Supplemenatry Material for Estimation of Spatial Deformation for Nonstationary Processes via Variogram Alignment

Ghulam A. Qadir[1], Ying Sun[1] and Sebastian Kurtek[2]

This file contains the proofs of Property 1 and 2, discussion on the robustness of the method under different divisions of the domain, some additional simulation studies and some additional results from the data application.

S1 Proofs of Properties 1 and 2

Property 1: For any $i \in \{1, 2, ..., k\}$, if $(\mathbf{s}, \mathbf{s}') \in \mathcal{G}_i \times \mathcal{G}_i \implies \mathcal{L}(\mathbf{s}, \mathbf{s}') = \{\mathcal{G}_i\} \implies \mathcal{P}(i, \mathbf{s}, \mathbf{s}') = \|\mathbf{s} - \mathbf{s}'\| \implies \mathcal{W}_i(\mathbf{s}, \mathbf{s}') = \frac{\|\mathbf{s}-\mathbf{s}'\|}{\|\mathbf{s}-\mathbf{s}'\|} = 1$ and $\mathcal{W}_j(\mathbf{s}, \mathbf{s}') = 0$ for all $j \neq i \implies \phi(\mathbf{s}, \mathbf{s}') = \phi_i(\|\mathbf{s} - \mathbf{s}'\|)$.

Property 2: If the process $\{X(\mathbf{s}) : \mathbf{s} \in \mathcal{G} \subset \mathbb{R}^{d_\mathcal{G}}\}$ is second-order stationary, then the regional distance warping functions are *identity functions* (i.e., $\phi_i(\|\mathbf{s} - \mathbf{s}'\|) = \|\mathbf{s} - \mathbf{s}'\|$)

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

as the sum of the weights $\mathcal{W}_i$ over $\mathcal{G}_i \in \mathcal{L}(\mathbf{s}, \mathbf{s}')$ is guaranteed to be 1.

S2 Extended Simulation: Continuously Varying Nonstationarity

In this section, we extend the simulation study presented in the main manuscript to explore the improvement in prediction performance by the proposed nonstationary method over the

[1]CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. E-mail: ghulam.qadir@kaust.edu.sa; ying.sun@kaust.edu.sa

[2]Department of Statistics, The Ohio State University, Columbus, OH 43210, USA. E-mail: kurtek.1@stat.osu.edu

stationary method when the true process departs from the locally stationary behavior to continuously varying nonstationarity. To this end, we repeat the exact same simulation study as presented in the main manuscript for different values of λ. In particular, we report prediction performance of both methods for $\lambda = 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50,$ and 0.60 in Figures 1, 2, 3, 4, 5, 6, 7, 8 and 9, respectively. Here, increasing the value of λ increases the continuity of the true spatially varying range parameter of the process, as shown in Figures 1(a), 2(a), 3(a), 4(a), 5(a), 6(a), 7(a), 8(a), and 9(a).

In terms of RMSE, NMSE and MAE, the stationary model provides lower scores than those provided by the nonstationary model for all the values of λ. However, the differences in scores are very small and the predicted values are very similar. On the other hand, the corresponding assessment metrics mCRPS, mLogS, G, as well as the accuracy and average width plots, show a varying degree of improvement by the nonstationary method over the stationary method for different values of λ. For smaller values of $\lambda = 0.15, 0.20, 0.25,$ and 0.30, where the true process behavior is not very far from local stationarity, the nonstationary method is clearly advantageous over the stationary method, in terms of a collective assessment of mCRPS, mLogS, G, and the accuracy and the average width plots (see Figures 1, 2, 3 and 4). As we proceed to $\lambda = 0.35$, where there is a significant departure from local stationarity, the nonstationary method still provides improved prediction performance in terms of mLogS, G and the accuracy plot. However, the magnitude of improvement is not very large (see Figure 5). For $\lambda = 0.35$, the average width of the $p-$PI estimated using the nonstationary method is slightly higher than the stationary method (see Figure 5(i)), but when considered in conjuction with G and the accuracy plot (see Figures 5(g) and 5(h)), the $p-$PI estimated using the nonstationary method is still preferred as there is a notable gain

in coverage accuracy. Similarly, for higher values of $\lambda = 0.40, 0.45$, and 0.50, improvements of the nonstationary method over the stationary method become less apparent (see Figures 6, 7 and 8). For the highest considered value of $\lambda = 0.6$, where the true process exhibits strong continuosly varying nonstationarity, the benefit of the nonstationary method over the stationary method becomes unclear as they seem to exhibit comparable performance (see Figure 9).

The above results suggests that the proposed nonstationary method, which is based on the regional stationarity assumption, can outperform the stationary method in terms of prediction when the true process closely exhibits local stationarity. Furthermore, the proposed nonstationary method is beneficial over the stationary method even when there is a mild departure from local stationarity to continously varying nonstationary, though the benefit tends to decrease as the degree of departure increases. When the process exhibits strong continuously varying nonstationarity, the proposed nonstationary method is not advantageous over the stationary method. In this case, a finer partition of the domain can be considered to better capture the nonstationarity, but improvement in prediction performance can only be expected when there is strong heterogeneity across different subregions.

S3 Simulation: Estimation of True Deformation Function

In this section, we simulate a deformation-based nonstationary process and evaluate how well the proposed method can approximate the true underlying deformed surface. In particular, we simulate a realization of a zero-mean Gaussian process X on a 70×70 regularly spaced grid in the spatial domain $\mathcal{G} = [0, 2]^2$. The nonstationary covariance function for the simulation is specified as: $\text{Cov}\{X(\mathbf{s}), X(\mathbf{s}')\} = \text{C}_{\mathcal{D}}\{\|\theta(\mathbf{s}) - \theta(\mathbf{s}')\|\}$, $\mathbf{s}, \mathbf{s}' \in \mathcal{G}$, where $\text{C}_{\mathcal{D}}(\cdot)$ is chosen

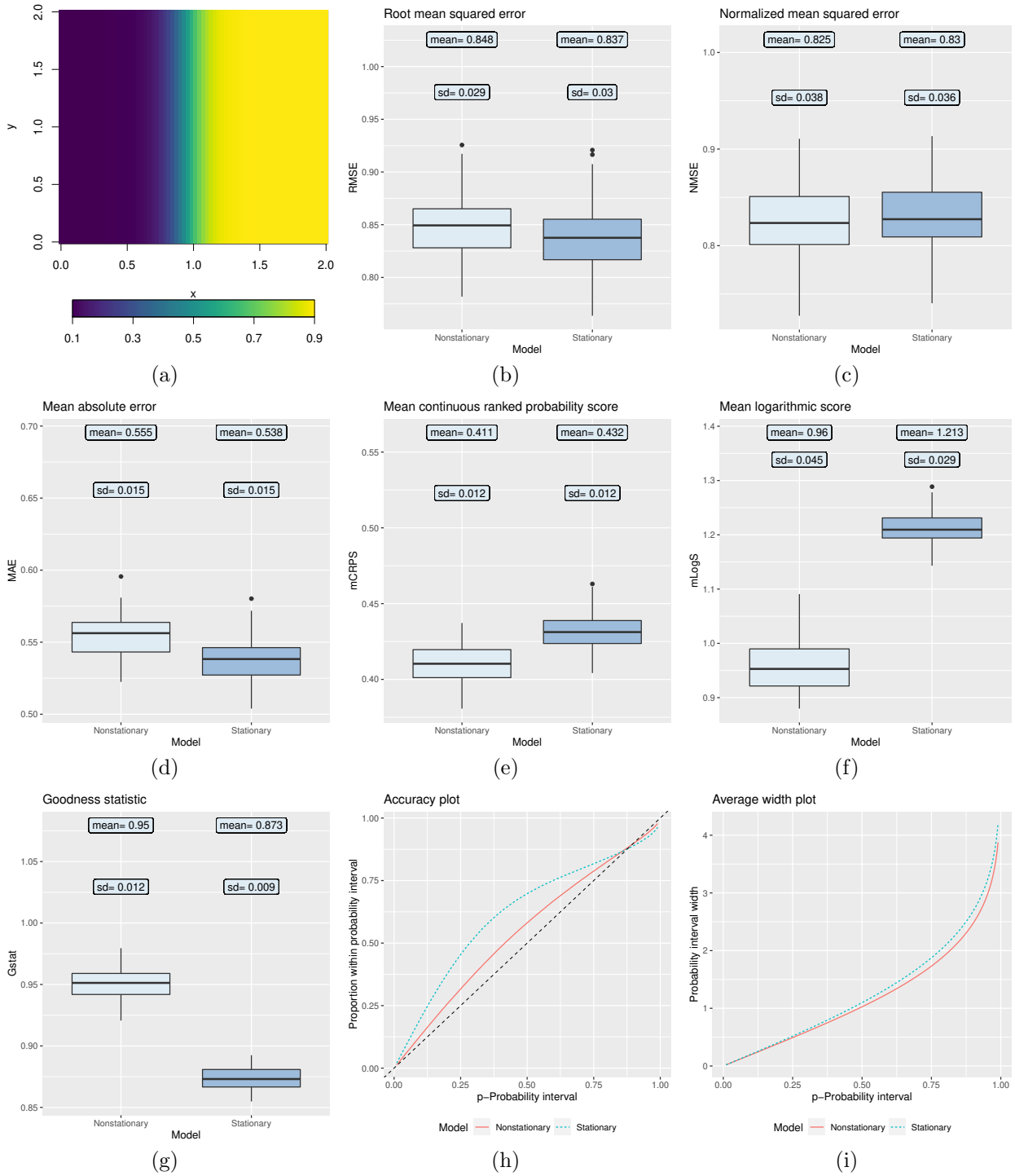


Figure 1: (a) True spatially varying range parameters (for $\lambda = 0.15$). Boxplots of (b) root mean squared error, (c) normalized mean squared error, (d) mean absolute error, (e) mean continuous ranked probability score, (f) mean logarithmic score, and (g) goodness statistic. (h) Accuracy plot. (i) Average width plot. The results are based on 100 simulation runs. Above the boxplots, we report the mean and standard deviations.

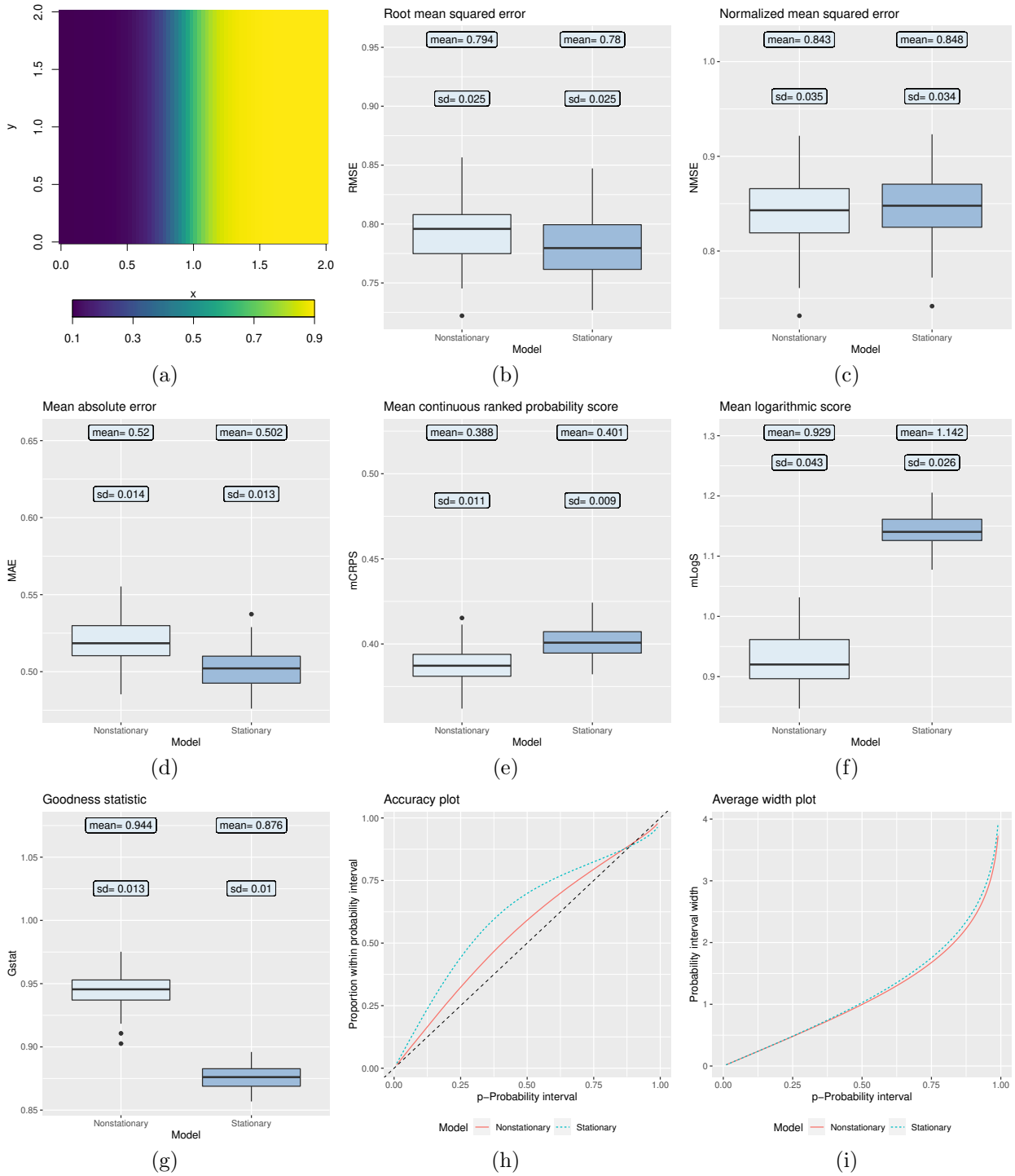


Figure 2: (a) True spatially varying range parameters (for $\lambda = 0.20$). Boxplots of (b) root mean squared error, (c) normalized mean squared error, (d) mean absolute error, (e) mean continuous ranked probability score, (f) mean logarithmic score, and (g) goodness statistic. (h) Accuracy plot. (i) Average width plot. The results are based on 100 simulation runs. Above the boxplots, we report the mean and standard deviations.

