COMPRESSION OF CLIMATE SIMULATIONS WITH A NONSTATIONARY GLOBAL SPATIO-TEMPORAL SPDE MODEL

By Geir-Arne Fuglstad[*] and Stefano Castruccio[†]

Norwegian University of Science and Technology[*] and University of Notre Dame[†]

Abstract Modern climate models pose an ever-increasing storage burden to computational facilities, and the upcoming generation of global simulations from the next Intergovernmental Panel on Climate Change will require a substantial share of the budget of research centers worldwide to be allocated just for this task. A statistical model can be used as a means to mitigate the storage burden by providing a stochastic approximation of the climate simulations. Indeed, if a suitably validated statistical model can be formulated to draw realizations whose spatio-temporal structure is similar to that of the original computer simulations, then the estimated parameters are effectively all the information that needs to be stored. In this work, we propose a new statistical model defined via a stochastic partial differential equation (SPDE) on the sphere and in evolving time. The model is able to capture nonstationarities across latitudes, longitudes and land/ocean domains for more than 300 million data points, while also overcoming the fundamental limitations of current global statistical models available for compression. Once the model is trained, surrogate runs can be instantaneously generated on a laptop by storing just 20 Megabytes of parameters as opposed to more than 6 Gigabytes of the original ensemble.

1. Introduction. Across the past decades, many fields of environmental and geophysical science have been increasingly relying on computer models which simulate the Earth system or parts thereof. These climate models allow a wide range of scientific investigations in Earth sciences: they detect and attribute climate change, spatially resolve information in poorly monitored areas, and provide physically consistent projections of future climate.

Advances in the understanding of the dynamics of the Earth system, along with the exponential increase in availability of computational power, have prompted the development of increasingly more realistic models, with integration of more physical processes, as well as higher spatial and temporal resolution. While these models are increasingly faithful, they are always an approximation of the Earth system which depends on multiple inputs, and every study relying on them is sensitive with respect to their choice. In order to achieve robust scientific conclusions, a collection (*ensemble*) of simulations is therefore necessary. While generating an ensemble is a computationally demanding task, the share of cost for storage has gradually become the most pressing issue for computing centers. The 2012 climate projections of the Intergovernmental

1

Panel on Climate Change (IPCC) Assessment Report 5 relied on an ensemble, the Coupled Model Intercomparison Project Phase 5 (CMIP5), whose size was in the range of 1-3 Petabytes (Taylor, Stouffer and Meehl, 2012). Nowadays, modern global models can generate as much as one Terabyte of data per computing day (Small et al., 2014), and the data volume for the upcoming Coupled Model Intercomparison Project Phase 6 (CMIP6, Meehl et al. (2014)), is expected to exceed 10 Petabytes (Eyring et al., 2016).

Compression for climate models has been investigated in recent work, and studies have been performed to assess the impact of traditional compression algorithms to model output (Baker et al., 2014). Lossy methods with compression rates of 5:1 to 20:1 (depending on the variable one wishes to compress) produce good results for one particular climate model (Baker et al., 2016), and do not impact the overall scientific conclusions as the compression effects are mitigated in the post-processing steps following the data generation. Baker et al. (2017) performed a comprehensive study of different compression methods for individual variables and argued that no single algorithm is ideal for all variables at all resolutions. The aforementioned works have acknowledged the value of compression for climate model output, as well as the need of higher compression rates to accommodate the size of current and future ensembles.

Only very recently have statistical models been formulated with the goal of compression. A *conditional approach* has been proposed, predicated on *storing what must be stored, and model the rest conditionally.* Guinness and Hammerling (2018) proposed a half spectral model (Stein, 2005) to store a spatial subset of low frequencies in time and their spatial dependence, and decompress conditionally on this information. The approach allows compression of daily temperatures with a rate comparable to traditional methods. To achieve higher compression rates, an *unconditional approach* has been proposed based on the assumption that no information other than the statistical parameters is to be stored, so that the goal is not to compress and decompress an individual simulation, but rather to characterize its distribution (Castruccio and Genton, 2016). While the application of a statistical model for compression is a new concept, the idea of using environmental data to estimate a sampling distribution and then produce more surrogate samples, in other words the use of *stochastic weather generators* (SWGs) for observational data or *emulators* for interpolation in some climate model input space, has a long history. The only substantial difference between classical approaches and the presented method is the emphasis on storage rather than availability of more realizations for SWGs, or computation bottleneck for emulators.

Statistical models for unconditional compression, or *stochastic generators* (Jeong et al. (2018, 2019), SGs henceforth) have been developed for atmospheric variables of global climate models (e.g. temperature, wind and precipitation) by proposing discrete spectral models across longitude and an autoregressive model across latitude for frequencies. By virtue of the spectral representation of the likelihood, sparsity is achieved for the spectral matrix, and inference can be performed for billions of data points.

While SGs have the potential to aid geoscientists to preserve space for future generations of climate models (Castruccio et al., 2019), their substantial reliance on discrete

spectral methods implies several fundamental limitations that hamper their usability to practitioners.

1. SGs currently rely on spectral methods only defined on a discrete latitude/longitude grid, with no continuous underlying process. Therefore, interpolation and simulation cannot be performed at a spatial scale smaller than the native grid of the climate model. In geoscience, local impact assessment is performed by linking in-situ monitoring data with coarse (global or regional) simulations interpolated at the same point (*statistical downscaling*), and without a continuous process, this can only be performed with *ad hoc* methods such as nearest neighbor attribution or inverse distance weighting.

2. The restriction of spectral models on a regular latitude × longitude grid limits its applicability to global variables in the atmospheric domain of a climate model. While adaptations of spectral methods to irregular grids have been proposed (Fuentes, 2007), they imply a loss in computational efficiency which cannot be acceptable in this setting given the considerable data size. Therefore, presently available SGs cannot be applied to variables solved on non-gridded domains such as sea surface temperature, land use or ice cover.

3. A Fourier transform across longitudes implies that spatial frequencies at different latitudes correspond to different physical distances, hence actual correlation as implied by the spectral model is forced to change across latitude. This is particularly problematic for data near the poles, where the physical distance is small and leads to instabilities (Castruccio and Genton, 2016). Further, every model expressed in latitude-longitude coordinate must provide more-or-less restrictive regularity conditions at the poles (Jun and Stein, 2007), and the spectral transformation across longitude allows for a wide range of flexible SGs which naturally account for non-stationarities occurring at a given latitude, but not across latitudes.

It is, in principle, possible to formulate a continuous version of the discrete spectral approach by extending the spectral density to higher frequencies and un-aliasing the frequencies defined at the observed wavenumbers. However, in order to allow a practically useful and computationally convenient solution for high compression rates, this work proposes a fundamental change in perspective by defining a broad and general class of SGs based on a stochastic partial differential equation (SPDE) approach (Lindgren, Rue and Lindström, 2011). The SPDE approach is a computationally affordable, scalable and very popular solution to approximate Gaussian processes via a discrete model on a triangulation over the domain. In this setting, the global process is described as a solution of an SPDE equation on the sphere with white noise forcing. Computational scalability is ensured by the use of finite element or finite volume methods for the discretization of the SPDE solution. The dependence only on neighboring triangles in the discretization implies sparsity of the precision matrix, and hence a Gaussian Markov random field (Rue and Held, 2005). A stationary version of the SPDE on the sphere has been used in Guinness and Hammerling (2018) for temporal frequencies in conditional compression, and several methods have been recently proposed to achieve

nonstationarity based on this approach. Most noticeably, stationarity has been relaxed by adding covariates (Ingebrigtsen, Lindgren and Steinsland, 2014; Ingebrigtsen et al., 2015), using spline penalties (Fuglstad et al., 2015a,b) or modifying the operator itself (Bolin and Lindgren, 2011). This work proposes a nonstationary covariance structure that is described through a spatially varying metric tensor (Fuglstad et al., 2019) over the sphere. This allows the dependence structure across land and ocean to vary in a manner that is reminiscent of the deformation method (Sampson and Guttorp, 1992) the Euclidean space.

In this work, we focus on a large ensemble of monthly global surface temperature, comprising a total of more than six Gigabytes. Monthly temperature is a variable with strong nonstationarities generated by the energy imbalance across latitude, land/ocean dynamics as well as local effects such as trade winds. Our work shows how the SPDE approach is able to capture all these spatial features in just 20 Megabytes of space, and can generate new surrogate simulates almost instantaneously on a common laptop. Further, the availability of an underlying continuous process allows a comparison between the global simulations with in-situ measurements: the model simulations for two locations in the United States will be compared with ground observations.

