

---

**MODELING NONSTATIONARY AND ASYMMETRIC  
MULTIVARIATE SPATIAL COVARIANCES VIA DEFORMATIONS**

Quan Vu, Andrew Zammit-Mangion, and Noel Cressie

*University of Wollongong, Australia*

**Supplementary Material**

**S1 Proofs**

**S1.1 Proof of Proposition 1**

Since  $\mathbf{C}_D^o(\cdot)$  is symmetric,  $C_{ij,D}^o(\cdot) = C_{ji,D}^o(\cdot)$ ,  $i, j = 1, \dots, p$ . Now consider any two locations  $\mathbf{s}, \mathbf{u} \in G$ . The covariance between  $Y_i(\mathbf{s})$  and  $Y_j(\mathbf{u})$  is given by

$$C_{ij,G}(\mathbf{s}, \mathbf{u}) = C_{ij,D}^o(\mathbf{f}(\mathbf{s}) - \mathbf{f}(\mathbf{u})) = C_{ji,D}^o(\mathbf{f}(\mathbf{s}) - \mathbf{f}(\mathbf{u})) = C_{ji,G}(\mathbf{s}, \mathbf{u}),$$

for  $i, j = 1, \dots, p$ , and therefore  $\mathbf{C}_G(\cdot, \cdot)$  is symmetric.

**S1.2 Proof of Proposition 2**

Consider any two locations  $\mathbf{s}, \mathbf{u}$  in  $G$ . The covariance between  $Y_i(\mathbf{s})$  and  $Y_j(\mathbf{u})$ ,  $i, j = 1, \dots, p$ , is

$$\text{cov}(Y_i(\mathbf{s}), Y_j(\mathbf{u})) = C_{ij,G}(\mathbf{s}, \mathbf{u}) = C_{ij,D}^o(\mathbf{f}_i(\mathbf{s}) - \mathbf{f}_j(\mathbf{u})),$$

while the covariance between  $Y_j(\mathbf{s})$  and  $Y_i(\mathbf{u})$  is

$$\begin{aligned} \text{cov}(Y_j(\mathbf{s}), Y_i(\mathbf{u})) &= C_{ji,G}(\mathbf{s}, \mathbf{u}) = C_{ji,D}^o(\mathbf{f}_j(\mathbf{s}) - \mathbf{f}_i(\mathbf{u})) \\ &= C_{ij,D}^o(\mathbf{f}_j(\mathbf{s}) - \mathbf{f}_i(\mathbf{u})), \end{aligned}$$

since  $\mathbf{C}_D^o(\cdot)$  is symmetric. Now, for  $i \neq j$ , there are many examples where  $\mathbf{f}_i(\mathbf{s}) - \mathbf{f}_j(\mathbf{u}) \neq \mathbf{f}_j(\mathbf{s}) - \mathbf{f}_i(\mathbf{u})$ , and hence  $\text{cov}(Y_i(\mathbf{s}), Y_j(\mathbf{u})) \neq \text{cov}(Y_j(\mathbf{s}), Y_i(\mathbf{u}))$  for  $i \neq j$ . That is, the cross-covariance matrix function  $\mathbf{C}_G(\cdot, \cdot)$  constructed through (3.2) is not necessarily symmetric.

### S1.3 Proof of Proposition 3

By (3.2), we have that for  $k = 1, \dots, n_i$ ,  $l = 1, \dots, n_j$ , and  $i, j = 1, \dots, p$ ,  $C_{ij,G}(\mathbf{s}_{ik}, \mathbf{s}_{jl}) = C_{ij,D}^o(\mathbf{f}_i(\mathbf{s}_{ik}) - \mathbf{f}_j(\mathbf{s}_{jl}))$ . Therefore, for  $i, j = 1, \dots, p$ , we have that  $\Sigma_{ij,G} = (C_{ij,D}^o(\mathbf{f}_i(\mathbf{s}_{ik}) - \mathbf{f}_j(\mathbf{s}_{jl})) : k = 1, \dots, n_i, l = 1, \dots, n_j)$ . Since  $\mathbf{C}_D^o(\cdot)$  is valid, we have that for any  $\mathbf{a} \in \mathbb{R}^N$ ,  $\mathbf{a} \neq \mathbf{0}$ ,  $\mathbf{a}'(C_{ij,D}^o(\mathbf{f}_i(\mathbf{s}_{ik}) - \mathbf{f}_j(\mathbf{s}_{jl})) : k = 1, \dots, n_i; l = 1, \dots, n_j; i, j = 1, \dots, p)\mathbf{a} \geq 0$ , and hence  $\mathbf{a}'\Sigma_G\mathbf{a} \geq 0$ . That is,  $\Sigma_G$  is nonnegative-definite.

### S1.4 Proof of Proposition 4

Since  $\mathbf{f}(\cdot)$  is the identity map,

$$C_{ij,G}(\mathbf{s}, \mathbf{u}) = C_{ij,D}^o(\mathbf{g}_i(\mathbf{s}) - \mathbf{g}_j(\mathbf{u})); \quad i, j \neq k,$$

$$C_{kk,G}(\mathbf{s}, \mathbf{u}) = C_{kk,D}^o(\mathbf{g}_k(\mathbf{s}) - \mathbf{g}_k(\mathbf{u})),$$

$$C_{ik,G}(\mathbf{s}, \mathbf{u}) = C_{ik,D}^o(\mathbf{g}_i(\mathbf{s}) - \mathbf{g}_k(\mathbf{u})); \quad i \neq k,$$

where  $k \in \{2, \dots, p\}$ . Write the affine transformation as  $\mathbf{g}_k(\mathbf{s}) = \tilde{\mathbf{A}}\mathbf{s} + \tilde{\mathbf{d}}$ , where  $\tilde{\mathbf{A}}$  is a  $d \times d$  matrix that is not generally equal to the identity matrix. Then,

$$C_{ij,G}(\mathbf{s}, \mathbf{u}) = C_{ij,D}^o(\mathbf{s} - \mathbf{u}); \quad i, j \neq k,$$

$$C_{kk,G}(\mathbf{s}, \mathbf{u}) = C_{kk,D}^o(\tilde{\mathbf{A}}\mathbf{s} - \tilde{\mathbf{A}}\mathbf{u}) = C_{kk,D}^o(\tilde{\mathbf{A}}(\mathbf{s} - \mathbf{u})),$$

$$C_{ik,G}(\mathbf{s}, \mathbf{u}) = C_{ik,D}^o(\mathbf{s} - \tilde{\mathbf{A}}\mathbf{u} - \tilde{\mathbf{d}}); \quad i \neq k.$$

As  $\mathbf{C}_D^o(\cdot)$  is stationary, it follows that  $C_{ij,G}(\mathbf{s}, \mathbf{u})$ ,  $i, j \neq k$  and  $C_{kk,G}(\mathbf{s}, \mathbf{u})$  are stationary, while  $C_{ik,G}(\mathbf{s}, \mathbf{u})$ ,  $i \neq k$ , is nonstationary unless  $\tilde{\mathbf{A}}$  is the identity matrix. That is, the cross-covariance function  $C_{ik,G}(\cdot, \cdot)$  is not necessarily stationary.

### S1.5 Proof of Theorem 1 and Corollary 1

#### Proof of Theorem 1:

By assumption,  $\mathbf{C}_G^{(1)}(\mathbf{s}, \mathbf{u}) = \mathbf{C}_G^{(2)}(\mathbf{s}, \mathbf{u})$ , for all  $\mathbf{s}, \mathbf{u} \in G$ , and hence by definition,  $\tilde{C}_{ij,D}^o(\|\mathbf{f}^{(1)}(\mathbf{s}) - \mathbf{f}^{(1)}(\mathbf{u})\|; a_{ij}^{(1)}) = \tilde{C}_{ij,D}^o(\|\mathbf{f}^{(2)}(\mathbf{s}) - \mathbf{f}^{(2)}(\mathbf{u})\|; a_{ij}^{(2)})$ , for all  $\mathbf{s}, \mathbf{u} \in G$  and  $i, j = 1, \dots, p$ . That is, for each  $\mathbf{s}, \mathbf{u} \in G$  and  $i, j =$

$1, \dots, p$ , we have that  $a_{ij}^{(1)} \|\mathbf{f}^{(1)}(\mathbf{s}) - \mathbf{f}^{(1)}(\mathbf{u})\| = a_{ij}^{(2)} \|\mathbf{f}^{(2)}(\mathbf{s}) - \mathbf{f}^{(2)}(\mathbf{u})\|$ , and hence  $a_{ij}^{(1)} \|\mathbf{f}^{(1)}(\mathbf{s}_l) - \mathbf{f}^{(1)}(\mathbf{s}_k)\| = a_{ij}^{(2)} \|\mathbf{f}^{(2)}(\mathbf{s}_l) - \mathbf{f}^{(2)}(\mathbf{s}_k)\|$ , for  $\mathbf{s}_l$  and  $\mathbf{s}_k$  two distinct points in  $G$ .

