Statistical Downscaling of Sea Surface Temperature Projections with a Multivariate Gaussian Process Model

Ayesha Ekanayaka
University of Cincinnati
Cincinnati, Ohio,USA
ekanaykk@mail.uc.edu

Emily Kang
University of Cincinnati
Cincinnati, OH, USA
kangel@ucmail.uc.edu

Peter Kalmus
Jet Propulsion Laboratory
Pasadena, CA, USA
peter.m.kalmus@jpl.nasa.gov

Amy Braverman
Jet Propulsion Laboratory
Pasadena, CA, USA
amy.j.braverman@jpl.nasa.gov

1 Introduction

General Circulation Models (GCMs) provide extremely informative future projections for the global climate system under different climate scenarios. These projections are made at relatively lower resolutions due to computational limitations. However, for the regional climate studies, there is an increased demand for GCM projections downscaled to much higher resolutions. Downscaling techniques are majorly categorized into two; Dynamical Downscaling (DD) and Statistical Downscaling (SD). Regional Climate Models (RCMs) with boundary conditions generated from GCMs are used to perform DD whereas in SD, a statistical relationship is built between high-resolution observational data and coarse resolution model outputs. DD is computationally expensive. Meanwhile, SD can be used to perform large-scale downscaling at a much less computational cost.

There are numerous previous SD methods found in the literature which are developed using a range of statistical tools. A SD method not always is a traditional statistical model-based approach. For example, deterministic tools such as Model Output Statistics (MOS), smoothing, and interpolation techniques are widely used to establish the empirical relationship [11],[12]. However, numerical methods as well as many machine learning approaches do not often produce uncertainty estimates which is a demanding skill for subsequent regional studies [4]. Methods based on probabilistic models such as regression and generalized linear models are capable of producing uncertainty estimates. But such SD methods can be used only in the presence of potential predictor variables [9],[1]. Moreover, downscaled results may vary with the choice of predictor variables [8]. An advanced spatio-temporal model directly relating coarse model projections to high-resolution observations is also lately been utilized effectively accounting for spatio-temporal dependencies [2]. But this method involves MCMC simulations which is not computationally desired for large-scale studies.

We propose a multivariate Gaussian process model to downscale low-resolution model outputs using high-resolution remote sensing data. Our proposed method is computationally efficient, accounts for spatio-temporal dependence, and provides associated uncertainty, that can be used in Earth science contexts with large data sets. Here, we demonstrate it by downscaling Sea Surface Temperature (SST) projections to the 1 km scale. We perform a representative case study and validation in the Great Barrier Reef (GBR) region.

2022 NeurIPS Workshop on Gaussian Processes, Spatiotemporal Modeling, and Decision-making Systems.

2 Data and model output

We use monthly averaged NASA/JPL Multiscale Ultrahigh Resolution (MUR, [6]) satellite SST data at 1 km resolution from June 2002 to December 2020 (a total of 223 months). We use monthly SST outputs from 19 Coupled Model Intercomparison Project Phase 6 (CMIP6) GCM models under the Shared Socioeconomic Pathways (SSPs) [10]. For the demonstration purpose, we chose SSP126 which is quite an optimistic climate scenario. The model time series are re-gridded to a common 1° grid, and run from June 2002 to December 2099 (1183 months). At each grid location, we take the mean of these model time series. The study area includes a total of 309,700 (N) 1 km MUR pixels and 35 (M) 1° coarse grid cells.