The work proceeds as follows. In Section 2 the data set is presented. The spatio-temporal model is introduced in Section 3, and inference is discussed in Section 4. In Section 5 we perform a comprehensive diagnostics of the new model assumptions, and compare it against well-established alternatives. Section 6 presents the application of the method to the data set, and discusses results and their implications. Section 7 concludes with a discussion.

2. Data. Throughout this work, we rely on the Large ENSemble experiment (LENS, Kay et al. (2015)), a collection of forty runs from the Community Earth System Model (CESM) version 1 developed at the National Center for Atmospheric Research. We use all runs for cross-validation, but our model is envisioned as a tool for compressing ensembles such as the CMIP5, which comprises only three to six runs per model (with fixed scenario). Therefore, we consider only a random sample of R runs, and provide a comprehensive discussion about the choice of $R = 5$ in the diagnostic in Section 5. We focus on monthly surface temperature (at a reference height of 2 meters above the ground level). Aggregation at this temporal resolution for this data set ensures Gaussianity, as shown in the diagnostic in Section 5. Atmospheric temperature is resolved on a regular $N \times M$ grid over the sphere, each grid cell being approximately 1 degree in size, where $N = 288$ and $M = 192$ is the number of longitude and latitude bands, respectively. The LENS consists of a historical part from 1920 to 2005, where all runs are identical except for a small perturbation applied to the initial temperature field for each ensemble member. The scenario forcing is the Representative Concentration Pathway 8.5 (RCP 8.5, van Vuuren et al. (2011)), a high greenhouse gases concentration trajectory from 2006 to 2100, for a total of $T = 95$ years. The label 8.5 represents the forcing of $8.5 W/m^2$ radiative forcing achieved by 2100. For this work, the historical part is not considered, so the data set comprises $N \times M \times 12T \times R \approx 315$ million data points for a total of 6.2 Gigabytes. In Section 6.2 we also consider observational

data for statistical downscaling. We consider monthly observations of temperatures from January 2008 to December 2018, provided by the NOAA's National Centers for Environmental Information (gis.ncdc.noaa.gov).

Throughout this work, we denote the LENS monthly temperature by $y_t^{(r)}(\boldsymbol{s}_i)$, for realization $r = 1, \ldots, R$ and time point $t = 1, \ldots, 12T$, where the location $\boldsymbol{s}_i \in \mathbb{S}^2 = \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\| = 1\}$. The index $i = 1, \ldots, MN$ denotes the spatial locations with column-wise stacking. Henceforth we also denote by $\boldsymbol{y}_t^{(r)} = \left(y_t^{(r)}(\boldsymbol{s}_1), \ldots, y_t^{(r)}(\boldsymbol{s}_{MN}) \right)^\top$ the vector of all spatial observations for realization r and time t, and let $\boldsymbol{y}^{(r)} = \left(\boldsymbol{y}_1^{(r)}, \ldots, \boldsymbol{y}_{12T}^{(r)} \right)$ and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{12T})$, where $\boldsymbol{\mu}_t = \mathbb{E}\left(\boldsymbol{y}_t^{(r)} \right)$ represents the temperature climatology.

3. The spatio-temporal model. The model assumes that the temperature is Gaussian, with a mean $\boldsymbol{\mu}$ independent of the realization. Therefore, we have that $\boldsymbol{y}^{(r)} = \boldsymbol{\mu} + \boldsymbol{a}^{(r)}$, where $\boldsymbol{a}^{(r)} = \left(\boldsymbol{a}_1^{(r)}, \ldots, \boldsymbol{a}_{12T}^{(r)} \right)$ is a centered spatio-temporal process for the Gaussian errors for realization r. Different subsections are dedicated to the description of the different parts of the model. The diagnostics are deferred to Section 5 and are not discussed here.

3.1. *Mean structure.* For every location $\mathbf{s}_i \in \mathbb{S}^2$, we first provide a model for the mean structure $\boldsymbol{\mu}(\mathbf{s}_i) = (\boldsymbol{\mu}_1(\mathbf{s}_i), \ldots, \boldsymbol{\mu}_{12T}(\mathbf{s}_i))$. A strong seasonal effect is expected due to intra-annual variability and a smooth, slow multi-decadal change is expected due to global heating from increased greenhouse gases in the RCP 8.5 scenario. We, therefore, decompose the mean with a periodic basis function on yearly scale with slowly varying amplitudes,

$$(3.1) \qquad \mu_t(\mathbf{s}_i) = \sum_{j=0}^{p_1} t^j \left(\alpha_{j,i} + \sum_{j'=1}^{p_2} \left\{ \gamma_{j,j',i} \cos(j'\omega t) + \delta_{j,j',i} \sin(j'\omega t) \right\} \right),$$

where $\delta_{j,6,i} = 0$ when $p_2 = 6$ to avoid co-linearity in the basis, as given 12 parameters for monthly temperature, every additional parameter would necessarily be a linear combination of the others. We also have $t = 1, 2, \ldots, 12T$, while $\omega = \pi/6$ denotes the annual frequency. The parameters $\gamma_{j,j',i}$ and $\delta_{j,j',i}$ capture the site-specific intra-annual seasonal variation with harmonics up to order p_2 in the polynomial trend. $p_2 = 0$ provides no intra-annual seasonal variation and $p_2 = 6$ provides full flexibility. Figure S5 in the supplementary material shows the map of the estimated parameters when $j = 0$: the strong variation in the map indicates how it is necessary to have a spatially varying mean. The parameter vector for the mean dependence $\boldsymbol{\theta}_{\text{mean}}$ comprises $N \times M$ vectors of $(p_1 + 1)(2p_2 + 1)$ site-specific parameters $\boldsymbol{\theta}_{\text{mean,i}}$, in other words $\boldsymbol{\theta}_{\text{mean}} = \{\boldsymbol{\theta}_{\text{mean,i}}, i = 1, \ldots, MN\}$, with $\boldsymbol{\theta}_{\text{mean,i}} = \{(\alpha_{j,i}, \gamma_{j,j',i}, \delta_{j,j',i}), j = 0, \ldots, p_1, j' = 1, \ldots, p_2\}$.

3.2. *Temporal structure.* We assume that the temporal evolution of the process can be described through a vector autoregressive process of order P (VAR(P)),

$$(3.2) \qquad \mathbf{y}_t^{(r)} - \boldsymbol{\mu}_t = \sum_{j=1}^{P} \boldsymbol{\Phi}_j(\mathbf{y}_{t-j}^{(r)} - \boldsymbol{\mu}_{t-j}) + \mathbf{S}\boldsymbol{\epsilon}_t^{(r)}, t = P+1, \ldots, 12T, r = 1, \ldots, R.$$

The matrices $\boldsymbol{\Phi}_j$ are $MN \times MN$ diagonal with the coefficients for each grid cell $\phi_{j,i}$ on the diagonal (see the diagnostics section for a complete discussion about this assumption), and $\mathbf{S}$ is an $MN \times MN$ diagonal matrix with diagonal elements given by the marginal standard deviations σ_i, for $i = 1, \ldots, MN$. In the diagnostics section we show that $P = 6$ is sufficient, i.e., there is partial autocorrelation of the detrended mean temperature up to six months lag. The innovations $\boldsymbol{\epsilon}_t^{(r)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ are independent and identically distributed both across r and t, and $\mathbf{C} = \mathbf{C}(\boldsymbol{\theta}_{\text{space}})$ is a $MN \times MN$ correlation matrix describing the spatial dependence structure of the innovations through the parameter vector $\boldsymbol{\theta}_{\text{space}}$ as will be described in the following section. In the supplementary material it is proved that the temporal model (3.2) is separable if and only if $\phi_{j,i} = \phi_j$, $i = 1, \ldots, NM$, i.e. if the temporal parameters are the same across all locations. The temporal dependence can then be described through $\boldsymbol{\theta}_{\text{time}} = \{\boldsymbol{\theta}_{\text{time,i}}\}, i = 1, \ldots, MN$, which consists of MN site-specific $P+1$ dimensional vectors $\boldsymbol{\theta}_{\text{time,i}} = \{(\phi_{1,i}, \ldots, \phi_{P,i}, \sigma_i)\}$.