Hence, from the definition of  $\mathbf{b}_1(\cdot)$  in Section 3.3,

$$\|\mathbf{b}_1 \circ \mathbf{f}^{(1)}(\mathbf{s}) - \mathbf{b}_1 \circ \mathbf{f}^{(1)}(\mathbf{u})\| = \|\mathbf{b}_1 \circ \mathbf{f}^{(2)}(\mathbf{s}) - \mathbf{b}_1 \circ \mathbf{f}^{(2)}(\mathbf{u})\|,$$

for all  $\mathbf{s}, \mathbf{u} \in G$ . Since  $\mathbf{b}_2(\cdot)$  and  $\mathbf{b}_3(\cdot)$  are distance-preserving transformations, we then have that

$$\|\mathbf{b}_0 \circ \mathbf{f}^{(1)}(\mathbf{s}) - \mathbf{b}_0 \circ \mathbf{f}^{(1)}(\mathbf{u})\| = \|\mathbf{b}_0 \circ \mathbf{f}^{(2)}(\mathbf{s}) - \mathbf{b}_0 \circ \mathbf{f}^{(2)}(\mathbf{u})\|.$$

Recall that three locations  $\mathbf{s}_k$ ,  $\mathbf{s}_l$ , and  $\mathbf{s}_m$  are chosen in  $G \subset \mathbb{R}^2$  such that  $\mathbf{f}^{(r)}(\mathbf{s}_k)$ ,  $\mathbf{f}^{(r)}(\mathbf{s}_l)$ , and  $\mathbf{f}^{(r)}(\mathbf{s}_m)$  are not colinear; and  $\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_k) = (0, 0)'$ ,  $\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_l) = (1, 0)'$ , and  $b_{0,2} \circ \mathbf{f}^{(r)}(\mathbf{s}_m) > 0$ , for  $r = 1, 2$ . Now, we have that for any two points  $\mathbf{s}, \mathbf{u} \in G$ , the distance  $\|\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}) - \mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{u})\|$  does not depend on  $r$ . Because the two points  $\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_k)$  and  $\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_l)$  are fixed in  $D_3$ , it follows that the distances  $\|\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_m) - \mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_k)\|$  and  $\|\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_m) - \mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_l)\|$  are also fixed. In  $\mathbb{R}^2$ , when the distances from a point  $\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_m)$  to two fixed points  $(0, 0)'$  and  $(1, 0)'$  are constant, there exist two possible points  $\mathbf{s}_m$ , assuming  $\mathbf{f}^{(r)}(\mathbf{s}_m)$  is not colinear with  $\mathbf{f}^{(r)}(\mathbf{s}_k)$  and  $\mathbf{f}^{(r)}(\mathbf{s}_l)$ : one where  $b_{0,2} \circ \mathbf{f}^{(r)}(\mathbf{s}_m) > 0$  and one where  $b_{0,2} \circ \mathbf{f}^{(r)}(\mathbf{s}_m) < 0$ . However, as we constrain  $b_{0,2} \circ \mathbf{f}^{(r)}(\mathbf{s}_m) > 0$ ,  $\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_m)$  is unique. Since we have three fixed points  $\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_k)$ ,  $\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_l)$ , and  $\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}_m)$ , and a fixed set of distances

$\|\mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{s}) - \mathbf{b}_0 \circ \mathbf{f}^{(r)}(\mathbf{u})\|$  for any two points  $\mathbf{s}, \mathbf{u}$ , then  $\mathbf{b}_0 \circ \mathbf{f}^{(1)}(\mathbf{s}) = \mathbf{b}_0 \circ \mathbf{f}^{(2)}(\mathbf{s})$ , for all  $\mathbf{s} \in G$ .

For the converse part of the proof, assume that  $\mathbf{b}_0 \circ \mathbf{f}^{(1)}(\mathbf{s}) = \mathbf{b}_0 \circ \mathbf{f}^{(2)}(\mathbf{s})$ , for all  $\mathbf{s} \in G$ . Then,

$$\|\mathbf{b}_1 \circ \mathbf{f}^{(1)}(\mathbf{s}) - \mathbf{b}_1 \circ \mathbf{f}^{(1)}(\mathbf{u})\| = \|\mathbf{b}_1 \circ \mathbf{f}^{(2)}(\mathbf{s}) - \mathbf{b}_1 \circ \mathbf{f}^{(2)}(\mathbf{u})\|,$$

for all  $\mathbf{s}, \mathbf{u} \in G$ , and therefore

$$\frac{\|\mathbf{f}^{(1)}(\mathbf{s}_l) - \mathbf{f}^{(1)}(\mathbf{s}_k)\|}{\|\mathbf{f}^{(1)}(\mathbf{s}_l) - \mathbf{f}^{(1)}(\mathbf{s}_k)\|} = \frac{\|\mathbf{f}^{(2)}(\mathbf{s}_l) - \mathbf{f}^{(2)}(\mathbf{s}_k)\|}{\|\mathbf{f}^{(2)}(\mathbf{s}_l) - \mathbf{f}^{(2)}(\mathbf{s}_k)\|}.$$

Because  $a_{ij}^{(1)} \|\mathbf{f}^{(1)}(\mathbf{s}_l) - \mathbf{f}^{(1)}(\mathbf{s}_k)\| = a_{ij}^{(2)} \|\mathbf{f}^{(2)}(\mathbf{s}_l) - \mathbf{f}^{(2)}(\mathbf{s}_k)\|$ , for  $i, j = 1, \dots, p$ , it follows that  $a_{ij}^{(1)} \|\mathbf{f}^{(1)}(\mathbf{s}) - \mathbf{f}^{(1)}(\mathbf{u})\| = a_{ij}^{(2)} \|\mathbf{f}^{(2)}(\mathbf{s}) - \mathbf{f}^{(2)}(\mathbf{u})\|$ , for all  $\mathbf{s}, \mathbf{u} \in G$  and  $i, j = 1, \dots, p$ . Therefore,  $\mathbf{C}_G^{(1)}(\mathbf{s}, \mathbf{u}) = \mathbf{C}_G^{(2)}(\mathbf{s}, \mathbf{u})$ , for all  $\mathbf{s}, \mathbf{u} \in G$ .

### Proof of Corollary 1:

Note that the Gaussian process model (3.5) is fully specified by its mean function and covariance function. Hence, its finite-dimensional distributions are solely a function of the mean and covariance-function parameters. Therefore, the log restricted likelihood function in (S2.1) where the mean-function parameters are profiled out, solely depends on covariance-function parameters and the data  $\mathbf{Z}$ . Now, suppose that two different sets each comprising a warping function and scale parameters,  $\{\mathbf{f}^{(1)}(\cdot), \{a_{ij}^{(1)}\}\}$  and  $\{\mathbf{f}^{(2)}(\cdot), \{a_{ij}^{(2)}\}\}$ , yield the same log restricted likelihood function, for any set of measurement loca-

tions  $\{\mathbf{s}_{ik} : k = 1, \dots, n_i; i = 1, \dots, p\} \subset G$ . Then, this necessarily means that  $\mathbf{C}_G^{(1)}(\mathbf{s}, \mathbf{u}) = \mathbf{C}_G^{(2)}(\mathbf{s}, \mathbf{u})$ , for all  $\mathbf{s}, \mathbf{u} \in G$  and, from the proof of Theorem 1 above, we see that this implies that  $\mathbf{b}_0 \circ \mathbf{f}^{(1)}(\mathbf{s}) = \mathbf{b}_0 \circ \mathbf{f}^{(2)}(\mathbf{s})$ , for all  $\mathbf{s} \in G$ , and that  $a_{ij}^{(1)} \|\mathbf{f}^{(1)}(\mathbf{s}_l) - \mathbf{f}^{(1)}(\mathbf{s}_k)\| = a_{ij}^{(2)} \|\mathbf{f}^{(2)}(\mathbf{s}_l) - \mathbf{f}^{(2)}(\mathbf{s}_k)\|$ , for  $i, j = 1, \dots, p$ . This in turn implies that the sets comprising the corresponding homogenized warping functions and transformed scale 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  $\tilde{a}_{ij}^{(r)} = a_{ij}^{(r)} \|\mathbf{f}^{(r)}(\mathbf{s}_l) - \mathbf{f}^{(r)}(\mathbf{s}_k)\|$ ,  $r = 1, 2$ , are identical. Further, since Theorem 1 establishes an if-and-only-if relation, there is no other set,  $\{\mathbf{b}_0 \circ \mathbf{f}^{(3)}(\cdot), \{\tilde{a}_{ij}^{(3)}\}\}$  say, that yields the same log restricted likelihood function, for any set of measurement locations  $\{\mathbf{s}_{ik} : k = 1, \dots, n_i; i = 1, \dots, p\} \subset G$ . This is because such a set would correspond to a different covariance function,  $\mathbf{C}_G^{(3)}(\mathbf{s}, \mathbf{u})$  say, where  $\mathbf{C}_G^{(3)}(\mathbf{s}, \mathbf{u}) \neq \mathbf{C}_G^{(1)}(\mathbf{s}, \mathbf{u})$  for some  $\mathbf{s}, \mathbf{u} \in G$ . Therefore, the set comprising a homogenized warping function and the transformed scale parameters,  $\{\mathbf{b}_0 \circ \mathbf{f}(\cdot), \{\tilde{a}_{ij}\}\}$ , where  $\tilde{a}_{ij} \equiv a_{ij} \|\mathbf{f}(\mathbf{s}_l) - \mathbf{f}(\mathbf{s}_k)\|$  for  $i, j = 1, \dots, p$ , has a unique log restricted likelihood function associated with it; this set is thus identifiable (see Kadane, 1974, for more discussion on identifiability).