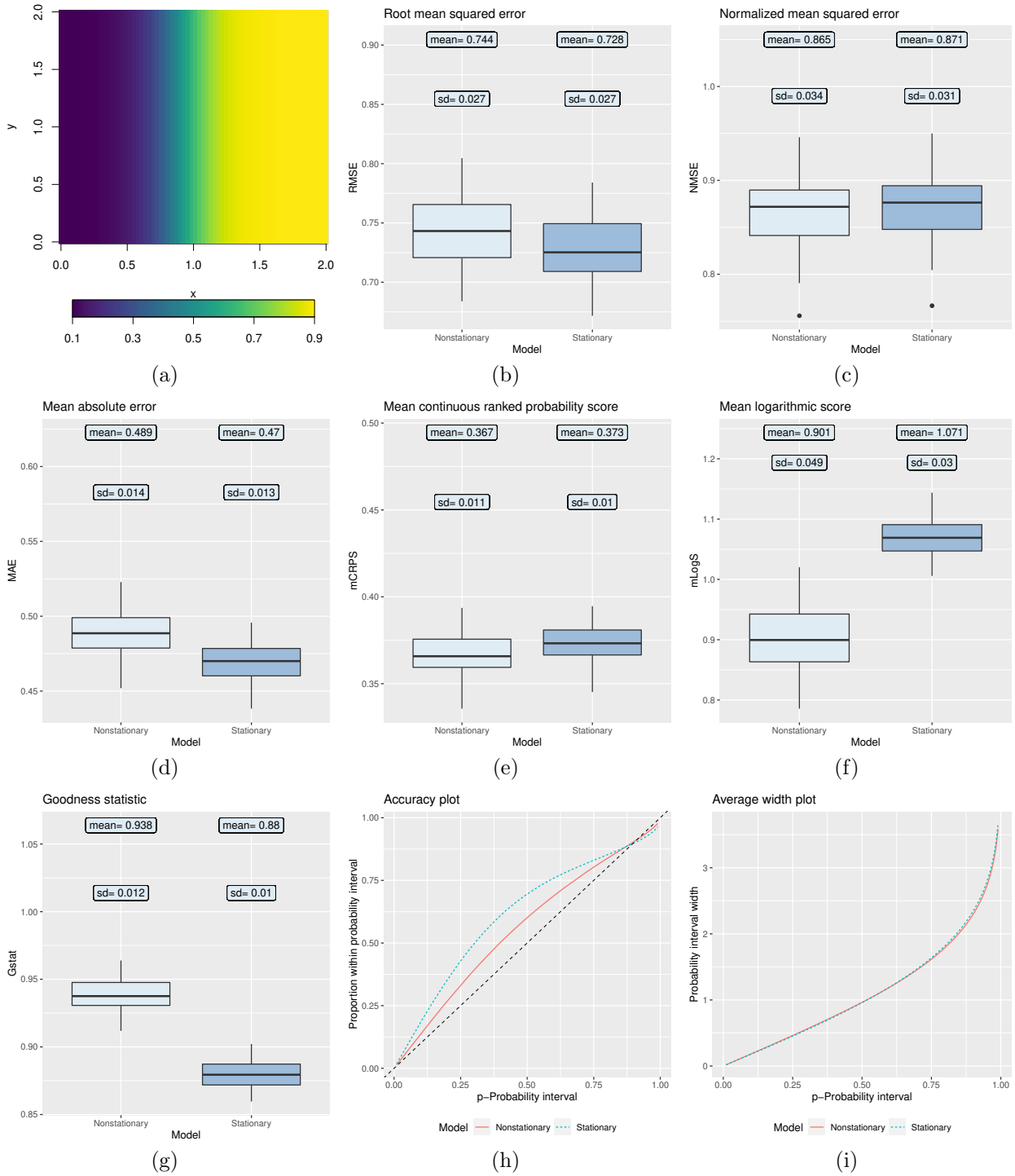


Figure 3: (a) True spatially varying range parameters (for $\lambda = 0.25$). Boxplots of (b) root mean squared error, (c) normalized mean squared error, (d) mean absolute error, (e) mean continuous ranked probability score, (f) mean logarithmic score, and (g) goodness statistic. (h) Accuracy plot. (i) Average width plot. The results are based on 100 simulation runs. Above the boxplots, we report the mean and standard deviations.

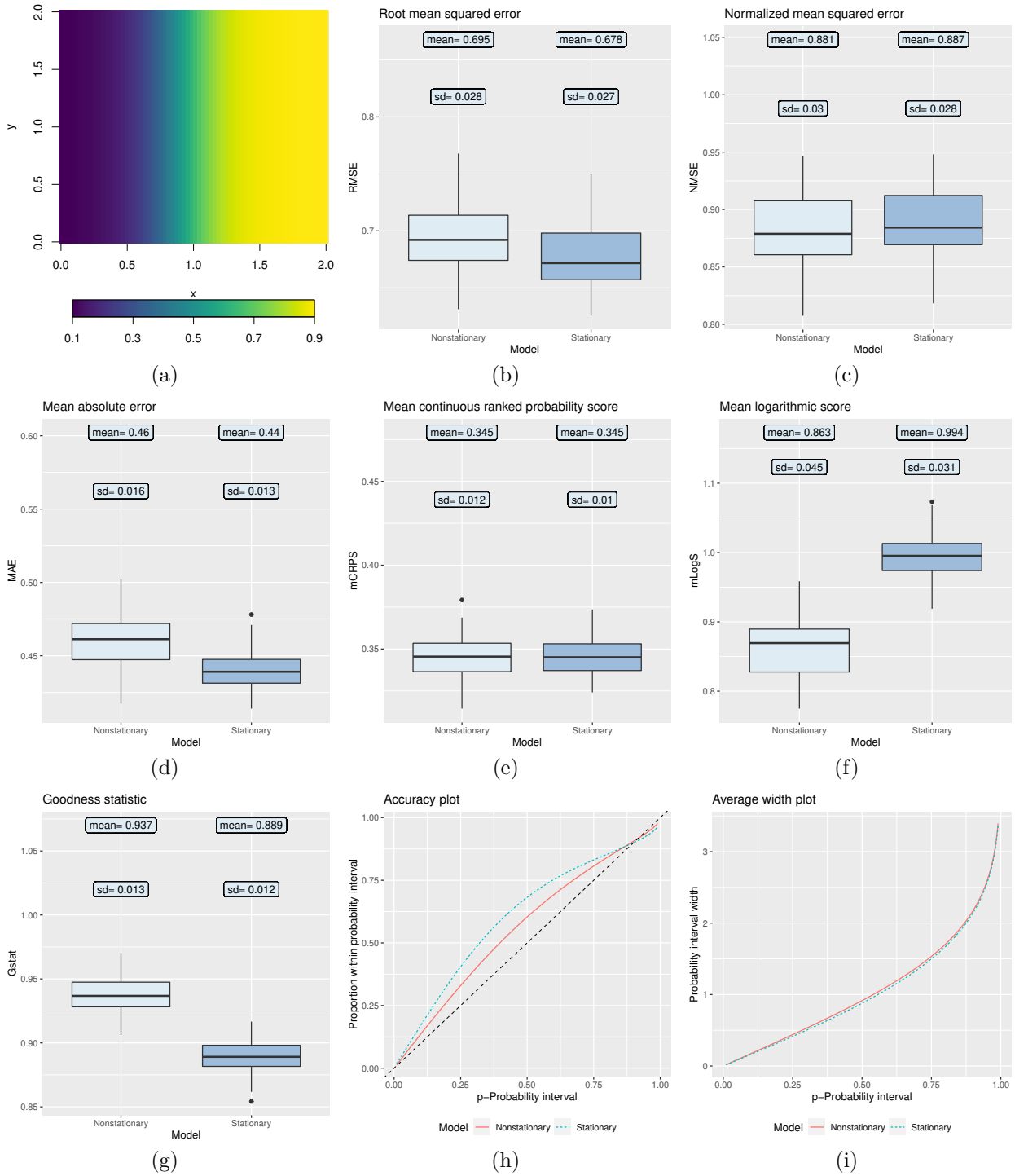


Figure 4: (a) True spatially varying range parameters (for $\lambda = 0.30$). Boxplots of (b) root mean squared error, (c) normalized mean squared error, (d) mean absolute error, (e) mean continuous ranked probability score, (f) mean logarithmic score, and (g) goodness statistic. (h) Accuracy plot. (i) Average width plot. The results are based on 100 simulation runs. Above the boxplots, we report the mean and standard deviations.

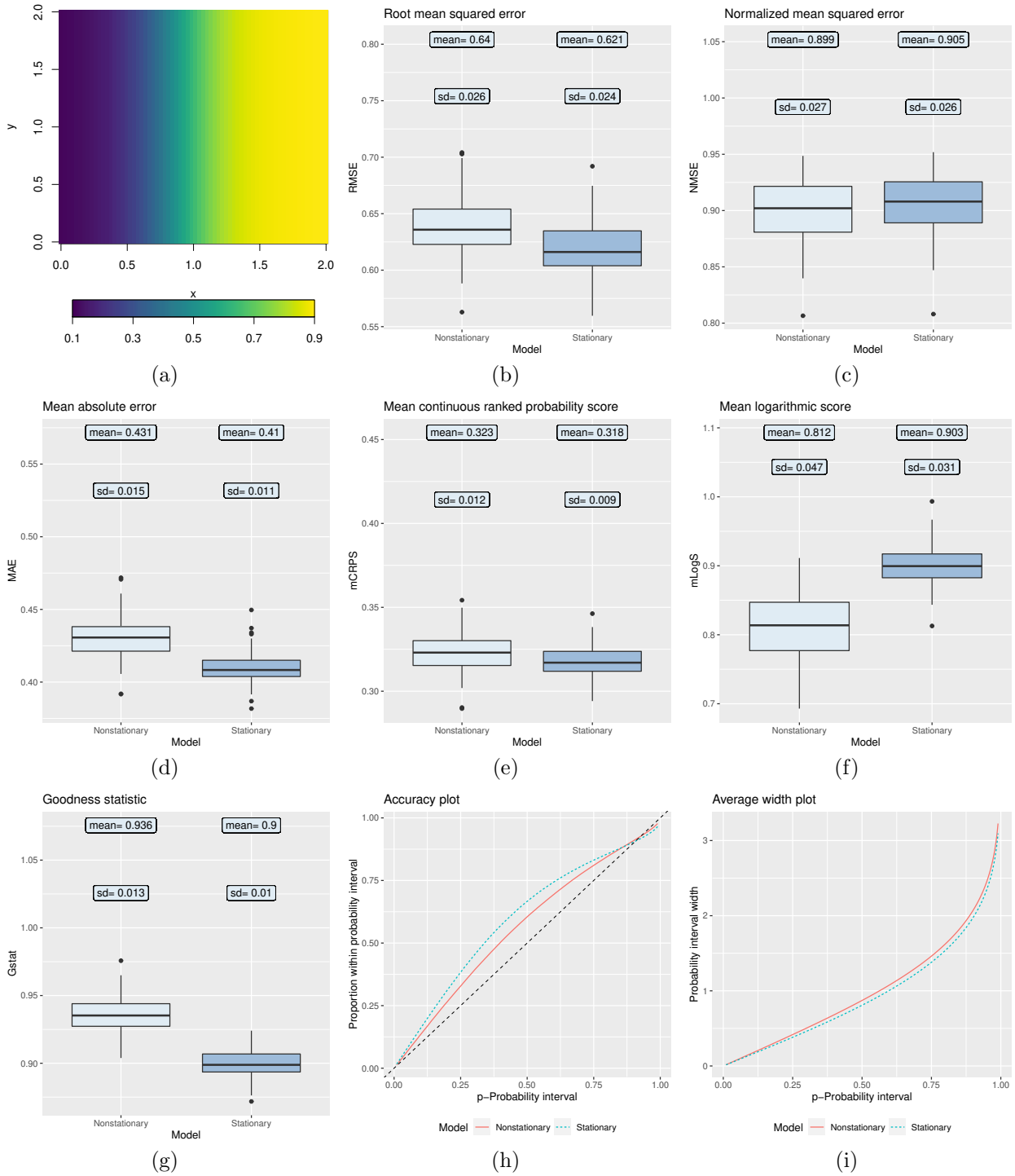


Figure 5: (a) True spatially varying range parameters (for $\lambda = 0.35$). Boxplots of (b) root mean squared error, (c) normalized mean squared error, (d) mean absolute error, (e) mean continuous ranked probability score, (f) mean logarithmic score, and (g) goodness statistic. (h) Accuracy plot. (i) Average width plot. The results are based on 100 simulation runs. Above the boxplots, we report the mean and standard deviations.

