Check for updates

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

Accepted Article

ARTICLE TYPE

A Stochastic Locally Diffusive Model with Neural Network-Based Deformations for Global Sea Surface Temperature

Wenjing Hu[1] | Geir-Arne Fuglstad[2] | Stefano Castruccio*[1]

[1] Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Indiana, USA

[2] Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence
*Stefano Castruccio, Crowley Hall, University of Notre Dame, Notre Dame, IN, USA. Email: scastruc@nd.edu

Summary

In this work, we propose a new approach to model large, irregularly distributed spatio-temporal global data via a locally diffusive Stochastic Partial Differential Equation (SPDE). The proposed model assumes a local deformation of the SPDE with nonlinear dependence on the covariates through a neural network. The proposed model can be fit in a computationally efficient manner using a triangulation over the sphere and sparsity of the precision matrix, as shown in an application with a large data set of simulated multi-decadal monthly sea surface temperature.

KEYWORDS:
Stochastic Partial Differential Equations, Neural Networks, Sea Surface Temperature, Locally Diffusive Model

1 | INTRODUCTION

Just as the majority of areas in science, engineering and beyond, environmental science has seen an exponential increase in volume, velocity and variety of data over the past decades. Globally monitored data have been particularly impacted by the 'Big Data' revolution, with improvement in satellite technology and Earth system model simulation which now allow multi-decadal simulations such as the Coupled Model Intercomparison Phase 6 (CMIP6, Eyring et al. (2016)) at 50km spatial resolution. Such enhanced data products are critical to improve our understanding of multi-decadal processes such as the El-Niño Southern Oscillation, as well as to assess the local impact of future climate.

The wide availability of global data calls for the development of appropriate statistical models on the spherical domain. The formulation of theoretically valid and practical models on a global scale represents a substantial challenge, as models on the Euclidean space cannot be directly applied to a spherical domain, see Gneiting (2013) for a detailed discussion on this topic. Additionally, for spherical data the simplifying assumption of an isotropic random field is not even approximately true (except for applications with very limited number of locations), as global variables could show change in spatial structure across latitude and longitude, as well as large geographical descriptors such as land and ocean, see Jeong, Jun, and Genton (2017) and Porcu, Alegría, and Furrer (2018) for a comprehensive review of the recent methodological developments.

Several approaches have been proposed in recent years to model non-isotropic global data. Jun and Stein (2007 2008) proposed a partial derivative approach to isotropic global data to achieve longitudinal stationarity (*axial symmetry*, Jones (1963)), and later Jun (2011) extended the approach to the multivariate setting. Castruccio (2016); Castruccio and Stein (2013) proposed a spectral approach to model large global gridded data with an axially symmetric model, and later work has extended this model to longitudinally non-stationary models with evolutionary spectrum approaches (Castruccio & Guinness 2017; Jeong, Castruccio, Crippa, & Genton 2018), three dimensional variables (Castruccio & Genton 2016), nonparametric coherence (Castruccio & Genton 2014) and in the multivariate setting (Edwards, Castruccio, & Hammerling 2019).

More recently, global models based on the use of Stochastic Partial Differential Equations (SPDEs) have been proposed. Such methods revolve around the assumption that the (Gaussian) process we aim to model is a solution of a global diffusive SPDE, which under some conditions can be approximated by a finite volumes solution whose weighting scheme has a spatial dependence informed by a Gaussian Markov Random Field (Lindgren, Rue, & Lindström 2011), thereby allowing fast likelihood evaluation. In particular, Guinness and Hammerling (2018) proposed a stationary

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/sta4.431

This article is protected by copyright. All rights reserved.

Accepted Article

global SPDE in the spectral domain for compressing climate model output, Ingebrigtsen, Lindgren, and Steinsland (2014); Ingebrigtsen, Lindgren, Steinsland, and Martino (2015) considered the effect of covariates, Fuglstad, Simpson, Lindgren, and Rue (2015) propose spline penalties, and Bolin and Lindgren (2011) formulated a model predicated on nested SPDE operators. More recently, Fuglstad and Castruccio (2020) proposed a nonstationary model based on the idea of locally diffusive SPDEs (Fuglstad, Simpson, Lindgren, & Rue 2019) with a changing dependence structure across land and ocean. The use of SPDEs for modeling global data has proven a very flexible tool, as this approach is not bound to any sampling geometry, and modifications of the differential operator naturally induce a controlled level of nonstationary while automatically preserving the theoretical validity of the process. The SPDE approach (at least in its original formulation) does not allow for a natural incorporation of covariates, whose influence in the variable we aim at modeling may be also nonlinear.

This work proposes a generalization of the local deformation approach in Fuglstad and Castruccio (2020) by allowing the covariates to control the degree of deformation of the locally diffusive SPDE. In order to model the dependence, we propose a feedforward, shallow neural network which allows a higher degree of flexibility than a linear model. We present an application to modeling Sea Surface Temperature (SST) as a function of surface wind from climate model simulations.

The work proceeds as follows. Section 2 describes the data set. Section 3 introduces our model and details the inferential approach. Section 4 presents the results. Section 5 concludes with a discussion. The code for this work can be found in the following GitHub repository: github.com/Env-an-Stat-group/21.Hu.Stat.

2 | DATA

We rely on the Community Earth System Model Large Ensemble (LENS, Kay et al. (2015)). A single model simulation from 1920 to 2005 is run under historical greenhouse gases concentrations, and on Jan 1st, 2006 35 small perturbations are applied to the atmospheric component of the model, resulting in a plume of 35 future projections until 2100 under the same scenario (Representative Concentration Pathway 8.5, van Vuuren et al. (2011), assuming a radiative forcing of 8.5Wm^{-2} by 2100) and physics, but differing initial conditions. Throughout this work we only fit the statistical model with R randomly chosen simulations. The ultimate aim is to show that the proposed statistical model, trained with this reduced training set, can reproduce the uncertainty generated from all the 35 runs, hence acting as a stochastic approximation of the climate model which can generate fast surrogates. A small training size R would yield unstable parameter estimates, an overly large R would defeat the purpose of an effective stochastic approximation of the original ensemble. In this work we set $R = 5$, as this was shown to be the smallest training size achieving stable parameter estimates in previous work (Fuglstad & Castruccio 2020; Hu & Castruccio 2021).