3 Methodology

3.1 Model Specification

Let $Y_{t,1}(s)$ denotes the monthly averaged observational SST at MUR location s on month t and $Y_{t,2}(s)$ to be the SST obtained from the state-of-art (standard) interpolation-based SD method [12]. Details on standard downscaling is provided in Appendix A. We assume that the bi-variate process $\mathbf{Y}_t(s) = (Y_{t,1}(s), Y_{t,2}(s))'$ can be additively modeled as,

$$\mathbf{Y}_t(s) = \boldsymbol{\mu}_t(s) + \boldsymbol{\nu}_t(s), \tag{1}$$

where $\boldsymbol{\mu}_t(s)$ denotes the mean of the process, explaining large-scale spatio-temporal variations; the second term $\boldsymbol{\nu}_t(s)$ is a multivariate Gaussian process (GP) to capture small-scale spatial variations. To alleviate computational difficulty associated with GP, we use basis representation:

$$\nu_{t,i}(s) = \sum_{k=1}^{K} \xi_i^k U^k(s) + \sum_{l=1}^{L} \eta_{t,i}^l S^l(s) + \epsilon_{t,i}(s) \tag{2}$$

where, ξ_i^k for $k = 1, \ldots, K$ are a set of fixed effects and $\eta_{t,i}^l$ for $l = 1, \ldots, L$ are a set of random effects to capture additional fine-scale spatial variations but assumed to be temporally independent. Here, $U^k(s)$ and $S^l(s)$ are known basis functions. The last term $\epsilon_{t,i}(s)$ in equation (2) is a Gaussian white noise process with zero mean and variance τ_i^2, independent with $\{\eta_{t,i}^l\}$. We assume further that the random vector $\boldsymbol{\eta}_t^l = (\eta_{t,1}^l, \eta_{t,2}^l)'$ at each level l follows $\boldsymbol{\eta}_t^l \sim N\left(0, \mathbb{Q}_l^{-1}\right)$. Motivated by multivariate Basis Graphical Lasso (BGL,[7]), we use empirical orthogonal functions as basis function in the model and assume that $\boldsymbol{\eta}_t^l$'s are independent across L levels [7]. Note that this assumption of independence across levels still allows spatial dependence within and across $Y_{t,1}(s)$ and $Y_{t,2}(s)$. The resulting precision matrix $\mathbb{Q}$ of the full vector $\boldsymbol{\eta}_t = (\boldsymbol{\eta}_t^1, .., \boldsymbol{\eta}_t^L)'$ is a sparse matrix of the form; $\mathbb{Q} = diag(\mathbb{Q}_1, .., \mathbb{Q}_L)$. A detailed theoretical justification for independence assumption when basis functions are orthogonal is discussed in [7].

3.2 Implementation and Inference

Let $\mathbf{U}(s) = \left(U^1(s), \ldots, U^K(s)\right)'$ and $\mathbf{S}(s) = \left(S^1(s), \ldots, S^L(s)\right)'$. Considering all available N fine-resolution observation locations, we re-write the model for the full data vector as below.

$$\begin{pmatrix} \mathbf{Y}_t(s_1) \\ \vdots \\ \mathbf{Y}_t(s_N) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_t(s_1) \\ \vdots \\ \boldsymbol{\mu}_t(s_N) \end{pmatrix} + \begin{pmatrix} \mathbf{U}'(s_1) \otimes \mathbb{I}_2 \\ \vdots \\ \mathbf{U}'(s_N) \otimes \mathbb{I}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\xi}^1 \\ \vdots \\ \boldsymbol{\xi}^K \end{pmatrix} + \begin{pmatrix} \mathbf{S}'(s_1) \otimes \mathbb{I}_2 \\ \vdots \\ \mathbf{S}'(s_N) \otimes \mathbb{I}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_t^1 \\ \vdots \\ \boldsymbol{\eta}_t^L \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_t(s_1) \\ \vdots \\ \boldsymbol{\epsilon}_t(s_N) \end{pmatrix} \tag{3}$$

where, $\boldsymbol{\xi}^k = (\boldsymbol{\xi}_1^k, \boldsymbol{\xi}_2^k)'$ for $k = 1, \ldots, K$ and $\boldsymbol{\eta}_t^l = (\eta_{t,1}^l, \eta_{t,2}^l)'$ for $l = 1, \ldots, L$. For simplicity we write this model as,

$$\underset{2N\times 1}{\mathbf{Y}_t} = \underset{2N\times 1}{\boldsymbol{\mu}_t} + \underset{2N\times 2K}{\mathbb{U}} \underset{2K\times 1}{\boldsymbol{\xi}} + \underset{2N\times 2L}{\mathbb{S}} \underset{2L\times 1}{\boldsymbol{\eta}_t} + \underset{2N\times 1}{\boldsymbol{\epsilon}_t}, \tag{4}$$