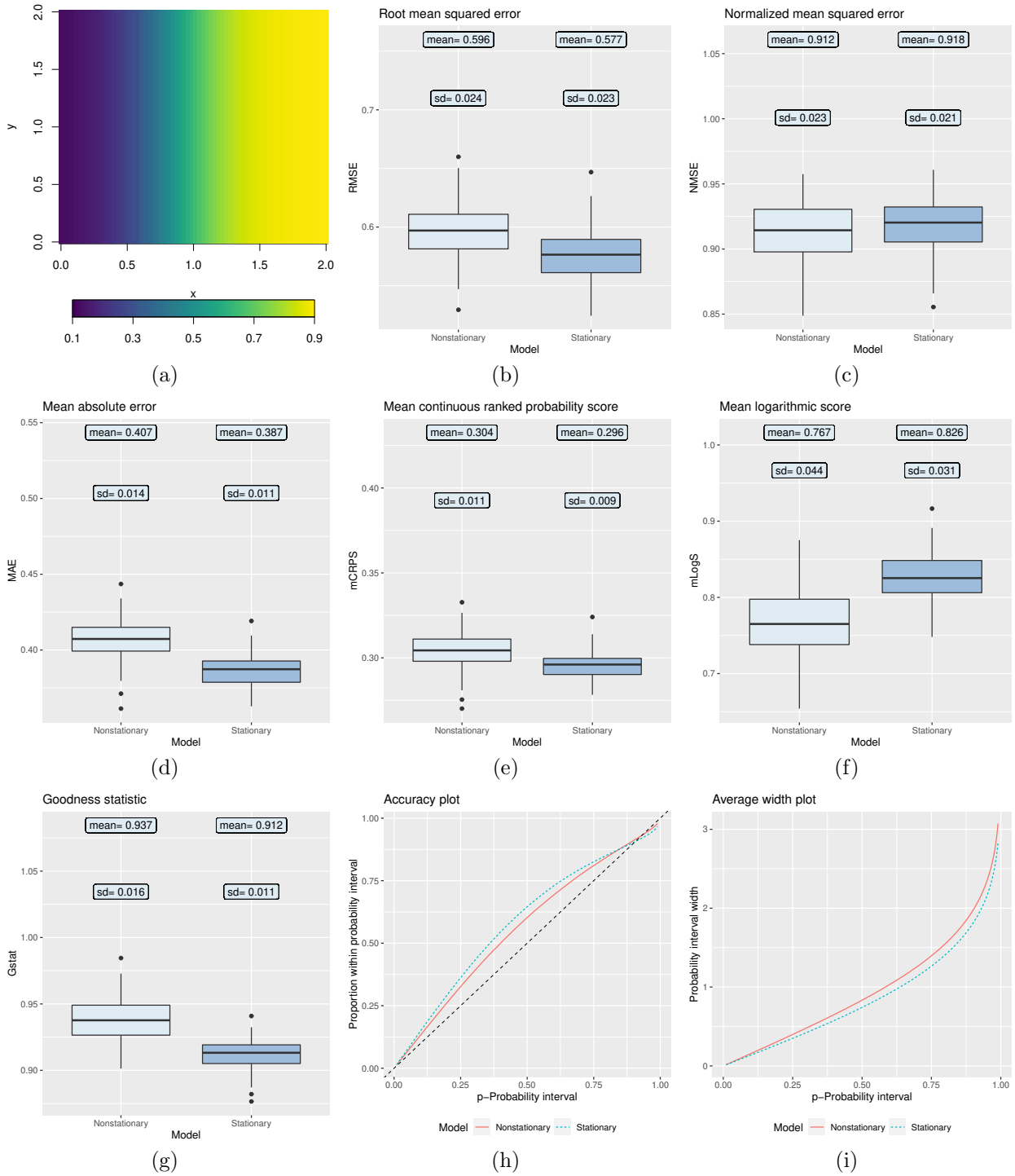


Figure 6: (a) True spatially varying range parameters (for $\lambda = 0.40$). Boxplots of (b) root mean squared error, (c) normalized mean squared error, (d) mean absolute error, (e) mean continuous ranked probability score, (f) mean logarithmic score, and (g) goodness statistic. (h) Accuracy plot. (i) Average width plot. The results are based on 100 simulation runs. Above the boxplots, we report the mean and standard deviations.

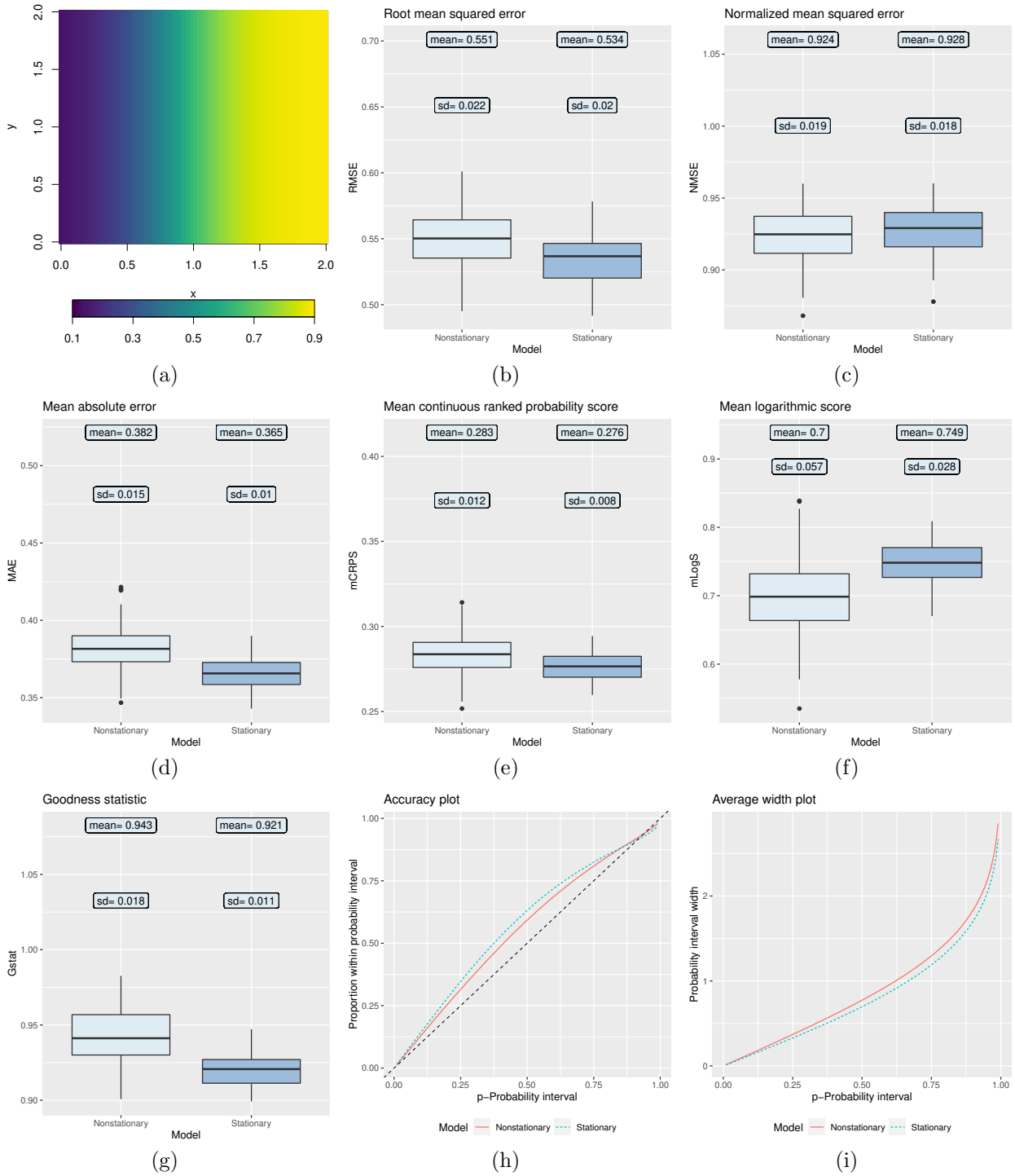


Figure 7: (a) True spatially varying range parameters (for $\lambda = 0.45$). Boxplots of (b) root mean squared error, (c) normalized mean squared error, (d) mean absolute error, (e) mean continuous ranked probability score, (f) mean logarithmic score, and (g) goodness statistic. (h) Accuracy plot. (i) Average width plot. The results are based on 100 simulation runs. Above the boxplots, we report the mean and standard deviations.

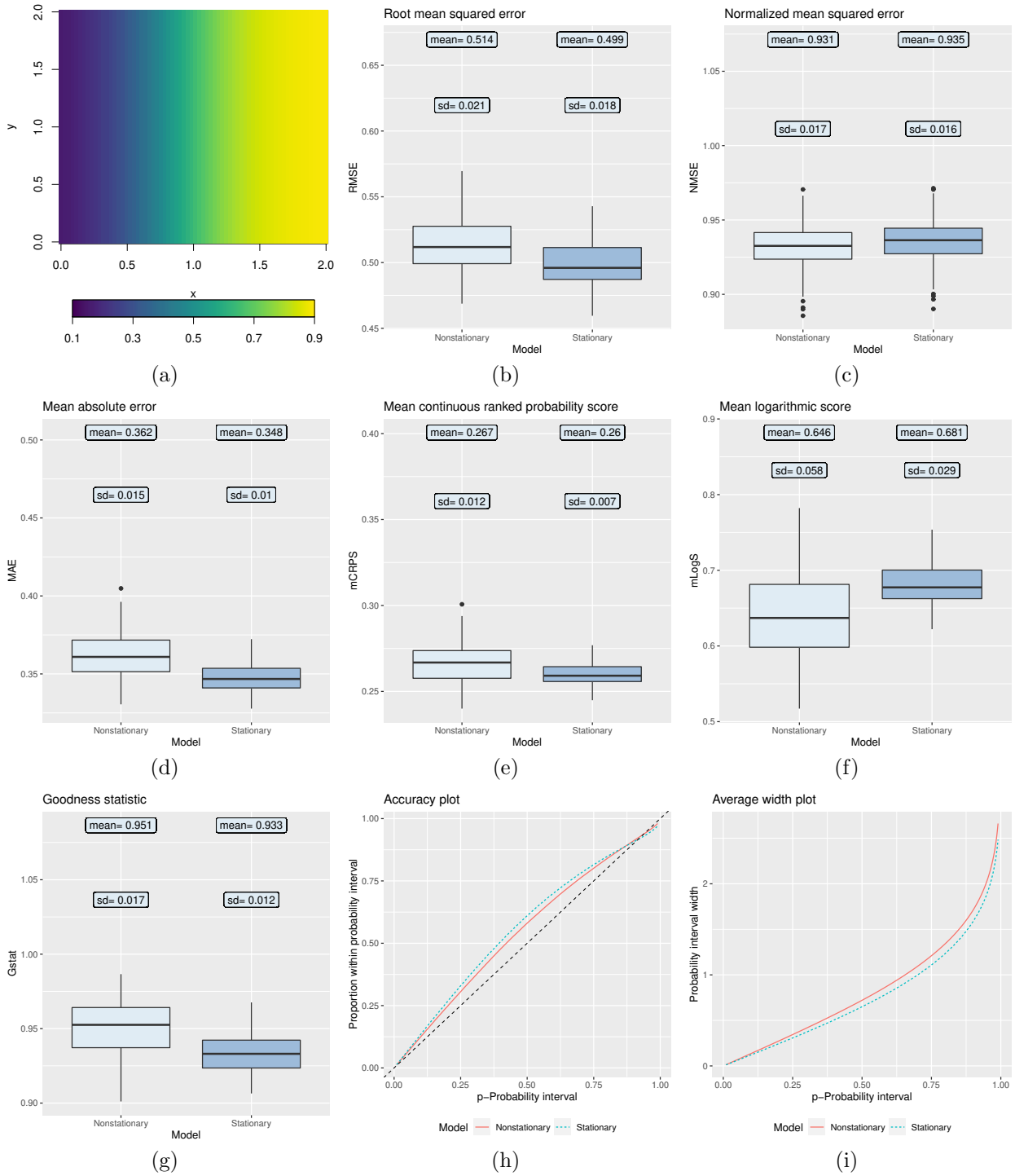


Figure 8: (a) True spatially varying range parameters (for $\lambda = 0.50$). Boxplots of (b) root mean squared error, (c) normalized mean squared error, (d) mean absolute error, (e) mean continuous ranked probability score, (f) mean logarithmic score, and (g) goodness statistic. (h) Accuracy plot. (i) Average width plot. The results are based on 100 simulation runs. Above the boxplots, we report the mean and standard deviations.

