementary Material Section S2, the proposed method is not very advantageous if the true nonstationarity departs too much form the locally stationary behavior toward continuously varying nonstationarity. The latter case requires a large number of subregions, which results in overfitting due to lack of enough data in each subregion (see Supplementary Material Section S4). To this end, the proposed method is best suited for data that exhibits nonstationarity with few regions with stationary dependence. Another limitation of the proposed method is that it can only partially exploit the parallel architecture of modern machines, as only Steps 1, 3, and 4 in Section 2.2.2 can be parallelized, whereas Steps 2 and 5 must be executed sequentially. Lastly, we have introduced a stepwise modeling approach, with separate steps for estimating the deformation and covariance function in the deformed space. It is desirable to develop a scheme for joint estimation of deformation and covariance function, which is another direction for future research.

Supplementary Materials

Section S1 provides proofs of Properties 1 and 2. Section S2 provides extended results from the simulation study presented in Section 3. Section S3 presents an additional simulation study. Section S4 provides a discussion on robustness of our method to different subdivisions of the spatial domain. Section S5 gives a quantitative assessment of CMDS. Section S6 provides the computational time for the estimation of the deformed space in the considered simulation studies and the data application. Section S7 provides some additional results from the data application. The code and data for reproducing the results are available at *https://github.com/ghulamabdul/Nonstat_Cov*.

Acknowledgments

The authors would like to thank the two reviewers, an associate editor, and the editor for constructive comments and helpful suggestions.

Funding

The work is partially supported by King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR) under

Award No: OSR-2019-CRG7-3800, the National Science Foundation (NSF) grants: DMS-1613054, CCF-1740761, CCF-1839252, DMS-2015226, and the National Institutes of Health (NIH) grant: R37-CA214955.

References

Anderes, E., and Chatterjee, S. (2009), "Consistent Estimates of Deformed Isotropic Gaussian Random Fields on the Plane," *The Annals of Statistics*, 37, 2324–2350. [1,2]

Anderes, E. B., and Stein, M. L. (2008), "Estimating Deformations of Isotropic Gaussian Random Fields on the Plane," *The Annals of Statistics*, 36, 719–741. [1,2]

Birchfield, S. T., and Subramanya, A. (2005), "Microphone Array Position Calibration by Basis-Point Classical Multidimensional Scaling," *IEEE Transactions on Speech and Audio Processing*, 13, 1025–1034. [4]

Bornn, L., Shaddick, G., and Zidek, J. V. (2012), "Modeling Nonstationary Processes Through Dimension Expansion," *Journal of the American Statistical Association*, 107, 281–289. [2]

Calder, C. A. (2008), "A Dynamic Process Convolution Approach to Modeling Ambient Particulate Matter Concentrations," *Environmetrics*, 19, 39–48. [1]

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley. [6]

Damian, D., Sampson, P. D., and Guttorp, P. (2001), "Bayesian Estimation of Semi-Parametric Non-Stationary Spatial Covariance Structures," *Environmetrics*, 12, 161–178. [1,2]

Deutsch, C. V. (1997), "Direct Assessment of Local Accuracy and Precision," *Geostatistics Wollongong*, 96, 115–125. [8]

Fouedjio, F. (2017), "Second-Order Non-Stationary Modeling Approaches for Univariate Geostatistical Data," *Stochastic Environmental Research and Risk Assessment*, 31, 1887–1906. [1]

Fouedjio, F., Desassis, N., and Rivoirard, J. (2016), "A Generalized Convolution Model and Estimation for Non-Stationary Random Functions," *Spatial Statistics*, 16, 35–52. [1]

Fouedjio, F., Desassis, N., and Romary, T. (2015), "Estimation of Space Deformation Model for Non-Stationary Random Functions," *Spatial Statistics*, 13, 45–61. [1,2]

Fouedjio, F., and Klump, J. (2019), "Exploring Prediction Uncertainty of Spatial Data in Geostatistical and Machine Learning Approaches," *Environmental Earth Sciences*, 78, 38. [8]

Fuentes, M. (2002), "Spectral Methods for Nonstationary Spatial Processes," *Biometrika*, 89, 197–210. [1]