In this work, we aim at characterizing the global spatio-temporal structure of the SST using wind as an explanatory variable. Both wind and temperature are considered at monthly resolution between 2006 to 2100, hence for a total of $T = 95$ years and $T \times 12 = 1,140$ months. SST is obtained from the ocean component of the climate model over an irregular grid covering the oceans comprising of a total of $S = 86,212$ spatial locations, see Figure 1a for a map the SST annual anomaly between 2100 and 2006. Surface wind is instead originally resolved in the atmospheric module of the climate model and later interpolated in the same grid as temperature. In total, each variable comprise of $S \times T \times \approx 491$ million data points.

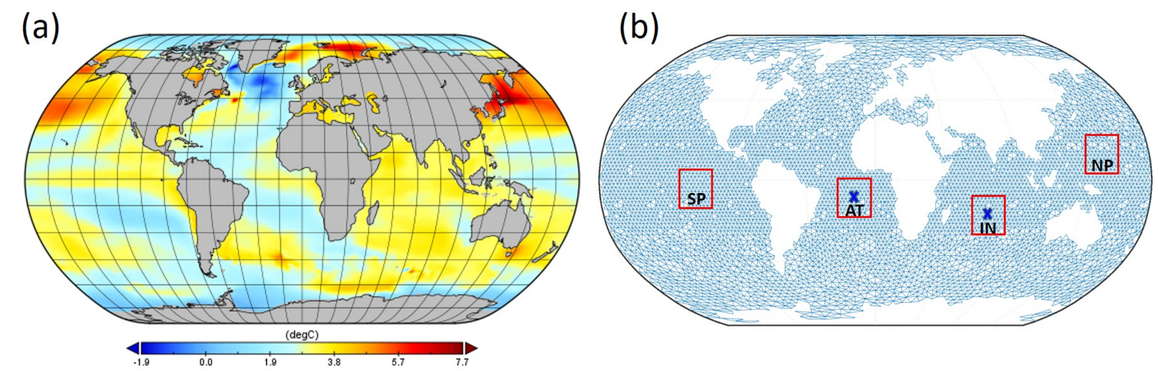

FIGURE 1 (a) Annual 2100-2006 SST anomaly averaged across ensemble members (in degrees Celsius); (b) Triangulation chosen for this work, along with the four regions (with names, SP=South Pacific, AT=Atlantic, IN=Indian Ocean, NP=North Pacific) chosen in the correlation comparison in Figure 3 and the cross-validation study in Table 2. The two black crosses in AT and IN represent the two locations referred to in Figure 2.

This article is protected by copyright. All rights reserved.

Accepted Article

Throughout this work, we denote with $\mathbf{y}_t^{(r)} = (y_t^{(r)}(\mathbf{s}_1), \dots, y_t^{(r)}(\mathbf{s}_N))$ the vector of SSTs at locations $\mathbf{s}_1, \dots, \mathbf{s}_N$ on the sphere for months $t = 1, \dots, 12T$ and realizations $r = 1, \dots, R$. The mean temperature is similarly denoted by $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{12T})$, where $\boldsymbol{\mu}_t = \mathrm{E}\left(\mathbf{y}_t^{(r)}\right)$.

3 | MODEL AND INFERENCE

We assume that the R simulations of the climate model are independent realizations of a Gaussian random field (GRF), which is a combination of a temporally evolving mean $\boldsymbol{\mu}_t$ and zero mean GRF $\boldsymbol{\eta}_t^{(r)}$, i.e., $\mathbf{y}_t^{(r)} = \boldsymbol{\mu}_t + \boldsymbol{\eta}_t^{(r)}$, $t = 1, \dots, 12T$ and $r = 1, \dots, R$. This assumption is ultimately justified by the chaotic nature of the primitive equation of the climate model (Lorenz 1963), which can be assumed to be independent (conditional on the mean) after a few days from the perturbation generating the ensemble.

3.1 | The mean and temporal model

We assume that the mean structure is

$$\mu_t(\mathbf{s}_i) = \sum_{j=0}^{J} t^j \left\{ \alpha_{i,j} + \sum_{h=1}^{H} \delta_{i,j,h} \cos\left(\frac{2\pi h t}{12}\right) + \delta'_{i,j,h} \sin\left(\frac{2\pi h t}{12}\right) \right\}, \tag{1}$$

where $\alpha_{i,j}, \delta_{i,j,h}, \delta'_{i,j,h}$ are parameters controlling the polynomial annual trend and the interannual behavior. This model allows different temporal evolution at each spatial location. The parameter vector for the mean $\boldsymbol{\theta}_{\text{mean}}$ comprises of N vectors of $(J+1)(2H+1)$ site-specific parameters. If the parameter vector for location i is denoted $\boldsymbol{\theta}_{\text{mean},i}$, the parameter vector of all locations is $\boldsymbol{\theta}_{\text{mean}}^\top = (\boldsymbol{\theta}_{\text{mean},i}^\top; i = 1 \dots, N)$, with $\boldsymbol{\theta}_{\text{mean},i}^\top = (\alpha_{i,j}, \delta_{j,i,h}, \delta'_{j,i,h}; j = 0, \dots, J; i = 1, \dots, N; h = 1, \dots, H)$, for a total of $N(J+1)(2H+1)$ elements in $\boldsymbol{\theta}_{\text{mean}}$. From diagnostics (not shown) and previous work (Fuglstad & Castruccio 2020), we choose $J = 2$ and $H = 6$.

The stochastic temporal evolution around the expected value is modeled through a vector autoregressive process of order P (VAR(P)),

$$\boldsymbol{\eta}_t^{(r)} = \sum_{j=1}^{P} \mathbf{A}_j \boldsymbol{\eta}_{t-j}^{(r)} + \mathbf{S}\boldsymbol{\epsilon}_t^{(r)}, \quad t = P+1, \dots, 12T, \quad r = 1, \dots, R. \tag{2}$$

The $N \times N$ matrix $\mathbf{S}$ is diagonal with elements $\sigma_i > 0$, $i = 1, \dots, N$, and for lag j, $\mathbf{A}_j$ is an $N \times N$ diagonal matrix with diagonal elements $a_{j,i}$ for $i = 1, \dots, N$. The innovations $\boldsymbol{\epsilon}_t^{(r)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ are independent and identically distributed both across r and t, where $\mathbf{C} = \mathbf{C}(\boldsymbol{\theta}_{\text{space}})$ is a $N \times N$ correlation matrix describing the spatial dependence structure of the innovations, as detailed in the next section. Let $\boldsymbol{\theta}_{\text{time}}^\top = (\boldsymbol{\theta}_{\text{time},i}^\top; i = 1, \dots, N)$, where location i has $P+1$ site-specific parameters $\boldsymbol{\theta}_{\text{time},i}^\top = (a_{1,i}, \dots, a_{P,i}, \sigma_i)$. There are then $(P+1)N$ elements in $\boldsymbol{\theta}_{\text{time}}$, and a total of $N\{(J+1)(2H+1) + P + 1\}$ parameters are needed to describe the mean and temporal structure.