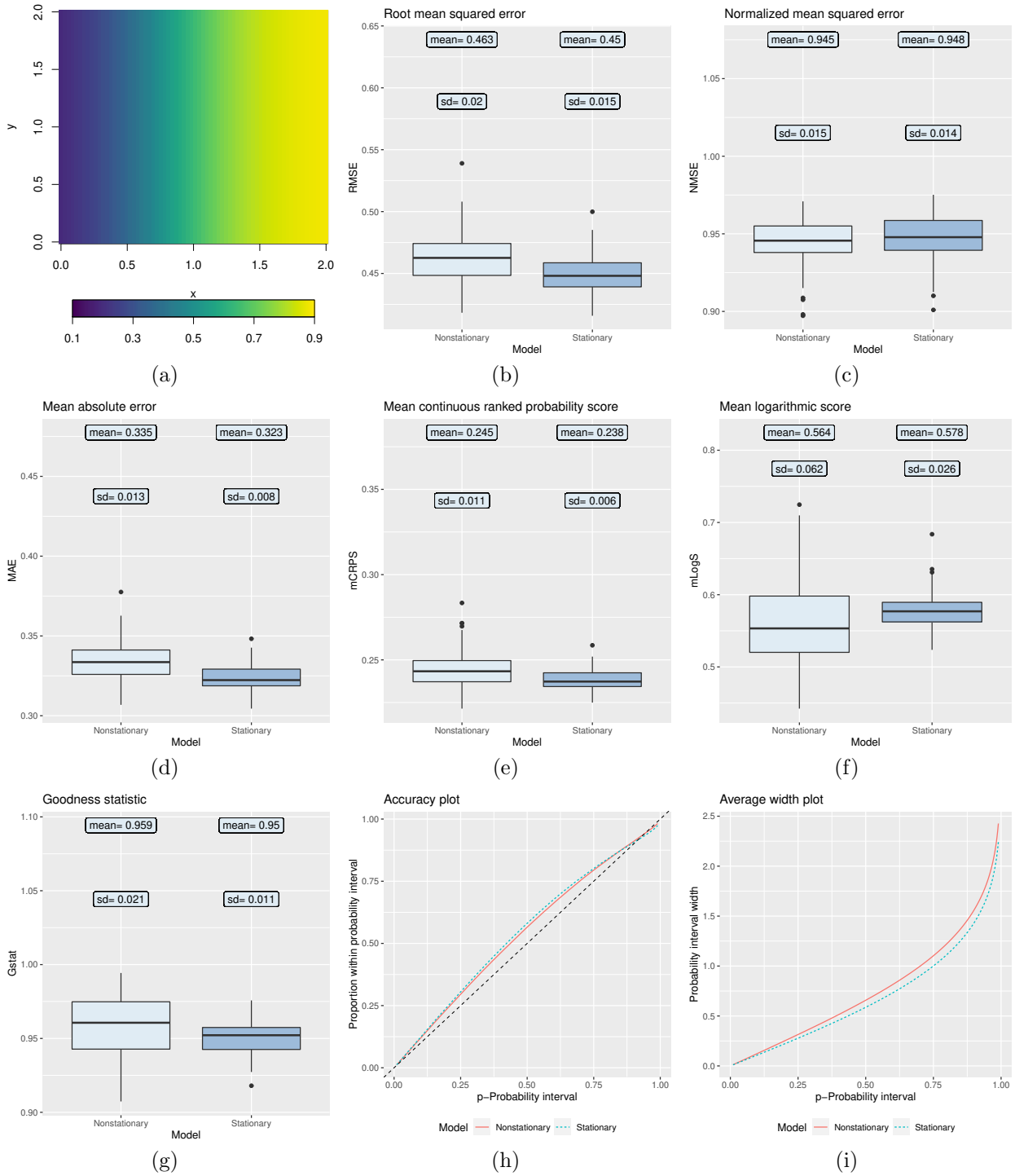
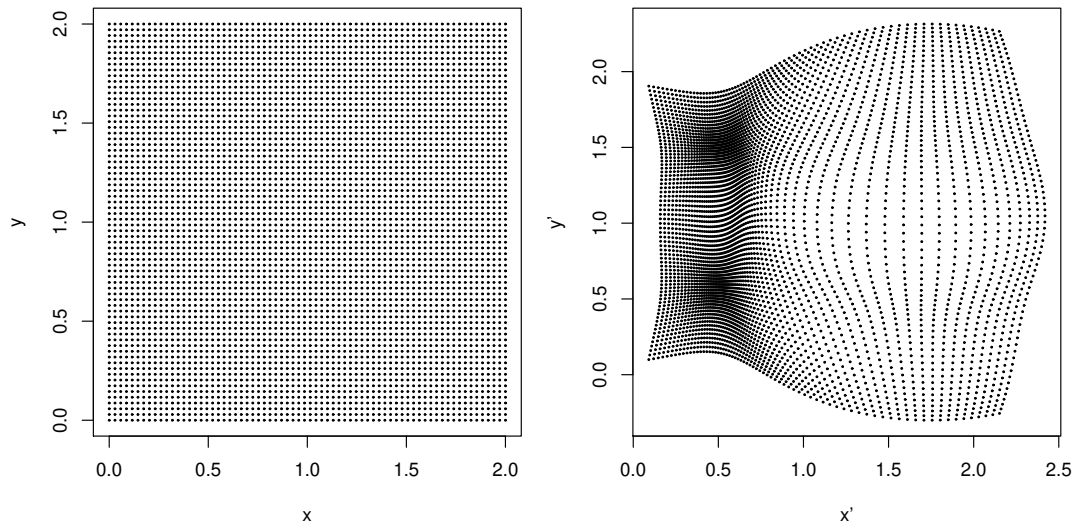


Figure 9: (a) True spatially varying range parameters (for $\lambda = 0.60$). Boxplots of (b) root mean squared error, (c) normalized mean squared error, (d) mean absolute error, (e) mean continuous ranked probability score, (f) mean logarithmic score, and (g) goodness statistic. (h) Accuracy plot. (i) Average width plot. The results are based on 100 simulation runs. Above the boxplots, we report the mean and standard deviations.

to be a stationary Matérn covariance function with its variance, smoothness and spatial range parameters fixed to 5, 0.8, and 0.16, respectively. The deformation functions θ is based on composition of elementary radial basis functions (Perrin and Monestiez, 1999). The functional form of the specified deformation function θ is $\theta(\mathbf{s}) = \theta_4 \circ \theta_3 \circ \theta_2 \circ \theta_1(\mathbf{s})$, where $\theta_i(\mathbf{s}) = \mathbf{c}_i + (\mathbf{s} - \mathbf{c}_i)\{1 + b_i \exp(-a_i \|\mathbf{s} - \mathbf{c}_i\|^2)\}$ is an elementary radial basis function, which is defined through parameters $\mathbf{c}_i \in \mathbb{R}^2$, $a_i > 0$, and $b_i \in (-1, \frac{1}{2}\exp\frac{3}{2})$. The parameters $\mathbf{c}_i$, a_i, b_i control the center, range and intensity of the deformation, respectively, where $b > 0$ leads to stretching and $b < 0$ leads to shrinkage. In our setting, we set the deformation parameters as $a_1 = a_2 = a_3 = a_4 = 2$, $b_1 = b_2 = 1$, $b_3 = b_4 = -0.5$, $\mathbf{c}_1 = (1.7, 0.5)$, $\mathbf{c}_2 = (1.7, 1.5)$, $\mathbf{c}_3 = (0.5, 0.5)$, and $\mathbf{c}_4 = (0.5, 1.5)$, which leads to a continuously varying nonstationarity that exhibits predominantly low and high range directional spatial dependence on the right and left side of $\mathcal{G}$, respectively. The true simulation grid $\mathbf{s} = (x, y) \in \mathcal{G}$ and the true deformed surface $\theta(\mathbf{s}) = (x', y')$ are shown in Figures 10(a) and10(b), respectively. The simulated realization of the nonstationary process X is shown in Figure 11.

The primary task is to recover the deformation function θ, shown in Figure 10(b), by applying the proposed method to the simulated realization. We randomly select a training sample of 1200 data points and use them to estimate the deformation function θ for all of the 4900 grid point locations in $\mathcal{G}$. While applying the proposed method, we divide $\mathcal{G}$ into two subregions $\mathcal{G}_1 = [0, 1.2) \times [0, 2]$ and $\mathcal{G}_2 = [1.2, 2] \times [0, 2]$, and keep all of the other methodological specifications the same as in Section 3 of the main manuscript. The comparison of the true and the estimated deformation functions (in the first two dimensions of maximum variation) are shown in Figure 12. The deformed space estimated using the proposed method, shown in Figure 12(b), reasonably recovers the large scale pattern of the true



(a) (b)

Figure 10: (a) Geographic space. (b) True deformed space.

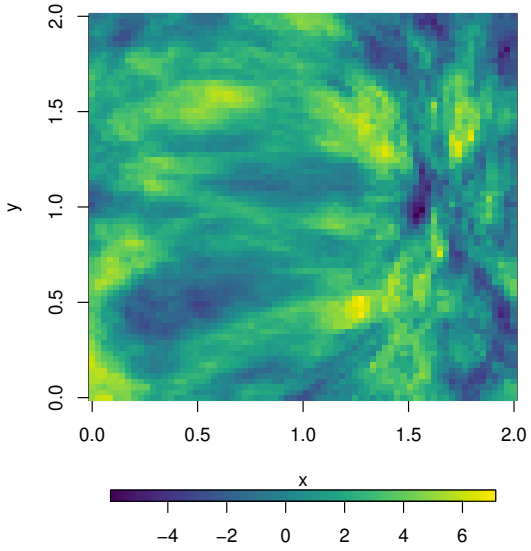


Figure 11: Simulated realization on the geographic space.

deformed space, shown in Figure 12(a); it mimics the local contractions in the left portion of $\mathcal{G}$ and expansions in the right portion of $\mathcal{G}$ of the true deformation function. However, the estimated deformation is not able to capture the small pattern of the true deformed space, which causes local anisotropies. As a result, the estimated deformed space does not exactly match the true deformed space. This is expected, because the true deformation function

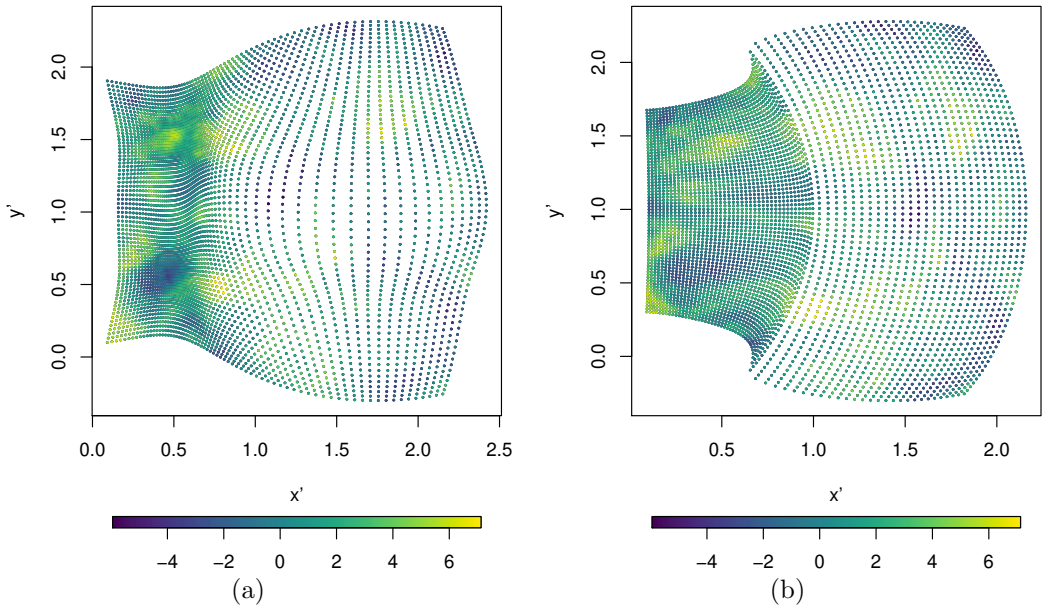


Figure 12: (a) True deformed space. (b) Estimated deformed space (first two dimensions of maximum variation).

leads to continuously varying nonstationarity, whereas the proposed method works under the assumption of regional stationarity; the estimated deformed space is a good "regionally stationary" approximation of the true "continuously varying nonstationary" deformation function.

We fit a stationary Matérn covariance function in the estimated deformed space via MLE using the training set, and visualize the comparison of true and estimated nonstationary correlations in Figure 13. We randomly select three reference locations from the grid points in $\mathcal{G}$, and show the true and estimated nonstationary correlations between the reference locations and all other grid point locations in $\mathcal{G}$ as heatmaps. The estimated nonstationary correlations effectively recover the varying strength of spatial dependence in different regions of $\mathcal{G}$. However, they ignore the varying direction of dependence (anisotropies) because of the underlying regional stationarity assumption. Overall, like the estimated deformed space, the estimated nonstationary correlations are reasonable "regionally stationary" approximations

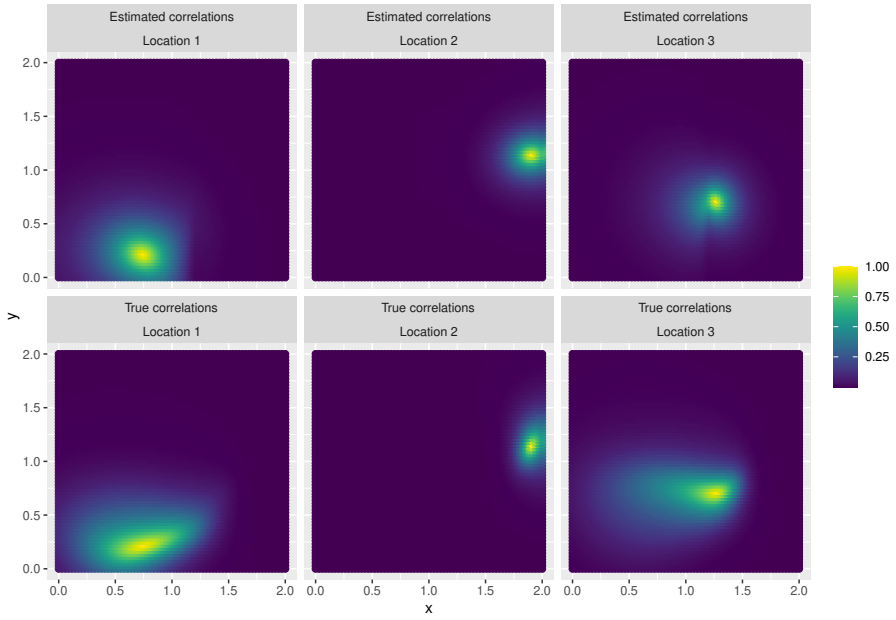


Figure 13: Estimated nonstationary correlations (top row) and true nonstationary correlations (bottom row) of three randomly selected reference locations on the simulation grid, $\mathbf{s} \in \mathcal{G}$, with every other location on a fine grid.

of the "continuously varying nonstationary" correlations.

S4 Robustness of Estimated Deformation to Domain Division

Let us consider a stochastic process $\{X(\mathbf{s}),\ \mathbf{s} \in \mathcal{G}\}$ with regional stationarity and global nonstationarity. Thus, it is possible to divide the entire domain $\mathcal{G}$ into subregions $\mathcal{G}_1, ..., \mathcal{G}_k$ such that $\bigcup_{i=1}^{k} \mathcal{G}_i = \mathcal{G}$, and each of the k regional processes $\{X(\mathbf{s}),\ \mathbf{s} \in \mathcal{G}_i\}$ are stationary processes admitting distinct isotropic variograms $\gamma_i(\|\cdot\|),\ i = 1, 2, ..., k$. Here, let the partitioning $\{\mathcal{G}_1, ..., \mathcal{G}_k\}$ be referred to as the **True Partitioning**, which is often unknown.

Let us assume that the regional variograms $\{\gamma_i(\|\cdot\|),\ i = 1, 2, ..., k\}$ are either known or can be estimated reasonably well enough. We align the k distinct variograms $\{\gamma_i(\|\cdot\|),\ i = 1, ..., k\}$ to estimate k distinct regional distance warping functions $\{\phi_i(\|\cdot\|),\ i = 1, ..., k\}$.