## S2 Log Restricted Likelihood and Prediction Formulas

The log restricted likelihood for our model under the assumption of Gaussianity for  $\mathbf{Y}$  can be written as (Cressie and Lahiri, 1993),

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{Z}) = -\frac{N-pq}{2} \log(2\pi) + \frac{1}{2} \log |\mathbf{X}'\mathbf{X}| - \frac{1}{2} \log |\boldsymbol{\Sigma}_Z| - \frac{1}{2} \log |\mathbf{X}'\boldsymbol{\Sigma}_Z^{-1}\mathbf{X}| - \frac{1}{2} \mathbf{Z}'\boldsymbol{\Pi}\mathbf{Z}, \quad (\text{S2.1})$$

where  $N = \sum_{i=1}^p n_i$ , and  $\boldsymbol{\Pi} = \boldsymbol{\Sigma}_Z^{-1} - \boldsymbol{\Sigma}_Z^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_Z^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_Z^{-1}$ .

The estimate of  $\boldsymbol{\beta}$ ,  $\hat{\boldsymbol{\beta}}$ , is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}_Z^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}_Z^{-1}\mathbf{Z}, \quad (\text{S2.2})$$

where  $\hat{\boldsymbol{\Sigma}}_Z$  denotes  $\boldsymbol{\Sigma}_Z$  evaluated at  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ .

Treating the plug-in REML estimates as known parameters, the joint distribution of the data  $\mathbf{Z}$  and the process  $\tilde{Y}_i(\cdot)$  evaluated at  $\mathbf{s}^*$  is

$$\begin{pmatrix} \mathbf{Z} \\ \tilde{Y}_i(\mathbf{s}^*) \end{pmatrix} \sim \text{Gau} \left( \begin{pmatrix} \mathbf{X} \\ \mathbf{x}_i^{*'} \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \boldsymbol{\Sigma}_Z & \boldsymbol{\sigma}^*(\mathbf{s}^*) \\ \boldsymbol{\sigma}^*(\mathbf{s}^*)' & C_{ii,G}(\mathbf{s}^*, \mathbf{s}^*) \end{pmatrix} \right), \quad (\text{S2.3})$$

where  $\mathbf{x}_i^* = (\mathbf{x}(\mathbf{s}^*)'I(j=i) : j=1, \dots, p)'$ , and  $\boldsymbol{\sigma}^*(\mathbf{s}^*) = (C_{1i,G}(\mathbf{s}_{11}, \mathbf{s}^*), \dots, C_{1i,G}(\mathbf{s}_{1n_1}, \mathbf{s}^*), \dots, C_{pi,G}(\mathbf{s}_{p1}, \mathbf{s}^*), \dots, C_{pi,G}(\mathbf{s}_{pn_p}, \mathbf{s}^*))'$ . From (S2.3), Gaussian conditioning yields

$$\begin{aligned} \text{E}(\tilde{Y}_i(\mathbf{s}^*) \mid \mathbf{Z}) &= \mathbf{x}_i^{*'}\boldsymbol{\beta} + \boldsymbol{\sigma}^*(\mathbf{s}^*)'\boldsymbol{\Sigma}_Z^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}), \\ \text{var}(\tilde{Y}_i(\mathbf{s}^*) \mid \mathbf{Z}) &= C_{ii,G}(\mathbf{s}^*, \mathbf{s}^*) - \boldsymbol{\sigma}^*(\mathbf{s}^*)'\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{\sigma}^*(\mathbf{s}^*). \end{aligned} \quad (\text{S2.4})$$

From (S2.4), it is also straightforward to make a probabilistic prediction of an observation at  $\mathbf{s}^*$ , say  $Z_i^*$ , since  $\text{E}(Z_i^* \mid \mathbf{Z}) = \text{E}(\tilde{Y}_i(\mathbf{s}^*) \mid \mathbf{Z})$ , and  $\text{var}(Z_i^* \mid \mathbf{Z}) =$

---


$$\text{var}(\tilde{Y}_i(\mathbf{s}^*) \mid \mathbf{Z}) + \tau_i^2.$$

### S3 Additional Tables and Figures

---

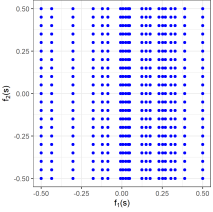
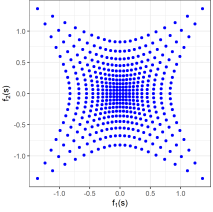
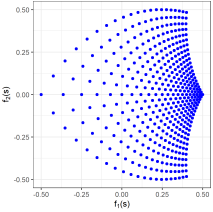
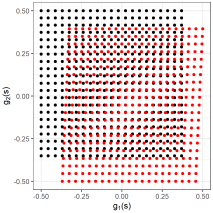
#### Algorithm 1: Parameter bootstrapping for uncertainty quantification

---

- 1 Find the REML estimate  $\hat{\boldsymbol{\theta}}$  by maximizing  $\mathcal{L}$  in (S2.1).
  - 2 Find the REML estimate  $\hat{\boldsymbol{\beta}}$  using (S2.2).
  - 3 Find the Cholesky factor  $\mathbf{L}$  of  $\hat{\boldsymbol{\Sigma}}_Z$ , that is, find  $\mathbf{L}$  such that  $\mathbf{L}\mathbf{L}' = \hat{\boldsymbol{\Sigma}}_Z$ .
  - 4 Decorrelate the data:  $\mathbf{Z}_0 = \mathbf{L}^{-1}(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}})$ .
  - 5 Generate an uncorrelated bootstrap sample  $\mathbf{Z}_{0,b}$  by sampling  $\mathbf{Z}_0$  with replacement.
  - 6 Create the correlated bootstrap sample  $\mathbf{Z}_b = \mathbf{L}\mathbf{Z}_{0,b} + \mathbf{X}\hat{\boldsymbol{\beta}}$ .
  - 7 Find the bootstrap estimate  $\hat{\boldsymbol{\theta}}_b$  and  $\hat{\boldsymbol{\beta}}_b$  from  $\mathbf{Z}_b$ .
  - 8 Repeat from step 5 for  $B$  times to create a bootstrap sample of size  $B$ . (For most problems,  $B \approx 1000$  should suffice.)
-



Table S1: Summary of the warping units. In each row, a warping unit is described. Parameters appearing in the functions composing the units are denoted using  $\theta$ .

Type of deformation function	Functional form	Usage	Visualization
Axial warping	$\tilde{\mathbf{f}}(\mathbf{s}) = \begin{pmatrix} \tilde{f}(s_1) \\ s_2 \end{pmatrix} \text{ or } \begin{pmatrix} s_1 \\ \tilde{f}(s_2) \end{pmatrix}, \text{ where}$ $\tilde{f}(s) = \sum_{i=1}^r w_i \phi_i(s), \text{ and where}$ $\phi_1(s) = s;$ $\phi_i(s) = \frac{1}{1 + \exp\{-\theta_1(s - \theta_2)\}}, i = 2, \dots, r.$	warp space along one of the axes	
Radial basis function	$\tilde{\mathbf{f}}(\mathbf{s}) = \mathbf{s} + w \exp\{-\theta_2 \ \mathbf{s} - \theta_1\ ^2\}(\mathbf{s} - \theta_1)$	expand or contract space locally around the center of the basis function	
Möbius transformation	$\tilde{\mathbf{f}}(\mathbf{s}) = \begin{pmatrix} \text{Re}(\phi(\mathbf{s})) \\ \text{Im}(\phi(\mathbf{s})) \end{pmatrix}, \text{ where}$ $\phi(\mathbf{s}) = \frac{\theta_1 z(\mathbf{s}) + \theta_2}{\theta_3 z(\mathbf{s}) + \theta_4};$ $z(\mathbf{s}) = s_1 + i s_2; \theta_1, \theta_2, \theta_3, \theta_4 \in \mathbb{C}$	move points around fixed points (usually in circular paths)	
Affine transformation	$\tilde{\mathbf{g}}(\mathbf{s}) = \mathbf{A}\mathbf{s} + \mathbf{d}$	align processes with respect to the first process (using shifts and rotations)	