3.2 | The spatial model

3.2.1 | A locally diffusive SPDE

We now focus on the innovations $\boldsymbol{\epsilon}_t^{(r)}$ in Equation (2), and since they are independent and identically distributed in time and realization, for simplicity of notation we drop subscript and superscript. The spatial structure of $\boldsymbol{\epsilon}$ is assumed to be Gaussian with a $N \times N$ correlation matrix controlled by parameters $\boldsymbol{\theta}_{\text{space}}$ with $N^2 = 122,880^2 \approx 1.5 \cdot 10^{10}$ elements for the SST dataset presented in Section 2. Likelihood evaluation with such a large amount of spatial data would imply storing and operating with a matrix $\mathbf{C}(\boldsymbol{\theta}_{\text{space}})$ of 900 Gigabytes, a task practically impossible for current computers. To circumvent this issue, in this work we rely on the SPDE approach introduced by Lindgren et al. (2011), which is based on the fact that a GRF with (a subclass of) one of the most popular covariance function (Matérn, Stein (1999)) is also a solution of a reaction diffusion SPDE (Whittle 1954). By solving the SPDE with finite volumes (hence reducing the dimensionality of the problem) and by virtue of an ‘explicit link’ between a discretized solution and a Gaussian Markov Random Field (Lindgren et al. 2011), inference can be achieved with sparse linear algebra. In the same work, Lindgren et al. (2011) also proposed to specify GRFs $\boldsymbol{\epsilon}$ with Matérn-like covariance structures on the sphere $\mathbb{S}^2$ by solving a reaction diffusion equation

$$(\kappa^2 - \Delta_{\mathbb{S}^2})\boldsymbol{\epsilon}(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{S}^2, \tag{3}$$

where $\kappa^2 > 0$ is a parameter, $\mathcal{W}$ is a spatial standard Gaussian white noise, and $\Delta_{\mathbb{S}^2}$ is the Laplacian (Laplace-Beltrami operator) on $\mathbb{S}^2$. The resulting covariance is close to the Matérn if the effective spatial range is small compared to the diameter of the sphere, although there is not a closed-form expression.

While solving (3) leads to fast and theoretically justified inference, the underlying model assumes the same local effect of the SPDE operator, and hence limits the analysis to stationary, isotropic models. In order to build more flexible models, the SPDE operator (3) can be used as a building block for more sophisticated approaches. Indeed, Fuglstad et al. (2019) proposed a local change of coordinates via a metric tensors to induce a

This article is protected by copyright. All rights reserved.

Accepted Article

non-stationary model. In this work we provide the basic details, a comprehensive and formal explanation with proofs can be found in Fuglstad and Castruccio (2020). Consider spherical coordinates $\mathbf{s} = (\mathsf{L}, \ell)$, where L is latitude and ℓ is longitude. We assume the length of a line element $\mathrm{d}\mathbf{s} = (\mathrm{dL}, \mathrm{d}\ell)^\top$ is described through $\|\mathrm{d}\mathbf{s}\| = \sqrt{\mathrm{d}\mathbf{s}^\top \mathbf{G}(\mathbf{s})\mathrm{d}\mathbf{s}}$, where $\mathbf{G}(\cdot)$ is a spatially varying positive definite 2×2 matrix. The spatially varying matrix $\mathbf{G}(\cdot)$ is generally called a metric tensor, and under the resulting changes in distance, we can write the SPDE in Equation (3) as

$$\{|\mathbf{G}(\mathbf{s})|^{1/2} - \nabla \cdot |\mathbf{G}(\mathbf{s})|^{1/2}\mathbf{G}(\mathbf{s})^{-1}\nabla\}\boldsymbol{\epsilon}(\mathbf{s}) = |\mathbf{G}(\mathbf{s})|^{1/4}\mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{S}^2, \tag{4}$$

where $\nabla = (\frac{\partial}{\partial \mathsf{L}}, \frac{\partial}{\partial \ell})$. In practice, we solve the SPDE on a triangulation of the sphere in local coordinates and there are no singularities at the poles.

The solution of the SPDE in equation (3) is a non-stationary GRF, whose spatial structure is controlled by the deformation of the differential operator. We use a spatially varying vector field $\mathbf{v}(\cdot) = (\mathsf{v}_1(\cdot), \mathsf{v}_2(\cdot))^\top$ and a positive-valued function $\rho(\cdot)$, and write the inverse metric tensor as

$$\mathbf{G}(\mathbf{s})^{-1} = \rho(\mathbf{s})^2 \frac{\mathbf{I}_2 + \mathbf{v}(\mathbf{s})\mathbf{v}(\mathbf{s})^\top}{\sqrt{1 + ||\mathbf{v}(\mathbf{s})||^2}}, \quad \mathbf{s} \in \mathbb{S}^2. \tag{5}$$

This implies that, at location $\mathbf{s}$, the original infinitesimal distance is multiplied by $1/\rho(\mathbf{s})$ and $\{\mathbf{I}_2 + \mathbf{v}(\mathbf{s})\mathbf{v}(\mathbf{s})^\top\}/\sqrt{1 + \|\mathbf{v}(\mathbf{s})\|^2}$ adds anisotropy. One can show that locally the effective range in the direction of $\mathbf{v}(\mathbf{s})$ is $\rho(\mathbf{s})\{1 + ||\mathbf{v}(\mathbf{s})||^2\}^{1/4}$ and the effective range in the direction orthogonal to $\mathbf{v}(\mathbf{s})$ is $\rho(\mathbf{s})/\{1 + ||\mathbf{v}(\mathbf{s})||^2\}^{1/4}$. This implies that the geometric average of the range in the direction of $\mathbf{v}(\mathbf{s})$ and its orthogonal direction is $\rho(\mathbf{s})$. Therefore, the spatially varying $\rho(\cdot)$ controls the geometric average of the strength of dependence in the two principal directions and $\mathbf{v}(\cdot)$ determines the direction and strength of the anisotropy.