Here, $\mathbb{U} = \mathbf{U} \otimes \mathbb{I}_2$, $\mathbb{S} = \mathbf{S} \otimes \mathbb{I}_2$ where $\mathbf{U}$ and $\mathbf{S}$ are basis matrices. Estimation is performed in two steps. First, we set $\mu_{t,1}(s) = \mu_{t,2}(s) = \mu_t(s)$ where, $\mu_t(s)$ is estimated using the 5-year moving average

of the second process $Y_{t,2}(s)$; i.e. $\hat{\mu}_t(s) = \frac{\sum_{t-5}^{t-1} Y_{t,2}(s)}{5}$ and subtract it from original processes. We calculate Empirical Orthogonal Functions (EOFs) and select a subset of EOFs as the basis functions such that the desired percentage of total variance is explained. To separate EOFs into $\mathbf{U}$ and $\mathbf{S}$ basis matrices and to estimate fixed effects, we fit a no intercept linear regression between de-trended data and EOFs [5]. Regression coefficients with comparatively larger magnitude are chosen to be the estimated fixed effects. Basis functions corresponds to those fixed effects are taken into the basis matrix $\mathbf{U}$ while the rest of the basis functions are included in the basis matrix $\mathbf{S}$. A detailed explanation on calculation and separation of EOFs is provided in Appendix C.Then we obtain the full detail residual vector $\mathbf{Z}_t = \mathbf{Y}_t - \hat{\boldsymbol{\mu}}_t - \mathbb{U}\hat{\boldsymbol{\xi}}$. Assuming the residuals from each observational month t is an independent realization of $\mathbf{Z}_t$ we can write joint negative log-likelihood up-to a normalizing constant as,

$$log(det(\mathbb{\Sigma})) + \frac{\sum_{t=1}^{T_o} \mathbf{Z}_t' \mathbb{\Sigma}^{-1} \mathbf{Z}_t}{T_o} = log(det(\mathbb{\Sigma})) + tr(\hat{\mathbb{\Sigma}}\mathbb{\Sigma}^{-1}) \tag{5}$$

where T_o is the number of observational months, $Var(\mathbf{Z}_t) = \mathbb{\Sigma}$, $\hat{\mathbb{\Sigma}} = \frac{\sum_{t=1}^{T_o} \mathbf{Z}_t' \mathbf{Z}_t}{T_o}$ is the sample covariance matrix. Note that here, $\mathbb{\Sigma} = \mathbb{S}\mathbb{Q}^{-1}\mathbb{S}' + \mathbb{D}$, where $\mathbb{D} = diag(\tau_1^2, \tau_2^2) \otimes \mathbb{I}_N$.

In the second step, we solve l_1-penalized maximum likelihood expression,

$$\hat{\mathbb{Q}} \in \underset{\mathbb{Q} \succeq 0}{\arg\min}\, log(det(\mathbb{S}\mathbb{Q}^{-1}\mathbb{S}' + \mathbb{D})) + tr(\hat{\mathbb{\Sigma}}(\mathbb{S}\mathbb{Q}^{-1}\mathbb{S}' + \mathbb{D})^{-1}) + P(\mathbb{Q}) \tag{6}$$

where,

$$P(\mathbb{Q}) = P(\mathbb{Q}_1, .., \mathbb{Q}_L) = \rho \sum_{l=1}^{L-1} \sum_{i \neq j} |(\mathbb{Q}_l)_{ij} - (\mathbb{Q}_{l+1})_{ij}|$$