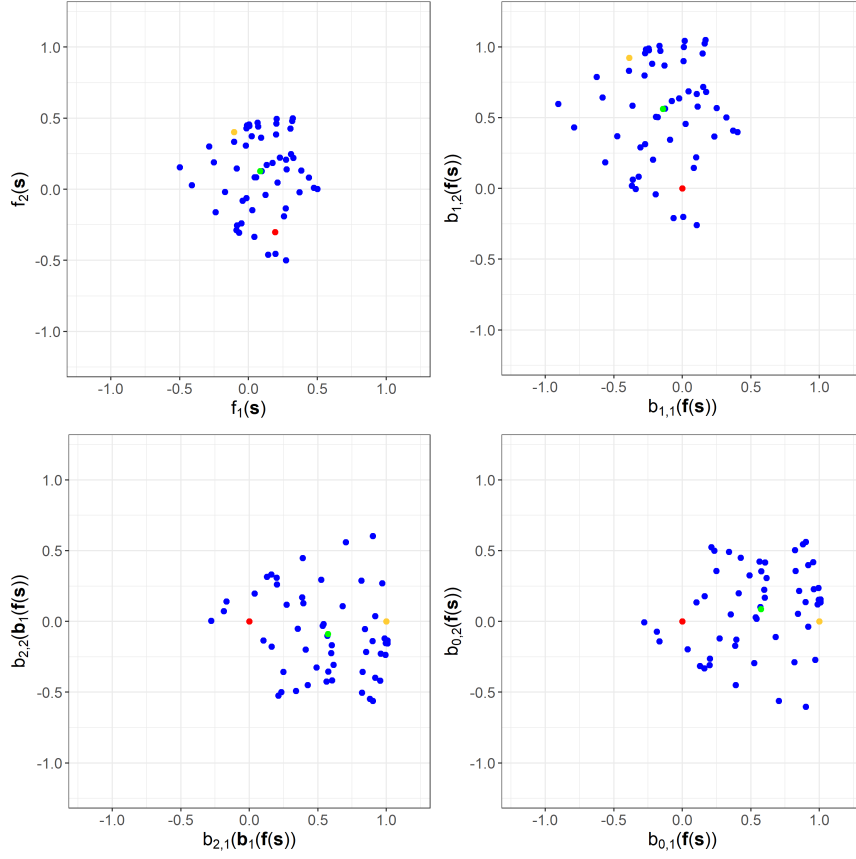


Figure S1: Illustration of the homogenizing function  $\mathbf{b}_0(\cdot)$ . Top left: A set of locations on the warped domain  $D$ , with the red, yellow, and green dots denoting  $\mathbf{f}(\mathbf{s}_k)$ ,  $\mathbf{f}(\mathbf{s}_l)$ , and  $\mathbf{f}(\mathbf{s}_m)$ , respectively. Top right: The locations on the scaled and shifted domain  $D_1$ . Bottom left: The locations on the scaled, shifted, and rotated domain  $D_2$ . Bottom right: The locations on the scaled, shifted, rotated, and reflected, domain  $D_3$ . This is the fixed frame of reference defined in Section 3.3.

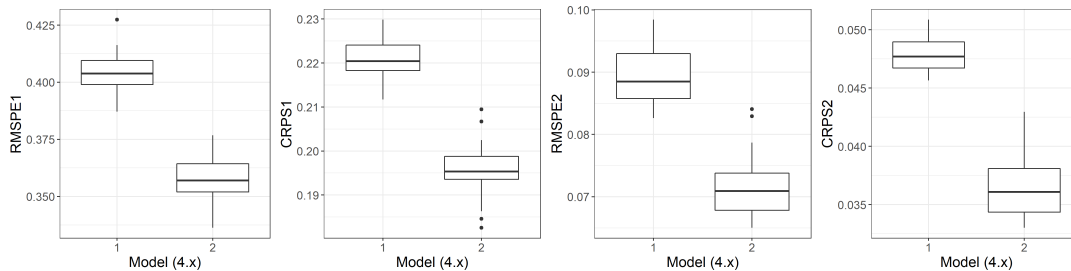


Figure S2: Boxplots of RMSPE and CRPS for both models across 30 simulations in Section 4.1. The left two panels correspond to  $\tilde{Y}_1(\cdot)$ , and the right two panels to  $\tilde{Y}_2(\cdot)$ .

Table S2: True values, estimates, and 95% bootstrap confidence intervals (CI), of the model parameters for the simulation study in Section 4.1, where the measurement locations are randomly sampled from  $G$ .

Parameters	True values	REML estimates	95% bootstrap CI
$\nu_{11}$	0.500	0.528	(0.235, 0.634)
$\nu_{22}$	1.500	1.226	(0.929, 1.664)
$\sigma_1$	1.000	1.040	(0.775, 1.329)
$\sigma_2$	0.900	0.932	(0.763, 1.388)
$\rho_{12}$	0.450	0.392	(0.321, 0.537)
$1/\tilde{a}$	0.329	0.412	(0.274, 0.556)
$\tau_1$	0.200	0.252	(0.178, 0.265)
$\tau_2$	0.100	0.097	(0.092, 0.108)
$\beta_{11}$	0.000	-0.232	(-0.947, 0.415)
$\beta_{21}$	0.000	0.047	(-0.762, 0.741)

Table S3: Five-fold cross-validation results, AIC, and the time required to fit, for the ocean temperatures at depths 0.5 meters and 318.1 meters for the study in Section 4.2, where data were missing at random.

	$T_{0.5}$		$T_{318.1}$		AIC	Time (s)
	RMSPE	CRPS	RMSPE	CRPS		
Model 4.2.1	0.0661	0.0306	0.0265	0.0129	-4791.1	1338.8
Model 4.2.2	0.0584	0.0188	0.0280	0.0136	-6246.4	2545.8
Model 4.2.3	0.0666	0.0198	0.0241	0.0123	-6403.4	4455.2

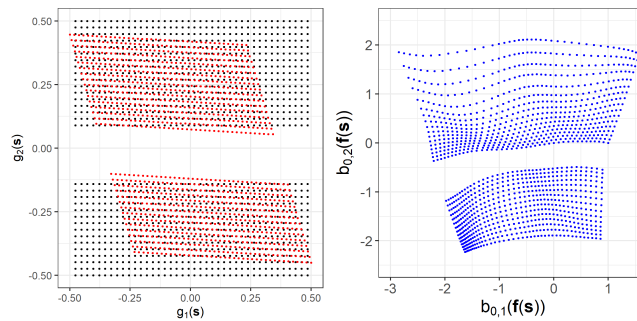


Figure S3: Measurement locations under the estimated aligning function (left panel) and the homogenized warping function (right panel) for the ocean-temperatures data set in Section 4.2, where data are missing in a block.

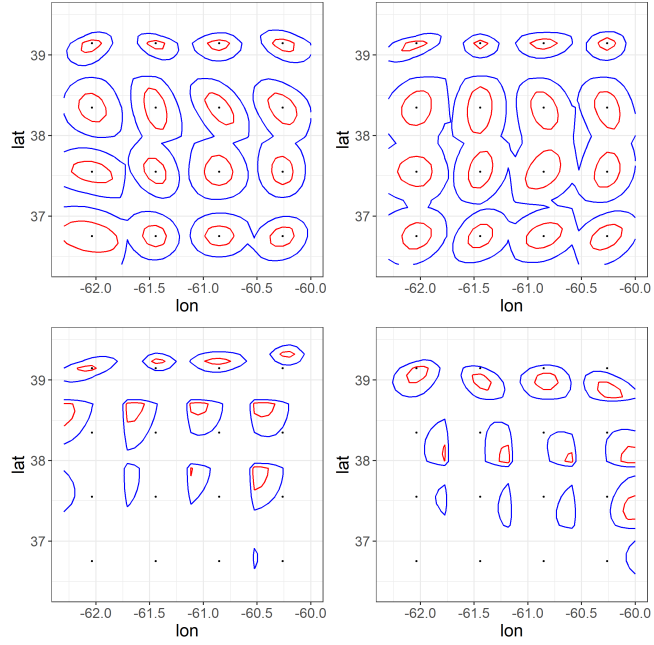


Figure S4: Contours illustrating the estimated covariance functions evaluated at a set of locations (black dots) for the data set in Section 4.2. Blue and red contours denote covariances equal to 0.4 and 0.8 times of  $\hat{\sigma}_{ij}^2$ , where  $\hat{\sigma}_{ij}^2 = \hat{\sigma}_i^2$ , for  $i = j$ , and  $\hat{\sigma}_{ij}^2 = \hat{\rho}_{ij}\hat{\sigma}_i\hat{\sigma}_j$ , for  $i \neq j$ , respectively. Top row: Marginal covariances of the first process,  $C_{11,G}(\cdot, \cdot)$  (left) and second process,  $C_{22,G}(\cdot, \cdot)$  (right). Bottom row: Cross-covariances of the first process with the second process,  $C_{12,G}(\cdot, \cdot)$  (left) and of the second process with the first process,  $C_{21,G}(\cdot, \cdot)$  (right). We see the estimated cross-covariances are asymmetric.