Fuentes, M., and Smith, R. L. (2001), "A New Class of Nonstationary Spatial Models," Technical Report, North Carolina State University, Department of Statistics. [2]

Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2015), "Exploring a New Class of Non-Stationary Spatial Gaussian Random Fields With Varying Local Anisotropy," *Statistica Sinica*, 25, 115–133. [2]

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2015), "Does Non-Stationary Spatial Data Always Require Non-Stationary Random Fields?," *Spatial Statistics*, 14, 505–531. [8]

Genton, M. G., and Kleiber, W. (2015), "Cross-Covariance Functions for Multivariate Geostatistics," *Statistical Science*, 30, 147–163. [4]

Gneiting, T., and Raftery, A. E. (2007), "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378. [6]

Goovaerts, P. (2001), "Geostatistical Modelling of Uncertainty in Soil Science," *Geoderma*, 103, 3–26. [8]

Guan, Y., Sampson, C., Tucker, J. D., Chang, W., Mondal, A., Haran, M., and Sulsky, D. (2019), "Computer Model Calibration Based on Image Warping Metrics: An Application for Sea Ice Deformation," *Journal of Agricultural, Biological and Environmental Statistics*, 24, 444–463. [3]

Guttorp, P., and Gneiting, T. (2006), "Studies in the History of Probability and Statistics XLIX: On the Matérn Correlation Family," *Biometrika*, 93, 989–995. [4]

Haas, T. C. (1990a), "Kriging and Automated Variogram Modeling Within a Moving Window," *Atmospheric Environment. Part A. General Topics*, 24, 1759–1769. [2,4]

——— (1990b), "Lognormal and Moving Window Methods of Estimating Acid Deposition," *Journal of the American Statistical Association*, 85, 950–963. [2,4]

Heaton, M. J., Christensen, W. F., and Terres, M. A. (2017), "Nonstationary Gaussian Process Models Using Spatial Hierarchical Clustering From Finite Differences," *Technometrics*, 59, 93–101. [2]

Higdon, D. (1998), "A Process-Convolution Approach to Modelling Temperatures in the North Atlantic Ocean," *Environmental and Ecological Statistics*, 5, 173–190. [1]

Higdon, D., Swall, J., and Kern, J. (1999), "Non-Stationary Spatial Modeling," *Bayesian Statistics*, 6, 761–768. [1]

Iovleff, S., and Perrin, O. (2004), "Estimating a Nonstationary Spatial Structure Using Simulated Annealing," *Journal of Computational and Graphical Statistics*, 13, 90–105. [1]

Ji, X., and Zha, H. (2004), "Sensor Positioning in Wireless Ad-Hoc Sensor Networks Using Multidimensional Scaling," in *IEEE INFOCOM 2004* (Vol. 4), IEEE, pp. 2652–2661. [4]

Kahle, D., and Wickham, H. (2013), "ggmap: Spatial Visualization With ggplot2," *The R Journal*, 5, 144–161. [10]

Kleiber, W. (2016), "High Resolution Simulation of Nonstationary Gaussian Random Fields," *Computational Statistics & Data Analysis*, 101, 277–288. [1]

Kurtek, S. A., Srivastava, A., and Wu, W. (2011), "Signal Estimation Under Random Time-Warpings and Nonlinear Signal Alignment," in *Advances in Neural Information Processing Systems*, pp. 675–683. [3]

Kurtek, S. A., Wu, W., Christensen, G. E., and Srivastava, A. (2013), "Segmentation, Alignment and Statistical Analysis of Biosignals With Application to Disease Classification," *Journal of Applied Statistics*, 40, 1270–1288. [3]

Lahiri, S., Robinson, D., and Klassen, E. (2015), "Precise Matching of PL Curves in $\mathbb{R}^N$ in the Square Root Velocity Framework," *Geometry, Imaging and Computing*, 2, 133–186. [3]

Li, Y., and Sun, Y. (2019), "Efficient Estimation of Nonstationary Spatial Covariance Functions With Application to High-Resolution Climate Model Emulation," *Statistica Sinica*, 29, 1209–1231. [1]

Lindgren, F., Rue, H., and Lindström, J. (2011), "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach," *Journal of the Royal Statistical Society*, Series B, 73, 423–498. [2]