Here, ρ is a penalty which penalize $\mathbb{Q}_l$ matrices at adjacent levels if the off-diagonals are not similar. Our penalization parameter here is similar to the fusion penalty introduced in [7] favouring the coherence of the SST processes to vary smoothly across the levels. An appropriate value for the penalty ρ can be chosen through a cross-validation procedure. Note that, evaluation of the likelihood in expression (6) requires an expensive Choleskey decomposition ($\mathcal{O}(p^3N^3)$) where p is the number of processes and N is the number of observational locations. Appendix B explains efficient computation of the likelihood in (6).

3.3 Prediction

Recall we assume a bi-variate normal distribution for the vector of random effects $\boldsymbol{\eta}_t^l = (\eta_{t,1}^l, \eta_{t,2}^l)'$ at each level l. Once $\hat{\mathbb{Q}}_l$ is estimated, it is straightforward to obtain the conditional distribution of $\eta_{t,1}^l | \eta_{t,2}^l$.

$$\eta_{t,1}^l | \eta_{t,2}^l \sim N\Big(\frac{\sigma_{12}}{\sigma_{22}} \eta_{t,2}^l, \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\Big) \tag{7}$$

Note that we can write the model only for the second process $\mathbf{Y}_{t,2} = \Big(Y_{t,2}(s_1), \ldots, Y_{t,2}(s_N)\Big)'$ as,

$$\mathbf{Y}_{t,2} = \boldsymbol{\mu}_{t,2} + \mathbf{U}\boldsymbol{\xi}_2 + \mathbf{S}\mathbf{A}\boldsymbol{\eta}_t + \boldsymbol{\epsilon}_{t,2} \tag{8}$$

where $\boldsymbol{\xi}_2 = \Big(\xi_2^1, \ldots, \xi_2^K\Big)'$ and $\mathbf{A}$ is a known linear transformation matrix such that $\mathbf{A}\boldsymbol{\eta}_t = \Big(\eta_{t,2}^1, \ldots, \eta_{t,2}^L\Big)'$. We have $\boldsymbol{\xi}_2$ already estimated and given the second process $\mathbf{Y}_{t,2}$ for a future month $t > T_o$, we can easily estimate $\boldsymbol{\mu}_{t,2}$ as described in section 3.2. Thus, we first obtain Generalized Least Squares (GLS) estimates; $\hat{\boldsymbol{\eta}}_{t,2} = \Big(\mathbf{S}\mathbf{\Sigma}_2^{-1}\mathbf{S}'\Big)^{-1}\mathbf{S}\mathbf{\Sigma}_2^{-1}\mathbf{Z}_{t,2}$ where, $\mathbf{\Sigma}_2 = Var(\mathbf{Z}_{t,2}) = Var(\mathbf{S}\mathbf{A}\boldsymbol{\eta}_t + \boldsymbol{\epsilon}_{t,2}) = \mathbf{S}\mathbf{A}\mathbb{Q}^{-1}\mathbf{A}'\mathbf{S}' + \mathbf{D}_2$ and $\mathbf{D}_2 = \tau_2^2\mathbb{I}_N$. Now, using the conditional distribution in equation (7) and the law of total expectation, we can calculate conditional expectation of $\eta_{t,1}^l$ as,

$$\hat{\eta}_{t,1}^l = Exp\Big[\eta_{t,1}^l | \mathbf{Z}_{t,2}\Big] = Exp\Big[Exp\Big[\eta_{t,1}^l | \eta_{t,2}^l, \mathbf{Z}_{t,2}\Big] | \mathbf{Z}_{t,2}\Big] \tag{9}$$