---

## S4 Additional Data Illustrations

### S4.1 Simulated Symmetric Nonstationary Data With A Missing Block

Following the simulation study in Section 4.1, we considered the case where the symmetric nonstationary data are missing in a block. This situation occurs often when observing environmental variables (for example, clouds could prevent a remote sensing instrument from collecting data over a large region). As in the study in Section 4.1, we sampled 1000 measurement locations at random 30 times, but this time on  $G \setminus G_0$ , where  $G_0 \equiv [-0.28, -0.08] \times [-0.48, -0.28]$  (i.e., the block of data was omitted for both processes). Model 4.1.1 and Model 4.1.2 were then fitted to the data. Figure S5 shows the true simulated fields, the predictions, and the prediction standard errors from the two models. From Figure S5, we see that the predictions from the DCSM recover the salient features in the true fields despite the relatively large gap. The DCSM also produces relatively lower prediction standard errors in the unobserved region than the stationary parsimonious Matérn model. Table S4 shows the RMSPE and CRPS from the two models when predicting the latent process at the grid locations in  $G_0$  and, again, it illustrates the improvement in RMSPE and CRPS that can be achieved when accounting for complex nonstationary properties of the process, even when the data have large gaps.

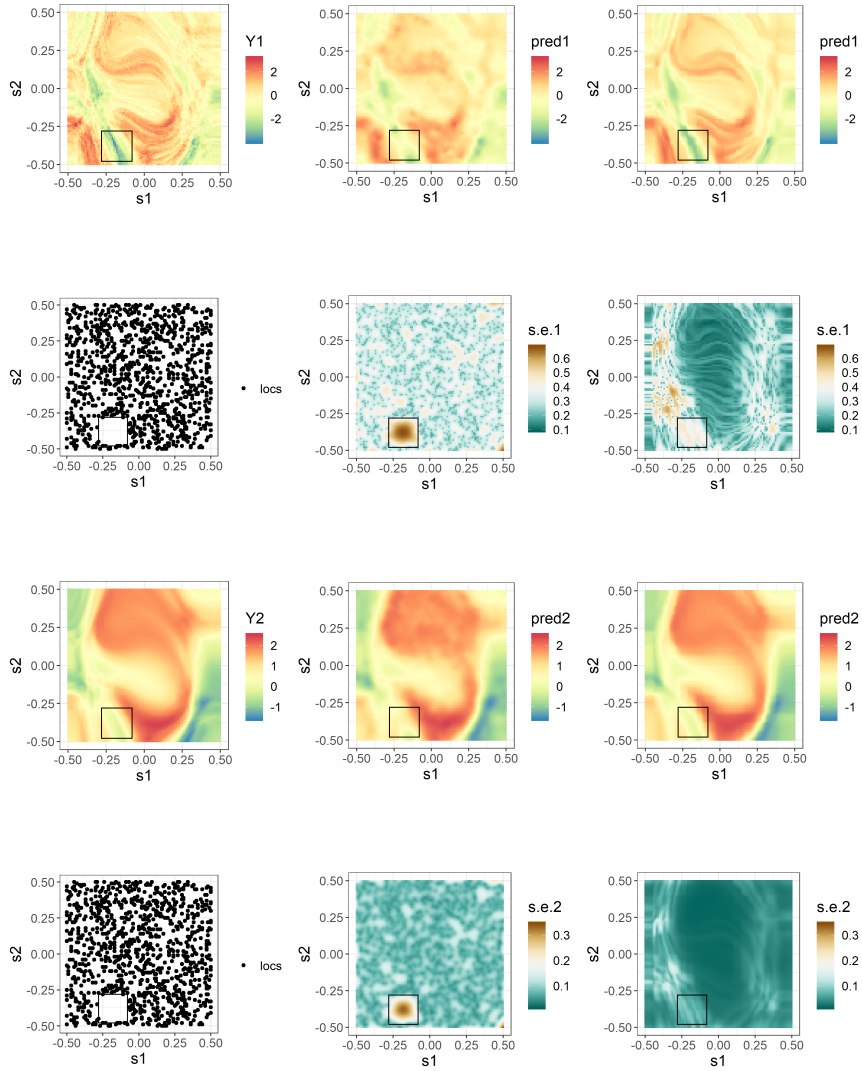


Figure S5: Same as Figure 1, but where the measurement locations are randomly sampled from  $G \setminus G_0$ , and where the hold-out region  $G_0$  is enclosed by the black square.

Table S4: Average hold-out validation results, AIC, and the time required to fit, for the simulation study in Section S4.1, where the measurement locations are randomly sampled 30 times from  $G \setminus G_0$ .

	$\tilde{Y}_1(\cdot)$		$\tilde{Y}_2(\cdot)$		AIC	Time (s)
	RMSPE	CRPS	RMSPE	CRPS		
Model 4.1.1	1.137	0.675	0.453	0.274	833.2	816.9
Model 4.1.2	0.713	0.424	0.261	0.167	310.4	1533.0

## S4.2 Simulated Asymmetric Nonstationary Data

We now demonstrate inference for a bivariate DCSM from simulated asymmetric nonstationary data. We simulated bivariate data from a bivariate Gaussian DCSM  $\tilde{\mathbf{Y}}(\cdot)$  with constant mean (i.e.,  $q = 1$  and  $\mathbf{x}(\cdot) = x(\cdot) = 1$ ) in (3.5), and cross-covariance matrix function as in (3.3). The data were simulated on the  $101 \times 101$  grid defined on the geographic domain  $G$ . Now the *shared warping function*  $\mathbf{f}(\cdot)$  is a composition of axial warping units, a single resolution radial basis function unit, and a Möbius transformation unit, while the *aligning function*  $\mathbf{g}_2(\cdot)$  is an affine transformation. (Recall from Section 3.1 that we fix  $\mathbf{g}_1(\cdot)$  to be the identity map.) On the warped domain, we use a stationary bivariate parsimonious Matérn model. As in Section 4.1, we randomly sampled without replacement 1000 locations from the grid and used them as measurement locations.

We compared the predictive performance of the following five models on the



$101 \times 101$  grid on  $G$ .

- Model S4.2.1: A bivariate, stationary, symmetric, parsimonious Matérn model.
- Model S4.2.2: A bivariate, marginally stationary, asymmetric model with  $\mathbf{f}(\cdot)$  the identity map and the aligning function  $\mathbf{g}_2(\cdot)$  an affine transformation (as described in Proposition 4), with Model S4.2.1 on the warped domain.
- Model S4.2.3: A univariate DCSM for each of the processes, with the warping function  $\mathbf{f}(\cdot)$  a composition of axial warping units, a single-resolution radial basis function, and a Möbius transformation unit, with the Matérn covariance model on the warped domain.
- Model S4.2.4: A bivariate symmetric DCSM, with the warping function as in Model S4.2.3, with Model S4.2.1 on the warped domain.
- Model S4.2.5: A bivariate asymmetric DCSM, with the aligning function  $\mathbf{g}_2(\cdot)$  as in Model S4.2.2, the warping function as in Model S4.2.3, and Model S4.2.1 on the warped domain. This is the model from which the data were simulated.

Figure S6 shows the true simulated fields and the predicted fields from Model S4.2.2, Model S4.2.4, and Model S4.2.5. From Figure S6, we can see that Model S4.2.2 smooths out certain features (similar to the symmetric case), while Model

S4.2.5 is able to reproduce sharper features than Model S4.2.3, illustrating that both nonstationarity and asymmetry could be important when modeling multivariate spatial processes. Figure S6 also shows the prediction standard errors for Model S4.2.2, Model S4.2.4, and Model S4.2.5. As in Section 4.1, we see that while there is no pattern in the prediction-standard-error map for Model S4.2.2, the DCSMs produce prediction standard errors that are reflective of the processes' local anisotropies and scales.

We can also compare, respectively, the estimated aligning function and the estimated warping function in Model S4.2.5 to the true aligning function and the true warping function in Model S4.2.5. Figure S7 shows the measurement locations under the true aligning and warping functions, and the measurement locations under the estimated aligning and warping functions. We see that the estimated aligning function generates a shift towards the east direction, which is similar to the true aligning function, which generates a shift towards the southeast direction. One can interpret that the second process needs to be shifted eastwards to align with the first process. The estimated warping function also retains important features of the true warping function, such as the contraction in the middle part of the domain. This can be interpreted that on the original domain  $G$ , the scale parameter in the middle region is smaller than the scale parameter in the boundary region.