We ensure separation between correlation structure and marginal variance exactly at each location $\mathbf{s}$ by considering the standardized process $\tilde{\boldsymbol{\epsilon}}(\mathbf{s}) = \boldsymbol{\epsilon}(\mathbf{s})/\sigma_{\mathsf{scale}}(\mathbf{s})$, where $\sigma_{\mathsf{scale}}(\mathbf{s})$ is calculated as the marginal standard deviations resulting from the choice of metric tensor.

3.2.2 | Parametrization of the metric tensor

We aim to describe the spatially varying metric tensor $\mathbf{G}(\cdot)$ through a combination of general functions and spatial covariates. For a generic sea location $\mathbf{s}$, we have

$$\log\{\rho(\mathbf{s})\} = \sum_{\mathsf{l}=0}^{\mathcal{L}} \sum_{\mathsf{m}=-\mathsf{l}}^{\mathsf{l}} \alpha_{\mathsf{ml}} \mathsf{Y}_{\mathsf{l}}^{\mathsf{m}}(\mathbf{s}), \tag{6a}$$

$$\mathbf{v}(\mathbf{s}) = \mathsf{f}(\mathbf{x}(\mathbf{s}'), \mathbf{s}' \in \partial\mathbf{s}) + \sum_{\mathsf{l}=1}^{\mathcal{L}} \sum_{\mathsf{m}=-\mathsf{l}}^{\mathsf{l}} \left\{\mathsf{E}_{\mathsf{lm}}^{(1)} \nabla \mathsf{Y}_{\mathsf{m}}^{\mathsf{l}}(\mathbf{s}) + \mathsf{E}_{\mathsf{lm}}^{(2)} \hat{\mathbf{r}}(\mathbf{s}) \times \nabla \mathsf{Y}_{\mathsf{m}}^{\mathsf{l}}(\mathbf{s})\right\}, \tag{6b}$$

where $\mathcal{L}$ is a non-negative integers, α_{ml} are real-valued coefficients and $\mathsf{Y}_{\mathsf{l}}^{\mathsf{m}}(\cdot)$ are Laplace's spherical harmonic of degree l and order m. Additionally, $\hat{\mathbf{r}}$ is the unit vector in the positive radial direction, and $\mathsf{E}_{\mathsf{lm}}^{(1)}$ and $\mathsf{E}_{\mathsf{lm}}^{(2)}$ are real coefficients. The covariates $\mathbf{x}(\cdot)$ are assumed to be spatially varying and it is assumed that their contribution to the vector field $\mathbf{v}(\mathbf{s})$ is modulated by a function f, which could also include covariates in the neighbor $\partial\mathbf{s}$ (by convention, in this set we also include the location $\mathbf{s}$ itself). Compared to a model with no covariates (Fuglstad & Castruccio 2020), the term f allows incorporate external information about the local variation not captured by the large scale general basis functions. Since the model needs to be defined for every point on the sphere $\mathbb{S}^2$, for a location $\mathbf{s}$ on land we use $\log(\rho(\mathbf{s})) = \alpha_0$ and $\mathbf{v}(\mathbf{s}) = \mathbf{0}$.

We consider two choices of f:

1. linear covariates at the same site $\mathbf{s}$. We assume a linear function with respect to a k-dimensional vector $\boldsymbol{\beta}_{\mathsf{LIN}}$

$$f(\mathbf{x}(\mathbf{s}'), \mathbf{s}' \in \partial\mathbf{s}) = f(\mathbf{x}(\mathbf{s})) = \sum_{i=1}^{k} \beta_{i;\mathsf{LIN}} \mathbf{x}^{(i)}(\mathbf{s}),$$

   where $\{\mathbf{x}^{(\mathsf{i})}\}_{\mathsf{i}=1}^{\mathsf{k}}$ is the collection of the covariates (in our application $\mathsf{k} = 2$ are the horizontal and vertical components of the surface wind field).

2. non-linear contributions of the covariates in the neighbor $\partial\mathbf{s}$ through a neural network. If we denote by $\mathbf{X}(\mathbf{s}) = \{\mathbf{x}(\mathbf{s}'), \mathbf{s}' \in \partial\mathbf{s}\}$ the $|\partial\mathbf{s}|\mathsf{k}$-dimensional vector comprising of the covariates observed in the neighborhood $\partial\mathbf{s}$, the function f is assumed to be a one layer (shallow) neural network

$$f(\mathbf{x}(\mathbf{s}'), \mathbf{s}' \in \partial\mathbf{s}) = g(\mathbf{X}(\mathbf{s})^\top \mathbf{W})\boldsymbol{\beta}_{\mathsf{NN}},$$

   where g is the activation function, a nonlinear function which could have different shapes, and in this work is a hypertangent (Goodfellow, Bengio, & Courville 2016). $\mathbf{W}$ is a $\mathsf{n}_\mathsf{h} \times |\partial|\mathsf{k}$ matrix of unknown entries, and $\boldsymbol{\beta}_{\mathsf{NN}}$ is a n_h-dimensional vector, where $|\partial|$ is the cardinality of $\partial\mathbf{s}$ (assumed constant in space). The choice is flexible and allows to capture non-linearity, but is also considerably more computationally expensive.

This article is protected by copyright. All rights reserved.

Accepted Article

TABLE 1 Models considered in this work. The second and third column refer to the parameters in equations (6a) and (6b).

Model	$\mathcal{L}$, L	$f(\mathbf{x}(\cdot))$
STAT	(0,0)	No
NSTAT-LIN	(0,0)	Linear
NSTAT-H	(4,4)	No
NSTAT-H-LIN	(4,4)	Linear
NSTAT-H-NN	(4,4)	Neural Network

We compare a total of five models: 1) a stationary model without covariates (STAT); 2) a non-stationary model with linear covariates (NSTAT-LIN); 3) a non-stationary model with harmonics (NSTAT-H); 4) a non-stationary model with harmonics and linear covariates (NSTAT-H-LIN); 5) A non-stationary model with harmonics and covariates modeled through a neural network (NSTAT-H-NN). These models are summarized in Table 1.