with $Var\Big[\eta_{t,1}^l | \mathbf{Z}_{t,2}\Big] = Exp\Big[Var\Big[\eta_{t,2}^l | \eta_{t,1}^l, \mathbf{Z}_{t,2}\Big] | \mathbf{Z}_{t,2}\Big] + Var\Big[Exp\Big[\eta_{t,1}^l | \eta_{t,2}^l, \mathbf{Z}_{t,2}\Big] | \mathbf{Z}_{t,2}\Big]$ conditional variance. Finally, we obtain the vector of downscaled SSTs $\hat{\mathbf{Y}}_{t,1}$ for a future month $t > T_o$ as,

$$\hat{\mathbf{Y}}_{t,1} = \hat{\boldsymbol{\mu}}_{t,1} + \mathbf{U}\hat{\boldsymbol{\xi}}_1 + \mathbf{S}\hat{\boldsymbol{\eta}}_{t,1} \tag{10}$$

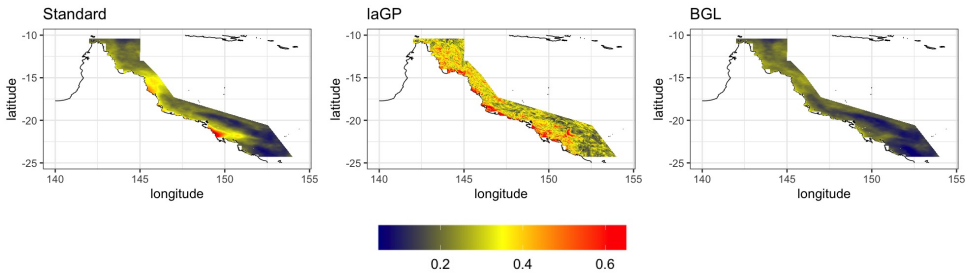

Figure 1:
MSE maps of standard(left), laGP(middle) and BGL(right) methods from MUR validation under ssp126.

Table 1: Averaged MSE separated by seasons from MUR validation

ssp126				
Season	GCM	Standard	laGP	BGL
Summer	0.396	0.305	0.419	**0.297**
Autumn	0.472	0.105	0.221	**0.088**
Winter	0.878	0.250	0.335	**0.163**
Spring	0.380	0.212	0.356	**0.153**
Overall	0.531	0.218	0.353	**0.175**

4 Results

We performed validation leaving out the three years from 2018 to 2020 to assess the performance of the proposed method. Downscaling for the GBR region (a total of 309,700 high-resolution pixels) was performed in Matlab on a Macbook Air with an 8-core 3.2GHz processor and 8 GB RAM. Given the second process $\mathbf{Y}_{t,2}$ for $t > T_o$, it only took 6.8 seconds to produce downscaled SSTs for a single future month.

We compare performance with the state-of-art (standard) SD method in literature for SSTs proposed using an interpolation-based method [12] and with another approach using local approximate Gaussian Process (laGP) [3]. In table 1, we compare Mean Squared Error(MSE) values averaged across locations. First column in the table shows MSE values obtained by simply interpolating row GCM data. Notice that our method has the lowest MSE values and the percentage reduction in overall MSE is about 20% when it is compared to the standard SD method. In figure 1, we compare the maps of MSE values averaged across seasons. Notice the striking improvement, especially along the coastal line. This improved accuracy is expected because, unlike the stat-of-art, our model is capable of successfully accounting for fine-scale spatial dependencies. We observe an unusual instability in the MSE map from the laGP method which is due its local structure.

5 Conclusion and Discussion

We have presented a novel statistical downscaling method. Our method is computationally feasible for large data sets, accounts for spatio-temporal dependencies, provides meaningful uncertainty estimates and has significantly reduced overall MSE. Therefore, it is suitable for a wide range of applications in Earth science and other fields, e.g. for accomplishing the SD of coarse-scale global climate model projections using fine-scale observational data. A possible extension is to generalize the current model to the framework of autoregressive co-kriging for multi-fidelity model output and then consider the observations, regional climate model output, and global climate model output as the high-, medium-, and low-fidelity data, respectively.