Lloyd, C. D., and Atkinson, P. M. (2000), "Interpolating Elevation With Locally-Adaptive Kriging," *Innovations in GIS*, 7, 241–253. [2,4]

——— (2002), "Non-Stationary Approaches for Mapping Terrain and Assessing Prediction Uncertainty," *Transactions in GIS*, 6, 17–30. [2,4]

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*, London: Academic Press. [2,4]

Marron, J. S., Ramsay, J. O., Sangalli, L. M., and Srivastava, A. (2015), "Functional Data Analysis of Amplitude and Phase Variation," *Statistical Science*, 30, 468–484. [3]

Matérn, B. (1986), *Spatial Variation* (2nd ed.), Berlin: Springer-Verlag. [4]

Meng, R., Saade, S., Kurtek, S., Berger, B., Brien, C., Pillen, K., Tester, M., and Sun, Y. (2017), "Growth Curve Registration for Evaluating Salinity Tolerance in Barley," *Plant Methods*, 13, 18. [3]

Moyeed, R. A., and Papritz, A. (2002), "An Empirical Comparison of Kriging Methods for Nonlinear Spatial Point Prediction," *Mathematical Geology*, 34, 365–386. [8]

Nychka, D., Hammerling, D., Krock, M., and Wiens, A. (2018), "Modeling and Emulation of Nonstationary Gaussian Fields," *Spatial Statistics*, 28, 21–38. [1]

Nychka, D., and Saltzman, N. (1998), "Design of Air-Quality Monitoring Networks," in *Case Studies in Environmental Statistics*, eds. D. Nychka, W. W. Piegorsch, and L. H. Cox, New York: Springer, pp. 51–76. [2]

Nychka, D., Wikle, C., and Royle, J. A. (2002), "Multiresolution Models for Nonstationary Spatial Covariance Functions," *Statistical Modelling*, 2, 315–331. [2]

Paciorek, C. J., and Schervish, M. J. (2006), "Spatial Modelling Using a New Class of Nonstationary Covariance Functions," *Environmetrics*, 17, 483–506. [1,2,5,8,9]

Reich, B. J., Eidsvik, J., Guindani, M., Nail, A. J., and Schmidt, A. M. (2011), "A Class of Covariate-Dependent Spatiotemporal Covariance Functions," *The Annals of Applied Statistics*, 5, 2265–2687. [1]

Risser, M. D. (2016), "Nonstationary Spatial Modeling, With Emphasis on Process Convolution and Covariate-Driven Approaches," arXiv no. 1610.02447. [1]

Robinson, D. T. (2012), *Functional Data Analysis and Partial Shape Matching in the Square Root Velocity Framework*, Ph.D. thesis, Florida State University. [3]

Samir, C., Kurtek, S., Srivastava, A., and Borges, N. (2016), "An Elastic Functional Data Analysis Framework for Preoperative Evaluation of Patients With Rheumatoid Arthritis," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp. 1–8. [3]

Sampson, P. D., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87, 108–119. [1,2]

Schmidt, A. M., and O'Hagan, A. (2003), "Bayesian Inference for Non-Stationary Spatial Covariance Structure via Spatial Deformations," *Journal of the Royal Statistical Society*, Series B, 65, 743–758. [1,2]

Srivastava, A., and Klassen, E. P. (2016), *Functional and Shape Data Analysis*, New York: Springer. [3]

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., and Marron, J. S. (2011), "Registration of Functional Data Using Fisher-Rao Metric," arXiv no. 1103.3817. [2,3]

Stephenson, J., Holmes, C., Gallagher, K., and Pintore, A. (2005), "A Statistical Technique for Modelling Non-Stationary Spatial Processes," in *Geostatistics Banff 2004*, eds. O. Leuangthong, and C. V. Deutsch, Dordrecht: Springer Netherlands, pp. 125–134. [2]

Torgerson, W. S. (1958), *Theory and Methods of Scaling*, New York: Wiley. [2,4]

Tucker, J. D. (2020), "fdasrvf: Elastic Functional Data Analysis," R Package Version 1.9.3. [3]

Tucker, J. D., Wu, W., and Srivastava, A. (2013), "Generative Models for Functional Data Using Phase and Amplitude Separation," *Computational Statistics & Data Analysis*, 61, 50–66. [3]