3.2.3 | Micro-scale variability and model summary

The spatial field will be assumed to be piece-wise constant on a triangulation of the sphere shown in Figure 1b. Since the triangulation is of coarser resolution than the distance among data locations, small-scale variation cannot be captured, and we include a nugget effect to absorb the data variability within each triangle when estimating the spatial structure. This gives one extra parameter, τ^2 denoting the nugget variance. Let n_{Cov} denote the number of parameters needed to describe the contribution of the covariates, which depends on the choice of linear ($n_{\mathrm{Cov}} = k$) or neural network ($n_{\mathrm{Cov}} = (|\partial|k + 1)n_h$) functional form for f in (6b).

The covariance matrix of $\boldsymbol{\epsilon}$ is therefore described through the parameters in the $(3(\mathcal{L}+1)^2 + 2 + n_{\mathrm{Cov}})$-dimensional vector,

$$\boldsymbol{\theta}_{\text{space}} = \left(\{\alpha_0\}, \{\alpha_{ml}, E^{(1)}_{lm}, E^{(2)}_{lm} \mid l = 0, \dots, \mathcal{L}, m = -l, \dots, l\}\}, \{\tau^2\}, \{\boldsymbol{\beta}_{\text{LIN}}\}\right),$$

in the linear case, and

$$\boldsymbol{\theta}_{\text{space}} = \left(\{\alpha_0\}, \{\alpha_{ml}, E^{(1)}_{lm}, E^{(2)}_{lm} \mid l = 0, \dots, \mathcal{L}, m = -l, \dots, l\}\}, \{\tau^2\}, \{\mathbf{W}\}, \{\boldsymbol{\beta}_{\text{NN}}\}\right),$$

in the neural network case, where $E^{(1,j)}_{0,m} = E^{(2,j)}_{0,m} = 0$ for $m = 0$, but are included for convenience of notation.

3.3 | Inference

Inference is performed sequentially and conditionally in three steps, in which we estimate 1) $\boldsymbol{\theta}_{\text{time}}$, 2) $\boldsymbol{\theta}_{\text{mean}}$ conditionally on $\hat{\boldsymbol{\theta}}_{\text{time}}$, and 3) $\boldsymbol{\theta}_{\text{space}}$ conditionally on $\hat{\boldsymbol{\theta}}_{\text{mean}}, \hat{\boldsymbol{\theta}}_{\text{time}}$. The stepwise approach is asympotically consistent (Edwards et al. 2019), and uncertainty and bias propagation have been shown not to play a significant role in the final estimates for similar models (Castruccio & Guinness 2017). Since there is no interaction between locations in either the mean or the temporal structure, as apparent from equations (1) and (2), the estimation of these two parts can be performed independently by multiple cores of a computer. We describe the two steps for a generic location i with spatial coordinate $\mathbf{s}_i$, and use the notation $\bar{y}_t(\mathbf{s}_i) = \frac{1}{R}\sum_{r=1}^{R} y_t^{(r)}(\mathbf{s}_i)$, $t = 1, \dots, 12T$.

Step 1: time

We use $P = 6$ in equation (2), and for each location $\mathbf{s}_i$ we independently use a restricted log-likelihood (technical details and proofs can be found in Castruccio and Stein (2013)). Let $\mathbf{Q}(\boldsymbol{\theta}_{\text{time},i})$ be the precision matrix arising from the autoregressive process of order P, then the restricted log-likelihood is

$$\ell(\boldsymbol{\theta}_{\text{time,i}} \mid \mathbf{D}^{(1)}(\mathbf{s}_i), \dots, \mathbf{D}^{(R)}(\mathbf{s}_i)) = \text{const} + (R-1)\log|\mathbf{Q}(\boldsymbol{\theta}_{\text{time},i})| - \sum_{r=1}^{R} \mathbf{D}^{(r)}(\mathbf{s}_i)^\top \mathbf{Q}(\boldsymbol{\theta}_{\text{time},i})\mathbf{D}^{(r)}(\mathbf{s}_i), \tag{7}$$

where $\mathbf{D}^{(r)}(\mathbf{s}_i) = (y_1^{(r)}(\mathbf{s}_i) - \bar{y}_1(\mathbf{s}_i), \dots, y_{12T}^{(r)}(\mathbf{s}_i) - \bar{y}_{12T}(\mathbf{s}_i))$, and const is a constant. We maximize Equation (7) to find estimates $\hat{a}_{j,i}$, for $j = 1, \dots, P$, and $\hat{\sigma}_i$ in equation (2).

Step 2: mean

For each location $\mathbf{s}_i$, we assume a mean with month-specific polynomials as described in equation (1). Let $\mathbf{X}$ be the design matrix generated by the basis functions. We perform inference conditionally to $\hat{\boldsymbol{\theta}}_{\text{time,i}}$, so that the precision matrix for the time series $\hat{\mathbf{Q}} = \mathbf{Q}(\hat{\boldsymbol{\theta}}_{\text{time,i}})$ is assumed to be

This article is protected by copyright. All rights reserved.

Accepted Article

fixed, and the mean is obtained via generalized least squares

$$\hat{\boldsymbol{\theta}}_{\text{mean,i}} = (\mathbf{X}^\top \hat{\mathbf{Q}} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{Q}} \bar{\mathbf{y}}(\mathbf{s}_i).$$

The vector of estimated means for all time points is then given by $\hat{\boldsymbol{\mu}}_{\text{i}} = \mathbf{X}\hat{\boldsymbol{\theta}}_{\text{mean,i}}$. Henceforth, we denote with $\hat{\boldsymbol{\mu}}_{\text{t,i}}$ the t-th element of $\mathbf{X}\hat{\boldsymbol{\theta}}_{\text{mean,i}}$, and $\hat{\boldsymbol{\mu}}_{\text{t}} = (\hat{\boldsymbol{\mu}}_{\text{t},1}, \ldots, \hat{\boldsymbol{\mu}}_{\text{t,N}})$. In Figure 2 we show the fitted mean along with the R SST time series for two locations as indicated in Figure 1b, one in the Atlantic and one in the Pacific Ocean, between 2030 to 2036. It is readily apparent how the fitted mean is able to capture the interannual variability of SST for these two points.