References

[1] S. Beecham, M. Rashid, and R. K. Chowdhury. Statistical downscaling of multi-site daily rainfall in a south australian catchment using a generalized linear model. *International Journal of Climatology*, 34(14):3654–3670, 2014.

[2] V. J. Berrocal, A. E. Gelfand, and D. M. Holland. A spatio-temporal downscaler for output from numerical models. *Journal of agricultural, biological, and environmental statistics*, 15(2):176–197, 2010.

[3] R. Gramacy. lagp: Large-scale spatial modeling via local approximate gaussian processes in r. *Journal of Statistical Software, Articles*, 72(1):1–46, 2016.

[4] S. F. Heron, J. A. Maynard, R. Van Hooidonk, and C. M. Eakin. Warming trends and bleaching stress of the world's coral reefs 1985–2012. *Scientific reports*, 6:38402, 2016.

[5] H.-C. Huang and N. Cressie. Deterministic/stochastic wavelet decomposition for recovery of signal from noisy data. *Technometrics*, 42(3):262–276, 2000.

[6] JPL MUR MEaSUREs Project. Ghrsst level 4 mur global foundation sea surface temperature analysis (v4.1), 2015.

[7] M. Krock, W. Kleiber, D. Hammerling, and S. Becker. Modeling massive multivariate spatial data with the basis graphical lasso. 01 2021.

[8] M. Lafaysse, B. Hingray, A. Mezghani, J. Gailhard, and L. Terray. Internal variability and model uncertainty components in future hydrometeorological projections: The alpine durance basin. *Water Resources Research*, 50(4):3317–3341, 2014.

[9] P. Mishra, D. Khare, A. Mondal, and S. Kundu. Multiple linear regression based statistical downscaling of daily precipitation in a canal command. 2014.

[10] B. C. O'Neill, E. Kriegler, K. Riahi, K. L. Ebi, S. Hallegatte, T. R. Carter, R. Mathur, and D. P. van Vuuren. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic change*, 122(3):387–400, 2014.

[11] J. Schmidli, C. Frei, and P. L. Vidale. Downscaling from gcm precipitation: a benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology*, 26(5):679–689, 2006.

[12] R. Van Hooidonk, J. A. Maynard, Y. Liu, and S.-K. Lee. Downscaled projections of caribbean coral bleaching that can inform conservation planning. *Global change biology*, 21(9):3389–3401, 2015.

A Standard Downscaling

This standard downscaling method is easy to perform and costs computationally less. But the method does not produce any uncertainty estimates and from the results in section 4, we prove that the SSTs produced from this method have higher prediction errors. To perform this downscaling, first model mean needs to be subtracted from GCM data. i.e if $W_t(B)$ is the GCM outputs at coarse grid cell B for month t, mean centered GCM projection $W_t^{centered}(B)$ is obtained as,

$$W_t^{centered}(B) = W_t(B) - \frac{\sum_{t=1}^{T_o} W_t(B)}{T_o} \tag{11}$$

where T_o is the total number of observational months. For example, if month t is a January, T_o is the total number of Januaries in the observational period. Then the resulting time series $\{W_t^{centered}\}$ for $t = 1, \ldots, T; T > T_o$ is interpolated to MUR pixels using bivariate interpolation. If $\{w_t^{centered}\}$ is the interpolated time series which is now available at MUR resolution, $Y_{t,2}(s)$ at MUR pixel s is finally obtained by adding $w_t^{centered}(s)$ to the observational SST average at MUR location s.

$$Y_{t,2}(s) = \frac{\sum_{t=1}^{T_o} Y_{t,1}(s)}{T_o} + w_t^{centered}(s) \tag{12}$$