3.3. *Spatial dependence.* The spatial correlation structure of $\boldsymbol{\epsilon}_t^{(r)}$, i.e., the correlation matrix $\mathbf{C} = \mathbf{C}(\boldsymbol{\theta}_{\text{space}})$, must be specified. Given the large size of the spatial data sets ($M \times N \approx 55,000$ points), an explicit definition of a correlation function would result in a matrix that could not be stored, and hence the likelihood could not be computed. Instead of focusing directly on the correlation, we use a model which is built using sparse conditional dependence structure. It can be seen from the spectral representation of a Matérn covariance function with smoothness $\nu = 1$ and range ρ that the resulting two dimensional Gaussian random field $u(\cdot)$ has a spatial Markov property, i.e., two locations are independent conditional on a set of locations that separates them so that they cannot be reached from each other (Rozanov, 1977). Whittle (1954) proved that $u(\cdot)$ can be regarded as the solution of the following SPDE:

$$(3.3) \qquad\qquad (1/\rho^2 - \Delta)u(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^2,$$

where $\Delta = \frac{\partial}{\partial x}^2 + \frac{\partial}{\partial y}^2$, ρ is the range of the correlation and $\mathcal{W}(\mathbf{s})$ is a Gaussian white noise process. More recently, Lindgren, Rue and Lindström (2011) proposed to use this link as a modeling tool for random fields. While this relationship was established over a two dimensional Euclidean space, the same work also proposed to adapt the aforementioned SPDE to $\mathbb{S}^2$ by finding a solution u of

$$(3.4) \qquad\qquad (1/\rho^2 - \Delta_{\mathbb{S}^2})u(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{S}^2,$$

where $\Delta_{\mathbb{S}^2}$ is the Laplacian (Laplace-Beltrami operator) on the two-dimensional sphere. The operator is inspired from the Matérn function, but an explicit form for the correlation function on the sphere is presently not available.

The SPDE approach is a powerful way of modeling Gaussian fields, but (3.4) is limited to stationary and isotropic processes. Fuglstad et al. (2019) introduced a more general construction in terms of metric tensors that induces a non-stationary model by allowing a spatially varying SPDE operator. Practically, we keep the same SPDE, but introduce a spatially varying metric tensor $\mathbf{G}(\cdot)$ that deforms the distance between points. Here we provide a similar constructive approach to the metric tensor adapted to global nonstationary data, and henceforth we work with its inverse $\mathbf{G}^{-1}(\cdot)$, as the range is related to the inverse of the distance. Assume a location on the sphere is described through (L, ℓ), where L is the angle with the x-axis in the xy-plane (latitude), and ℓ is the angle with negative z-axis (longitude). For any location which is neither of the poles, the derivative is described with respect to these coordinates in the directions of increasing L and ℓ. We then introduce a spatially varying vector field $\mathbf{v}(\cdot) = (v_1(\cdot), v_2(\cdot))^\top$ described in the same coordinate system and a positive-valued function $\rho(\cdot)$, and define the inverse tensor metric as

$$(3.5) \qquad \mathbf{G}(\mathbf{s})^{-1} = \rho(\mathbf{s})^2 \frac{\mathbf{I}_2 + \mathbf{v}(\mathbf{s})\mathbf{v}(\mathbf{s})^{\mathrm{T}}}{\sqrt{1 + ||\mathbf{v}(\mathbf{s})||^2}}, \quad \mathbf{s} \in \mathbb{S}^2.$$

The tensor $\mathbf{G}(\cdot)$ is a spatially varying positive-definite matrix-valued function that scales local distances, so that the length of a line element on the sphere $\mathrm{d}\mathbf{l} = (\mathrm{d}L, \mathrm{d}\ell)$ at location $\mathbf{s}$ is

$$\mathrm{d}\mathbf{l}^\top \mathrm{d}\mathbf{l} = \begin{bmatrix} \mathrm{d}L & \mathrm{d}\ell \end{bmatrix} \mathbf{G}(\mathbf{s}) \begin{bmatrix} \mathrm{d}L \\ \mathrm{d}\ell \end{bmatrix} = \frac{1}{\rho(\mathbf{s})^2} \begin{bmatrix} \mathrm{d}L & \mathrm{d}\ell \end{bmatrix} \left(\frac{\mathbf{I}_2 + \mathbf{v}(\mathbf{s})\mathbf{v}(\mathbf{s})^{\mathrm{T}}}{\sqrt{1 + ||\mathbf{v}(\mathbf{s})||^2}} \right)^{-1} \begin{bmatrix} \mathrm{d}L \\ \mathrm{d}\ell \end{bmatrix}.$$

This implies that, at location $\mathbf{s}$, the original infinitesimal distance is multiplied by $1/\rho(\mathbf{s})$ and the non-diagonal term

$$\left(\frac{\mathbf{I}_2 + \mathbf{v}(\mathbf{s})\mathbf{v}(\mathbf{s})^{\mathrm{T}}}{\sqrt{1 + ||\mathbf{v}(\mathbf{s})||^2}} \right)^{-1},$$

adds anisotropy to the metric tensor. One can show that the distances along the direction of $\mathbf{v}(\mathbf{s})$ are scaled by $1/(\rho(\mathbf{s})(1 + ||\mathbf{v}(\mathbf{s})||^2)^{1/4}$, so that the effective range would be $\rho(\mathbf{s})(1 + ||\mathbf{v}(\mathbf{s})||^2)^{1/4}$. In the direction orthogonal to $\mathbf{v}(\mathbf{s})$, instead, the distance is scaled by $(1 + ||\mathbf{v}(\mathbf{s})||^2)^{1/4}/\rho(\mathbf{s})$, so that the effective range would be $\rho(\mathbf{s})/(1 + ||\mathbf{v}(\mathbf{s})||^2)^{1/4}$. Moreover, 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.

Once the vector field $\mathbf{v}(\cdot)$ and $\rho(\cdot)$ are specified, the spatially varying metric tensor $\mathbf{G}(\cdot)$ can be obtained from (3.5), and the SPDE in (3.3) no longer describes a stationary spatial field, but rather a nonstationary process which is the solution of the following SPDE

$$(3.6) \qquad [|\mathbf{G}(\mathbf{s})|^{1/2} - \nabla \cdot |\mathbf{G}(\mathbf{s})|^{1/2}\mathbf{G}(\mathbf{s})^{-1}\nabla]u(\mathbf{s}) = |\mathbf{G}(\mathbf{s})|^{1/4}\mathcal{W}(\mathbf{s}), \quad \boldsymbol{s} \in \mathbb{S}^2,$$

The details of the derivation are deferred to the supplementary material, and it can be shown that the stationary case in (3.4) is retrieved by assuming $\mathbf{G}(\cdot) \equiv 1/\rho^2 \mathbf{I}_2$, where $\rho > 0$ is a constant. The choice of defining a new SPDE by deforming distances in the original space automatically translates into spatially varying coefficients for the SPDE, but drastically improves interpretability of the parameters compared to directly parametrizing coefficients in an SPDE. The above SPDE induces some spatial variation in the marginal standard deviations, but for slowly changing $\rho(\cdot)$ and $\mathbf{v}(\cdot)$, the marginal standard deviation is close but not exactly 1. We make the separation between correlation structure and marginal variance exact at each location $\mathbf{s}$ by considering $\tilde{u}(\boldsymbol{s}) = u(\boldsymbol{s})/\sigma_{\text{scale}}(\boldsymbol{s})$, where $\sigma_{\text{scale}}(\mathbf{s})$ is calculated as the marginal standard deviations resulting from the choice of metric tensor. The resulting process $\tilde{u}(\cdot)$ has marginal variance 1.

3.3.1. *Parametrization of the metric tensor.* The fields $\rho(\cdot)$ and $\mathbf{v}(\cdot)$ can be described through some (scalar or vector) basis decomposition such as spherical harmonics and spherical vector harmonics, respectively. This approach does not however account for different behavior in the spatial structure across land and ocean. Indeed, simple diagnostics (not shown, see Castruccio and Guinness (2017) for a complete discussion) shows a strong spatial dependence across ocean, and only a weak correlation across land. Allowing for a changing dependence structure across these two geographical descriptors while still preserving positive definiteness of the implied covariance and parameter interpretability is a modelling challenge. Castruccio and Guinness (2017) approached the problem from a spectral perspective by allowing an evolutionary spectrum across the two domains.

In this work, we propose a buffer area around the coastlines, using a triangulated mesh which will be discussed in the inference section (see Figure S3 in the supplementary materials), with a separate parameter that defines a multiplicative drop in the strength of correlation in the buffer area. The buffering approach was proposed in a different context by Bakka et al. (2019), and, in the context, of this paper leads to

$$(3.7) \qquad \log\{\rho^j(\mathbf{s})\} = \sum_{l=0}^{\mathcal{L}} \sum_{m=-l}^{l} \alpha_{ml}^j Y_l^m(\mathbf{s}),$$

where α_{ml}^j are real-valued coefficients and $Y_l^m(\mathbf{s})$ are Laplace's spherical harmonic of degree l and order m. Moreover

$$(3.8) \qquad \mathbf{v}^j(\mathbf{s}) = \sum_{l=1}^{\mathcal{L}} \sum_{m=-l}^{l} \left\{ E_{lm}^{(1,j)} \nabla Y_m^l(\mathbf{s}) + E_{lm}^{(2,j)} \hat{\mathbf{r}}(\mathbf{s}) \times \nabla Y_m^l(\mathbf{s}) \right\},$$

where $\hat{\mathbf{r}}$ is the unit vector in the positive radial direction, $E_{lm}^{(1,j)}$ and $E_{lm}^{(2,j)}$ are real coefficients, and $\mathcal{L}$ is the highest order included in the bases. Finally, $j = \{\text{land}, \text{sea}\}$ depending on whether $\mathbf{s}$ is over land or sea.

