Statistica Sinica Preprint No: SS-2020-0156	
Title	Modeling Nonstationary and Asymmetric Multivariate Spatial Covariances via Deformations
Manuscript ID	SS-2020-0156
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0156
Complete List of Authors	Quan Vu, Andrew Zammit-Mangion and Noel Cressie
Corresponding Author	Quan Vu
E-mail	quanv@uow.edu.au

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Statistica Sinica

MODELING NONSTATIONARY AND ASYMMETRIC MULTIVARIATE SPATIAL COVARIANCES VIA DEFORMATIONS

Quan Vu, Andrew Zammit-Mangion, and Noel Cressie

University of Wollongong, Australia

Abstract: Multivariate spatial-statistical models are often used when modeling environmental and socio-demographic processes. The most commonly used models for multivariate spatial covariances assume both stationarity and symmetry for the cross-covariances, but these assumptions are rarely tenable in practice. In this article, we introduce a new and highly flexible class of nonstationary and asymmetric multivariate spatial covariance models that are constructed by modeling the simpler and more familiar stationary and symmetric multivariate covariances on a warped domain. Inspired by recent developments in the univariate case, we propose modeling the warping function as a composition of a number of simple injective warping functions in a deep-learning framework. Importantly, covariance-model validity is guaranteed by construction. We establish the types of warpings that allow for cross-covariance symmetry and asymmetry, and we use likelihood-based methods for inference that are computationally efficient. The utility of this new class of models is shown through two data illustrations: a simulation study on nonstationary data, and an application to ocean temperatures at two different depths.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Key words and phrases: Cross-Covariance, Deep Learning, Gaussian Process, Spatial Statistics, Warping

1. Introduction

Multivariate spatial-statistical models are used to jointly model two or more variables that are spatially indexed. They find widespread use in several application domains, such as the environmental sciences and the social sciences, where spatial processes interact. The utility of multivariate models lies in the concept of "borrowing strength," where in this setting, information on one process (obtained, for example, through observation) imparts information on the other processes being jointly modeled, but that may or may not be directly observed. Multivariate spatial models need to adequately model both the marginal behavior of the spatial processes and the joint dependence between the processes. Often, the central object of interest when constructing a multivariate spatial model is the cross-covariance matrix function. This function encodes the marginal covariances and cross-covariances of the spatial processes, and its use improves inference over one-at-a-time univariate analyses of each process.

Typically, the two simple assumptions of stationarity and symmetry are made when modeling the marginal behavior of, and the joint dependence between, multiple processes. For example, the popular linear model of coregionalization (LMC) (e.g., Goulard and Voltz, 1992; Wackernagel, 2003; Gu and Shen,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

2020) assumes both of these properties, as does a recent multivariate model constructed using spectral representations (Qadir and Sun, 2020) and the multivariate Matérn model (Gneiting et al., 2010; Apanasovich et al., 2012). In the multivariate Matérn model, the elements of the cross-covariance matrix function are all Matérn covariance functions. However, although this has proven to be considerably more flexible than the traditional LMC, covariance nonstationarity and asymmetry are present in many scientific applications and should be modeled. For example, the rates of ice loss in Antarctica are clearly nonstationary, because more ice loss occurs in regions of high ice-stream velocity, which are at the boundary of the continent (Zammit-Mangion et al., 2015b). In addition, in the inversion of a trace-gas, the cross-covariance between the flux field and the mole-fraction field is asymmetric because of atmospheric transport (Zammit-Mangion et al., 2015a). Asymmetry of the cross-covariances is clearly present in the ocean temperature data in Section 4.2, due to diffusive and advective oceanographic processes.

Nonstationarity in a univariate setting has been addressed using spatial deformations (e.g., Sampson and Guttorp, 1992; Damian et al., 2001; Schmidt and O'Hagan, 2003; Fouedjio et al., 2015; Porcu et al., 2020); basis functions (e.g., Cressie and Johannesson, 2008); stochastic partial differential equations (SPDEs) (e.g., Lindgren et al., 2011; Fuglstad et al., 2015); and process convolution with a spatially varying kernel, which leads to spatially varying model

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

parameters (e.g., Higdon et al., 1999; Paciorek and Schervish, 2006). Addressing nonstationarity in a multivariate setting is more problematic, because one needs to ensure validity, that is, nonnegative-definiteness, of all possible covariance matrices constructed through a proposed cross-covariance matrix function. Valid spatial multivariate modeling approaches that account for nonstationarity include those of Gelfand et al. (2004) and Kleiber and Nychka (2012), who extended the LMC and multivariate Matérn model, respectively, to contain spatially varying parameters. Other approaches consider basis functions (Nguyen et al., 2017) or systems of SPDEs (Hu and Steinsland, 2016; Hildeman et al., 2019).

Some multivariate models also model asymmetry. For example, Li and Zhang (2011) built on the bivariate example given by Ver Hoef and Cressie (1993) and extended the general stationary multivariate model to the asymmetric case. Apanasovich and Genton (2010) used latent dimensions to model asymmetric cross-covariances, and Cressie and Zammit-Mangion (2016) used a noncentered kernel to introduce asymmetry in the joint dependence structure; see also Ver Hoef and Barry (1998) and Majumdar and Gelfand (2007).

An attractive way to introduce nonstationarity and asymmetry in multivariate spatial-statistical models is through a generalization of the univariate deformation approach of Sampson and Guttorp (1992). It is well known that nonstationarity can be modeled by deforming space; specifically, a stationary

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

process on a warped domain can induce a highly nonstationary process on the original (geographic) domain. In the multivariate case, one may apply a common deformation to all of the processes, or separate deformations to each process. As we demonstrate here, using a common deformation function enforces symmetry and constrains the nonstationary behavior (i.e., the local anisotropies and scales) to be common for each process. However, process-specific deformation functions allow for distinct nonstationary behavior and practically guarantee asymmetry. Multivariate models built using spatial deformations bring with them several advantages over some of the other models mentioned above. In particular, they are able to capture complex nonstationary and asymmetric behavior. Their cross-covariance functions are valid by construction. Furthermore, if deep compositional warping functions are used, they are computationally efficient to fit and predict with.

The remainder of this paper is organized as follows. In Section 2, we provide some background on multivariate spatial models and univariate deep compositional spatial models (DCSMs). In Section 3, we introduce a multivariate generalization of the univariate DCSM, and show how asymmetry can be induced in a simple manner through what we call "aligning functions." In Section 4, we show results from two data illustrations. First, we show the utility of multivariate DCSMs for modeling symmetric nonstationary simulated data. Then, we show its utility in an application on modeling Atlantic Ocean temperatures at

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

two different depths. In these illustrations of our methodology, we show, through cross-validation and visualization, that spatial predictions from the multivariate DCSMs are generally superior to those from conventional multivariate spatial models. In Section 5, we summarize our conclusions. Additional material is provided in the Supplementary Material.

2. Background

The multivariate DCSM that we construct in Section 3 requires the specification of a conventional symmetric, stationary, possibly isotropic, multivariate covariance model and a deep warping function. In this section, we briefly review these two building blocks.

2.1 Multivariate Spatial Covariance Models

Consider a p-variate spatial process $\mathbf{Y}(\mathbf{s}) \equiv (Y_1(\mathbf{s}), \ldots, Y_p(\mathbf{s}))'$, $\mathbf{s} \in G$, where we refer to $G \subset \mathbb{R}^d$ as the *geographic* domain in d-dimensional Euclidean space. We assume that $\text{var}(Y_i(\mathbf{s})) < \infty$, for all $\mathbf{s} \in G$ and all $i = 1, \ldots, p$. Therefore, this multivariate process has finite expectation $\boldsymbol{\mu}(\cdot) \equiv (\mu_1(\cdot), \ldots, \mu_p(\cdot))'$, and a valid cross-covariance matrix function $\mathbf{C}_G(\cdot\,,\cdot) \equiv (C_{ij,G}(\cdot\,,\cdot) : i, j = 1, \ldots, p)$, where $C_{ij,G}(\mathbf{s}, \mathbf{u}) = \text{cov}(Y_i(\mathbf{s}), Y_j(\mathbf{u}))$; $\mathbf{s}, \mathbf{u} \in G$. For $i, j = 1, \ldots, p$, the covariance function $C_{ii,G}(\cdot\,,\cdot)$ is the covariance function of the process $Y_i(\cdot)$ and, for $i \neq j$, $C_{ij,G}(\cdot\,,\cdot)$ is the cross-covariance function of $(Y_i(\cdot), Y_j(\cdot))'$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

In some cases, the cross-covariance matrix function depends only on $\mathbf{h} \equiv \mathbf{s} - \mathbf{u}$. That is, for $i, j = 1, \ldots, p$, $C_{ij,G}(\mathbf{s}, \mathbf{u}) \equiv C^o_{ij,G}(\mathbf{h}); \mathbf{s}, \mathbf{u} \in G$, where now each $C^o_{ij,G}(\cdot)$ is a function of displacement. In this case, we say that $\mathbf{C}^o_G(\cdot) = (C^o_{ij,G}(\cdot) : i, j = 1, \ldots, p)$ is a stationary cross-covariance matrix function. A cross-covariance matrix function is said to be symmetric if, for $i, j = 1, \ldots, p$, $C_{ij,G}(\mathbf{s}, \mathbf{u}) = C_{ji,G}(\mathbf{s}, \mathbf{u})$, for $\mathbf{s}, \mathbf{u} \in G$. In the stationary case, symmetry is given by $C^o_{ij,G}(\mathbf{h}) = C^o_{ji,G}(\mathbf{h})$, for $\mathbf{h} = \mathbf{s} - \mathbf{u}$ and $\mathbf{s}, \mathbf{u} \in G$.