In a similar manner to Section 4.1, we repeated the procedure of randomly

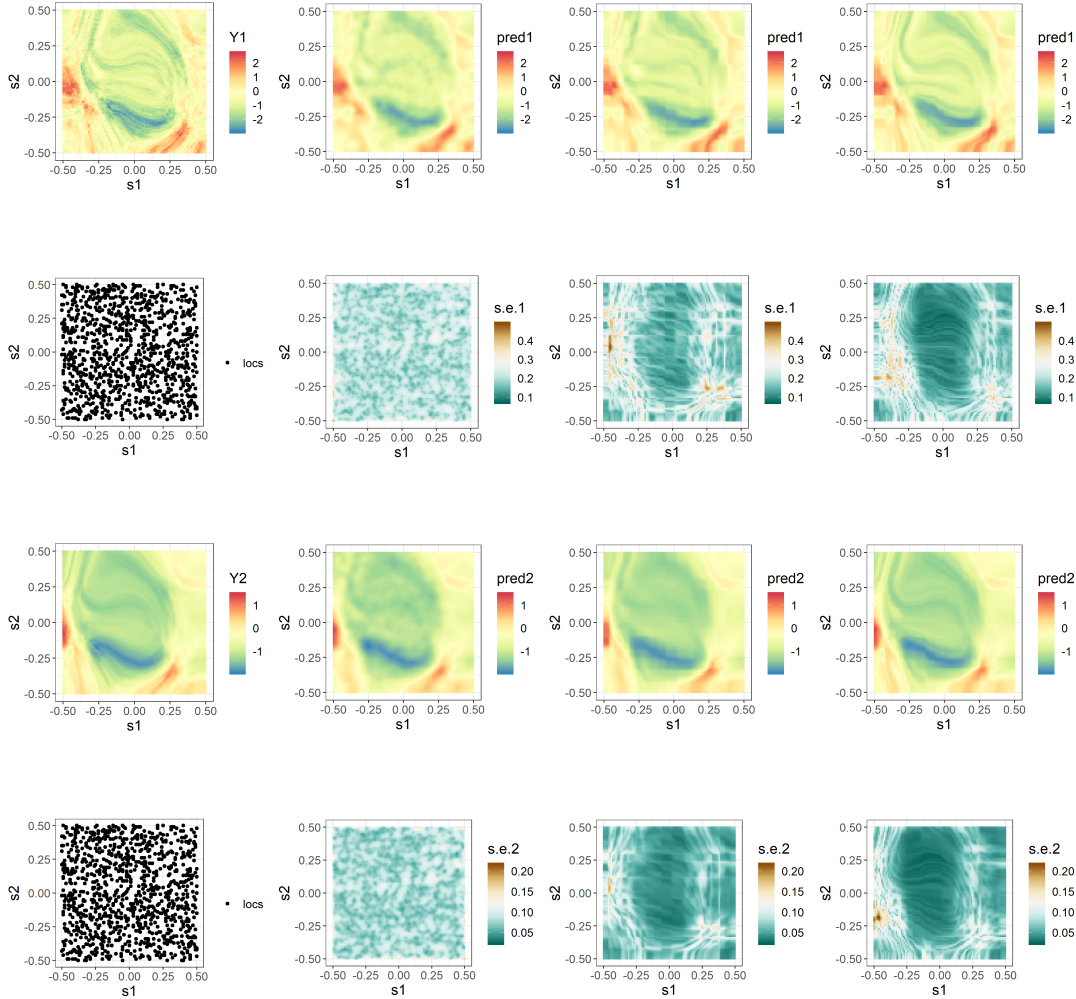


Figure S6: Comparison of predictions and prediction standard errors when using an asymmetric, stationary, parsimonious Matérn model (Model S4.2.2); a symmetric DCSM (Model S4.2.4); and an asymmetric DCSM (Model S4.2.5) in the study of Section S4.2, where measurement locations were randomly sampled without replacement from  $G$ . First row: The process  $\tilde{Y}_1(\cdot)$  (first panel); predictions obtained using Model S4.2.2 (second panel), Model S4.2.4 (third panel) and Model S4.2.5 (fourth panel). Second row: Locations of the measurement of  $\tilde{Y}_1(\cdot)$  (first panel); prediction standard errors obtained when using Model S4.2.2 (second panel), Model S4.2.4 (third panel) and Model S4.2.5 (fourth panel). Third and fourth rows: Analogous to the first and second rows, respectively, for the process  $\tilde{Y}_2(\cdot)$ .

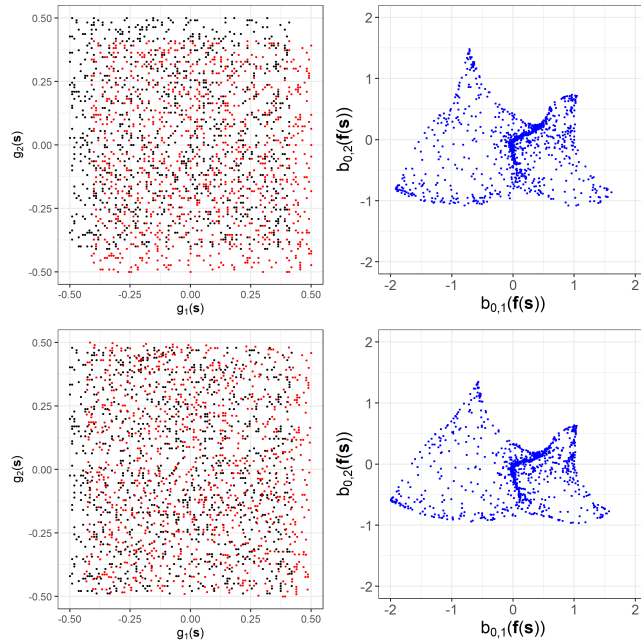


Figure S7: Measurement locations under the aligning functions and the homogenized warping functions for the data set in Section S4.2. Top row: Measurement locations under the true aligning function (left panel) and the true warping function (right panel). Bottom row: Measurement locations under the estimated aligning function (left panel) and the estimated warping function (right panel) using Model S4.2.5.

sampling 1000 locations 30 times from  $G$  and computing predictive diagnostics. Table S5 displays the results from the cross-validation study, and Figure S8 shows the boxplots of the RMSPE and CRPS for the models across the 30 simulations. Model S4.2.5, which considers both nonstationarity and asymmetry, produces the best predictions in terms of lowest RMSPE and CRPS, while the symmetric nonstationary model (Model S4.2.4) as well as the asymmetric stationary model (Model S4.2.2) are seen to yield slight improvements over the conventional symmetric, stationary model (Model S4.2.1). Surprisingly, even

Table S5: Average hold-out validation results, AIC, and the time required to fit, for the simulation study in Section S4.2, where the measurement locations are randomly sampled 30 times from  $G$ .

	$\tilde{Y}_1(\cdot)$		$\tilde{Y}_2(\cdot)$		AIC	Time (s)
	RMSPE	CRPS	RMSPE	CRPS		
Model S4.2.1	0.304	0.169	0.091	0.048	386.3	815.3
Model S4.2.2	0.291	0.162	0.086	0.046	195.4	1188.6
Model S4.2.3	0.314	0.177	0.088	0.046	44.2	2025.7
Model S4.2.4	0.287	0.159	0.087	0.045	-3.8	1530.5
Model S4.2.5	0.269	0.149	0.080	0.041	-242.5	2654.1

when accounting for nonstationarity in each process, the decoupled univariate model (Model S4.2.3) yields the worst predictions, showing the importance of a multivariate model.

### S4.3 Simulated Data with Misspecified Warping Functions

In this section, we consider cases where the deformations are misspecified. Specifically, we present two cases: First, we consider data simulated from a bivariate, symmetric, stationary covariance function. Second, we consider data simulated from a bivariate, asymmetric, nonstationary covariance function with a misspecified warping function.

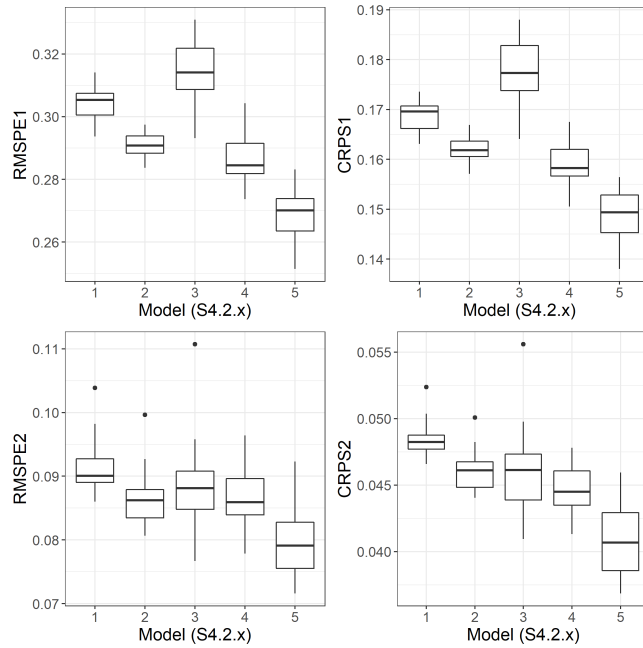


Figure S8: Boxplots of RMSPE and CRPS for the five models across 30 simulations in Section S4.2. The top row corresponds to  $\tilde{Y}_1(\cdot)$ , and the bottom row to  $\tilde{Y}_2(\cdot)$ .

### S4.3.1 Simulated Symmetric Stationary Data

In this section, we demonstrate the use of a bivariate DCSM with simulated symmetric stationary data. We simulated bivariate data from a bivariate, stationary, symmetric, parsimonious Matérn model. The data were simulated on a  $101 \times 101$  grid defined on the geographic domain  $G$ . We randomly sampled without replacement 1000 locations from the grid and used them as measurement locations.