If $\mathbf{s}$ is instead in the buffer zone, we have a parameter d such that $\rho(\mathbf{s}) = d\rho^j(\mathbf{s})$, whereas $\mathbf{v}(\mathbf{s}) = \mathbf{v}^j(\mathbf{s})$. The parameter d acts a multiplicative drop in the land correlation and allows for a controlled transition between the two domains. The proposed construction requires two locally defined processes over land and ocean, linked with a buffer zone which assumes a multiplicative drop in the land correlation. This construction allows for a well-defined Gaussian process as the SPDE construction guarantees the existence of a solution even in the spatially varying metric tensor case.

3.3.2. *Subgrid variation.* The spatial field will be assumed to be piece-wise constant on a triangulation of the sphere, as is shown in Section 4. 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. We use separate variances for the nugget effect on land and sea since the drastically different spatial ranges induce different levels of subgrid variation. This means two extra parameters, τ_{land}^2 and τ_{sea}^2, denoting the nugget variances on land and sea, respectively.

In summary, the covariance matrix of $\boldsymbol{\epsilon}_t^{(r)}$ is described through the parameters in the $6(\mathcal{L}+1)^2 + 3$ dimensional vector

$$
\begin{aligned}
\boldsymbol{\theta}_{\text{space}} &= \left(\{\alpha_{ml}^j, E_{lm}^{(1,j)}, E_{lm}^{(2,j)} \mid l = 0, \ldots, \mathcal{L}, m = -l, \ldots, l, j = \{\text{land}, \text{sea}\}\} , \right. \\
&\qquad \left. \{\tau_j^2, j = \{\text{land}, \text{sea}\}\}, d \right),
\end{aligned}
$$

where $E_{0,m}^{(1,j)} = E_{0,m}^{(2,j)} = 0$ for $m = 0$ and $j = 1, 2$, but are included for convenience of notation.

4. Inference. Our model encourages sparsity in the spatio-temporal dependence structure and is therefore computationally efficient, but performing inference for such a large data set simultaneously over the entire parameter space is not a feasible task. Instead, we rely on inference conditionally in a step-wise fashion. We initially estimate $\boldsymbol{\theta}_{\text{time}}$, then $\boldsymbol{\theta}_{\text{mean}}$, and finally the spatial parameters $\boldsymbol{\theta}_{\text{space}}$. This stepwise approach results in an asympotically consistent estimation (Edwards, Hammerling and Castruccio, 2019). Further, while uncertainty and bias propagation have been shown not to play a significant role in the final estimates for similar models (Castruccio and Guinness, 2017), the same work discussed intermediate steps to control for both that will not be implemented here. Further, more recent work on (spatially aggregated) monthly temperature of the LENS (Castruccio et al., 2019) showed that assuming parameters fixed does not produce any sensible change in the results. Our discussion on inference is divided into the three subsections describing each of the steps.

4.1. *Step 1: Estimating the temporal structure.* Inference on the temporal structure is performed separately at each site assuming an $\text{AR}(P)$ structure, as implied by (3.2), with $P = 6$ (see diagnostics Section 5). $\boldsymbol{\theta}_{\text{time,i}}$ can be estimated without specifying the mean, given the availability of R independent realizations. Castruccio and Stein (2013)
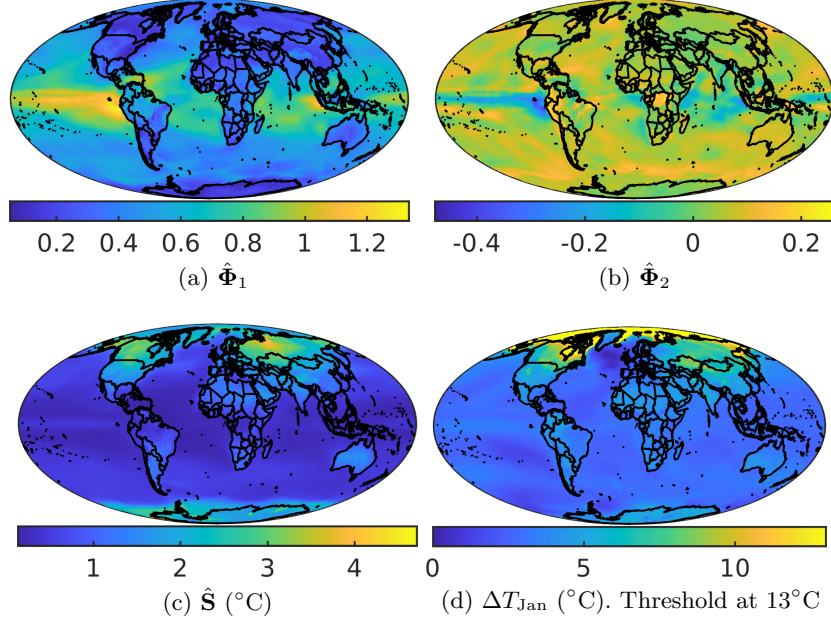
Figure 1: Global maps of estimates of (a-c) temporal parameters from (3.2) and (d) ΔT_{Jan}, defined as the difference between the expected value for January 2100 and the average values of the expected values of Januaries in 2013–2043 in °C, the reference time period used in past work related to the same data set.

showed that the restricted likelihood can be written in terms of the difference field $\mathbf{D}^{(r)}(\mathbf{s}_i) = \mathbf{y}^{(r)}(\mathbf{s}_i) - \bar{\mathbf{y}}(\mathbf{s}_i)$, where $\bar{\mathbf{y}}(\mathbf{s}_i) = \frac{1}{R} \sum_{r=1}^{R} \mathbf{y}^{(r)}(\mathbf{s}_i)$. If we denote by $\mathbf{D}(\mathbf{s}_i) = \left(\mathbf{D}^{(1)}(\mathbf{s}_i), \ldots, \mathbf{D}^{(R)}(\mathbf{s}_i) \right)$ and by $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\theta}_{\text{time,i}})$ the (sparse) precision matrix implied by the $\text{AR}(P)$ model, then the restricted loglikelihood can be written as

$$(4.1) \qquad \ell(\boldsymbol{\theta}_{\text{time,i}} \mid \mathbf{D}(\mathbf{s}_i)) = \text{const} + (R-1)\log|\mathbf{Q}| - \sum_{r=1}^{R} \mathbf{D}^{(r)}(\mathbf{s}_i)^{\top} \mathbf{Q} \mathbf{D}^{(r)}(\mathbf{s}_i),$$

where const is a generic constant. Inference can be performed by maximizing (4.1), independently (in parallel) for every location. Henceforth, we denote by $\hat{\boldsymbol{\Phi}}_j$ the $MN \times MN$ diagonal matrix with entries $\hat{\phi}_{j,i}$, for $j = 1, \ldots, P$ and $\hat{\mathbf{D}}$ the $MN \times MN$ diagonal matrix with entries $\hat{\sigma}_i$.

Figure 1 shows the map of the estimated temporal parameters. In panels (a-b) we show $\hat{\boldsymbol{\Phi}}_1, \hat{\boldsymbol{\Phi}}_2$, respectively, as indicated in (3.2) (the parameter estimates for $j = 3, \ldots, P = 6$ are shown in Figure S1). In panel (c) we show results for $\hat{\mathbf{S}}$, and in panel (d) we show the expected temperature anomaly in January 2100 against the January average of 2013–2043, the same time window used in the reference publication of the LENS (Kay et al., 2015). $\hat{\boldsymbol{\Phi}}_1$ is larger over ocean than land, as the ocean warming rate is slower due to the ocean heat uptake and deep oceanic convection. $\hat{\boldsymbol{\Phi}}_2$ is particularly

large and negative in the Equatorial Pacific, and is partly attributable to the El-Niño oscillation. $\hat{\mathbf{S}}$ is highest at the poles, and lower over ocean than land. The January anomaly highlights how the winter warming is expected to be more dominant at the poles and across land.