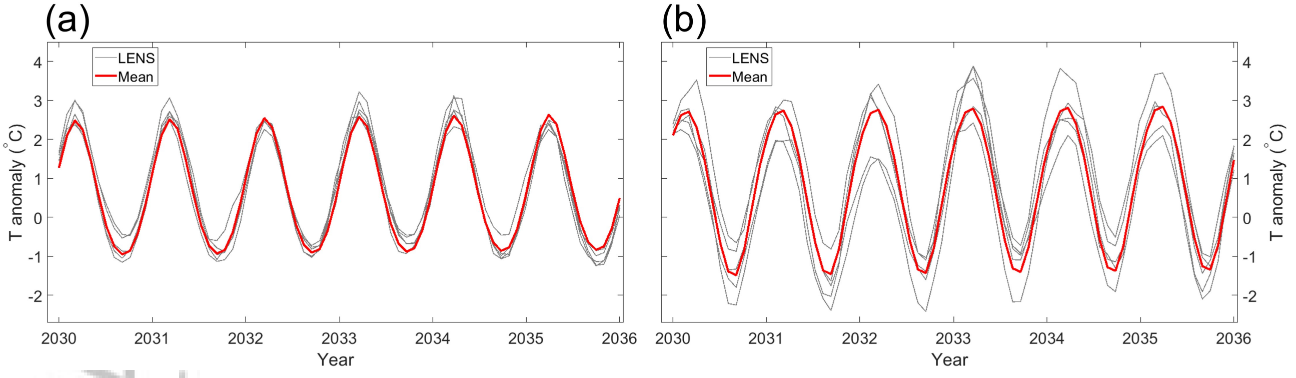

FIGURE 2 LENS simulations (in grey) and estimated mean according to the functional form in equation (1) (in red). The two locations are indicated by the blue crosses in Figure 1b, namely (a) Atlantic and (b) Indian Ocean.

Step 3: space

Conditionally on $\hat{\boldsymbol{\theta}}_{\text{mean}}, \hat{\boldsymbol{\theta}}_{\text{time}}$, we can estimate the innovations by inverting equation (2). If we denote by $\hat{\boldsymbol{\eta}}_{\text{t}}^{(\text{r})} = \mathbf{y}_{\text{t}}^{(\text{r})} - \hat{\boldsymbol{\mu}}_{\text{t}}$, we have:

$$\hat{\boldsymbol{\varepsilon}}_t^{(r)} = \hat{\mathbf{S}}^{-1/2} \left\{ \hat{\boldsymbol{\eta}}_t^{(r)} - \sum_{j=1}^{P} \hat{\mathbf{A}}_j \hat{\boldsymbol{\eta}}_{t-j}^{(r)} \right\}, \quad t = P+1, \ldots, 12T. \tag{8}$$

We estimate $\boldsymbol{\theta}_{\text{space}}$ through maximum likelihood based on the independent realizations of the innovations $\hat{\boldsymbol{\varepsilon}}_1^{(1)}, \ldots, \hat{\boldsymbol{\varepsilon}}_{\text{12T}}^{(\text{R})} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta}_{\text{space}}))$. Inference can be performed efficiently by solving the locally diffusive SPDE (4) with a finite volume approach, assuming constant values on each triangle (Fuglstad & Castruccio 2020). The triangulation on the sphere used in this work, comprising of 15,392 triangles, is shown in Figure 1b. It can be shown that the corresponding precision matrix is sparse, similarly to the stationary SPDE case (Lindgren et al. 2011). We do not repeat the proofs here, which are identical to the model with no covariates and can be found in Fuglstad and Castruccio (2020).

4 | RESULTS

In this section, we evaluate the performance of the proposed models against a stationary SPDE model. Throughout this section, the covariate field $\mathbf{x}(\mathbf{s})$ in equation (6b) is assumed to be the two dimensional wind field as introduced in Section 2.

Figure 3 shows the contour plot of the correlation function for the decorrelated residuals in time from (8), for the four reference regions highlighted in Figure 1b. Panel (a) of Figure 3 shows the empirical correlation across all times and realizations of the ensemble, which is regarded as the "true" correlation. Panels (b) and (c) show the correlation implied by the stationary (STAT) and non-stationary model with covariates expressed though neural networks (NSTAT-H-NN), respectively. It can be easily seen that STAT fails to capture the directional effects due to Ocean circulation in IN, AT and especially SP, as this model by construction assumes a unchanged dependence structure in space. The model NSTAT-H-NN, instead, is flexible enough to able to capture these patterns with harmonic basis functions and especially wind fields.

A more formal comparison among the models introduced in Table 1 is performed via cross-validation. Data from the four regions in Figure 1b are removed, and the models' different predictive ability in terms of both Root Mean Squared Error (RMSE) and Continuous Ranked Probability Score (CRPS) is shown in Table 2. Unsurprisingly, the stationary model has considerably worse predictive performance (both in terms of RMSE and CRPS) than any other model, given its inability to articulate a spatially varying dependence structure. Once non-stationarity is assumed, either through

This article is protected by copyright. All rights reserved.

Accepted Article

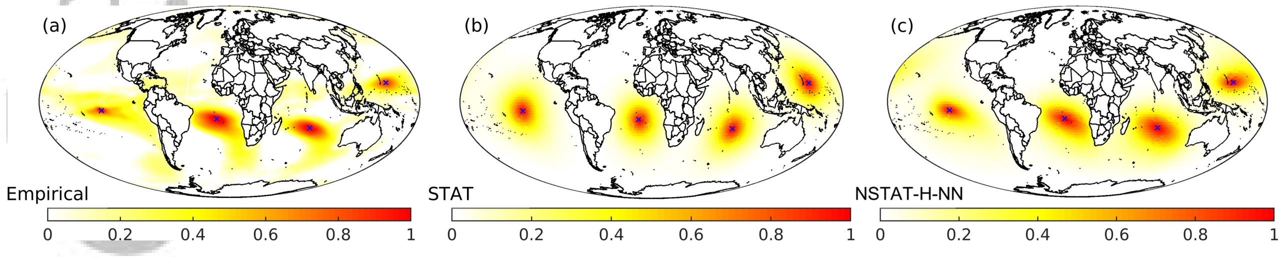

FIGURE 3 Correlation between reference points (marked with blue crosses, center of the squares in Figure 1) and all nearby locations. The empirical estimates from LENS (panel (a)) are compared with the models (b) STAT and (c) NSTAT-H-NN as described in Table 1. All locations are shown in the same plot and negative values have been set to 0 for the ease of visualization.