B Efficient Likelihood Computation

Using the Sherman-Morrison-Woodbury formula, likelihood expression in (6) can be re-written reducing likelihood evaluation to $\mathcal{O}(p^3L^3)$ as,

$$log(det(\mathbb{Q} + \mathbf{S}'\mathbb{D}^{-1}\mathbf{S})) - log(det(\mathbb{Q})) - \\ tr(\mathbf{S}'\mathbb{D}^{-1}\hat{\mathbf{\Sigma}}\mathbb{D}^{-1}\mathbf{S}(\mathbb{Q} + \mathbf{S}'\mathbb{D}^{-1}\mathbf{S})^{-1}) + P(\mathbb{Q}) \tag{13}$$

The block-diagonal structure of $\mathbb{Q}$ further reduces matrix computation of size $pL \times pL$ to L computations of $p \times p$ matrices. However, the likelihood in expression (13) is still non-smooth and non-convex with respect to $\mathbb{Q}$. Thus, estimation is performed using difference-of-convex (DC) algorithm where the next guess of $\mathbb{Q}$ is obtained by solving a convex optimization problem with the concave part linearized at the previous guess $\mathbb{Q}^{(j)}$ [7].

C EOFs Calculation and Separation

In the current study, we only have a limited number of observational months. Thus, to have a reasonable number of EOFs per model, we group months into seasons and fit four different models for the four seasons. For example, to fit the model for the summer season, we combine data from December,January and February months. Therefore, the total number of observational months for summer season $T_O = \sum_{j=1}^{3} T_{o_j}$; where $j = 1, 2, 3$ for December, January and February. EOFs are calculated combining de-trended data from the respective months into $N \times p \times T_O$ data matrix where N is the total number of MUR locations and p is the number of processes. Then we take economic form of the Singular Value Decomposition (SVD) of the constructed matrix.

EOFs only with large eigen values are chosen such that the desired percentage of the total variance is explained. Then we follow the idea presented in [5] to separate EOFs into $\mathbb{U}$ and $\mathbb{S}$ basis matrices. That is, we use the QQ-plot of regression coefficients obtained by fitting a no-intercept regression between $\mathbf{Y}_2 - \hat{\boldsymbol{\mu}}_2$ (or $\mathbf{Y}_1 - \hat{\boldsymbol{\mu}}_1$) and EOFs. Here, $\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i = \Big[(\mathbf{Y}_{1,i} - \hat{\boldsymbol{\mu}}_{1,i}), \ldots, (\mathbf{Y}_{T_o,i} - \hat{\boldsymbol{\mu}}_{T_o,i})\Big]$ where $\mathbf{Y}_{t,i} = \Big(Y_{t,i}(s_1), \ldots, Y_{t,i}(s_N)\Big)'$ and $\hat{\boldsymbol{\mu}}_{t,i} = \Big(\hat{\mu}_{t,i}(s_1), \ldots, \hat{\mu}_{t,i}(s_N)\Big)'$. Let $\{\omega_m\}$ where $m = 1, \ldots, K + L$ be the vector of OLS estimates and $||\omega_m||^p$ be the p-quantile of the vector of $\{|\omega_m|\}$. Then if q^p is the p-quantile of the standard Gaussian distribution, a slope $\hat{\tau}$ is estimated as follows.

$$\hat{\tau} = ||\omega_m||^{1-2\alpha}/q^{1-\alpha} \tag{14}$$

Here α is chosen such that the QQ-line with slope $\hat{\tau}$ fits the estimated coefficients that are small or moderate in absolute value. Finally, A large absolute coefficient is defined as a coefficient which satisfies: $|\omega_m| > \hat{\tau} \times q_{max}$ where, $q_{max} \equiv max\{q^{(1-p)} : |\omega_m|^{(1-2p)} < \hat{\tau}q^{(1-p)}, p = 1/2l, 2/2l, ..., 1/2\}$. See the figure 2. Based on the QQ-plot we choose the basis functions for matrix $\mathbb{U}$ whose coefficients are significantly large while the rest of the basis functions go into basis matrix $\mathbb{S}$.