4.2. *Step 2: Estimating the mean structure.* The mean structure is estimated by further smoothing the ensemble mean at each site $\bar{\mathbf{y}}(\mathbf{s}_i)$ (as defined in the previous section) with the basis function proposed in (3.1). We denote by $\mathbf{X}$ the design matrix generated by (3.1) at each location with $p_1 = 3$ and $p_2 = 6$ (see diagnostics section for the choice of these parameters). We estimate $\boldsymbol{\theta}_{\text{mean,i}}$ conditional on the inference in the previous step, i.e., on $\hat{\boldsymbol{\theta}}_{\text{time,i}}$. Once the temporal parameters have been estimated, the precision matrix for the time series at location $\mathbf{s}_i$ is computed by $\hat{\mathbf{Q}} = \mathbf{Q}(\hat{\boldsymbol{\theta}}_{\text{time,i}})$. The mean parameters are then obtained by the generalized least squares

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

As in the previous step, this estimation can be performed independently for each site, hence allowing an easily parallelizable inference. The vector of estimated means for all time points for location $\mathbf{s}_i$ is given by $\mathbf{X}\hat{\boldsymbol{\theta}}_{\text{mean,i}}$, and henceforth we denote by $\hat{\boldsymbol{\mu}}_{t,i}$ the t-th element of $\mathbf{X}\hat{\boldsymbol{\theta}}_{\text{mean,i}}$. Further, we let $\hat{\boldsymbol{\mu}}_t = (\hat{\boldsymbol{\mu}}_{t,1}, \ldots, \hat{\boldsymbol{\mu}}_{t,MN})$.

4.3. *Step 3: Estimation of spatial covariance structure.* Given $\hat{\boldsymbol{\theta}}_{\text{mean}}, \hat{\boldsymbol{\theta}}_{\text{time}}$, the spatial correlation matrix $\mathbf{C}$ of (3.2) can be estimated by considering the innovations $\boldsymbol{\varepsilon}_t^{(r)}$ of the VAR(P) process in (3.2). We first standardize, de-trend and temporally de-correlate:

$$(4.2) \qquad \hat{\boldsymbol{\varepsilon}}_t^{(r)} = \hat{\mathbf{S}}^{-1/2} \left\{ \boldsymbol{y}_t^{(k)} - \hat{\boldsymbol{\mu}}_t - \sum_{j=1}^{P} \hat{\boldsymbol{\Phi}}_j (\boldsymbol{y}_{t-j}^{(k)} - \hat{\boldsymbol{\mu}}_{t-j}) \right\}.$$

For the SPDE model, we must estimate 2 nugget parameters, 1 buffer parameter, 73 parameters for non-stationarity on sea, and 73 parameters for non-stationarity on land, for a total of 149 parameters. These parameters are estimated by maximum likelihood, assuming a triangulation of the globe as described in Lindgren, Rue and Lindström (2011) and attributing every location to the corresponding triangle. In this paper we use a triangulation with 15392 triangles, with a land/ocean buffer to allow for changes across these two domains, as shown in Figure S3. We do not use the finite-element approach for solving the SPDE on the sphere as in Lindgren, Rue and Lindström (2011), but a finite volume approach on each triangle, similar to what was done in Fuglstad et al. (2015a,b) for a regular grid. The formal derivation of the sparse precision matrix $\mathbf{C}^{-1}$ is provided in the supplementary material, here we provide the general idea. The SPDE (3.6) is integrated over a generic triangle T_i and we require equality in distribution jointly for all triangles,

$$\left[\int_{T_i} \frac{1}{\rho(\mathbf{s})^2} - \nabla \cdot \mathbf{H}(\mathbf{s}) \nabla \right] u(\mathbf{s}) \, dV \overset{d}{=} \int_{T_i} \frac{1}{\rho(\mathbf{s})} \mathcal{W}(\mathbf{s}) \, dV,$$
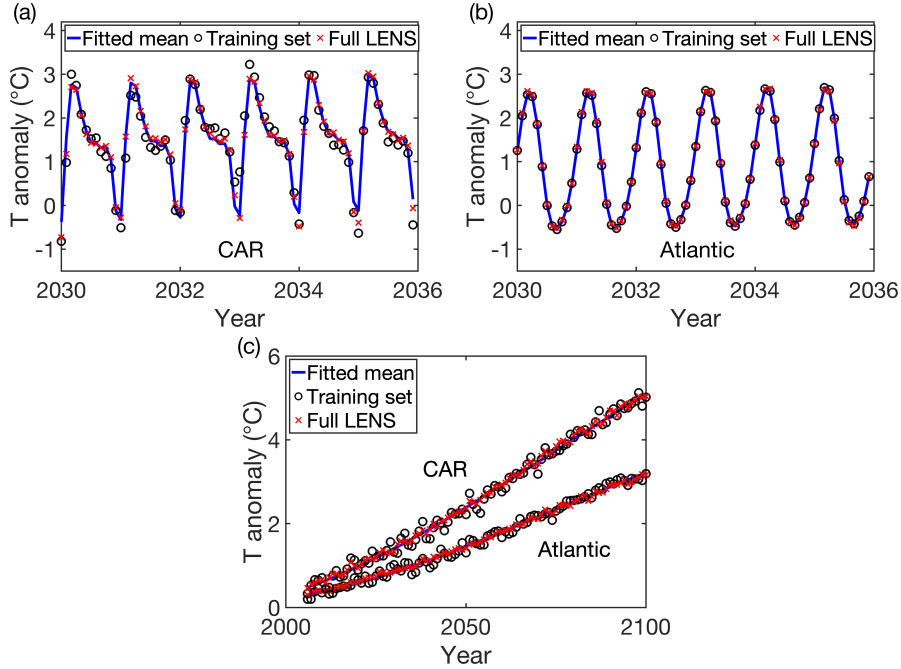
Figure 2: Comparison of the SG (fitted mean), the LENS based on $R = 5$ (training set) and 40 runs (full LENS) for (a-b) intra- and (c) inter-annual variability. The comparison is performed for the locations *CAR* and *Atlantic* in panels (a) and (b), respectively, while panel (c) shows the results for both. The locations are mapped in Figure 3.

where dV is the surface measure on the triangles. The use of several approximations of this equation and the divergence theorem ensure that the previous equation leads to a set of linear equations for a Gaussian vector assumed constant across each triangle.

5. Diagnostics. The proposed model relies on assumptions that this section will explore and discuss in detail. Henceforth, we refer to four locations which exemplify different dependence structure across the globe: one in the Atlantic Ocean (*Atlantic*, 15°W, 8.95°S), one in the Pacific Ocean (*Pacific*, 111.25°W, 7.07°S), one in Oklahoma, US (*USA*, 97.5°W, 35.34°N), and one in the Central African Republic (*CAR*, 17.5°E, 4.24°N), as indicated by the crosses in Figure 3. Here we report the results for diagnostics of the training set size, mean structure and spatial structure. In the supplementary material, we also validate independence conditional on climate, temporal structure and Gaussianity.

5.1. *Training set size and mean structure.* The training set size R must be chosen so that the precision of the parameter estimates is balanced with the implied benefits of the SG. A small R would lead to unstable parameter estimates, while a large R would imply a large ensemble, an impractical scenario in many applications. Indeed, for the vast majority of ensembles such as the CMIP5 mentioned in the introduction, R ranges

between 3 and 6 for the end-of-century, multi-decadal projections. The mean structure in (3.1) is also spatially dependent, as discussed in the supplementary material and also shown in Figure S5, and requires a choice of p_1 and p_2. Both coefficients must be large enough to capture the temporal and site-specific evolution of the climatology, but also small to reduce the amount of parameters to be stored. In the supplementary material, we provide a comprehensive discussion about how to assess the fit, and how does the fit changes as a function of R, p_1 and p_2. The different panels of Figure S6 show a diminished return in goodness of fit in increasing R and (p_1, p_2), and justify the choice of $R = 5$, $p_1 = 3$ and $p_2 = 6$.

The fit can be also qualitatively assessed by comparing the LENS mean with the estimated mean from the SG. Figure 2 shows this comparison for two sites, CAR and *Atlantic*, each representative of land and ocean behavior, in terms of intra- and inter-annual variability. In panels (a-c), the fitted mean structure in (3.1), the ensemble mean for the R members of the training set, as well as the full LENS mean is compared. Panels (a-b) represent the intra-annual comparison for CAR over land point, and *Atlantic* over the ocean, for a sample time period between 2030 and 2036, expressed as anomaly with respect to the average over 2013-2043. For CAR in panel (a), it is apparent how the training set mean and the ensemble mean are generally well aligned, and the fitted model is able to capture the overall monthly climatology well. For *Atlantic* in panel (b), the same comparison is performed, and it is apparent how the annual cycle resembles the seasonal behavior of the ocean temperatures. In this setting, both ensemble means are generally well aligned, and the mean structure in (3.1) is able to capture the pattern. Panel (c) presents a comparison of the multi-decadal variability for the entire simulation period and the same two locations. Ensemble annual mean for both training set and full LENS is compared with the implied annual structure from the model. *Atlantic* shows an overall slower warming rate against CAR, as expected by higher ocean heat capacity. The assumed polynomial increase with temperature is flexible enough to capture the different rates of increase, resulting in a satisfactory fit. Figure S7 in the supplementary material also shows similar results for the two points at the Arctic and Antarctic pole.