Stationarity and symmetry are strong assumptions in practice, but they remain popular because they facilitate the construction of valid cross-covariance matrix functions with a relatively small set of parameters. Among the most popular stationary symmetric multivariate covariance models is the multivariate Matérn model (Gneiting et al., 2010), where the marginal covariance functions and cross-covariance functions are all Matérn covariance functions. In this model, every process may have a different degree of smoothness, thus circumventing a key limitation of the LMC, where, for $i = 1, \ldots, p$, the smoothness of each $Y_i(\cdot)$ is generally the same by construction.

The isotropic Matérn *correlation* function is given by $\mathcal{M}(\mathbf{h}|\nu, a) = \frac{2^{1-\nu}}{\Gamma(\nu)}(a\|\mathbf{h}\|)^\nu \mathrm{K}_\nu(a\|\mathbf{h}\|)$, where ν is the smoothness parameter, $\mathrm{K}_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν, a is the scale parameter, and $\Gamma(\cdot)$ is the gamma function. A multivariate spatial-statistical process $\mathbf{Y}(\cdot)$ has a multivariate Matérn cross-covariance matrix function if, for $i, j = 1, \ldots, p$, and

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

$\mathbf{s}, \mathbf{u} \in G$,

$$\begin{aligned} C_{ii,G}(\mathbf{s},\mathbf{u}) &\equiv \text{cov}(Y_i(\mathbf{s}), Y_i(\mathbf{u})) = \sigma_i^2 \mathcal{M}(\mathbf{h}|\nu_{ii}, a_{ii}), \\ C_{ij,G}(\mathbf{s},\mathbf{u}) &\equiv \text{cov}(Y_i(\mathbf{s}), Y_j(\mathbf{u})) = \rho_{ij}\sigma_i\sigma_j \mathcal{M}(\mathbf{h}|\nu_{ij}, a_{ij}), \quad \text{for } i \neq j, \end{aligned} \tag{2.1}$$

where $\{a_{ij}\}$ are scale parameters, $\{\nu_{ij}\}$ are smoothness parameters, $\{\sigma_i^2\}$ are variance parameters, and $\{\rho_{ij}\}$ are cross-correlation parameters. From (2.1), we can see that multivariate Matérn cross-covariance functions are stationary, symmetric, and isotropic cross-covariance functions. In order to ensure validity, some constraints must be placed on the parameters of the multivariate Matérn covariance models shown in (2.1). The parsimonious Matérn covariance models have even stricter constraints than the more general multivariate Matérn models do (see also Apanasovich et al., 2012), but they have been shown to be flexible enough to model several environmental processes of interest (see Gneiting et al., 2010, for more details). We shall use the multivariate parsimonious Matérn covariance model in Section 4 to construct multivariate DCSMs for two bivariate spatial data sets.

2.2 Deep Compositional Spatial Models

The univariate deep compositional spatial modeling approach of Zammit-Mangion et al. (2021) uses injective warpings to construct nonstationary covariance models from simple covariance models. The idea of using deformations (or warpings) to modify the properties of a process stems from the work of

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Sampson and Guttorp (1992); see also Meiring et al. (1997), Sampson et al. (2001), Schmidt and O'Hagan (2003), Calandra et al. (2016), and the references therein. In this article, we extend the univariate deep compositional approach to the important multivariate case.

Consider, for the moment, a univariate process $Y(\cdot)$ with $\text{var}(Y(\mathbf{s})) < \infty$, $\mathbf{s} \in G$, and with nonstationary covariance function $C_G(\cdot\,,\cdot)$. After warping the space G, suppose that $C_G(\cdot\,,\cdot)$ can be expressed as a simpler stationary covariance function, $C_D(\cdot\,,\cdot)$, on a deformed space D, using a warping function $\mathbf{f} : G \to D$. Specifically, $C_G(\mathbf{s},\mathbf{u}) \equiv C_D(\mathbf{f}(\mathbf{s}),\mathbf{f}(\mathbf{u}))$, for $\mathbf{s},\mathbf{u} \in G$, where $C_D(\cdot\,,\cdot)$ is a familiar (stationary) covariance function. In DCSMs, the warping function $\mathbf{f}$ is constrained to be smooth and injective in order to preclude the possibility of space-folding; see also Perrin and Monestiez (1999). In particular, it is expressed as the composition $\mathbf{f}(\cdot) \equiv \mathbf{f}_{[L]} \circ \mathbf{f}_{[L-1]} \circ \cdots \circ \mathbf{f}_{[1]}(\cdot)$, where $\mathbf{f}_{[1]}(\cdot), \ldots, \mathbf{f}_{[L]}(\cdot)$ are simple elemental injective functions, and L is the number of warpings (or layers). This compositional construction is very flexible in that it can model highly nonstationary spatial processes, yet it is simple enough to facilitate parameter estimation from relatively sparse data. Zammit-Mangion et al. (2021) call the functions $\mathbf{f}_{[1]}(\cdot), \ldots, \mathbf{f}_{[L]}(\cdot)$ warping units, and propose three types: axial warping units, radial basis function units, and Möbius transformation units. In this article, we also use these three types of warping units; see Table S1 in the Supplementary Material for more details. For example, in Section 4.1, the warping

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

function $\mathbf{f}(\cdot)$ is a composition of $L = 4$ warping units, where $\mathbf{f}_{[1]}(\cdot)$ and $\mathbf{f}_{[2]}(\cdot)$ are two axial warping units (one for each spatial dimension), $\mathbf{f}_{[3]}(\cdot)$ is a radial basis function unit, and $\mathbf{f}_{[4]}(\cdot)$ is a Möbius transformation unit.

Zammit-Mangion et al. (2021) modeled a low-rank univariate process that was approximately stationary on the warped domain. Here, in a multivariate setting, we construct valid flexible models for covariances and cross-covariances on a geographic domain G by considering a stationary and symmetric cross-covariance matrix function on the warped domain D. We use the warping functions to model the nonstationary and asymmetric behavior of the multivariate spatial process.

3. Multivariate DCSMs

3.1 Model

We now construct multivariate DCSMs by extending the univariate construction of Zammit-Mangion et al. (2021). In the univariate case, one warping function is all that is required; however, in the p-variate case, we could use p different warping functions, one for each process.

We start off with the special case where a single warping function is used for all of the p processes. In this case, for $i, j = 1, \ldots, p$, we have

$$C_{ij,G}(\mathbf{s}, \mathbf{u}) = C_{ij,D}(\mathbf{f}(\mathbf{s}), \mathbf{f}(\mathbf{u})) = C^{o}_{ij,D}(\mathbf{f}(\mathbf{s}) - \mathbf{f}(\mathbf{u})); \quad \mathbf{s}, \mathbf{u} \in G, \tag{3.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

where $\mathbf{C}_D^o(\cdot) \equiv (C_{ij,D}^o(\cdot) : i, j = 1, \ldots, p)$ is a stationary nonnegative-definite cross-covariance matrix function.

Proposition 1. *If* $\mathbf{C}_D^o(\cdot)$ *is symmetric, then the cross-covariance matrix function* $\mathbf{C}_G(\cdot\,, \cdot)$ *defined in* (3.1)*, is symmetric.*

Proof: See Section S1.1 of the Supplementary Material.

Consider now the case where p warpings, one for each process, are used to construct the cross-covariance matrix function of the p-variate process. In this case, for $i, j = 1, \ldots, p$,

$$C_{ij,G}(\mathbf{s}, \mathbf{u}) = C_{ij,D}(\mathbf{f}_i(\mathbf{s}), \mathbf{f}_j(\mathbf{u})) = C_{ij,D}^o(\mathbf{f}_i(\mathbf{s}) - \mathbf{f}_j(\mathbf{u})); \quad \mathbf{s}, \mathbf{u} \in G, \tag{3.2}$$

where $\{\mathbf{f}_i(\cdot) : i = 1, \ldots, p\}$ are process-specific warping functions and, as in Proposition 1, the valid stationary cross-covariance matrix function $\mathbf{C}_D^o(\cdot)$ is symmetric.