Further, we define the GDWF for the **True Partitioning** as

$$\phi(\mathbf{s},\mathbf{s}') = \sum_{\mathcal{G}_i \in \mathcal{L}(\mathbf{s},\mathbf{s}')} \mathcal{W}_i(\mathbf{s},\mathbf{s}')\phi_i(\|\mathbf{s}-\mathbf{s}'\|),\ (\mathbf{s},\mathbf{s}') \in \mathcal{G}\times\mathcal{G}, \tag{S4.1}$$

where $\mathcal{L}(\mathbf{s},\mathbf{s}')$ is the set of subregions $\mathcal{G}_i$ such that the line segment joining the locations $\mathbf{s}$ and $\mathbf{s}'$ passes through all of the subregions in this set, and $\mathcal{W}_i(\mathbf{s},\mathbf{s}')$ are the location-dependent weights for the i^{th} regional distance warping function. We define the weights as $\mathcal{W}_i(\mathbf{s},\mathbf{s}') = \frac{\mathcal{P}(i,\mathbf{s},\mathbf{s}')}{\|\mathbf{s}-\mathbf{s}'\|}$, where $\mathcal{P}(i,\mathbf{s},\mathbf{s}')$ is the length of the segment joining $\mathbf{s}$ and $\mathbf{s}'$ in $\mathcal{G}_i$.

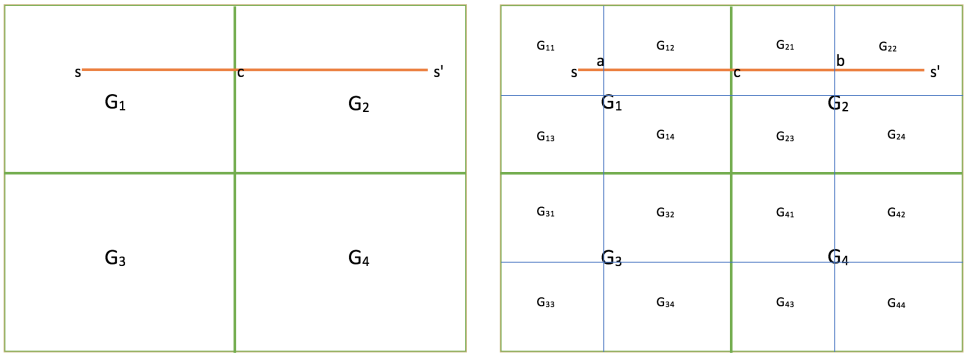
Now, let us consider finer divisions of the domain where the i^{th} subregion $\{\mathcal{G}_i,\ i=1,...,k\}$ is further divided into m_i subregions $\mathcal{G}_{i1},...,\mathcal{G}_{im_i}$ such that

$$\bigcup_{j=1}^{m_i}\mathcal{G}_{ij} = \mathcal{G}_i,\ i=1,2,...,k \text{ and } \bigcup_{i=1}^{k}\bigcup_{j=1}^{m_i}\mathcal{G}_{ij} = \mathcal{G}. \tag{S4.2}$$

Here, we refer to the partitioning $\{\mathcal{G}_{ij},\ i=1,...,k,\ j=1,...,m_i\}$ as the **Guessed Partitioning**. Then, the $\sum_{i=1}^{k} m_i$ regional processes $\{X(\mathbf{s}),\ \mathbf{s}\in\mathcal{G}_{ij}\}$, $i=1,2,...,k,\ j=1,2,...,m_i$ are also stationary processes admitting the variograms $\gamma_{ij}(\|\cdot\|),\ i=1,2,...,k,\ \ j=1,2,...,m_i$, respectively. We align the $\sum_{i=1}^{k} m_i$ regional variograms $\{\gamma_{ij}(\|\cdot\|),\ \ i=1,2...,k,\ \ j=1,2,...,m_i\}$ to obtain $\sum_{i=1}^{k} m_i$ regional distance warping functions $\{\phi_{ij}(\|\cdot\|,\ i=1,2...,k,\ j=1,2,...,m_i\}$. The GDWF for the **Guessed Partitioning** is given by

$$\phi(\mathbf{s},\mathbf{s}') = \sum_{\mathcal{G}_{ij} \in \mathcal{L}(\mathbf{s},\mathbf{s}')} \mathcal{W}_{ij}(\mathbf{s},\mathbf{s}')\phi_{ij}(\|\mathbf{s}-\mathbf{s}'\|),\ (\mathbf{s},\mathbf{s}') \in \mathcal{G}\times\mathcal{G}, \tag{S4.3}$$

where $\mathcal{L}(\mathbf{s},\mathbf{s}')$ is the set of **guessed** subregions $\mathcal{G}_{ij}$ such that the line segment joining the locations $\mathbf{s}$ and $\mathbf{s}'$ passes through all of the subregions in this set, and $\mathcal{W}_{ij}(\mathbf{s},\mathbf{s}')$ are the location-dependent weights for the $(ij)^{th}$ regional distance warping function. We define the weights as $\mathcal{W}_{ij}(\mathbf{s},\mathbf{s}') = \frac{\mathcal{P}(ij,\mathbf{s},\mathbf{s}')}{\|\mathbf{s}-\mathbf{s}'\|}$, where $\mathcal{P}(ij,\mathbf{s},\mathbf{s}')$ is the length of the line segment joining $\mathbf{s}$ and $\mathbf{s}'$ that lies in subregion $\mathcal{G}_{ij}$.



(a) (b)

Figure 14: (a) True Partitioning. (b) Guessed Partitioning.

Note that $\gamma_{ij}(\|\cdot\|) = \gamma_i(\|\cdot\|),\ j = 1, ..., m_i,\ i = 1, 2..., k$ because the sub-processes $\{X(\mathbf{s}),\ \mathbf{s} \in \mathcal{G}_{ij}\}$ will have the same spatial dependence structure as that of the parent stationary process $\{X(\mathbf{s}),\ \mathbf{s} \in \mathcal{G}_i\}$, and consequently, we will obtain many identical regional distance warping functions such that $\phi_{ij}(\|\cdot\|) = \phi_i(\|\cdot\|),\ j = 1, ..., m_i,\ i = 1, 2..., k$. Therefore, because $\sum_{j=1}^{m_i} \mathcal{P}(ij, \mathbf{s}, \mathbf{s'}) = \mathcal{P}(i, \mathbf{s}, \mathbf{s'}),$ for all $i = 1, ..., k$, Equation S4.3 becomes

$$\begin{aligned}
\phi(\mathbf{s}, \mathbf{s}') &= \sum_{\mathcal{G}_i \in \mathcal{L}(\mathbf{s},\mathbf{s}')} \sum_{j=1}^{m_i} \mathcal{W}_{ij}(\mathbf{s}, \mathbf{s}')\phi_i(\|\mathbf{s} - \mathbf{s}'\|),\ (\mathbf{s}, \mathbf{s}') \in \mathcal{G} \times \mathcal{G}, \\
&= \sum_{\mathcal{G}_i \in \mathcal{L}(\mathbf{s},\mathbf{s}')} \sum_{j=1}^{m_i} \frac{\mathcal{P}(ij, \mathbf{s}, \mathbf{s'})}{\|\mathbf{s} - \mathbf{s}'\|}\phi_i(\|\mathbf{s} - \mathbf{s}'\|),\ (\mathbf{s}, \mathbf{s}') \in \mathcal{G} \times \mathcal{G} \\
&= \sum_{\mathcal{G}_i \in \mathcal{L}(\mathbf{s},\mathbf{s}')} \frac{\mathcal{P}(i, \mathbf{s}, \mathbf{s'})}{\|\mathbf{s} - \mathbf{s}'\|}\phi_i(\|\mathbf{s} - \mathbf{s}'\|),\ (\mathbf{s}, \mathbf{s}') \in \mathcal{G} \times \mathcal{G} \\
&= \sum_{\mathcal{G}_i \in \mathcal{L}(\mathbf{s},\mathbf{s}')} \mathcal{W}_i(\mathbf{s}, \mathbf{s}')\phi_i(\|\mathbf{s} - \mathbf{s}'\|),\ (\mathbf{s}, \mathbf{s}') \in \mathcal{G} \times \mathcal{G},
\end{aligned}$$

leading to the same GDWF as given in Equation S4.1. This indicates robustness of our approach under finer subdivisions of the domain. Generally, the true partitioning $\{\mathcal{G}_1, ..., \mathcal{G}_k\}$ is unknown. However, depending on the size of the data, we can divide the entire domain

into very fine guessed subregions $\{\mathcal{G}_{ij},\ i = 1, ..., k,\ j = 1, ..., m_i\}$ such that Equation S4.2 approximately holds true.

The above result is further illustrated using a simple pictorial example. In Figure 14(a), the true partitioning is given by $\mathcal{G} = \bigcup_{i=1}^{4} \mathcal{G}_i,\ k = 4$. We let $\mathbf{s} \in \mathcal{G}_1$ and $\mathbf{s'} \in \mathcal{G}_2$. Then, $\mathcal{L}(\mathbf{s}, \mathbf{s'}) = \{\mathcal{G}_1, \mathcal{G}_2\}$, $\mathcal{P}(1, \mathbf{s}, \mathbf{s'}) = \|\mathbf{s} - \mathbf{c}\|$ and $\mathcal{P}(2, \mathbf{s}, \mathbf{s'}) = \|\mathbf{c} - \mathbf{s'}\|$, resulting in weights $\mathcal{W}_1(\mathbf{s}, \mathbf{s'}) = \frac{\|\mathbf{s}-\mathbf{c}\|}{\|\mathbf{s}-\mathbf{s'}\|}$ and $\mathcal{W}_2(\mathbf{s}, \mathbf{s'}) = \frac{\|\mathbf{c}-\mathbf{s'}\|}{\|\mathbf{s}-\mathbf{s'}\|}$. This gives the GDWF $\phi(\mathbf{s}, \mathbf{s'}) = \frac{\|\mathbf{s}-\mathbf{c}\|\phi_1(\|\mathbf{s}-\mathbf{s'}\|)+\|\mathbf{c}-\mathbf{s'}\|\phi_2(\|\mathbf{s}-\mathbf{s'}\|)}{\|\mathbf{s}-\mathbf{s'}\|}$. Next, we compute the GDWF for a finer guessed partitioning of the domain as shown in Figure 14(b): $\mathcal{G} = \bigcup_{i=1}^{4} \bigcup_{j=1}^{m_i} \mathcal{G}_{ij},\ k = 4,\ m_i = 4,\ i$. We let $\mathbf{s} \in \mathcal{G}_{11}$ and $\mathbf{s'} \in \mathcal{G}_{22}$. Then, $\mathcal{L}(\mathbf{s}, \mathbf{s'}) = \{\mathcal{G}_{11}, \mathcal{G}_{12}, \mathcal{G}_{21}, \mathcal{G}_{22}\}$, $\mathcal{P}(11, \mathbf{s}, \mathbf{s'}) = \|\mathbf{s} - \mathbf{a}\|$, $\mathcal{P}(12, \mathbf{s}, \mathbf{s'}) = \|\mathbf{a} - \mathbf{c}\|$, $\mathcal{P}(21, \mathbf{s}, \mathbf{s'}) = \|\mathbf{c} - \mathbf{b}\|$ and $\mathcal{P}(22, \mathbf{s}, \mathbf{s'}) = \|\mathbf{b} - \mathbf{s'}\|$, resulting in weights $\mathcal{W}_{11}(\mathbf{s}, \mathbf{s'}) = \frac{\|\mathbf{s}-\mathbf{a}\|}{\|\mathbf{s}-\mathbf{s'}\|}$, $\mathcal{W}_{12}(\mathbf{s}, \mathbf{s'}) = \frac{\|\mathbf{a}-\mathbf{c}\|}{\|\mathbf{s}-\mathbf{s'}\|}$, $\mathcal{W}_{21}(\mathbf{s}, \mathbf{s'}) = \frac{\|\mathbf{c}-\mathbf{b}\|}{\|\mathbf{s}-\mathbf{s'}\|}$ and $\mathcal{W}_{22}(\mathbf{s}, \mathbf{s'}) = \frac{\|\mathbf{b}-\mathbf{s'}\|}{\|\mathbf{s}-\mathbf{s'}\|}$. This gives the GDWF $\phi(\mathbf{s}, \mathbf{s'}) = \frac{\|\mathbf{s}-\mathbf{a}\|\phi_{11}(\|\mathbf{s}-\mathbf{s'}\|)+\|\mathbf{a}-\mathbf{c}\|\phi_{12}(\|\mathbf{s}-\mathbf{s'}\|+\|\mathbf{c}-\mathbf{b}\|\phi_{21}(\|\mathbf{s}-\mathbf{s'}\|)+\|\mathbf{b}-\mathbf{s'}\|\phi_{22}(\|\mathbf{s}-\mathbf{s'}\|)}{\|\mathbf{s}-\mathbf{s'}\|}$. Applying $\phi_{ij}(\|\cdot\|) = \phi_i(\|\cdot\|),\ i = 1, ..., k,\ j = 1, ..., m_i$, we obtain

$$\begin{aligned}\phi(\mathbf{s}, \mathbf{s'}) &= \frac{1}{\|\mathbf{s} - \mathbf{s'}\|} \times \{\|\mathbf{s} - \mathbf{a}\|\phi_1(\|\mathbf{s} - \mathbf{s'}\|) + \|\mathbf{a} - \mathbf{c}\|\phi_1(\|\mathbf{s} - \mathbf{s'}\|) \\ &\qquad + \|\mathbf{c} - \mathbf{b}\|\phi_2(\|\mathbf{s} - \mathbf{s'}\|) + \|\mathbf{b} - \mathbf{s'}\|\phi_2(\|\mathbf{s} - \mathbf{s'}\|)\} \\ &= \frac{1}{\|\mathbf{s} - \mathbf{s'}\|}\{\|\mathbf{s} - \mathbf{c}\|\phi_1(\|\mathbf{s} - \mathbf{s'}\|) + \|\mathbf{c} - \mathbf{s'}\|\phi_2(\|\mathbf{s} - \mathbf{s'}\|)\},\end{aligned}$$

which is the same as before.