TABLE 2 RMSE (CRSP) for hold-out an entire area The unit is 10^{-1}. Areas are shown in Figure 1b.

Model	AT	IN	SP	NP
STAT	4.24 (2.37)	4.63 (2.58)	5.46 (3.04)	4.76 (2.68)
NSTAT-LIN	3.45 (1.94)	3.93 (2.18)	**4.83** (2.66)	4.11 (2.30)
NSTAT-H	3.46 (1.92)	3.81 (2.11)	4.87 (2.66)	4.12 (2.31)
NSTAT-H-LIN	3.39 (1.89)	3.79 **(2.10)**	4.86 (2.66)	**3.95 (2.22)**
NSTAT-H-NN	**3.34 (1.88)**	**3.78 (2.10)**	4.85 **(2.65)**	**3.95 (2.22)**

harmonics or with the wind vector as covariate, the results are substantially improved. For three regions (AT, IN and NP) out of four, the model with neural network dependence shows improved predictive ability against all other models, hence underlying the added value in considering a nonlinear of SST with the neighboring wind field. In SP, the neural network model is the best in terms of CRPS, and very close to the best result for RMSE.

Once properly validated against the other statistical models, surrogate simulations from NSTAT-H-NN (which we recall has been trained with only five LENS members, see Section 2) are compared with simulations from the entire LENS. The goal is to assess whether the statistical model is able to produce surrogate SST simulations that resemble the LENS, and hence is able to approximate the LENS uncertainty using only a fraction of the simulations as training set. If that is the case, then the statistical model can be regarded as a good stochastic approximation (a *stochastic generator*, SG (Castruccio, Hu, Sanderson, Karspeck, & Hammerling 2019; Hu & Castruccio 2021)) of the LENS, which would only require a small number of simulations and hence would save computational time and storage space.

Figure 4(a-b) compared one SST simulation from the SG and one from a LENS simulation (not in the training set) for July 2030. It is readily apparent how the two maps are similar, with physical patterns such as colder temperatures near the poles, and warmer temperatures at low to mid latitudes, and sensibly higher temperatures in the northern hemisphere as expected by the Boreal summer. While this comparison confirms the SG's ability to provide physically consistent realizations, a comparison of individual simulations cannot asses the SG's ability to approximate the LENS uncertainty. To this end, Figure 4(c-d) compares the uncertainty in the monthly temperature anomaly from 2006-2030 from all thirty-five LENS simulations against thirty-five SG simulations for the Atlantic and Indian locations, as indicated in Figure 1b. Across both locations, the SG is able to capture not just the expected interannual variability, with warmer temperatures in the Boreal summer, but also the relatively small uncertainty around the point estimates, uniformly across months.

5 | CONCLUSION

In this work we have introduced a locally diffusive SPDE model for global data, with covariates controlling the level of deformation of the differential operator via neural networks, and we have shown how this model can be solved with finite volumes by providing the weighting scheme of the basis function with a GMRF which implies sparsity in the precision matrix, thereby dramatically reducing the computation burden and allowing inference for data sets of hundred of millions of data points.

This article is protected by copyright. All rights reserved.

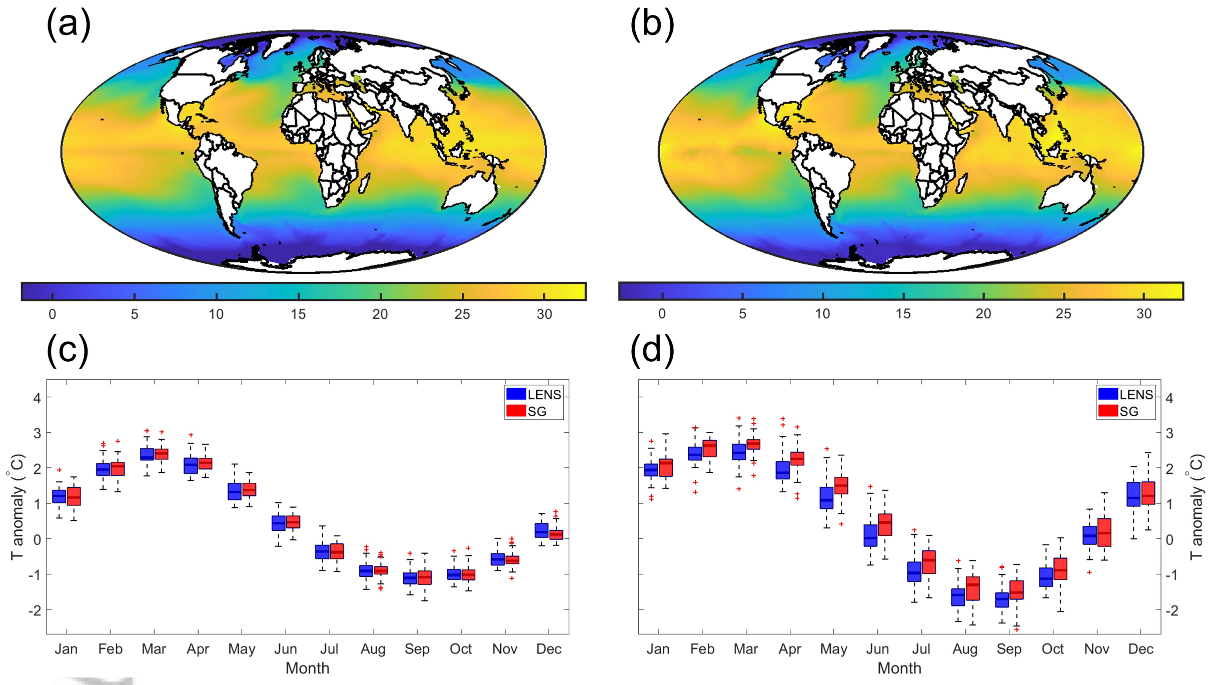

FIGURE 4 (a-b) Comparison of SST maps in in July 2030 from (a) one LENS member (not in the training set) (a) and one statistical model (SG) simulation. (c-d) Comparison of the internal variability for the monthly temperature anomaly from 2006-2030 for all thirty-five LENS and thirty five SG simulations, expressed as blue and red boxplots, respectively. Two locations are chosen: AT and IN in panels c) and d), respectively, as indicated by the blue crosses in Figure 1b.