5.2. *Spatial structure.* The use of spatially varying coefficients in the SPDE operator, as well as the use of buffers to allow for different land/ocean behavior allows for an increase in flexibility. There is a clear, strong difference in spatial structure between land and sea (diagnostics not shown), and in order to assess the extent of the improvement of our non-stationary SG, we compare it a SG which is stationary across the both domains, i.e. with constant but different SPDE coefficients over land and sea. Figure 3 shows the contour plot of the correlation function for the decorrelated residuals in time as defined in (4.2) for the four reference points. Panel (a) shows the LENS correlation, i.e., the empirical correlation across all times and realizations of the ensemble, and compares it with the correlation implied by the stationary SG in panel (b), as well as the proposed non-stationary SG in panel (c). The limitation of the stationary SG are easily detectable, as both the *Atlantic* and *Pacific* location show a long range correlation which cannot be captured by a model with no preferred direction of dependence. The proposed non-stationary model is instead able to capture this dependence
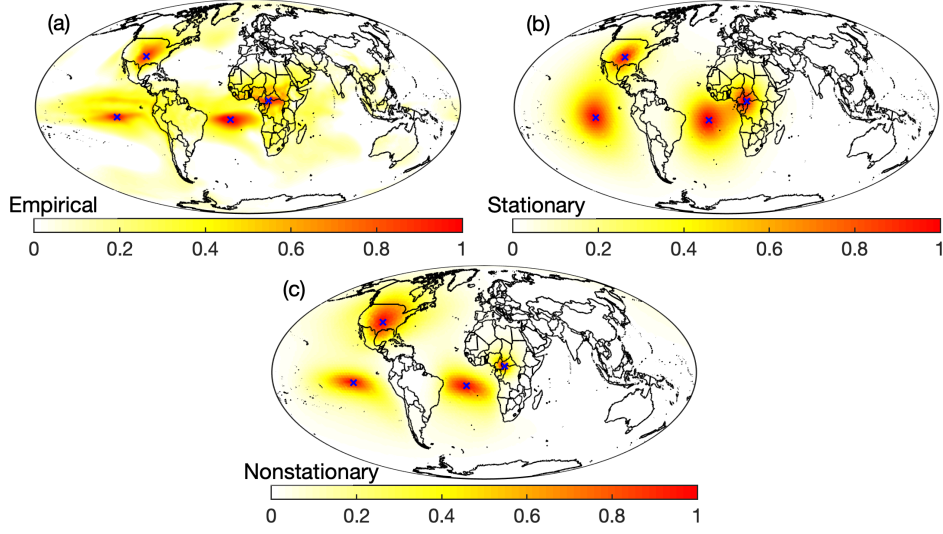
Figure 3: Estimated correlation between reference points marked with blue crosses and all other locations. Multiple locations are shown in the same plot. (a) LENS estimate (b) stationary SG, and (c) non-stationary SG. Negative values have been set to 0 for the ease of visualization.

considerably more adequately, as the contour map is deformed to match the shape of the LENS correlation.

To quantify the extent of the improvement of the proposed non-stationary SG, we also perform a simulation study to test the interpolating ability of the model. We focus on four regions centered in the reference points (see Figure S11), and after temporally detrending, we fit the stationary and non-stationary SG, and compute the difference in BIC between them. We also perform cross-validation to each time and each of the R runs, and average the results. We choose two sampling schemes for cross-validation: 1) leave out the full areas shown in Figure S11, and 2) leave out only one point for each of the areas. In both cases RMSE as well as Continuous Ranked Probability Score (CRPS) is computed. The results are shown in Table S1. The results confirm the visual intuition of Figure 3: *Atlantic* and *Pacific* show a larger improvement when considering a non-stationary SG in terms of BIC and better interpolation performances for both sampling schemes and measures. *CAR* and *USA* show an overall smaller BIC improvement, and a consistently better performance only when the entire area is removed for cross-validation.

The assumption of a constant spatial dependence structure across time was investigated through independent fits for ten-year windows (diagnostics not shown has demonstrated robustness with respect to the choice of the length of the time window). We observed that parameters of the stationary SG display some variability, which compared to their absolute value is negligible; see Figure S13. This behavior is also displayed by the non-stationary model. Indeed, Figure S14 shows the contour plot of the correlation

for the four reference points, across the ten temporal windows. It can be seen how the contours are visually indistinguishable across time, thus further indicating that the spatial structure does not sensibly change across years. We have further evaluated the contour plots of the correlation structure for each month in figure S15, and while overall the spatial structure was still found to be constant, we observe some small changes due to seasonality in the *Atlantic* reference point for the months of August-November.

6. Unconditional compression and downscaling of the Large Ensemble. Once inference is performed for the $R = 5$ LENS members in the training set (out of the forty members) as detailed in Section 4, we apply this model to act both as a SG, i.e., a means to unconditionally compress the climate model ensemble, as well as a statistical downscaler by performing spatial interpolation.

6.1. *Unconditional compression.* The model's ability to capture the LENS variability implies that only five runs are necessary for the monthly mean temperature. Therefore, the full LENS, which is 6.2 Gb, can be effectively replaced by 20Mb of disk space for the estimated parameters, and allows the SG to be downloaded from a common laptop. The simulations are obtained following a three-steps approach: 1) generate $\varepsilon_t^{(r)}$ from the SPDE model (3.6) 2) temporally correlate the error with the VARMA(P) model (3.2) and 3) add the trend (3.1). A single realization can be generate in less than 30 seconds per simulation, and would scale linararly with the number of simulations.

In this section, we simulate 40 new runs from the SG, and compare them with the LENS members in terms of their generated uncertainty. Since the SG aims at characterizing the distribution generating the data and not the data themselves, we do not perform a comparison in terms of RMSE or other pointwise measures. Instead, a fair comparison focuses on the ability of the SG to capture the uncertainty implied by the LENS. Since the different ensemble members only differ by an atmospheric perturbation and not by the forcing or physics, the uncertainty is usually referred as *unforced* or *internal variability* (Yoshimori et al., 2005; Nikiema and Laprise, 2011; Torn, 2016).

Figure 4 shows a comparison of the internal variability for the four reference points in the diagnostic Section 5. The red and blue boxplots represent the variability according to the SG and LENS, respectively. In panel (a), the temperature change from 2013–2043 (the reference period for the LENS diagnostics in Kay et al. (2015)) for all four points is shown, and the overall mean, as well as the uncertainty is well reproduced by the SG. The four locations have very different variability depending on latitude and domain, e.g. the CAR temperature is expected to be relatively stable due to its geographical position in the tropics, whereas the climate of the USA location is mostly continental and more subject to variability due to possible changes in intensity and occurrence of jet streams. Panel (b) focuses on the CAR location from 2013 to 2050, and shows the uncertainty in the temperature anomaly across seven representative years. The SG is overall able to capture the mean trend in time, as well as the internal variability. Therefore, the parameters of the SG are able to generate approximately the same uncertainty for multi-decadal temperature change than the full LENS with less
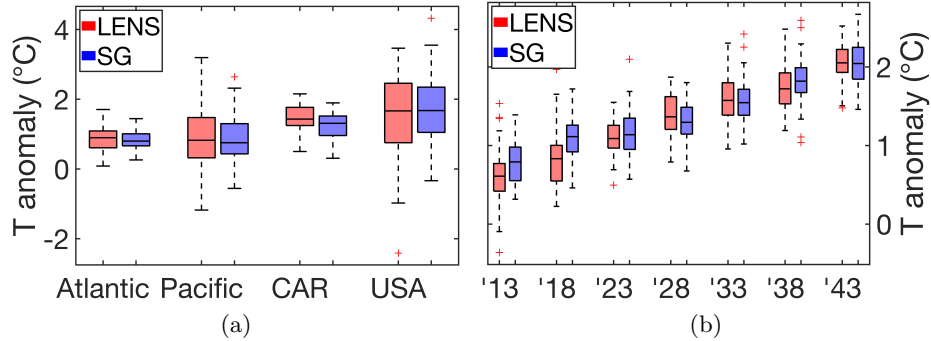
Figure 4: Comparison of LENS and SG internal variability, expressed as red and blue boxplots, respectively. (a) Temperature anomaly from 2013-2043 for the four reference locations mapped in Figure 3 (coordinates are provided in Section 5). (b) Temperature anomaly of seven representative years for *CAR*.

than 1% of the storage space required.