However, in practice, we cannot estimate the variograms exactly even if the true processes share an identical spatial dependence. Therefore, while replacing $\gamma_{ij}(\|.\|),\ i = 1, ..., k,\ j = 1, ..., m_i$ with the estimated variograms $\hat{\gamma}_{ij}(\|.\|),\ i = 1, ..., k,\ j = 1, ..., m_i$, the exact equality $\gamma_{ij}(\|\cdot\|) = \gamma_i(\|\cdot\|),\ j = 1, ..., m_i,\ i = 1, 2..., k$, becomes $\hat{\gamma}_{ij}(\|\cdot\|) \approx \hat{\gamma}_i(\|\cdot\|),\ j = 1, ..., m_i,\ i = 1, 2..., k$. Consequently, the estimated regional distance warping functions follow $\hat{\phi}_{ij}(\|\cdot\|) \approx$

$\hat{\phi}_i(\|\cdot\|),\ j = 1, ..., m_i,\ i = 1, 2..., k$ instead of $\phi_{ij}(\|\cdot\|) = \phi_i(\|\cdot\|),\ j = 1, ..., m_i,\ i = 1, 2..., k$. This approximation leads to slightly different deformations for the true partitioning and the guessed partitioning. Thus, accuracy of the estimation of regional variograms affects the robustness of our method under finer subdivisions of the domain.

There is a clear trade-off between the number of subregions and the amount of available data for estimation within each subregion. The guessed partitioning can be forced to consist of very small subregions such that Equation S4.2 approximately holds true. However, very small subregions lead to fewer data points per subregion and poor estimation of regional variograms. This is consequently reflected in the estimation of the regional distance warping functions and the final deformation. Therefore, the number of subregions must be driven by the extensiveness of available spatial data. This point is further demonstrated in the following simulation study.

S4.1 Simulation: Variogram Estimation Under Finer Subdivisions of the Domain

We consider a zero-mean Gaussian process X over a domain $\mathcal{G} = [0, 8]^2$, with the following nonstationary Matérn covariance function (Paciorek and Schervish, 2006):

$$C^{NS}(\mathbf{s}_i, \mathbf{s}_j : \tilde{\eta}) = \sigma(\mathbf{s}_i)\sigma(\mathbf{s}_j)\frac{|\Sigma(\mathbf{s}_i)|^{1/4}|\Sigma(\mathbf{s}_j)|^{1/4}}{2^{\nu-1}\Gamma(\nu)}\left|\frac{\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j)}{2}\right|^{-1/2}(2\sqrt{\nu Q_{ij}})^{\nu}K_{\nu}(2\sqrt{\nu Q_{ij}}), \tag{S4.4}$$

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

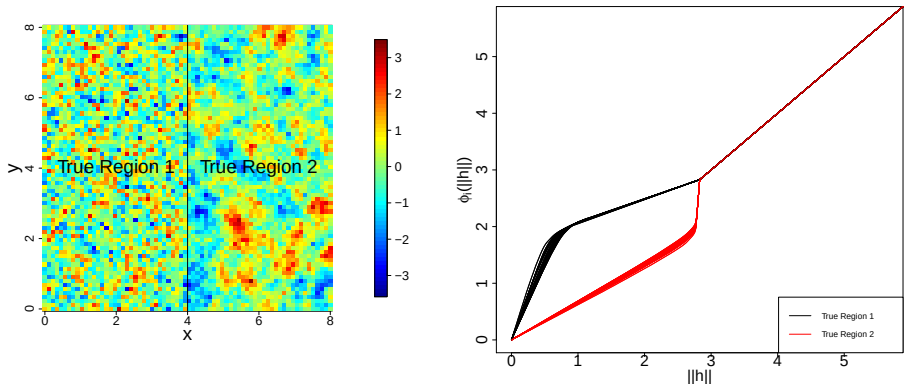
To obtain realizations from a regionally stationary process with nonstationarity only in

the spatial range, we simulate 50 realizations of X on a regular grid of 70×70 points on $\mathcal{G}$, with $\nu = 0.5$, $\{\sigma(\mathbf{s}) = 1,\ \mathbf{s} \in \mathcal{G}\}$ and

$$\Sigma(\mathbf{s})_{2\times 2} = \begin{cases} \text{diag}(0.1^2, 0.1^2), & \text{if } x \leq 4 \\ \text{diag}(0.45^2, 0.45^2), & \text{if } x > 4. \end{cases}$$

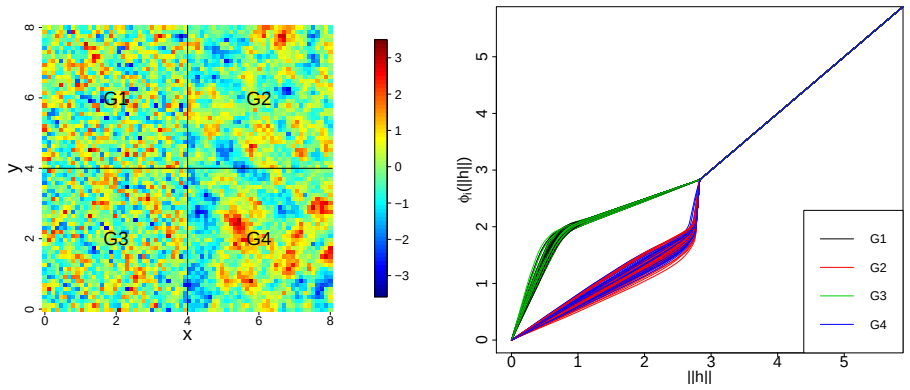
For this simulation, the true partitioning consists of two disjoint regions, i.e., "True Region 1"$=[0,4] \times [0,8]$ and "True Region 2"$=(4,8] \times [0,8]$. We consider three cases for domain partitioning, starting with the true partitioning and progressing to finer subdivisions. For the true partitioning case we estimate the isotropic Matérn variogram for "True Region 1" and "True Region 2" for each of the 50 simulated realizations using Maximum Likelihood Estimation (MLE). During estimation, we fix the smoothness parameter $\nu = 0.5$ to avoid identifiability issues (Zhang, 2004). We then estimate the regional distance warping functions for "True Region 1" and "True Region 2" by aligning their respective estimated regional variograms. The estimated regional variograms become numerically constant for distances greater than $\sqrt{8}$, and therefore, we set the value $\|\mathbf{h}_t\| = \sqrt{8}$ and assign identity regional distance warping functions for $\|\mathbf{h}\| > \sqrt{8}$. Figure 15(a) shows one realization of the simulated process, with a solid black line indicating the true partitioning. The estimated regional distance warping functions are shown in Figure 15(b). The 50 pairs of regional distance warping functions are nearly identical for each run indicating that enough data is available per subregion to efficiently estimate the regional variograms.

For the second case, we consider a finer subdivision of the domain into four equal subregions $\{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4\}$. Following the same estimation procedure as described for the true partitioning, we estimate the isotropic Matérn regional variograms and regional distance warping functions for the four subregions. Figure 16(a) shows one realization of the simulated

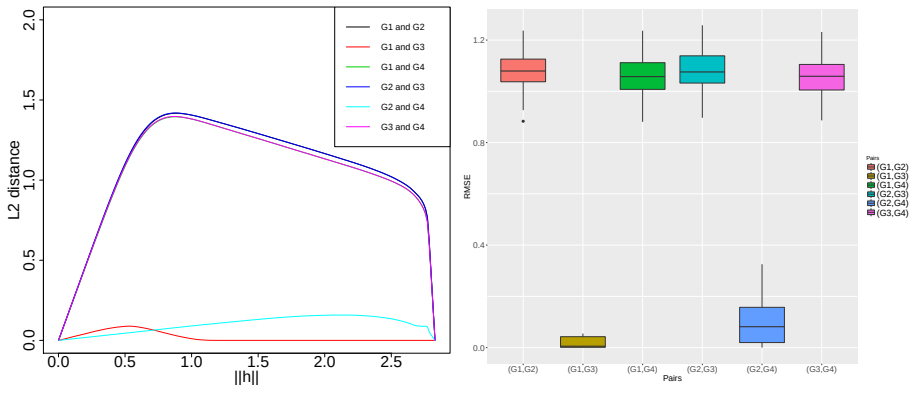


(a) (b)

Figure 15: (a) One realization of the simulated process, with a solid black line depicting the true partitioning. (b) Estimated regional distance warping functions for the two subregions.

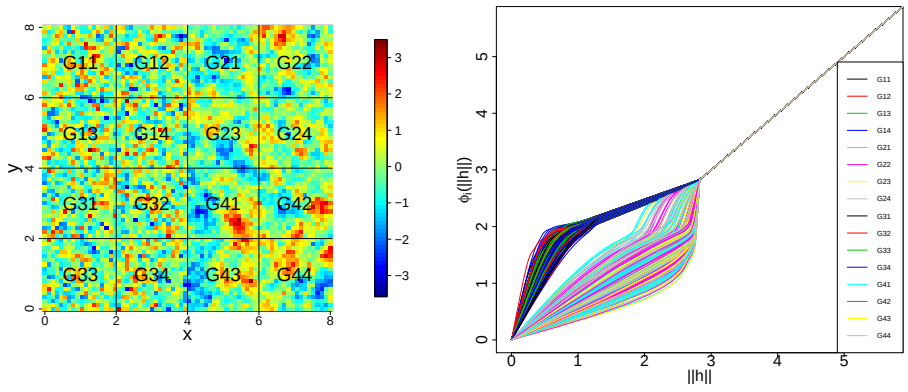


(a) (b)



(c) (d)

Figure 16: (a) One realization of the simulated process, with solid black lines depicting the partitioning. (b) Estimated regional distance warping functions. (c) Pointwise $\mathbb{L}^2$ distances between regional distance warping functions. (d) Boxplots of pairwise RMSE values.



(a) (b)

Figure 17: (a) One realization of the simulated process, with solid black lines depicting partitioning. (b) Estimated regional distance warping functions.

process, with solid black lines depicting the partitioning. The estimated regional distance warping functions are shown in Figure 16(b) for each of the 50 runs; they show very similar behavior within subregions $(\mathcal{G}_1, \mathcal{G}_3)$ and $(\mathcal{G}_4, \mathcal{G}_4)$, which is expected because of the common true underlying variograms for the pair of sub-processes $(\{X(\mathbf{s}), s \in \mathcal{G}_1\}, \{X(\mathbf{s}), s \in \mathcal{G}_3\})$ and $(\{X(\mathbf{s}), s \in \mathcal{G}_2\}, \{X(\mathbf{s}), s \in \mathcal{G}_4\})$. The very similar pairs of regional distance warping functions in Figure 16(b), as well as the closeness of Figure 15(b) and Figure 16(b) in terms of their shapes, demonstrate that the finer partitioning and the true partitioning lead to nearly identical deformations.

Figure 16(c) shows the pointwise $\mathbb{L}^2$ distance between the regional distance warping functions for every pair of subregions in the finer partitioning, averaged over 50 runs and evaluated up to $\|\mathbf{h}_t\|$. Figure 16(d) shows the boxplot for the root mean squared error (RMSE) between every pair of the regional distance warping functions evaluated up to $\|\mathbf{h}_t\|$. The $\mathbb{L}^2$ distances and the boxplots of RMSE values for the pairs $(\mathcal{G}_1, \mathcal{G}_3)$ and $(\mathcal{G}_2, \mathcal{G}_4)$ are concentrated near 0, which provides quantitative validation of the pairwise proximity of the corresponding regional distance warping functions.

For the third case, we consider an even finer subdivision of the domain into sixteen subregions $\{\mathcal{G}_{ij},\ i,j = 1,...,4.\}$. Figure 17(a) shows one realization from the simulated process, with solid black lines depicting this partitioning. Similarly to the other two cases considered previously, we estimate the regional distance warping functions and show them in Figure 17(b). The estimated regional distance warping functions show more variability in this case due to fewer data points in each subregion, resulting in an inefficient estimation of the regional variograms. The sub-processes $\{X(\mathbf{s}), \mathbf{s} \in \mathcal{G}_{ij},\ j = 1,...,4\}$ have a common true underlying variogram for each $i = 1,\ldots,4$, but their poor estimation results in regional distance warping functions that differ from those displayed in Figure 15(b) and Figure 16(b). Consequently, the estimated deformations in this case will look different than in the previous two cases. This shows the trade-off between the number of subregions chosen to partition the original domain and the accuracy of estimation of the true regional variograms.

S5 Effectiveness of Classical Multidimensional Scaling

Our method relies on mapping the transformed distance matrix $\Delta_{n\times n} = \{\hat{\phi}(\mathbf{s}_i, \mathbf{s}_j)\}_{i,j=1}^{n}$ to the deformed coordinates $\hat{\theta}(\mathbf{s}_i),\ i = 1,...,n$, (in some finite dimensional space), which we achieve using classical multidimensional scaling (CMDS). The algorithmic step-by-step description of CMDS is given as follows:

Step 1: Compute $\Delta^{(2)}$, which is the Hadamard square of Δ.

Step 2: Compute $B = -\frac{1}{2}F\Delta^{(2)}F$, where $F = I - \frac{1}{n}11'$ is the centering matrix.

Step 3: For a specified dimension $d_{\mathcal{D}}$, compute $d_{\mathcal{D}}$ largest eigenvalues $\lambda_1,\ldots,\lambda_{d_{\mathcal{D}}}$, and the corresponding $d_{\mathcal{D}}$ eigenvectors $e_1,\ldots,e_{d_{\mathcal{D}}}$.

Step 4: Compute $K = E_{d_{\mathcal{D}}}\Lambda_{d_{\mathcal{D}}}^{1/2}$, where $E_{d_{\mathcal{D}}}$ is the matrix of the $d_{\mathcal{D}}$ eigenvectors and $\Lambda_{d_{\mathcal{D}}}$

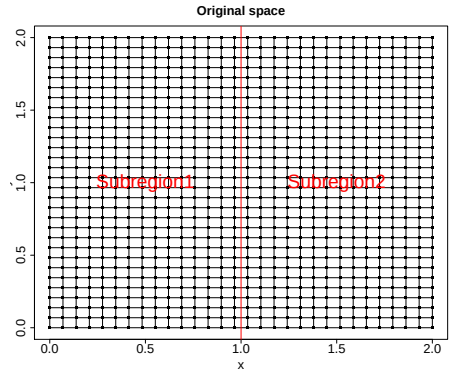


Figure 18: Original space with two subregions.

is the diagonal matrix of $d_{\mathcal{D}}$ eigenvalues. Then, K is the desired coordinate matrix. The CMDS can be easily performed in R using the function `cmdscale` from the `stats` package.