Proposition 2. *If* $\mathbf{C}_D^o(\cdot)$ *is symmetric, then the cross-covariance matrix function* $\mathbf{C}_G(\cdot, \cdot)$ *defined in* (3.2) *is not necessarily symmetric.*

Proof: See Section S1.2 of the Supplementary Material.

The validity of the cross-covariance matrix function constructed using warping functions is established through the following proposition.

Proposition 3. *Assume that* $\mathbf{C}_D^o(\cdot)$ *is a valid, stationary cross-covariance matrix function. Consider the spatial locations* $\{\mathbf{s}_{i1}, \ldots, \mathbf{s}_{in_i}\}$*, where* $n_i > 0$*, for*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

$i = 1, \ldots, p$. Let $\boldsymbol{\Sigma}_G = (\boldsymbol{\Sigma}_{ij,G} : i, j = 1, \ldots, p)$, where $\boldsymbol{\Sigma}_{ij,G} = (C_{ij,G}(\mathbf{s}_{ik}, \mathbf{s}_{jl}) : k = 1, \ldots, n_i,\ l = 1, \ldots, n_j)$, $N = \sum_{i=1}^{p} n_i$, and $C_{ij,G}(\cdot, \cdot)$ is given by (3.2). Then, $\boldsymbol{\Sigma}_G$ is nonnegative-definite.

Proof: See Section S1.3 of the Supplementary Material.

Summarizing the results of Proposition 1–3, we see that if $\mathbf{C}_D^o(\cdot)$ is a valid stationary cross-covariance matrix function, then the cross-covariance matrix function $\mathbf{C}_G(\cdot, \cdot)$ constructed through (3.2) is valid (i.e., nonnegative-definite). Furthermore, if $\mathbf{f}_i(\cdot) \neq \mathbf{f}_j(\cdot)$, for any $i, j = 1, \ldots, p$, then the cross-covariance matrix function is not necessarily symmetric.

Using p general warpings, as in (3.2), will yield a highly flexible parameterized model, but one that may be prone to over-fitting. In practice, any asymmetry present is likely to be simple and dominated by global shifts and rotations. Hence, to model asymmetry, we propose expressing each $\mathbf{f}_i(\cdot)$ as a composition of a shared warping function $\mathbf{f}(\cdot)$ and a process-specific "aligning function" $\mathbf{g}_i(\cdot)$, for $i = 1, \ldots, p$. That is, for $i, j = 1, \ldots, p$, we let

$$\begin{aligned} C_{ij,G}(\mathbf{s}, \mathbf{u}) &= C_{ij,D}(\mathbf{f} \circ \mathbf{g}_i(\mathbf{s}), \mathbf{f} \circ \mathbf{g}_j(\mathbf{u})) \\ &= C_{ij,D}^o(\mathbf{f} \circ \mathbf{g}_i(\mathbf{s}) - \mathbf{f} \circ \mathbf{g}_j(\mathbf{u})); \quad \mathbf{s}, \mathbf{u} \in G, \end{aligned} \tag{3.3}$$

where $\mathbf{g}_i(\cdot)$, for $i = 1, \ldots, p$, are simple transformations commonly used to align spatial fields and that can include translations and rotations (e.g., Wiens et al., 2020). It can be seen from Propositions 2 and 3 that the effect of the align-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

ing functions $\{\mathbf{g}_i\}$ is to introduce asymmetry, while preserving nonnegative-definiteness. Note that a common frame of reference for the aligning functions needs to be chosen when aligning fields in this way. Therefore, without loss of generality, we fix $\mathbf{g}_1(\cdot)$ to be the identity map, in which case, $\mathbf{f}_1(\cdot) = \mathbf{f} \circ \mathbf{g}_1(\cdot)$ is simply the shared warping function $\mathbf{f}(\cdot)$. The cross-covariance model in (3.3) is a generalization of the asymmetric cross-covariance model of Li and Zhang (2011), where the shared warping function $\mathbf{f}(\cdot)$ is the identity map and the aligning functions are translations, that is, where $\mathbf{g}_i(\mathbf{s}) = \mathbf{s} + \mathbf{d}_i$, for some $\mathbf{d}_i \in \mathbb{R}^2$, $i = 2, \ldots, p$.

Under the cross-covariance-matrix model (3.3), nonstationarity can be introduced through both the shared warping function $\mathbf{f}(\cdot)$ and the aligning functions $\{\mathbf{g}_i(\cdot) : i = 1, \ldots, p\}$. Perhaps not surprisingly, the aligning functions can induce nonstationarity in the cross-covariance functions even when $\mathbf{f}(\cdot)$ is the identity map, as we demonstrate in the following proposition.

Proposition 4. *Consider the p-variate cross-covariance matrix model* (3.3)*, where $\mathbf{f}(\cdot)$ is the identity map; one of the aligning functions $\mathbf{g}_k(\cdot)$, for some $k \in \{2, \ldots, p\}$, is an affine transformation, and $\{\mathbf{g}_i(\cdot), i \neq k\}$ are identity maps. Then, $C_{ik,G}(\cdot, \cdot)$, for $i \neq k$, is not necessarily stationary.*

Proof: See Section S1.4 of the Supplementary Material.

Proposition 4 represents one simple way to introduce nonstationarity. More

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

generally, when one has p warping functions $\{\mathbf{f}_i(\cdot) : i = 1, \ldots, p\}$, nonstationarity of $\mathbf{C}_{ij,G}(\cdot, \cdot)$ is obtained by choosing $\mathbf{f}_i(\cdot)$ and $\mathbf{f}_j(\cdot)$ such that $\mathbf{f}_i(\mathbf{s}) - \mathbf{f}_j(\mathbf{u})$ is not a function of $\mathbf{s} - \mathbf{u}$, for $\mathbf{s}, \mathbf{u} \in G$.

3.2 Parameter Estimation

Assume now that we have observations $\{Z_{ik} : k = 1, \ldots, n_i;\ i = 1, \ldots, p\}$ of a p-variate Gaussian process $\tilde{\mathbf{Y}}(\cdot)$, where

$$Z_{ik} = \tilde{Y}_i(\mathbf{s}_{ik}) + \epsilon_{ik}; \quad k = 1, \ldots, n_i, \quad i = 1, \ldots, p. \tag{3.4}$$

In (3.4), $\{\epsilon_{ik} : k = 1, \ldots, n_i; i = 1, \ldots, p\}$ are independent Gaussian measurement errors that satisfy $\epsilon_{ik} \sim \text{Gau}(0, \tau_i^2)$ for $k = 1, \ldots, n_i$ and $i = 1, \ldots, p$, and $\tau_1^2, \ldots, \tau_p^2$ are measurement-error variances that are assumed to be process-specific, and hence all potentially different. We model the p-variate Gaussian process $\tilde{\mathbf{Y}}(\cdot) = (\tilde{Y}_1(\cdot), \ldots, \tilde{Y}_p(\cdot))'$ to have first moment that is linear in covariates $\mathbf{x}(\cdot) \equiv (x_1(\cdot), x_2(\cdot), \ldots, x_q(\cdot))'$. That is,

$$\tilde{Y}_i(\cdot) = \mathbf{x}(\cdot)'\boldsymbol{\beta}_i + Y_i(\cdot), \quad i = 1, \ldots, p, \tag{3.5}$$

where $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p \in \mathbb{R}^q$ are vectors of unknown coefficients that need to be estimated, and now $\mathbf{Y}(\cdot) \equiv (Y_1(\cdot), \ldots, Y_p(\cdot))'$ is a zero-mean second-order nonstationary multivariate Gaussian process on the geographic domain G.

Let $\tilde{\mathbf{Y}}_i \equiv (\tilde{Y}_i(\mathbf{s}_{i1}), \ldots, \tilde{Y}_i(\mathbf{s}_{in_i}))'$, $\tilde{\mathbf{Y}} \equiv (\tilde{\mathbf{Y}}_1', \ldots, \tilde{\mathbf{Y}}_p')'$, $\mathbf{Y}_i \equiv (Y_i(\mathbf{s}_{i1}), \ldots, Y_i(\mathbf{s}_{in_i}))'$, $\mathbf{Y} \equiv (\mathbf{Y}_1', \ldots, \mathbf{Y}_p')'$, $\mathbf{X}_i \equiv (\mathbf{x}(\mathbf{s}_{i1}), \ldots, \mathbf{x}(\mathbf{s}_{in_i}))'$, $\mathbf{X} =$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

bdiag$(\mathbf{X}_1, \ldots, \mathbf{X}_p)$, where bdiag$(\cdot)$ returns a block diagonal matrix from its arguments, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_p')'$. Then, (3.5) can be written compactly as

$$\tilde{\mathbf{Y}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Y}. \tag{3.6}$$

The covariance matrix of $\mathbf{Y}$, $\boldsymbol{\Sigma}_G \equiv \text{cov}(\mathbf{Y})$, is given by $(\boldsymbol{\Sigma}_{ij,G} : i, j = 1, \ldots, p)$, where $\boldsymbol{\Sigma}_{ij,G} \equiv (C_{ij,G}(\mathbf{s}_{ik}, \mathbf{s}_{jl}) : k = 1, \ldots, n_i; l = 1, \ldots, n_j)$. Furthermore, $\text{cov}(\tilde{\mathbf{Y}}) = \text{cov}(\mathbf{Y})$.