D laGP Model

In this model setup we assume,

$$\underbrace{\begin{pmatrix} Y_{t,1}(s_1) \\ \vdots \\ Y_{t,1}(s_N) \end{pmatrix}}_{\mathbf{Y}_{t,1}} = \underbrace{\begin{pmatrix} \mu_{t,1}(s_1) \\ \vdots \\ \mu_{t,1}(s_N) \end{pmatrix}}_{\boldsymbol{\mu}_{t,1}} + \underbrace{\begin{pmatrix} \epsilon_{t,1}(s_1) \\ \vdots \\ \epsilon_{t,1}(s_N) \end{pmatrix}}_{\boldsymbol{\epsilon}_{t,1}} \tag{15}$$

where, $\boldsymbol{\mu}_{t,1}$ is the vector of mean component and assume $\boldsymbol{\epsilon}_{t,1} \sim GP\Big(\boldsymbol{\mu}(.), \mathbf{C}(.,.)\Big)$. GP is a popular modeling tool in spatial statistics and computer experiments because of its ability to handle non-stationary processes through flexible covariance structures. But, as modeling with GP becomes computationally infeasible with large data, we used local approximate Gaussian Process (laGP) which

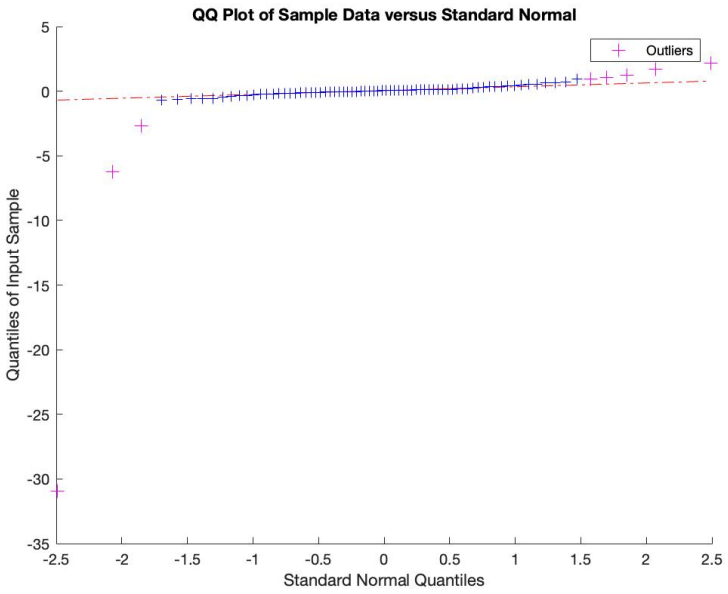

Figure 2:
QQ-plot of estimated coefficients of basis functions for the summer season.

is an efficient approximation to full GP [3]. The method is implemented in R package laGP. To make a prediction at a location x_o, laGP() function takes two inputs called the design matrix ($\mathbb{X}$) and the response vector ($\mathbf{Z}$). We set $\mathbf{Z} = \left(\mathbf{Z}'_{1,1}, \ldots, \mathbf{Z}'_{T_o,1}\right)'$ where $\mathbf{Z}_{t,1} = \mathbf{Y}_{t,1} - \hat{\boldsymbol{\mu}}_{t,1}$. Again, $\hat{\boldsymbol{\mu}}_{t,1}$ here is the estimated mean as described in section 3.2. We used three input variables for the design matrix $\mathbb{X}$ namely; longitude, latitude and the variable $\mathbf{Z}_{t,2} = \mathbf{Y}_{t,2} - \hat{\boldsymbol{\mu}}_{t,2}$. For a future month $t > T_o$, $\hat{\boldsymbol{\epsilon}}_{t,1}$ will be the predictive mean with the respective predictive variance from laGP. Finally, the vector of downscaled SSTs $\hat{\mathbf{Y}}_{t,1}$ for $t > T_o$ is,

$$\hat{\mathbf{Y}}_{t,1} = \hat{\boldsymbol{\mu}}_{t,1} + \hat{\boldsymbol{\epsilon}}_{t,1} \tag{16}$$