In this simulation, we investigate the effectiveness of CMDS in mapping the transformed distance matrix to deformed coordinates. We consider 30×30 regularly spaced points in the domain $[0, 2]^2$, shown in Figure 18, and divide it into two subregions using the line $x = 1$. We consider two different types of parametric regional distance warping functions:

1. Case 1:

$$\phi_1(\|\mathbf{h}\|) = \begin{cases} \sqrt{8}\frac{e^{a\|\mathbf{h}\|/\sqrt{8}}-1}{e^a-1}, & \text{if } a \neq 0 \\ \|\mathbf{h}\|, & \text{if } a = 0 \end{cases}$$

$$\phi_2(\|\mathbf{h}\|) = \begin{cases} \sqrt{8}\frac{e^{-a\|\mathbf{h}\|/\sqrt{8}}-1}{e^{-a}-1}, & \text{if } a \neq 0 \\ \|\mathbf{h}\|, & \text{if } a = 0 \end{cases}$$

2. Case 2:

$$\phi_1(\|\mathbf{h}\|) = \sqrt{8}\mathbf{B}^C(\|\mathbf{h}\|/\sqrt{8}|\alpha_1, \beta_1) \quad \phi_2(\|\mathbf{h}\|) = \sqrt{8}\mathbf{B}^C(\|\mathbf{h}\|/\sqrt{8}|\alpha_2, \beta_2),$$

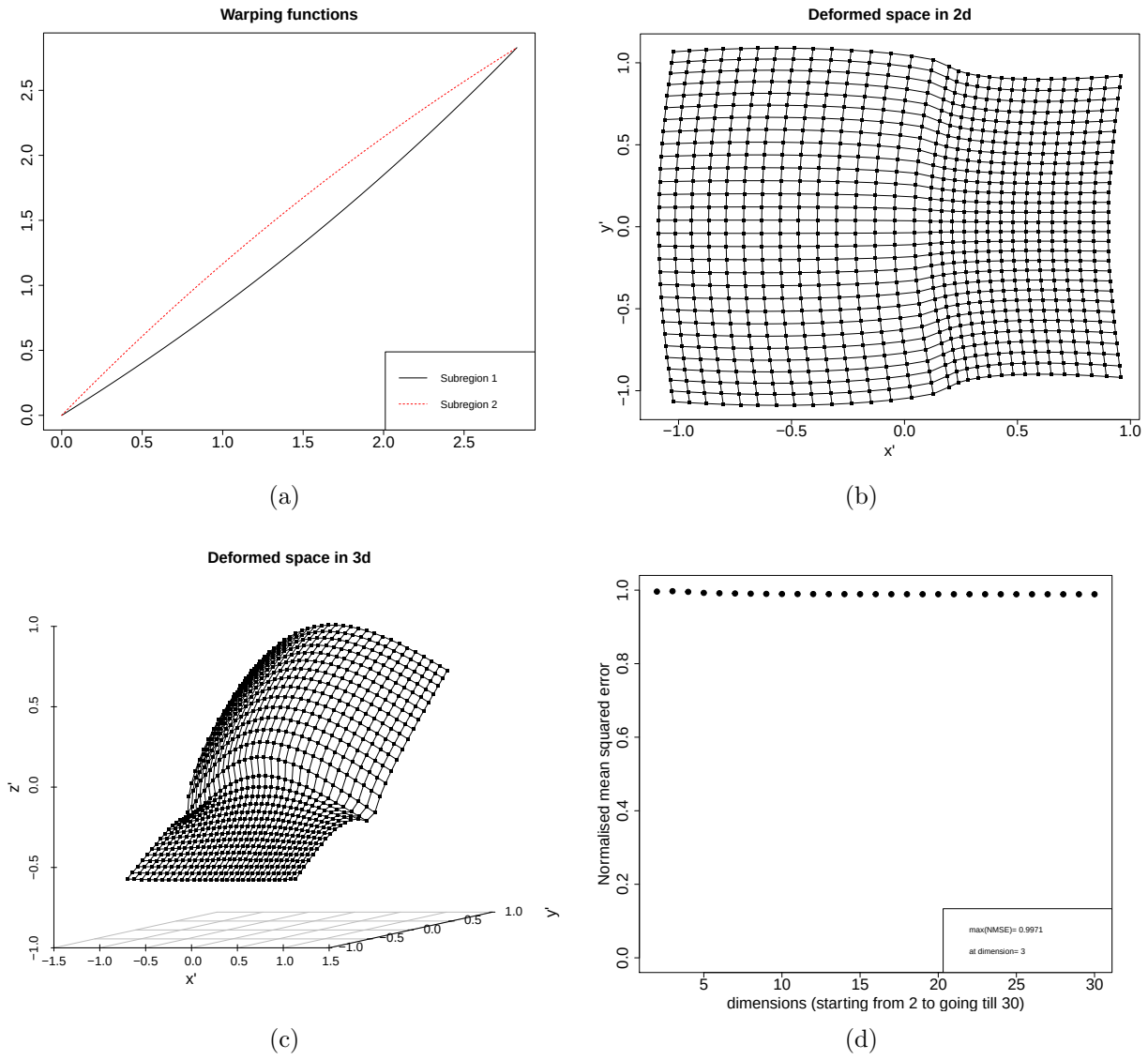
where $\mathbf{B}^C(\cdot|\alpha, \beta)$ is the cumulative distribution function of the Beta distribution for given shape parameters α and β. The parameter $|a|$ controls the intensity of warping in Case 1, whereas the parameters $(\alpha_1, \beta_1, \alpha_2, \beta_2)$ control the shape of the regional distance warping

Case	Parameter setting	Max NMSE	Dimension for Max NMSE
1	(1)	1	2
1	(2)	0.9971	3
1	(3)	0.9791	3
1	(4)	0.946	3
2	(1)	0.9428	3
2	(2)	0.9682	2
2	(3)	0.681	30
2	(4)	0.8258	30

Table 1: Quantitative assessment of the estimated deformed space based on CMDS for different types of regional distance warping functions.

functions in Case 2. We consider four settings for these parameters in each case: Case 1 $|a|$ = (1) 0 (identity warping), (2) 0.5, (3) 1.5, (4) 2.5; Case 2 $(\alpha_1, \beta_1, \alpha_2, \beta_2)$ = (1) $(0.7, 1.5, 1, 1/2.5)$, (2) $(1, 1.4, 1, 1/2.5)$, (3) $(0.25, 1.4, 8, 2)$, (4) $(0.25, 1.4, 2, 1)$. For each of the parameter settings, we compute the GDWF $\phi(\mathbf{s}, \mathbf{s'})$ using the proposed method, and estimate the deformed space in dimensions $d_{\mathcal{D}} = 2, ..., 30$. We then compute the normalized mean squared error (NMSE) between the transformed distance matrix $\Delta_{900\times900} = \{\phi(\mathbf{s}_i, \mathbf{s}_j)\}_{i,j=1}^{900}$ and the distance matrix of the estimated deformed space under different values of $d_{\mathcal{D}}$. A value of NMSE equal to 1 indicates perfect mapping of distances to the coordinates using CMDS.

Table 1 reports the maximum NMSE and the dimension $d_{\mathcal{D}}$ at which the maximum NMSE is attained for all of the eight simulated situations. Figures 19-22 show the summary of results for Case 1 (2) and (4) and Case 2 (1) and (3), respectively (we do not show figures for the other parameter settings for brevity). We observe that, for Case 1, as we increase the value of $|a|$ (i.e., the intensity of regional distance warping functions), the performance of CMDS decreases. However, even in Case 1 (4) with considerable amount of warping, the NMSE is 0.946 for a deformed space of dimension 3, a value very close to 1 indicating a very

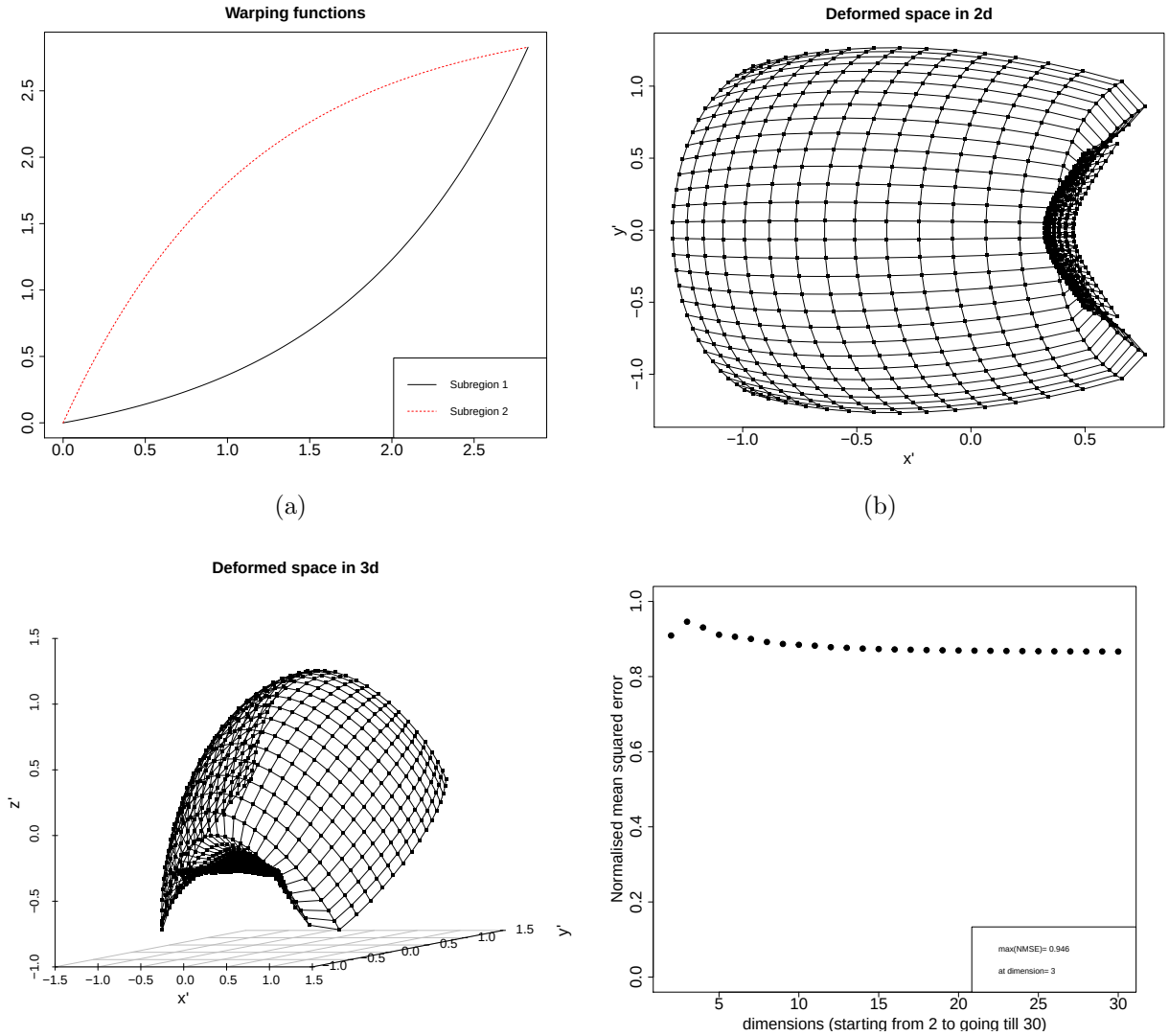


(a) (b)

(c) (d)

Figure 19: Case 1 (2): (a) Regional distance warping function (black: warping for subregion 1, red: warping for subregion 2). (b) Deformed space in 2d. (c) Deformed space in 3d. (d) Plot of NMSE vs. dimension of deformed space.