Let $\mathbf{Z}_i \equiv (Z_{i1}, \ldots, Z_{in_i})'$, $\mathbf{Z} \equiv (\mathbf{Z}_1', \ldots, \mathbf{Z}_p')'$, $\boldsymbol{\epsilon}_i \equiv (\epsilon_{i1}, \ldots \epsilon_{in_i})'$, and $\boldsymbol{\epsilon} \equiv (\boldsymbol{\epsilon}_1', \ldots, \boldsymbol{\epsilon}_p')'$. Then, from (3.4) and (3.6), we have, $\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Y} + \boldsymbol{\epsilon}$, where the covariance matrix of $\boldsymbol{\epsilon}$, $\mathbf{V} \equiv \text{cov}(\boldsymbol{\epsilon})$, is diagonal. The model for the observations, $\mathbf{Z}$, is therefore

$$\mathbf{Z} \sim \text{Gau}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_Z), \tag{3.7}$$

where $\boldsymbol{\Sigma}_Z = \boldsymbol{\Sigma}_G + \mathbf{V}$.

Likelihood-based inference can be used to estimate the parameters (including the warping parameters) in (3.7) (Zammit-Mangion et al., 2021). Here, we use restricted maximum likelihood (REML) in which the parameters in $\boldsymbol{\beta}$ have been marginalized out to estimate the parameters in $\boldsymbol{\Sigma}_Z$. REML is known to provide less biased estimators of variance-component parameters (Cressie and Lahiri, 1996), and can help avoid degenerate predictions (e.g., Gu et al., 2018). Let $\boldsymbol{\theta}$ be the vector containing all parameters appearing in the covariance matrix $\boldsymbol{\Sigma}_Z$, which includes the unknown parameters appearing in the warping function

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

$\mathbf{f}(\cdot)$ (and the aligning functions $\{\mathbf{g}_i(\cdot)\}$, if present), the parameters in the cross-covariance matrix function of the process on D (i.e., the scale parameter a, smoothness parameters $\{\nu_{ij}\}$, variance parameters $\{\sigma_i^2\}$, and cross-correlation parameters $\{\rho_{ij}\}$ for the parsimonious Matérn covariance function), and the measurement-error variances. The restricted maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is found by maximizing the log-restricted likelihood $\mathcal{L}(\boldsymbol{\theta}; \mathbf{Z})$ with respect to $\boldsymbol{\theta}$, for some given $\mathbf{Z}$. After obtaining $\hat{\boldsymbol{\theta}}$, the associated estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is found through generalized least squares. The REML parameter estimates are "plugged in" and allow spatial predictions of the hidden processes at an unobserved location $\mathbf{s}^*$. For more details on fitting and prediction, see Section S2 of the Supplementary Material.

The computational time complexity of evaluating the restricted likelihood is the sum of that of evaluating the deformation function and that of factorizing the full joint covariance matrix on the deformed space. The computational complexity of evaluating the aligning functions $\{\mathbf{g}_i(\cdot)\}$, when these are affine transformations, is $O(\sum_{i=2}^{p} n_i)$. The computational complexity of evaluating the shared warping-function layer $\mathbf{f}_{[l]}(\cdot)$ is $O(Nr_l)$, where $N = \sum_{i=1}^{p} n_i$ is the total number of observations for all processes, and r_l is the number of basis functions in $\mathbf{f}_{[l]}(\cdot)$. Hence, the total complexity for evaluating the deformation function is $O(\sum_{i=2}^{p} n_i + N \sum_{l=1}^{L} r_l)$. The complexity of factorizing the covariance matrix on the deformed space is $O(N^3)$. Usually, we choose $r_l \ll N$, for $l = 1, \ldots, L$,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

so that the computational complexity is dominated by the factorization of the covariance matrix on the deformed space. The actual runtime also changes with the number of iterations used to optimize the parameter estimates, which needs to be larger when the model is more complex. Parameter estimation was done using gradient-based optimization via the `R` package `tensorflow` (Allaire and Tang, 2019), which computes gradients using automatic differentiation, and can be run on a graphics processing unit (GPU).

3.3 Fixing the Frame of Reference

While setting $\mathbf{g}_1(\cdot)$ to be the identity map establishes a common frame of reference for the aligning functions, the shared warping function $\mathbf{f}(\cdot)$ and any scale parameters appearing in the cross-covariance matrix function are themselves non-identifiable unless this common frame of reference is fixed. Non-identifiability occurs when there exist at least two distinct parameters, $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ say, for which $\mathcal{L}(\boldsymbol{\theta}_1; \cdot) = \mathcal{L}(\boldsymbol{\theta}_2; \cdot)$ (Kadane, 1974). Note that this notion of non-identifiability of parameters differs from the identifiability problem of consistent estimation under infill asymptotics described by Zhang (2004). In our case, if we use a stationary symmetric cross-covariance matrix function on the warped domain that is also isotropic, the likelihood is invariant to translation, rotation, or reflection of $\mathbf{f}(\cdot)$. Because we also allow $\mathbf{f}(\cdot)$ to stretch and contract the geographic domain, any scale parameter a_{ij} associated with the cross-covariance function $C^o_{ij,D}(\cdot)$ is also

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

non-identifiable (see Anderes and Stein, 2008; Anderes and Chatterjee, 2009, for more details). While this invariance and lack of identifiability does not pose a problem for prediction, it does mean that we cannot make inference on certain properties of the warping function, such as stretches/contractions and rotations, without further assumptions. As we discuss next, it helps to use a function $\mathbf{b}_0(\cdot)$, which we call a *homogenizing function*, to place the estimates of $\mathbf{f}(\cdot)$ in a fixed frame of reference and to obtain transformations of the scale parameters that are identifiable.

We illustrate our methodology on the two-dimensional Euclidean space with $d = 2$. Specifically, we establish a fixed frame of reference (which can be easily generalized for $d > 2$) when we assume isotropy on D, as follows. Consider three locations, say $\mathbf{s}_k$, $\mathbf{s}_l$, and $\mathbf{s}_m$, in $G \subset \mathbb{R}^2$ such that $\mathbf{f}(\mathbf{s}_k)$, $\mathbf{f}(\mathbf{s}_l)$, and $\mathbf{f}(\mathbf{s}_m)$ are not colinear. Then, we use the homogenizing function to shift, scale, rotate, and reflect the warped domain D such that $\mathbf{b}_0 \circ \mathbf{f}(\mathbf{s}_k) = (0,0)'$, $\mathbf{b}_0 \circ \mathbf{f}(\mathbf{s}_l) = (1,0)'$, and $b_{0,2} \circ \mathbf{f}(\mathbf{s}_m) > 0$, where $b_{0,2}(\cdot)$ refers to the second element of $\mathbf{b}_0(\cdot)$. A homogenizing function that accomplishes these transformations is given by

$$\mathbf{b}_0(\cdot) \equiv \mathbf{b}_3 \circ \mathbf{b}_2 \circ \mathbf{b}_1(\cdot), \tag{3.8}$$

where $\mathbf{b}_1(\cdot)$ shifts and scales, $\mathbf{b}_2(\cdot)$ rotates around the origin, and $\mathbf{b}_3(\cdot)$ reflects around the horizontal axis. Figure S1 in the Supplementary Material illustrates the effect of the homogenizing function $\mathbf{b}_0(\cdot)$ on points in D.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Denote $\tilde{\mathbf{s}}_k \equiv \mathbf{f}(\mathbf{s}_k), \tilde{\mathbf{s}}_l \equiv \mathbf{f}(\mathbf{s}_l)$, and $\tilde{\mathbf{s}}_m \equiv \mathbf{f}(\mathbf{s}_m)$. The shifting and scaling is done through the function $\mathbf{b}_1(\mathbf{s}) \equiv \frac{1}{\|\tilde{\mathbf{s}}_l - \tilde{\mathbf{s}}_k\|}(\mathbf{s} - \tilde{\mathbf{s}}_k); \quad \mathbf{s} \in D$. Denote the scaled and shifted domain as D_1, where $D_1 \equiv \{\mathbf{b}_1(\mathbf{s}) : \mathbf{s} \in D\}$. Note that the distance between $\mathbf{b}_1(\tilde{\mathbf{s}}_l)$ and $\mathbf{b}_1(\tilde{\mathbf{s}}_k)$ is fixed to one in D_1. The operation that rotates $\mathbf{b}_1(\tilde{\mathbf{s}}_l)$ to the point $(1, 0)'$ is given by

$$\mathbf{b}_2(\mathbf{s}) \equiv \begin{pmatrix} \cos\psi_l & \sin\psi_l \\ -\sin\psi_l & \cos\psi_l \end{pmatrix} \mathbf{s}; \quad \mathbf{s} \in D_1,$$

where $\psi_l = \text{atan2}(b_{1,2}(\tilde{\mathbf{s}}_l), b_{1,1}(\tilde{\mathbf{s}}_l))$ is the angle of $\mathbf{b}_1(\tilde{\mathbf{s}}_l)$, and $b_{1,i}(\cdot)$ refers to the ith element of $\mathbf{b}_1(\cdot)$. Denote the scaled, shifted, and rotated domain as D_2, where $D_2 \equiv \{\mathbf{b}_2(\mathbf{s}) : \mathbf{s} \in D_1\}$. Finally, the reflection operation that ensures that $b_{0,2}(\tilde{\mathbf{s}}_m) > 0$ is given by

$$\mathbf{b}_3(\mathbf{s}) \equiv \begin{pmatrix} 1 & 0 \\ 0 & g_m \end{pmatrix} \mathbf{s}; \quad \mathbf{s} \in D_2,$$

where $g_m \equiv \text{sign}(\mathbf{b}_2 \circ \mathbf{b}_1(\tilde{\mathbf{s}}_m))$, equal to -1 if a reflection around the horizontal axis is needed, and equal to $+1$ otherwise. The fixed frame of reference is defined to be the domain $D_3 \equiv \{\mathbf{b}_3(\mathbf{s}) : \mathbf{s} \in D_2\}$.