A comparison of the internal variability between the SG and LENS among sites justifies the use of a marginal model, but does not however justify the need of a spatial model. Indeed, if one is only interested in site-specific or aggregated temperature, similar results can be obtained with a simple time series model (Castruccio et al., 2019). A spatial model is however necessary in order to produce physically consistent maps. Indeed, a SG with no spatial dependence would produce nonsensical maps, as the simulated monthly temperature evolution at one site would not be influenced by nearby values. Movie S1 in the supplementary material compares the same simulations as in Figure 4 with a model with the same temporal structure as in this work but with spatially independent innovations for 2006-2009, the first four years of simulations. A spatially independent simulation is clearly unsuitable for any type of scientific investigation. Figure 5 shows a comparison for one LENS and one SG simulation for January and July, a (Boreal) winter and a summer month, in 2050. The maps have the same large scale patterns, as is apparent from the figures, and local behavior has already been compared in the diagnostics section 5, where the non-stationary SG has been shown to have good interpolation performance.

6.2. *Statistical downscaling.* The availability of an underlying continuous model allows to generate SG simulations at an arbitrarily fine resolution. Figure 6 shows an example of the 2050 temperature maps in January and July according to one LENS member (a,c) against one downscaled SG (b,d), with a 288×192 resolution in the domain shown. The coarse features of the LENS are readily apparent at this level of detail, for both the Boreal winter and summer. Instead, the downscaled SG allows for smoother spatial fields and can simulate temperature at any spatial location, hence allowing a fair comparisons with in-situ measurements. The extent of the improvement in smoothness from the LENS to the downscaled SG can be measured by regridding
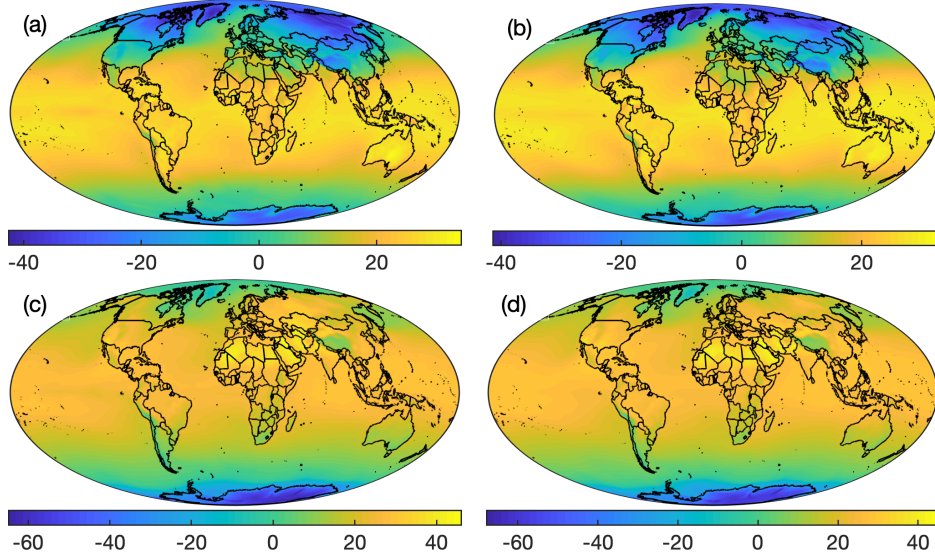
Figure 5: Comparison of one LENS and one SG simulation in January (a-b) and July 2050 (c-d). (a,c) Show the temperature maps for the LENS, while (b,d) show the same maps for the SG.

the LENS at high resolution via a nearest neighborhood approach, and compare the roughness of the map, defined as $R^{(j)}(t) = \int \|\Delta f^{(j)}(t)\|^2 \mathrm{d}x \mathrm{d}y$, where $j = 1, 2$ for the LENS and the downscaled SG, respectively at time t, with the integral over the domain in Figure 6. If we define by $\bar{R}^{(j)} = \frac{1}{12T} \sum_{t=1}^{12T} R^{(j)}(t)$ the average roughness in time, the downscaled SG resulted in a ten-fold decrease in roughness, i.e., $\frac{\bar{R}^{(2)} - \bar{R}^{(1)}}{\bar{R}^{(1)}} = -0.91$. The decrease in roughness also 93% and 79% for January (a,c) and July (b,d), respectively.

We choose two locations indicated by the red crosses in panels (a-d): one near Chicago (Chicago Aurora Municipal Airport) and one near San Antonio (Hondo Municipal Airport). We consider monthly temperatures from January 2008 to December 2018, provided by the NOAA's National Centers for Environmental Information (NCEI, gis.ncdc.noaa.gov). The two sites where chosen to exemplify two different types of climate (continental and subtropical) and also based on the availability of a continuous ten-year record of observations. We perform $R = 100$ SG simulations on the US domain in panels (a-d) and consider the uncertainty of the simulated temperature profiles against the observations. Panels (e-f) show the internal variability at the two sites for the entire time period in the form of a functional boxplot (Sun and Genton, 2011). The purple band represents the functional interquartile range, while the blue lines are the outlying bands, and the dashed black line represents the observational data. The intra-annual variability of the station in Chicago (e) is very strong due to continental climate, and a sharp winter-summer transition. The downscaled SG highlights a larger seasonal uncertainty in these two seasons compared to the relatively short springs and falls in the Midwest. Noticeably, the observational record is well within the functional
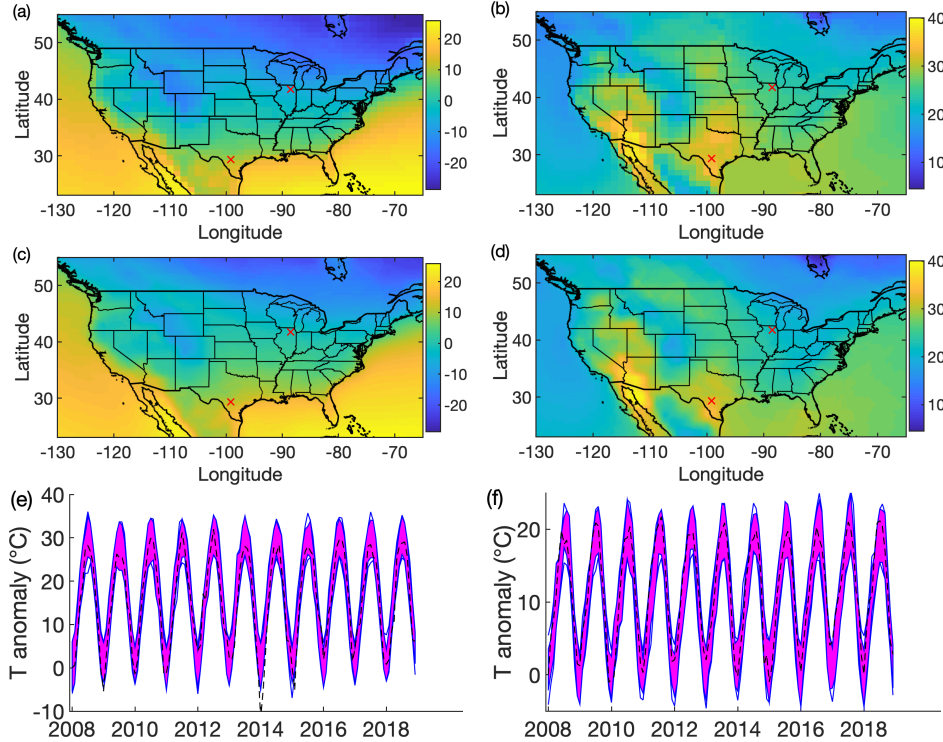
Figure 6: (a,c) January and (b,d) July 2050 monthly US temperatures for one LENS member (a,b) and one downscaled SG simulation (c,d). The two red crosses are the sites near Chicago and San Antonio chosen for comparison. (e-f) Functional boxplot of the internal variability for $R = 100$ downscaled SG of monthly temperature (from January 2008 to December 2018) against the ground observations near Chicago (e) and San Antonio (f).

interquartile range for all months except January 2014, where a polar vortex across the Canada, the Midwest and the Great Plains generated temperatures ten degrees Celsius below average for a few days, thus resulting in a smaller monthly average than the usual climate.

For the station near San Antonio, the humid subtropical climate, as well as its adjacency with the Gulf of Mexico mitigates the intra-annual variability and results in milder springs and falls. The aforementioned polar vortex recorded in January 2014 in Chicago did not impact the temperatures at this latitude, so the observations are all well within the functional interquartile range, with no noticeable outlier.