While showing improvement against previous models, the flexibility in the use of neural networks necessarily implies an increase in the number of parameters and hence nontrivial computations. If even more flexible models such as deep neural networks are sought, the use of efficient iterative methods for gradient calculation (*backpropagation*, Goodfellow et al. (2016)) would need to be embedded in our maximum likelihood algorithm. Alternatively, stochastic approximation of neural networks via random weight matrices could be considered to reduce the parameter space (Bonas & Castruccio 2021; Huang, Castruccio, & Genton 2021). More recent and sophisticated neural networks are also possible to use, such as spiking neural networks, a biologically-inspired approach which would translate the covariates into spike trains and provide a network for them (Maass 1997). In terms of the SPDE, possible extensions range from the use of non-Gaussian models for high-resolution global variables, possibly (but not necessarily) within the latent Gaussian model framework as well as multivariate SPDE to model data and covariates jointly instead of conditionally such as in this work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in as part of the Large Ensemble project at the National Center for Atmospheric Research at www.earthsystemgrid.org.

ACKNOWLEDGEMENTS

This work is partially supported by NSF grant DMS-2014166.

References

Bolin, D., & Lindgren, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone

Accepted

This article is protected by copyright. All rights reserved.

mapping. *Annals of Applied Statistics*, *5*(1), 523–550.

Bonas, M., & Castruccio, S. (2021). *Calibration of spatial forecasts from citizen science urban air pollution data with sparse recurrent neural networks.* arXiv:2105.02971.

Castruccio, S. (2016). Assessing the spatio-temporal structure of annual and seasonal surface temperature for cmip5 and reanalysis. *Spatial Statistics*, *18*, 179-193.

Castruccio, S., & Genton, M. G. (2014). Beyond axial symmetry: An improved class of models for global data. *Stat*, *3*(1), 48-55.

Castruccio, S., & Genton, M. G. (2016). Compressing an Ensemble with Statistical Models: An Algorithm for Global 3D Spatio-Temporal Temperature. *Technometrics*, *58*(3), 319-328.

Castruccio, S., & Guinness, J. (2017). An evolutionary spectrum approach to incorporate large-scale geographical descriptors on global processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *66*(2), 329-344.

Castruccio, S., Hu, Z., Sanderson, B., Karspeck, A., & Hammerling, D. (2019). Reproducing internal variability with few ensemble runs. *Journal of Climate*, *32*(24).

Castruccio, S., & Stein, M. L. (2013). Global space–time models for climate ensembles. *Annals of Applied Statistics*, *7*(3), 1593–1611.

Edwards, M., Castruccio, S., & Hammerling, D. (2019). A multivariate global spatiotemporal stochastic generator for climate ensembles. *Journal of Agricultural, Biological and Environmental Sciences*, *24*, 464–483.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958.

Fuglstad, G.-A., & Castruccio, S. (2020). Compression of climate simulations with a nonstationary global SpatioTemporal SPDE model. *The Annals of Applied Statistics*, *14*(2), 542 – 559.

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2015). Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics*, *14*, 505–531.

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2019). Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, *114*, 445–452.

Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, *19*(4), 1327–1349.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* [Book]. MIT Press.

Guinness, J., & Hammerling, D. (2018). Compression and conditional emulation of climate model output. *Journal of the American Statistical Association*, *113*(521), 56-67.

Hu, W., & Castruccio, S. (2021). Approximating the internal variability of bias-corrected global temperature projections with spatial stochastic generators. *Journal of Climate*. in press.

Huang, H., Castruccio, S., & Genton, M. G. (2021). *Forecasting high-frequency spatio-temporal wind power with dimensionally reduced echo state networks.* arXiv:2102.01141.

Ingebrigtsen, R., Lindgren, F., & Steinsland, I. (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, *8*, 20–38.

Ingebrigtsen, R., Lindgren, F., Steinsland, I., & Martino, S. (2015). Estimation of a non-stationary model for annual precipitation in southern norway using replicates of the spatial field. *Spatial Statistics*, *14*, *Part C*, 338–364.

Jeong, J., Castruccio, S., Crippa, P., & Genton, M. G. (2018). Reducing Storage of Global Wind Ensembles with Stochastic Generators. *Annals of Applied Statistics*, *12*(1), 490-509.

Jeong, J., Jun, M., & Genton, M. G. (2017). Spherical Process Models for Global Spatial Statistics. *Statistical Science*, *32*(4), 501 – 513.

Jones, R. (1963). Stochastic processes on a sphere. annals of mathematical statistics. *Annals of Mathematical Statistics*, *34*, 213–218.

Jun, M. (2011). Non-stationary cross-covariance models for multivariate processes on a globe. *Scandinavian Journal of Statistics*, *38*(4), 726–747.

Jun, M., & Stein, M. L. (2007). An approach to producing space–time covariance functions on spheres. *Technometrics*, *49*(4), 468-479.

Jun, M., & Stein, M. L. (2008). Nonstationary covariance models for global data. *The Annals of Applied Statistics*, *2*(4), 1271 – 1289.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., … Vertenstein, M. (2015). The community earth system model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333 - 1349.

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(4), 423–498.

Lorenz, E. (1963). Deterministic non-periodic flow. *Geoscientific Model Development*, *20*(2), 130-141.

Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, *10*(9), 1659-1671.

Porcu, E., Alegría, A., & Furrer, R. (2018). Modeling temporally evolving and spatially globally dependent data. *International Statistical Review*, *86*(2), 344-377.

Accepted Article

This article is protected by copyright. All rights reserved.

Stein, M. (1999). *Statistics for spatial data: Some theory for kriging*. New York.

van Vuuren, D. P., Jae Edmonds, M. K., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., … Rose, S. K. (2011). The representative concentration pathways: an overview. *Climatic Change*, *109*(5).

Whittle, P. (1954). On Stationary Processes in the Plane. *Biometrika*, *41*(3-4), 434-449.

□

Accepted Article

This article is protected by copyright. All rights reserved.