Fixing the frame of reference can be useful when, for example, one is bootstrapping to do uncertainty quantification of the warped locations, because these warped locations are non-identifiable otherwise. Importantly, we have the following result when the covariance functions in the deformed space are solely functions of (scaled) distances.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Theorem 1. *Assume that the cross-covariance functions on the warped domain,* $\tilde{C}^{o}_{ij,D}(\mathbf{h}; a_{ij})$*, where* $\{a_{ij}\}$ *are scale parameters, are solely functions of* $a_{ij}\|\mathbf{h}\|, \mathbf{h} \in \mathbb{R}^2, a_{ij} > 0$*. Consider two cross-covariance matrix functions* $\mathbf{C}^{(1)}_G(\cdot,\cdot)$ *and* $\mathbf{C}^{(2)}_G(\cdot,\cdot)$*, where* $C^{(r)}_{ij,G}(\mathbf{s},\mathbf{u}) \equiv \tilde{C}^{o}_{ij,D}(\left\|\mathbf{f}^{(r)}(\mathbf{s}) - \mathbf{f}^{(r)}(\mathbf{u})\right\|; a^{(r)}_{ij})$*, for* $r = 1, 2$*, and* $\mathbf{s}, \mathbf{u} \in G$*. If* $\mathbf{C}^{(1)}_G(\cdot,\cdot) = \mathbf{C}^{(2)}_G(\cdot,\cdot)$*, then* $\mathbf{b}_0 \circ \mathbf{f}^{(1)}(\cdot) = \mathbf{b}_0 \circ \mathbf{f}^{(2)}(\cdot)$*, where* $\mathbf{b}_0(\cdot)$ *is given by* (3.8)*, and* $a^{(1)}_{ij}\left\|\mathbf{f}^{(1)}(\mathbf{s}_l) - \mathbf{f}^{(1)}(\mathbf{s}_k)\right\| = a^{(2)}_{ij}\left\|\mathbf{f}^{(2)}(\mathbf{s}_l) - \mathbf{f}^{(2)}(\mathbf{s}_k)\right\|$*. Conversely, if* $\mathbf{b}_0 \circ \mathbf{f}^{(1)}(\cdot) = \mathbf{b}_0 \circ \mathbf{f}^{(2)}(\cdot)$ *and* $a^{(1)}_{ij}\left\|\mathbf{f}^{(1)}(\mathbf{s}_l) - \mathbf{f}^{(1)}(\mathbf{s}_k)\right\| = a^{(2)}_{ij}\left\|\mathbf{f}^{(2)}(\mathbf{s}_l) - \mathbf{f}^{(2)}(\mathbf{s}_k)\right\|$*, for* $i, j = 1, \ldots, p$*, then* $\mathbf{C}^{(1)}_G(\cdot,\cdot) = \mathbf{C}^{(2)}_G(\cdot,\cdot)$*.*

Proof: See Section S1.5 of the Supplementary Material.

Theorem 1 shows that, after homogenization using (3.8), locations warped using functions that yield the same cross-covariance matrix functions on the geographic domain must coincide. This result can be used to obtain a visual appreciation of the uncertainty in the estimated warping function when bootstrapping the warping parameters. Informally, after homogenization, two covariance functions that are similar should yield points that are in close proximity to one another, and vice versa. We use such a visual diagnostic in our simulation study in Section 4.1.

Theorem 1 also reveals that there is a one-to-one correspondence between the cross-covariance matrix function on the geographic domain and the scale parameters in the warped domain after homogenization. Specifically, $\tilde{a}_{ij} = a^{(1)}_{ij}\left\|\mathbf{f}^{(1)}(\mathbf{s}) - \mathbf{f}^{(1)}(\mathbf{u})\right\| = a^{(2)}_{ij}\left\|\mathbf{f}^{(2)}(\mathbf{s}) - \mathbf{f}^{(2)}(\mathbf{u})\right\|$, for $i, j = 1, \ldots, p$, if and only if

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

$\mathbf{C}_G^{(1)}(\mathbf{s},\mathbf{u}) = \mathbf{C}_G^{(2)}(\mathbf{s},\mathbf{u})$, for all $\mathbf{s},\mathbf{u} \in G$. This leads to the following corollary, which shows that consistent inference of a transformation of the different processes' scale parameters in the warped domain can be made after homogenizing the warpings to a fixed frame of reference. This can be useful for validating our methods when the true warping function is known, as it is in the simulation study presented in Section 4.1.

Corollary 1. *Assume the conditions of Theorem 1 hold, and define* $\tilde{a}_{ij} \equiv a_{ij}\|\mathbf{f}(\mathbf{s}_l) - \mathbf{f}(\mathbf{s}_k)\|$, *for* $i,j = 1,\ldots,p$. *Then, the set comprising the homogenized warping function and transformed scale parameters,* $\{\mathbf{b}_0 \circ \mathbf{f}(\cdot), \{\tilde{a}_{ij}\}\}$, *is identifiable. That is, two sets of parameters* $\{\mathbf{b}_0 \circ \mathbf{f}^{(1)}(\cdot), \{\tilde{a}_{ij}^{(1)}\}\}$ *and* $\{\mathbf{b}_0 \circ \mathbf{f}^{(2)}(\cdot), \{\tilde{a}_{ij}^{(2)}\}\}$, *where* $\tilde{a}_{ij}^{(r)} = a_{ij}^{(r)}\left\|\mathbf{f}^{(r)}(\mathbf{s}_l) - \mathbf{f}^{(r)}(\mathbf{s}_k)\right\|$, *for* $r = 1,2$, *yield the same log-restricted likelihood function* $\mathcal{L}^{(1)}(\cdot;\cdot)$ *and* $\mathcal{L}^{(2)}(\cdot;\cdot)$ *if and only if they are identical. That is,* $\{\mathbf{b}_0 \circ \mathbf{f}(\cdot), \{\tilde{a}_{ij}\}\}$ *is identifiable in the sense of Kadane (1974).*

Proof: See Section S1.5 of the Supplementary Material.

Fixing the frame of reference allows us to do uncertainty quantification on any warping function parameters and transformed scale parameters. While under certain conditions, REML estimators are asymptotically Gaussian (Cressie and Lahiri, 1996), we are not aware of an analytical form of the asymptotic distribution of the REML estimators for a nonstationary covariance model constructed through deformation. Hence, we use bootstrapping to make inference

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

on these parameters. Bootstrapping with spatial data needs to be done with care, because the data are correlated; see Solow (1985) and Olea and Pardo-Iguzquiza (2011) for more discussion. A bootstrapping algorithm for quantifying the uncertainties of the parameters in model (3.7) is shown in Algorithm 1 in Section S3 of the Supplementary Material (Olea and Pardo-Iguzquiza, 2011). We use Algorithm 1 to visualize uncertainties on warped locations, and we use it for uncertainty quantification of parameter estimates in the simulation study of Section 4.1.

4. Data Illustrations

In this section, we use two illustrations to show the potential benefit of using multivariate DCSMs over conventional ones. In Section 4.1, we show the results from a study using data simulated from a symmetric nonstationary bivariate-covariance model. In Section 4.2, we show the results from a study using North Atlantic Ocean temperatures at two different depths. Section S4 of the Supplementary Material contains additional data illustrations using data simulated from an asymmetric nonstationary bivariate covariance model, from models with misspecified warping functions, and from a trivariate covariance model. It also contains an experiment using real maximum-and-minimum-temperature data in the United States. The code and data for reproducing the results from all our data illustrations are available from

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

`https://github.com/quanvu17/deepspat_multivar`.