We first examined the ability of the DCSM to retrieve the true warping function (in this case, the identity function). We used the same architecture in

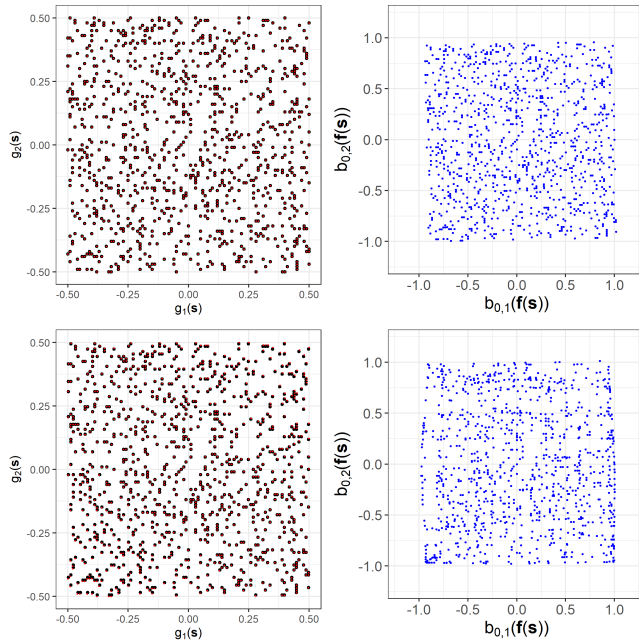


Figure S9: Measurement locations under the aligning functions and the homogenized warping functions for the data set in Section S4.3.1. Top row: Measurement locations under the true aligning function (left panel) and the true warping function (right panel). Bottom row: Measurement locations under the estimated aligning function (left panel) and the estimated warping function (right panel).

the DCSM as in Model S4.2.5 in Section S4.2. Figure S9 shows the measurement locations under the true aligning and warping function, and the measurement locations under the estimated aligning and warping function. We observe that both the aligning function and the warping function have been correctly estimated to be approximately the identity functions.

We also compared the predictive performance of the DCSM with the true model.

- Model S4.3.1.1: The bivariate, stationary, symmetric, parsimonious Matérn

Table S6: Hold-out validation results, AIC, and the time required to fit, for the simulation study in Section S4.3.1, where the measurement locations are randomly sampled from  $G$ .

	$\tilde{Y}_1(\cdot)$		$\tilde{Y}_2(\cdot)$		AIC	Time (s)
	RMSPE	CRPS	RMSPE	CRPS		
Model S4.3.1.1	0.334	0.188	0.097	0.054	304.1	840.6
Model S4.3.1.2	0.346	0.195	0.108	0.060	437.1	2714.9

model. This is the model from which the data were simulated.

- Model S4.3.1.2: The same bivariate DCSM as Model S4.2.5 in Section S4.2.

Table S6 shows the cross-validation results of the simulation study. As expected, Model S4.3.1.2 produces worse RMSPE and CRPS than Model S4.3.1.1, but it does not perform much worse. Given that we use very complex deformations in Model S4.3.1.2, this shows that the DCSM is quite robust to overfitting.

### S4.3.2 Simulated Asymmetric Nonstationary Data with a Misspecified Warping Function

In this section, we demonstrate the use of a DCSM with simulated asymmetric nonstationary data with a misspecified warping function. We simulated bivariate data from a bivariate Gaussian DCSM in a manner similar to Section S4.2, but now we use a different warping function,  $\mathbf{f}(\mathbf{s}) = \mathbf{o} + (\mathbf{s} - \mathbf{o}) \|\mathbf{s} - \mathbf{o}\|$  (Fouedjio et al., 2015), where  $\mathbf{o} = (0, 0)'$ . The data were simulated on a  $101 \times 101$  grid defined on the geographic domain  $G$ . We randomly sampled without replacement 1000



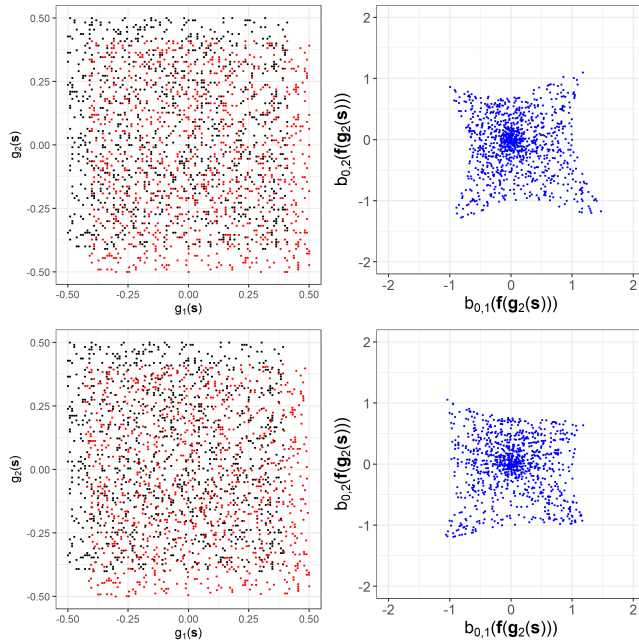


Figure S10: Measurement locations under the aligning functions and the homogenized warping functions for the data set in Section S4.3.2. Top row: Measurement locations under the true aligning function (left panel) and the true warping function (right panel). Bottom row: Measurement locations under the estimated aligning function (left panel) and the estimated warping function (right panel).

locations from  $G$  and used them as measurement locations.

As in Section S4.3.1, we examined the ability of the DCSM to retrieve the true warping function using the same model as Model S4.2.5 in Section S4.2. Figure S10 shows the measurement locations under the true aligning and warping function, and the measurement locations under the estimated aligning and warping function. We observe that the DCSM has correctly estimated the south-eastern shift in the aligning function, and the contraction in the middle region of the warped domain.

#### S4.4 Simulated Trivariate Symmetric Nonstationary Data

In the previous data illustrations, we considered bivariate spatial data. In this section, we consider trivariate spatial data. We simulated trivariate data from a Gaussian DCSM,  $\tilde{\mathbf{Y}}(\cdot)$ , with constant mean. The data were simulated on an equally spaced  $51 \times 51$  grid of the geographic domain,  $G \equiv [-0.5, 0.5] \times [-0.5, 0.5]$  resulting in 2601 data. 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. On the warped domain, we modeled the covariances using a trivariate stationary, isotropic parsimonious Matérn model. As in Section 4.1, we randomly sampled 1000 locations from the grid and used these as measurement locations.

We compared the predictive performance of the trivariate stationary parsimonious Matérn model (Model S4.4.1) to those of the trivariate DCSM (Model S4.4.2) by calculating the cross-validated predictive performance at the remaining 1601 locations using the RMSPE and the CRPS. Table S7 summarizes the results from this study. From this table, it is clear that, similar to the bivariate case in Section 4.1, there is a large improvement in RMSPE and CRPS when using the DCSM (Model S4.4.2) over the stationary parsimonious Matérn model (Model S4.4.1). The visualization of the estimated warping function is given in Figure S11. We can see that, similar to the bivariate case, the estimated warping function has retrieved important features such as the contraction in the middle

Table S7: Hold-out validation results, AIC, and the time required to fit, for the simulation study in Section S4.4, where the measurement locations are randomly sampled from  $G = [-0.5, 0.5] \times [-0.5, 0.5]$ .

	$\tilde{Y}_1(\cdot)$		$\tilde{Y}_2(\cdot)$		$\tilde{Y}_3(\cdot)$		AIC	Time (s)
	RMSPE	CRPS	RMSPE	CRPS	RMSPE	CRPS		
Model S4.4.1	0.309	0.171	0.099	0.053	0.035	0.019	-1924.0	1385.4
Model S4.4.2	0.276	0.148	0.077	0.040	0.027	0.015	-2865.0	9047.0

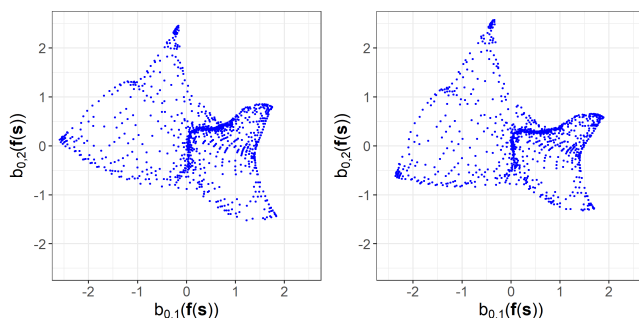


Figure S11: Measurement locations under the true (left panel) and estimated (right panel) homogenized warping functions for the data set in Section S4.4.

part of the domain. However, we find that more iterations are needed, thus more computing time, to train the DCSM in the trivariate-data example.

### S4.5 Modeling Maximum and Minimum Temperatures in Southwestern United States

In this section, we consider monthly maximum and minimum temperatures (which are calculated by averaging daily maximum and minimum temperatures

over the month) in July 2018 at 909 weather stations over four southwestern states in the United States: Utah, Colorado, Arizona, and New Mexico. The data were extracted from the Global Summary of the Month data set of the National Oceanic and Atmospheric Administration<sup>1</sup>.