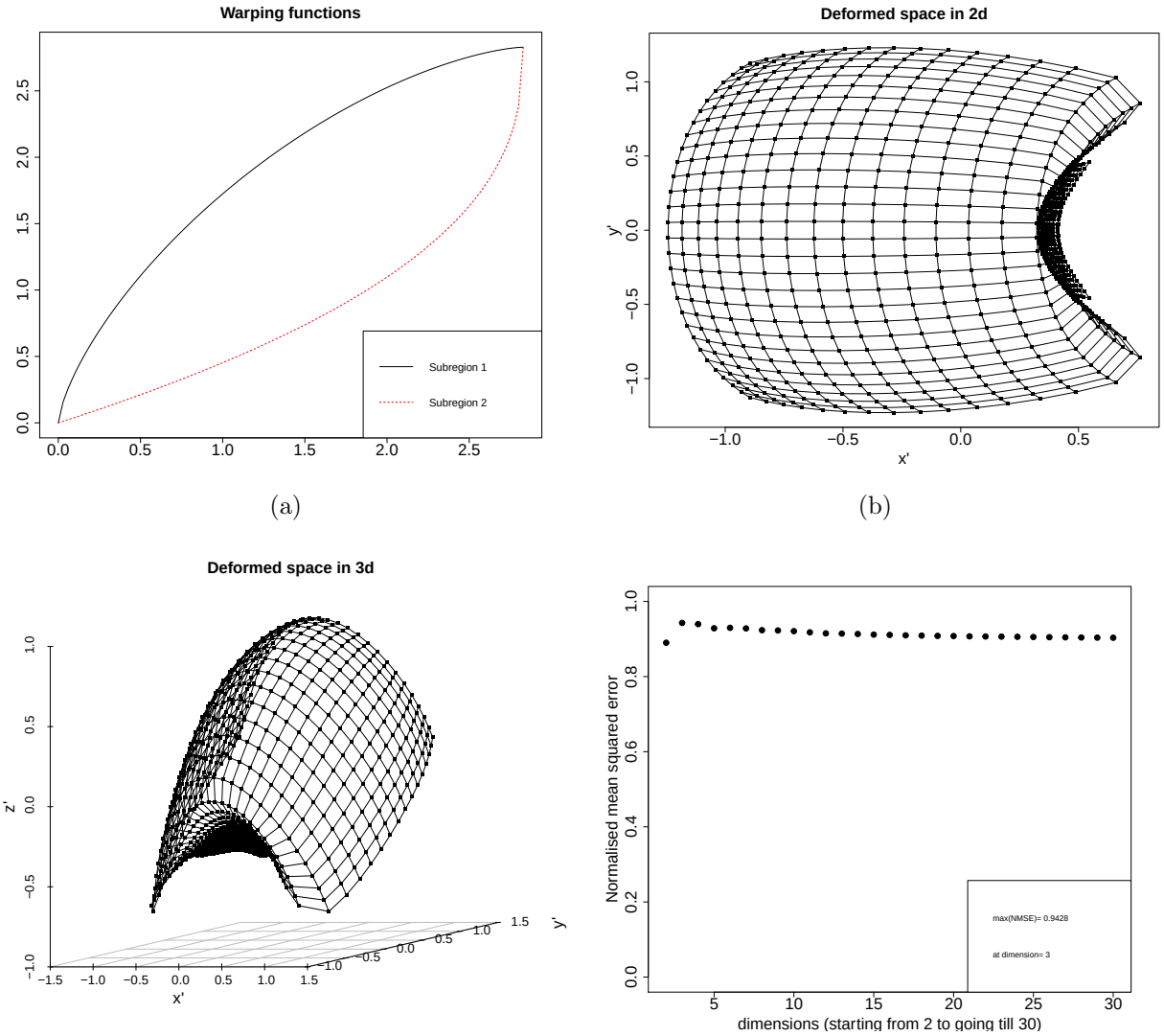
good approximation. For Case 2 (1) and (2), CMDS performs very well and produces NMSE values close to 1. However, for Case 2 (3) and (4) with extreme regional warping functions (in parts of the subdomain the warping functions become nearly vertical or horizontal), CMDS does not perform well. These settings correspond to extreme deformations of the geographic space, which we do not expect in realistic applications.



(a) (b)

(c) (d)

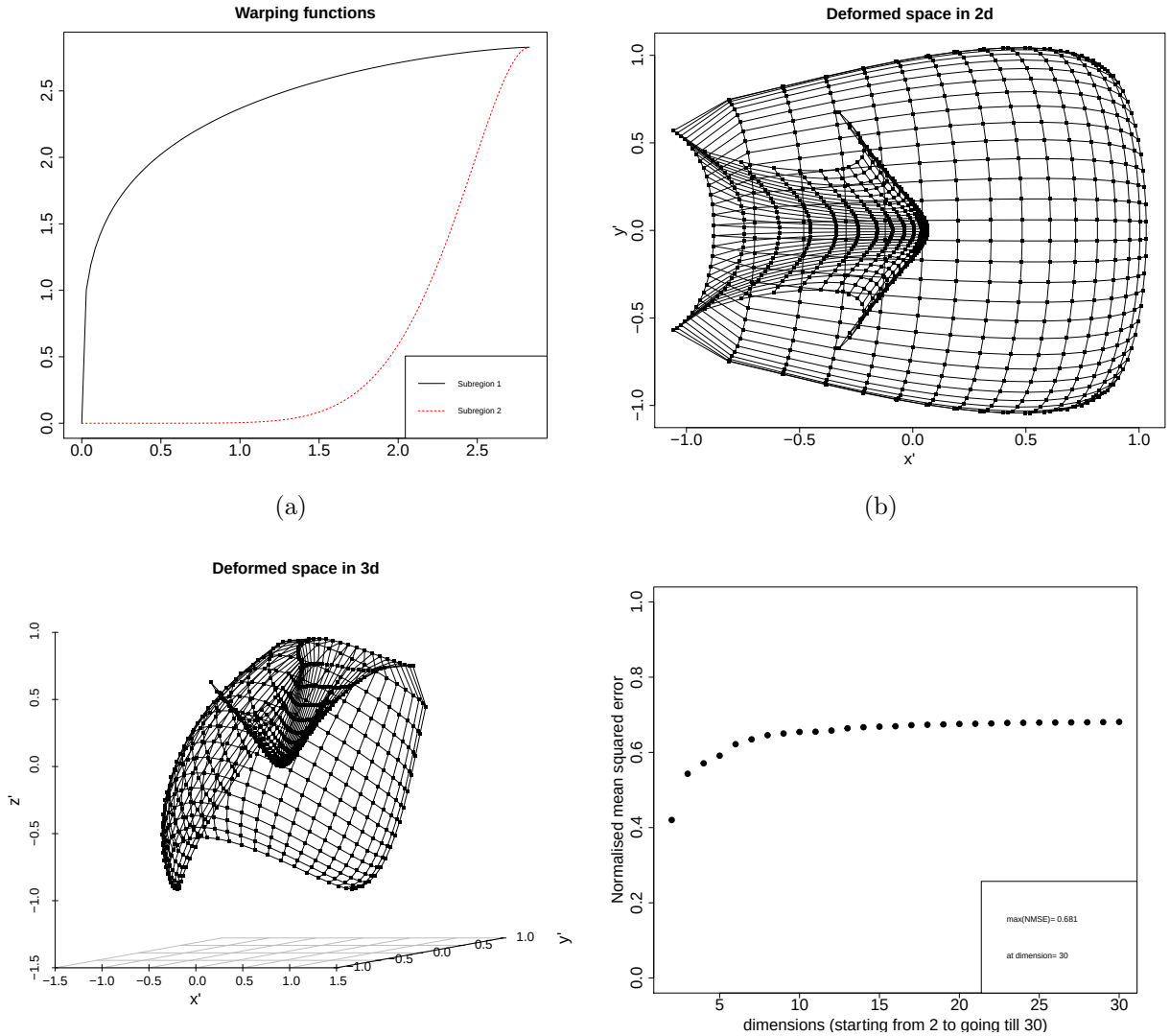
Figure 20: Case 1 (4): (a)-(d) Same as in Figure 19.



(a) (b)

(c) (d)

Figure 21: Case 2 (1): (a)-(d) Same as in Figure 19.



(a) (b)

(c) (d)

Figure 22: Case 2 (3): (a)-(d) Same as in Figure 19.

S6 Computation Timings

In this section, we provide a brief discussion of the computational cost of the proposed approach for the simulation studies and the data application shown in the main paper as well as in the supplementary material. Since the proposed method of estimation of deformation consists of various steps, and each step requires some level of tuning by the user, it is not easy to generalize the computational complexity of the proposed approach. For instance, the user can choose from (1) least square method, (2) weighted least square method, (3) maximum likelihood estimation, etc., for the estimation of regional variograms and the computational cost would be different based on this choice. Similarly, the choice of the shape of subregions might also affect the computational cost while computing the weights in the GDWF. Additionally, there are costs associated with the alignment algorithm and CMDS.

Section	No. of training points	No. of points in the deformed space	Time to estimate the deformed space
4 (main paper)	254	2754	203.21s
3 (main paper)	1200	4900	$\approx$ 500s (per simulation run)
S2 (supplementary)	1200	4900	$\approx$ 500s (per simulation run)
S3 (supplementary)	1200	4900	302.53s

Table 2: Run times to estimate the deformed space in Section 4 of the main paper (first row), Section 3 of the main paper (second row), Section S2 of the supplementary material (third row) and Section S3 of the supplementary material (fourth row). Note that Section 3 of the main paper and Section S2 of the supplementary material involve multiple simulation runs, and therefore, the reported time is the approximate time per simulation run. The timings are based on computations in R (R Core Team, 2020) on a MacBook Pro laptop with 2.3 GHz Intel Core i9 processor and 16 GB of 2667MHz DDR4 RAM.

However, to provide some perspective on the computational cost, we report the computational timings in Table 2 for the simulation studies and the data application considered. The reported timings reflect the computational cost as a function of the size of the training data and the number of points to be estimated in the deformed space, when regional variograms

are estimated via MLE and subregions share a linear boundary.

S7 Additional Figures from the Data Application

In this section, we present some additional results from the Section 4 of the main manuscript. The alignment results based on the entire data are shown in Figure 23. The estimated regional Matérn variogram models, standardized using their respective regional variances are shown in Figure 23(b). The estimated variance parameter varies slightly for the two subregions; however, due to their negligible difference, we chose to ignore this insignificant nonstationarity in variance. The estimated regional variograms show long range spatial dependence for the Eastern subregion and relatively shorter range spatial dependence for the Western subregion. Figure 23(d) shows the two estimated regional distance warping functions obtained via registration of the standardized estimated regional variograms. The large deviation of regional distance warping functions from identity warping confirms a high degree of nonstationarity when the entire region is considered. Based on these regional distance warping functions, we can infer that modeling the two subregions with a common stationary variogram in the geographic space is clearly an imprecise approach to describe the spatial dependence of this process.

The entire data-based estimated deformed space for locations corresponding to the observed data, and the fine grid of locations chosen for interpolation, are shown in Figure 24 for the first three dimensions of maximum variation; the optimal value of ψ is chosen to be 15. We observe that the deformation leads to a very tight configuration of highly correlated points corresponding to the Eastern subregion and a highly sparse configuration corresponding to the Western subregion. This results in an approximately constant spatial

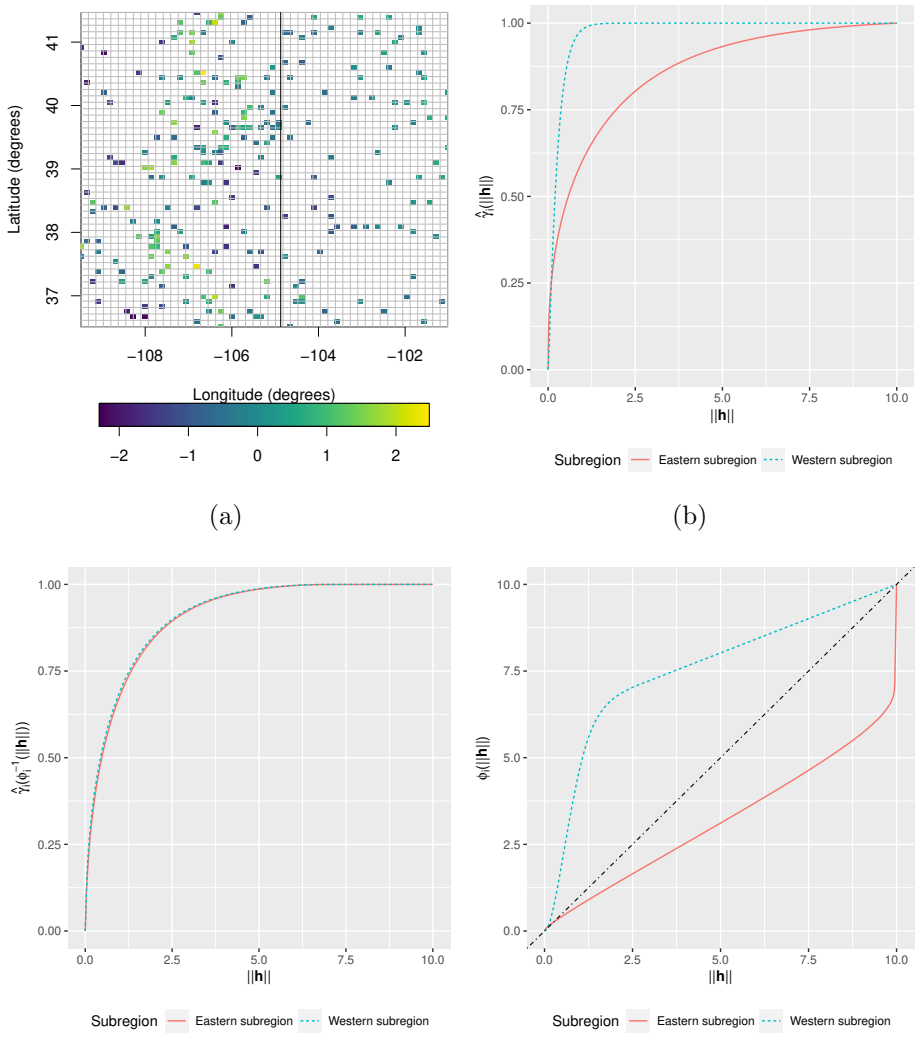


Figure 23: (a) Standardized log-transformed total annual precipitation data with the solid black line depicting the chosen partitioning. The grey colored lines depict the additional grid locations for the interpolation. (b) Estimated standardized regional variograms. (c) Registered variograms. (d) Regional distance warping functions.

range and smoothness over the entire deformed space, making it suitable for modeling using a stationary model.

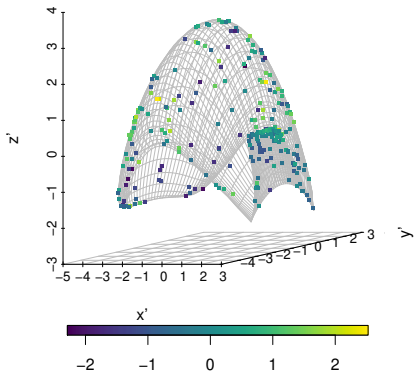


Figure 24: Standardized log-transformed total annual precipitation data in the estimated deformed space (first three dimensions of maximum variation).

References

Paciorek, C. J. and M. J. Schervish (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics 17*(5), 483–506.

Perrin, O. and P. Monestiez (1999). Modelling of non-stationary spatial structure using parametric radial basis deformations. In J. Gómez-Hernández, A. Soares, and R. Froidevaux (Eds.), *geoENV II — Geostatistics for Environmental Applications*, Dordrecht, pp. 175–186. Springer Netherlands.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association 99*(465), 250–261.