4.1 Simulated Symmetric Nonstationary Data

We first demonstrate the use of multivariate DCSMs on data simulated using a symmetric nonstationary bivariate covariance model. We simulated the bivariate data from a Gaussian multivariate DCSM, $\tilde{\mathbf{Y}}(\cdot)$, with constant mean (i.e., $q = 1$ and $\mathbf{x}(\cdot) = x_1(\cdot) = 1$ in (3.5), so that there are two intercepts, β_{11} and β_{21}, that need to be estimated). The data were simulated on an equally spaced 101×101 grid of the geographic domain $G \equiv [-0.5, 0.5] \times [-0.5, 0.5]$. The warping function we used was a composition of axial warping units, followed by a single-resolution radial basis function unit, followed by a Möbius transformation unit; see Zammit-Mangion et al. (2021) for a detailed description of these warping units. On the warped domain, we modeled the covariances using a stationary isotropic multivariate parsimonious Matérn model. We randomly sampled 1000 locations from the grid, and used these as measurement locations.

We compared the predictions of the stationary parsimonious Matérn model (Model 4.1.1) with those of the multivariate DCSM (Model 4.1.2) in order to gauge the loss in prediction performance when the nonstationarity arising from the warping is ignored. After fitting Model 4.1.1 and Model 4.1.2 to the observations at the 1000 locations, we computed the predictions and prediction standard errors of the latent processes on the 101×101 grid. Figure 1 shows the true

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

simulated fields, predictions, and prediction standard errors, from both models. As observed in the univariate case (Zammit-Mangion et al., 2021), we see that the DCSM can predict sharp features in the spatial fields, while the stationary parsimonious Matérn model smooths out such features. Furthermore, while the stationary parsimonious Matérn model produces prediction standard errors that are mostly unrelated to the process behavior (due to the stationarity assumption), the DCSM produces prediction standard errors that are highly reflective of the processes' local anisotropies and scales. These visualizations illustrate the advantages of using a multivariate DCSM over a stationary multivariate model when the underlying processes are highly nonstationary.

To compare the predictive performance of the two models quantitatively, we calculated the predictive performance at the remaining 9201 locations using two commonly used scoring rules, namely the root mean squared prediction error (RMSPE) and the continuous-ranked probability score (CRPS) (Gneiting and Raftery, 2007). We repeated the procedure of random sampling 1000 locations and accessing the predictive performance 30 times. Table 1 summarizes the results, with averages from the 30 cross-validation studies, and also gives the average Akaike information criterion (AIC) from these studies. Figure S2 in the Supplementary Material shows box plots of the RMSPE and CRPS for both models across the 30 simulations. From the table and figure, it is clear that there is a large improvement in RMSPE and CRPS when using the DCSM

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

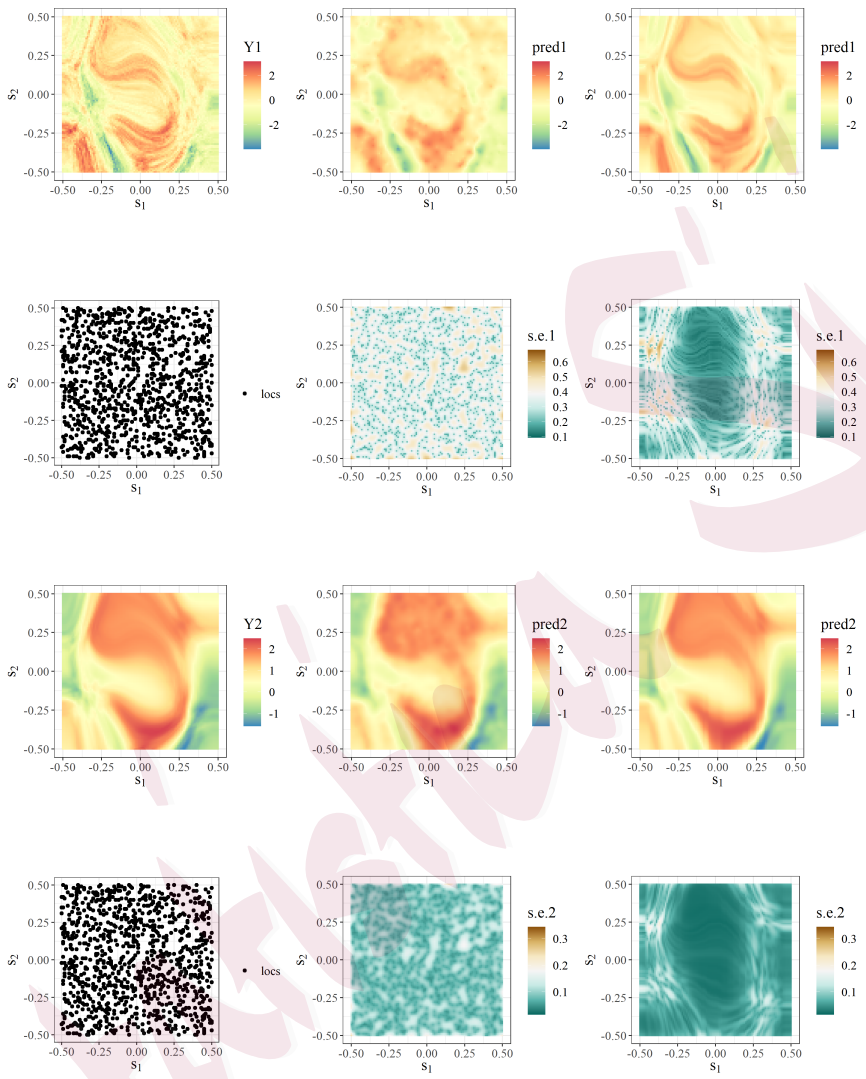

Figure 1: Comparison of predictions and prediction standard errors when using a bivariate stationary parsimonious Matérn model (Model 4.1.1) and a bivariate DCSM (Model 4.1.2) in the study of Section 4.1, where measurement locations were randomly sampled from G. First row: The process $\tilde{Y}_1(\cdot)$ (left panel), and the predictions obtained using Model 4.1.1 (center panel) and Model 4.1.2 (right panel). Second row: Locations of the measurement of $\tilde{Y}_1(\cdot)$ (left panel), and the prediction standard errors obtained when using Model 4.1.1 (center panel) and Model 4.1.2 (right panel). Third and fourth rows: Analogous to the first and second rows, respectively, for the process $\tilde{Y}_2(\cdot)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Table 1: Average hold-out validation results, AIC, and the time required to fit for the simulation study in Section 4.1, where the measurement locations are randomly sampled 30 times from G.

	$\tilde{Y}_1(\cdot)$		$\tilde{Y}_2(\cdot)$			
	RMSPE	CRPS	RMSPE	CRPS	AIC	Time (s)
Model 4.1.1	0.404	0.221	0.089	0.048	887.0	823.5
Model 4.1.2	0.358	0.196	0.072	0.037	257.0	1545.4

(Model 4.1.2) over the stationary parsimonious Matérn model (Model 4.1.1). This improvement was expected because the data were generated from the highly nonstationary process. Using the more sophisticated nonstationary model comes at some computational cost. It took 1545.4 seconds, on average, to fit Model 4.1.2, almost twice as long as the 823.5 seconds it took to fit Model 4.1.1.

We next used a bootstrap to examine the ability of the DCSM to recover the true parameters. We bootstrapped in a fixed frame of reference (via the homogenizing function) with 1000 bootstrap samples to quantify uncertainties on the model parameters using the method outlined in Section 3.3. Figure 2 shows the measurement locations in G, the measurement locations under the true warping function and homogenization, the measurement locations under the estimated warping function and homogenization, and the bootstrap samples

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

of the warped locations. We see that the estimated warped locations and the bootstrap samples of the warped locations are similar to the warped locations under the true warping function. Specifically, important features, such as the contraction in the middle part of the domain, are recovered. Table S2 in the Supplementary Material lists the true cross-covariance matrix function parameters, along with their estimates and their 95% bootstrap confidence intervals. The REML estimates are relatively close to the true parameter values, and all of the 95% bootstrap confidence intervals of the model parameters contain the true values.

We also considered the case where the data are missing in a block, which is shown in Section S4.1 in the Supplementary Material.