We compared the performance of the bivariate DCSMs to that of bivariate stationary Matérn models. Elevation is a potential covariate when modeling temperature. We considered two trend models, namely one with only an intercept (in which case  $Y_1(\cdot)$  and  $Y_2(\cdot)$  are highly nonstationary) and one with an intercept and elevation as a covariate (in which case  $Y_1(\cdot)$  and  $Y_2(\cdot)$  can be expected to be nonstationary but less so). Also, the maximum and minimum temperatures in a given month (here July 2018) can be treated as symmetric spatial processes since their interaction can be expected to be highly co-located. Thus, asymmetry of the cross-covariance matrices was not modeled in this data illustration. We considered the following four models:

- Model S4.5.1: A bivariate, stationary, parsimonious Matérn covariance model, and with only an intercept in the trend.
- Model S4.5.2: A bivariate, symmetric DCSM with the parsimonious bivariate Matérn covariance on the warped domain, and with only an intercept in the trend
- Model S4.5.3: A bivariate, stationary, parsimonious Matérn covariance

---

<sup>1</sup><https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00946>

model, and with an intercept and elevation as trend.

- Model S4.5.4: A bivariate, symmetric DCSM with the parsimonious bivariate Matérn covariance on the warped domain, and with an intercept and elevation as trend.

We used the same general architecture for the warping function of the bivariate DCSM as was used in the simulation study in Section 4.1.

We first examined the predictive performance of the four models when the data were missing at random by performing a five-fold cross-validation. We randomly chose 905 stations from the 909 stations for the study, and randomly divided the 905 stations into five groups of 181 stations each in order to carry out a five-fold cross-validation. Table S8 shows the results of this five-fold cross-validation study. We see a slight improvement in the predictive performance of the DCSMs over the corresponding stationary covariance models, on the order of 3–5% for RMSPE and CRPS. We also see a substantial improvement when elevation is included in the trend model, showing that addressing nonstationarity in the mean function is helpful in this application.

We next considered hold-out validation, where the data are missing in a block, and we held out 131 stations lying between 36°N–39°N and 104°W–108°W from all the 909 stations. The western part of the hold-out region is a mountainous area that extends into the non-hold-out region, while the eastern part of the hold-out region is an area with lower elevation. We then fitted the four models

Table S8: Five-fold cross-validation results, AIC, and the time required to fit, for the maximum and minimum temperature data in the southwestern USA for the study in Section S4.5, where data were missing at random.

	$T_{max}$		$T_{min}$		AIC	Time (s)
	RMSPE	CRPS	RMSPE	CRPS		
Model S4.5.1	3.17	1.74	2.44	1.33	7140.8	359.4
Model S4.5.2	3.09	1.69	2.31	1.26	7003.5	791.8
Model S4.5.3	1.07	0.58	2.02	1.14	5475.5	413.2
Model S4.5.4	1.04	0.57	1.98	1.11	5398.5	766.3

to the remaining 778 stations. Table S9 displays the hold-out validation results for predicting the maximum and minimum temperatures at the 131 hold-out stations, and Figure S12 shows the maps of predictions and prediction standard errors for Model S4.5.1 and Model S4.5.2.

From Table S9, the predictive performances of the bivariate DCSMs show a huge improvement in RMSPE and CRPS, on the order of 30% over those of the bivariate stationary covariance models when predicting maximum temperature. The visualization in Figure S12 further illustrates the utility of using DCSMs: With the intercept-only structure (i.e., constant mean) in the mean, the bivariate DCSM can predict lower temperatures on the western part of the hold-out region (that is, the mountainous areas), whereas the stationary model reverts to the constant mean. The DCSM also produces lower prediction standard errors over

Table S9: Hold-out validation results, AIC, and the time required to fit, for the maximum and minimum temperature data in the southwestern USA for the study in Section S4.5, where data were missing in a block shown in Figure S12.

	$T_{max}$		$T_{min}$		AIC	Time (s)
	RMSPE	CRPS	RMSPE	CRPS		
Model S4.5.1	4.66	2.59	3.90	2.20	7610.1	418.7
Model S4.5.2	3.84	2.19	2.57	1.51	7474.6	915.0
Model S4.5.3	1.54	0.87	2.11	1.22	5858.4	475.0
Model S4.5.4	1.06	0.59	1.99	1.16	5787.2	898.5

the hold-out region than the stationary model. These results corroborate those from the simulation study in Section 4.1, which showed that bivariate DCSMs can be useful even when data are missing over a large region. Figure S13 shows the estimated warping function in Model S4.5.2.

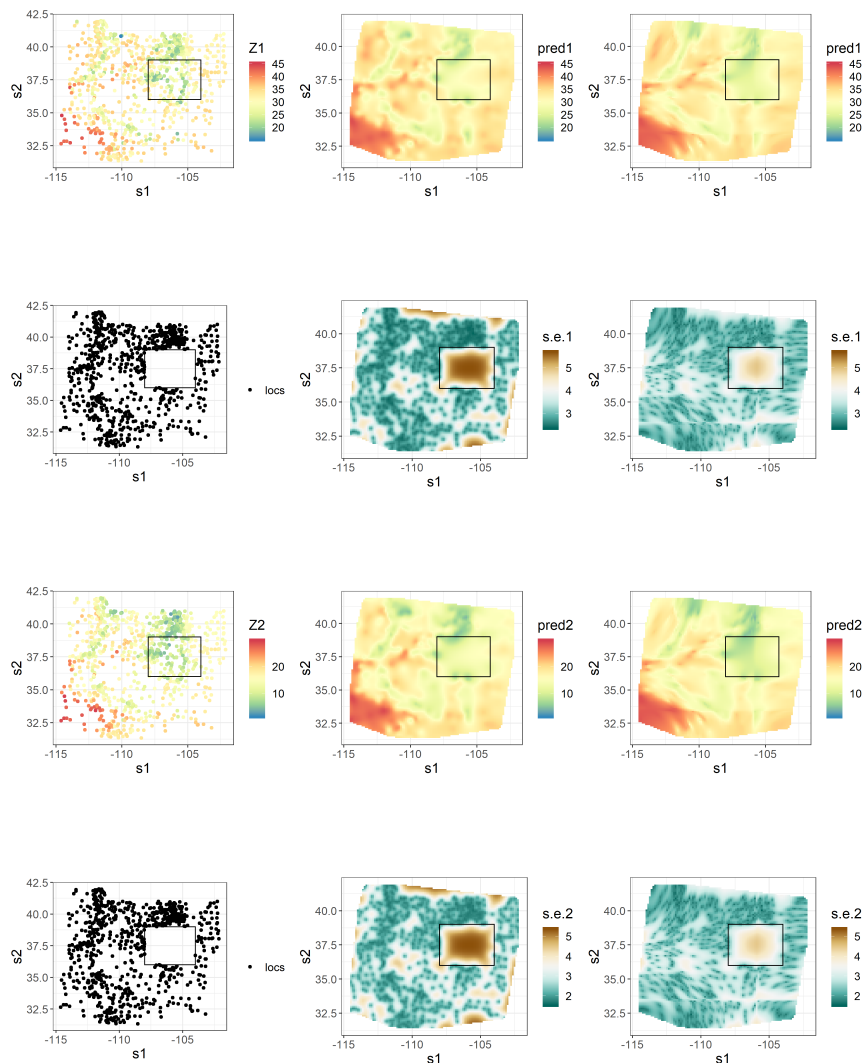


Figure S12: Comparison of predictions and prediction standard errors when using a bivariate symmetric, stationary, parsimonious Matérn model with constant mean (Model S4.5.1), and a bivariate symmetric DCSM with constant mean (Model S4.5.2). Data were left out of the region enclosed by the black rectangle. First row: Maximum temperature observations,  $Z_1$  (left panel); predictions obtained using Model S4.5.1 (center panel), and Model S4.5.2 (right panel). Second row: Locations of the measurement of  $Z_1$  (left panel); prediction standard errors obtained when using Model S4.5.1 (center panel), and Model S4.5.2 (right panel). Third and fourth rows: Analogous to the first and second rows, respectively, for the minimum temperature,  $Z_2$ .



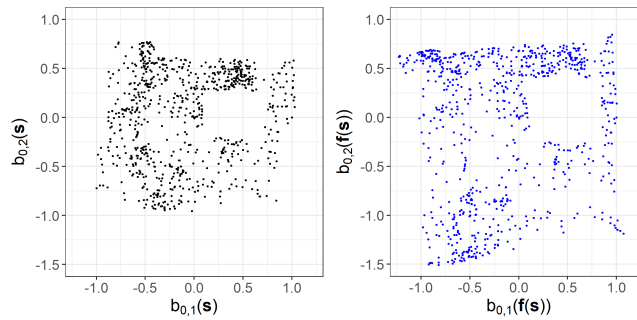


Figure S13: Measurement locations on the geographical domain and under the estimated homogenized warping function for the data set in Section S4.5.

## References

- Cressie, N. and Lahiri, S. N. (1993). The asymptotic distribution of REML estimators. *Journal of Multivariate Analysis*, 45:217–233.
- Fouedjio, F., Desassis, N., and Romary, T. (2015). Estimation of space deformation model for non-stationary random functions. *Spatial Statistics*, 13:45–61.
- 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.