4.2 Modeling Temperatures in the North Atlantic Ocean at Two Different Depths

We next consider sea temperatures in the North Atlantic Ocean at two very different depths: 0.5 m and 318.1 m. The data were obtained from the Copernicus Marine Environment Monitoring Service (CMEMS)[1]. We analyzed temperatures on July 1, 2018, between 36.3°N–39.6°N and 60.0°W–63.3°W, with 1600 measurements on a 40×40 grid. Panels (1,1) and (3,1) in Figure 3 show the

[1] http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=GLOBAL_ANALYSIS_FORECAST_PHY_001_024

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

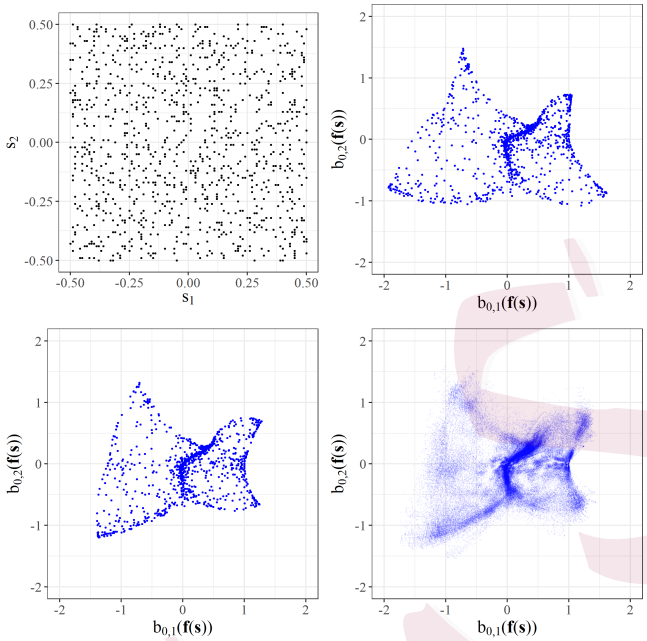

Figure 2: Measurement locations on the original domain G and the warped domain after homogenization, D_3. Top row: Measurement locations on the original domain (left panel); true warped measurement locations (right panel). Bottom row: REML estimate of the warped measurement locations (left panel); bootstrap distribution of the warped measurement locations (right panel; for visualization purposes, only 100 bootstrap samples are shown).

temperatures at the two depths, where we can see that there is a small amount of misalignment in the temperature processes, suggesting that the modeling of cross-covariance asymmetry may be important.

We considered the following models:

- Model 4.2.1: A bivariate stationary symmetric parsimonious Matérn model with only an intercept in the trend.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

- Model 4.2.2: A bivariate symmetric DCSM, where the warping function $\mathbf{f}(\cdot)$ is a composition of axial warping units, a single-resolution radial basis function unit, and a Möbius transformation unit, and with Model 4.2.1 on the warped domain.

- Model 4.2.3: A bivariate asymmetric DCSM, where the aligning function $\mathbf{g}_2(\cdot)$ is an affine transformation (as described in Proposition 4) and the warping function is as in Model 4.2.2, and with Model 4.2.1 on the warped domain.

The predictive performance of these three models was first examined using a five-fold cross-validation study, where we randomly divided the 1600 measurement locations into five groups. The results are summarized in Table S3 in the Supplementary Material. We see that allowing for model nonstationarity and/or asymmetry does indeed result in improved predictions, but the observed improvement is not substantial.

We also considered the case where the data are missing in a block. Specifically, we assumed that we have all the measurements on the grid, except those between 37.5°N–38.2°N. Figure 3 shows the true fields and the predictions using Model 4.2.1, Model 4.2.2, and Model 4.2.3, while Table 2 shows the diagnostic results when predicting the temperature at the missing locations. The improvement of the bivariate DCSMs over the stationary symmetric Model 4.2.1 is

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

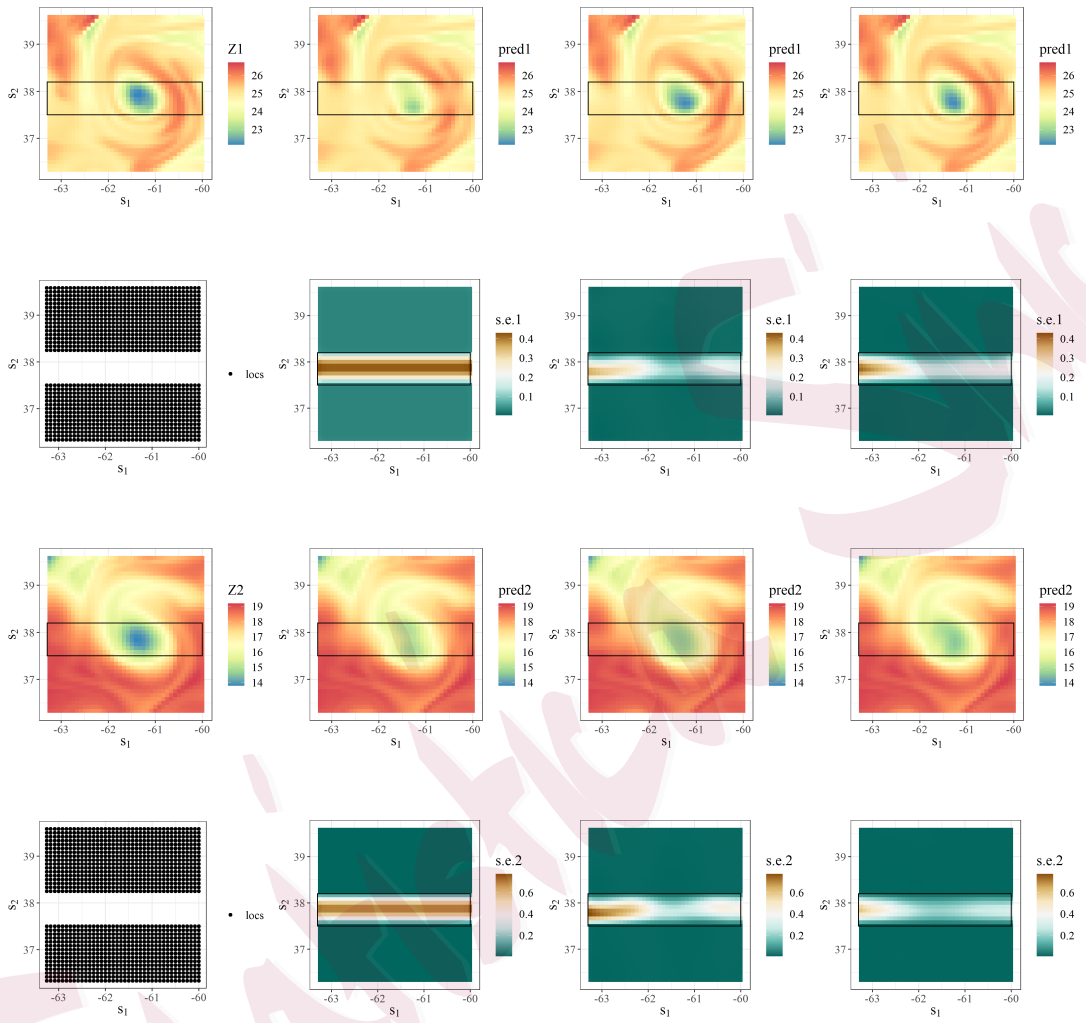

Figure 3: Comparison of predictions when using a symmetric stationary parsimonious Matérn model (Model 4.2.1), a bivariate symmetric DCSM (Model 4.2.2), and a bivariate asymmetric DCSM (Model 4.2.3). First row: Original temperature observations at depth 0.5 m, Z_1 (first panel); predictions obtained using Model 4.2.1 (second panel), Model 4.2.2 (third panel), and Model 4.2.3 (fourth panel). Second row: Locations of the retained measurements (first panel); prediction standard errors obtained when using Model 4.2.1 (second panel), Model 4.2.2 (third panel), and Model 4.2.3 (fourth panel). Third and fourth rows: Analogous to the first and second rows, respectively, for Z_2, the temperature observations at depth 318.1 m.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Table 2: Hold-out validation results, AIC, and the time required to fit for the ocean temperature data at depths 0.5 m and 318.1 m for the study in Section 4.2, where data were missing in the white block shown in the second and fourth rows of Figure 3.

	$T_{0.5}$		$T_{318.1}$			
	RMSPE	CRPS	RMSPE	CRPS	AIC	Time (s)
Model 4.2.1	0.450	0.226	0.463	0.236	-5274.7	1311.1
Model 4.2.2	0.234	0.136	0.301	0.176	-6891.7	2491.0
Model 4.2.3	0.228	0.126	0.301	0.166	-7892.1	4343.2

evident in this case. Observe that the asymmetric version of the DCSM (Model 4.2.3) produces slightly better predictive diagnostics than those of the symmetric version (Model 4.2.2), illustrating the importance of being able to model asymmetry. Visualizations of the nonstationary and asymmetric structure of estimated cross-covariance matrix function are given in Figure S3 and Figure S4 in the Supplementary Material.

5. Conclusion

We have introduced a new class of cross-covariance matrix functions that are valid by construction and that capture both nonstationarity and asymmetry.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Specifically, using p warping functions, each of which is constructed as a composition of elementary injective warping units, we model p-variate spatial processes that have nonstationary and asymmetric covariances on the geographic domain. These are modeled in terms of processes with stationary, symmetric, and possibly isotropic covariances on a warped domain. We also consider a special case where the same warping function is used for all p processes, resulting in a symmetric cross-covariance matrix function on the geographic domain. We show the benefit of using multivariate DCSMs over classical stationary models, such as the multivariate parsimonious Matérn model, through illustrations based on simulated data and real-world data.

There are a number of avenues that can be considered for future development of the proposed models. First, the models we propose do not consider nonstationarity in the variance parameters or in the cross-correlation parameters. Spatially varying variance parameters and cross-correlation parameters can be introduced, as in Kleiber and Nychka (2012) and Messick et al. (2017). Second, we use only the parsimonious multivariate Matérn model on the warped domain, but indeed any model could be used (e.g., one based on the cross-variogram). Third, we present data examples for two-dimensional space ($d = 2$), but our model can also be used in higher-dimensional space. The challenge is to find warping functions that are injective in a higher-dimensional domain. Furthermore, because the computational complexity of evaluating the likelihood function in the model

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

is $O(N^3)$ for a data set of size N, the model needs to be modified when modeling large spatial data sets. Specifically, to deal with very large nonstationary spatial data sets, one would need to extend the model in such a way that it has a scalable structure on the deformed space. Several models that can deal with very large spatial data sets are summarized in Heaton et al. (2019). Finally, when using our multivariate spatial models, several decisions need to be made on the architecture (e.g., the number of layers of warping units, the order of the warping units, etc.), and further work needs to be carried out to determine how these decisions affect the predictive performance.

In conclusion, we show that multivariate DCSMs are easy to construct and fit from simple injective warping functions. We also show that they can provide superior predictive performance compared with that of conventional stationary models, particularly when data are missing over large regions.

Supplementary Material

Section S1 contains the proofs of Propositions 1–4, Theorem 1, and Corollary 1; Section S2 gives the log-restricted likelihood and prediction formulae; Section S3 contains additional tables and figures; and Section S4 contains additional data illustrations.

Acknowledgments

Quan Vu was supported by a University Postgraduate Award from the Uni-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

versity of Wollongong, Australia. Andrew Zammit-Mangion's research was supported by an Australian Research Council Discovery Early Career Research Award (DECRA) DE180100203 and by Discovery Project DP190100180. Noel Cressie's research was supported by Australian Research Council Discovery Projects DP150104576 and DP190100180, and by NSF grant SES-1132031 funded through the NSF-Census Research Network (NCRN) program. Cressie's and Zammit-Mangion's research was also supported by NASA ROSES grant 17-OCO2-17-0012. The authors are grateful to the associate editor and three referees for their constructive comments and suggestions. They also thank Matt Moores for discussions related to this research.

References

Allaire, J. J. and Tang, Y. (2019). *tensorflow: R Interface to 'TensorFlow'*. Online: Available from `https://github.com/rstudio/tensorflow`.

Anderes, E. and Chatterjee, S. (2009). Consistent estimates of deformed isotropic Gaussian random fields on the plane. *The Annals of Statistics*, 37:2324–2350.

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.

Apanasovich, T. V. and Genton, M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika*, 97:15–30.

Apanasovich, T. V., Genton, M. G., and Sun, Y. (2012). A valid Matérn class of cross-covariance

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107:180–193.

Calandra, R., Peters, J., Rasmussen, C. E., and Deisenroth, M. P. (2016). Manifold Gaussian processes for regression. In *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, pages 3338–3345. IEEE, Vancouver, BC, Canada.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B*, 70:209–226.

Cressie, N. and Lahiri, S. N. (1996). Asymptotics for REML estimation of spatial covariance parameters. *Journal of Statistical Planning and Inference*, 50:327–341.

Cressie, N. and Zammit-Mangion, A. (2016). Multivariate spatial covariance models: A conditional approach. *Biometrika*, 103:915–935.

Damian, D., Sampson, P. D., and Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics*, 12:161–178.

Fouedjio, F., Desassis, N., and Romary, T. (2015). Estimation of space deformation model for non-stationary random functions. *Spatial Statistics*, 13:45–61.

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.

Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13:263–312.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105:1167–1177.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102:359–378.

Goulard, M. and Voltz, M. (1992). Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24:269–286.

Gu, M. and Shen, W. (2020). Generalized probabilistic principal component analysis of correlated data. *Journal of Machine Learning Research*, 21:13.

Gu, M., Wang, X., and Berger, J. O. (2018). Robust Gaussian stochastic process emulation. *Annals of Statistics*, 46:3038–3066.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., and Zammit-Mangion, A. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24:398–425.

Higdon, D., Swall, J., and Kern, J. (1999). Non-stationary spatial modeling. *Bayesian Statistics*, 6:761–768.

Hildeman, A., Bolin, D., and Rychlik, I. (2019). Joint spatial modeling of significant wave height and wave period using the SPDE approach. *arXiv preprint, arXiv:1906.00286*.

Hu, X. and Steinsland, I. (2016). Spatial modeling with system of stochastic partial differential equations. *Wiley Interdisciplinary Reviews: Computational Statistics*, 8:112–125.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Kadane, J. B. (1974). The role of identification in Bayesian theory. In Fienberg, S. E. and Zellner, A., editors, *Studies in Bayesian Econometrics and Statistics*, pages 175–191. Amsterdam, The Netherlands.

Kleiber, W. and Nychka, D. (2012). Nonstationary modeling for multivariate spatial processes. *Journal of Multivariate Analysis*, 112:76–91.

Li, B. and Zhang, H. (2011). An approach to modeling asymmetric multivariate spatial covariance structures. *Journal of Multivariate Analysis*, 102:1445–1453.

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.

Majumdar, A. and Gelfand, A. E. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Mathematical Geology*, 39:225–245.

Meiring, W., Monestiez, P., Sampson, P., and Guttorp, P. (1997). Developments in the modelling of nonstationary spatial covariance structure from space-time monitoring data. In Baafi, E. Y. and Schofield, N. A., editors, *Geostatistics Wollongong '96*, pages 162–173. Kluwer, Dordrecht, The Netherlands.

Messick, R. M., Heaton, M. J., and Hansen, N. (2017). Multivariate spatial mapping of soil water holding capacity with spatially varying cross-correlations. *Annals of Applied Statistics*, 11:69–92.

Nguyen, H., Cressie, N., and Braverman, A. (2017). Multivariate spatial data fusion for very large remote sensing datasets. *Remote Sensing*, 9:142.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Olea, R. A. and Pardo-Iguzquiza, E. (2011). Generalized bootstrap method for assessment of uncertainty in semivariogram inference. *Mathematical Geosciences*, 43:203–228.

Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17:483–506.

Perrin, O. and Monestiez, P. (1999). Modelling of non-stationary spatial structure using parametric radial basis deformations. In Gomez-Hernandez, J., Soares, A., and Froidevaux, R., editors, *geoENV II—Geostatistics for Environmental Applications*, pages 175–186. Springer, New York, NY.

Porcu, E., Senoussi, R., Mendoza, E., and Bevilacqua, M. (2020). Reduction problems and deformation approaches to nonstationary covariance functions over spheres. *Electronic Journal of Statistics*, 14:890–916.

Qadir, G. A. and Sun, Y. (2020). Semiparametric estimation of cross-covariance functions for multivariate random fields. *Biometrics*, in press, DOI:`https://doi.org/10.1111/biom.13323`.

Sampson, P., Damian, D., and Guttorp, P. (2001). Advances in modeling and inference for environmental processes with nonstationary spatial covariance. In Monestiez, P., Allard, D., and Froidevaux, R., editors, *GeoENV III–Geostatistics for Environmental Applications*, pages 17–32. Springer, New York, NY.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87:108–119.

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

structure via spatial deformations. *Journal of the Royal Statistical Society: Series B*, 65:743–758.

Solow, A. R. (1985). Bootstrapping correlated data. *Mathematical Geology*, 17:769–775.

Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69:275–294.

Ver Hoef, J. M. and Cressie, N. (1993). Multivariable spatial prediction. *Mathematical Geology*, 25:219–240.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Springer, Berlin.

Wiens, A., Kleiber, W., Barnhart, K. R., and Sain, D. (2020). Surface estimation for multiple misaligned point sets. *Mathematical Geosciences*, 52:527–542.

Zammit-Mangion, A., Cressie, N., Ganesan, A. L., O'Doherty, S., and Manning, A. J. (2015a). Spatio-temporal bivariate statistical models for atmospheric trace-gas inversion. *Chemometrics and Intelligent Laboratory Systems*, 149:227–241.

Zammit-Mangion, A., Ng, T. L. J., Vu, Q., and Filippone, M. (2021). Deep compositional spatial models. *Journal of the American Statistical Association*, in press, DOI:`https://doi.org/10.1080/01621459.2021.1887741`.

Zammit-Mangion, A., Rougier, J., Schön, N., Lindgren, F., and Bamber, J. (2015b). Multivariate spatio-temporal modelling for assessing Antarctica's present-day contribution to sea-level rise. *Environmetrics*, 26:159–177.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0156

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261.

Quan Vu, Andrew Zammit-Mangion, and Noel Cressie

School of Mathematics and Applied Statistics, University of Wollongong, Australia

E-mail: quanv@uow.edu.au, azm@uow.edu.au, ncressie@uow.edu.au