7. Discussion. We proposed a new approach to climate model compression with a space-time statistical model whose global structure is derived from a solution of a SPDE. The proposed approach is able to store the information for a climate model ensemble of more than 6 Gigabytes in less than 20 Megabytes, and the generation of new simulations is an instantaneous task that can be performed on a laptop. Our method

overcomes the fundamental limitations of spectral-based approaches, and represents a significant advance towards the development of a usable, simple-to-use tool based on statistical models. The SPDE model is able to jointly capture different land/ocean behaviors by allowing a buffer zone between the two domains, as well as a multiplicative coefficient describing the drop in correlation. Once the model parameters are estimated, they are stored and used to simulate surrogate climate model runs in order to assess the internal variability of the ensemble, and to compare the downscaled output with observational data. While we focused on a global model at ∼1 degree resolution, the induced sparsity in the precision matrix would make this model scalable to considerably higher spatial resolutions, e.g. ∼25km as in the Coupled Model Intercomparison Project Phase 6 (CMIP6, Eyring et al. (2016)).

The SPDE approach represents a considerable improvement from currently available methods. Firstly, the proposed model allows for an underlying continuous process, hence allowing for both conditional simulation and statistical downscaling at arbitrarily high spatial resolution. In contrast, discrete spectral methods can only be used for simulation at the native grid, a limitation which considerably hampers its practical use. Secondly, the SPDE approach does not depend on the sampling geometry, and it can be applied to any variable, even the ones whose observational domain is not a regular grid, such as sea surface temperature. Thirdly, as also observed by Guinness and Hammerling (2018) for an SPDE stationary model, the triangulation implies no numerical stability issues at the poles due to latitude/longitude parametrization and sampling geometry (i.e., more points at very high or low latitudes), and allows to naturally model nonstationarity across latitudes. Finally, the use of a metric tensor with spatially varying deformation in the SPDE framework allows to retain a well-defined model with an intuitive interpretation of the parameters. The proposed approach is also computationally convenient. We can achieve inference for more than 300 million data points by means of the sparsity in the precision matrix implied by the finite volume approach used to discretize the SPDE and a vector autoregressive structure in time.

The main limitation of the proposed method is its focus on Gaussian processes. While this is not a significant shortcoming for multi-decadal ensembles with end-of-century projections, this could be an issue to practitioners interested in compressing ensembles of short, high resolution and possibly regional simulations where hourly and daily output is of interest. Our proposed approach can however be generalized to the non-Gaussian setting by proposing a latent Gaussian model. More specifically, the SPDE approach can be used as a latent process in a hierarchical Bayesian model setting with a non-Gaussian distribution at data level.

SUPPLEMENTARY MATERIAL

Supplement A: Supplementary document
(). A document providing additional technical details and supplementary details.

Supplement B: Example code
(). Example code for fitting the non-stationary global covariance structure to a reduced number of realizations.

Supplement C: Movie

(). A movie demonstrating the need for spatial structure when generating surrogate climate model runs.

References.

BAKER, A. H., XU, H., DENNIS, J. M., LEVY, M. N., NYCHKA, D., MICKELSON, S. A., EDWARDS, J., VERTENSTEIN, M. and WEGENER, A. (2014). A Methodology for Evaluating the Impact of Data Compression on Climate Simulation Data. In *Proceedings of the 23rd international symposium on High-performance parallel and distributed computing* 203–214. ACM HPDC '14.

BAKER, A. H., HAMMERLING, D. M., MICKELSON, S. A., XU, H., STOLPE, M. B., NAVEAU, P., SANDERSON, B., EBERT-UPHOFF, I., SAMARASINGHE, S., DE SIMONE, F., CARBONE, F., GENCARELLI, C. N., DENNIS, J. M., KAY, J. E. and LINDSTROM, P. (2016). Evaluating Lossy Data Compression on Climate Simulation Data within a large ensemble. *Geoscientific Model Development* **9** 4381-4403.

BAKER, A. H., XU, H., HAMMERLING, D. M., LI, S. M. and CLYNE, J. P. (2017). Toward a multi-method approach: Lossy data compression for climate simulation data. In *High Performance Computing. ISC High Performance 2017* 30–42. J. Kunkel, R. Yokota, M. Taufer, and J. Shalf (Eds.).

BAKKA, H., VANHATALO, J., ILLIAN, J. B., SIMPSON, D. and RUE, H. (2019). Non-stationary Gaussian models with physical barriers. *Spatial Statistics* **29** 268 - 288.

BOLIN, D. and LINDGREN, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Annals of Applied Statistics* **5** 523–550.

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

CASTRUCCIO, S. and 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** 329–344.

CASTRUCCIO, S. and STEIN, M. L. (2013). Global space–time models for climate ensembles. *Annals of Applied Statistics* **7** 1593–1611.

CASTRUCCIO, S., HU, Z., SANDERSON, B., KARSPECK, A. and HAMMERLING, D. (2019). Reproducing Internal Variability with Few Ensemble Runs. *Journal of Climate* **32** 8511-8522.

EDWARDS, M., HAMMERLING, D. and CASTRUCCIO, S. (2019). Marginally Parametrized Spatio-Temporal Models and Stepwise Maximum Likelihood Estimation. arXiv:1806.11388.

EYRING, V., BONY, S., MEEHL, G. A., SENIOR, C. A., STEVENS, B., STOUFFER, R. J. and TAYLOR, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9** 1937–1958.

FUENTES, M. (2007). Approximate Likelihood for Large Irregularly Spaced Spatial Data. *Journal of the American Statistical Association* **102** 321-331.

FUGLSTAD, G.-A., LINDGREN, F., SIMPSON, D. and RUE, H. (2015a). Exploring a New Class of Non-stationary Spatial Gaussian Random Fields with Varying Local Anisotropy. *Statistica Sinica* **25** 115–133.

FUGLSTAD, G.-A., SIMPSON, D., LINDGREN, F. and RUE, H. (2015b). Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics* **14** 505–531.

FUGLSTAD, G.-A., SIMPSON, D., LINDGREN, F. and RUE, H. (2019). Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association* **114** 445–452.

GUINNESS, J. and HAMMERLING, D. (2018). Compression and Conditional Emulation of Climate Model Output. *Journal of the American Statistical Association* **113** 56-67.

INGEBRIGTSEN, R., LINDGREN, F. and STEINSLAND, I. (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics* **8** 20–38.

INGEBRIGTSEN, R., LINDGREN, F., STEINSLAND, I. and 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. and GENTON, M. G. (2018). Reducing Storage of Global Wind Ensembles with Stochastic Generators. *Annals of Applied Statistics* **12** 490-509.

JEONG, J., YAN, Y., CASTRUCCIO, S. and GENTON, M. G. . (2019). A Stochastic Generator of Global Monthly Wind Energy with Tukey g-and-h Autoregressive Processes. *Statistica Sinica* **29** 1105-1126.

JUN, M. and STEIN, M. L. (2007). An Approach to Producing Space–Time Covariance Functions on Spheres. *Technometrics* **49** 468-479.

KAY, J. E., DESER, C., PHILLIPS, A., MAI, A., HANNAY, C., STRAND, G., ARBLASTER, J. M., BATES, S. C., DANABASOGLU, G., EDWARDS, J., HOLLAND, M., KUSHNER, P., LAMARQUE, J. F., LAWRENCE, D., LINDSAY, K., MIDDLETON, A., MUNOZ, E., NEALE, R., OLESON, K., POLVANI, L. and 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** 1333–1349.

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

MEEHL, G. A., MOSS, R., TAYLOR, K. E., EYRING, V., STOUFFER, R. J., BONY, S. and STEVENS, B. (2014). Climate Model Intercomparisons: Preparing for the Next Phase. *Eos, Transactions American Geophysical Union* **95** 77-78.

NIKIEMA, O. and LAPRISE, R. (2011). Budget study of the internal variability in ensemble simulations of the Canadian Regional Climate Model at the seasonal scale. *Journal of Geophysical Research: Atmospheres* **116**.

ROZANOV, J. A. (1977). Markov random fields and stochastic partial differential equations. *Mathematics of the USSR-Sbornik* **32** 515.

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability* **104**. Chapman & Hall, London.

SAMPSON, P. D. and GUTTORP, P. (1992). Nonparametric Estimation of Nonstationary Spatial Covariance Structure. *Journal of the American Statistical Association* **87** 108-119.

SMALL, R. J., BACMEISTER, J., BAILEY, D., BAKER, A., BISHOP, S., BRYAN, F., CARON, J., DENNIS, J., GENT, P., HSU, H.-M., JOCHUM, M., LAWRENCE, D., MUÑOZ, E., DINEZIO, P., SCHEITLIN, T., TOMAS, R., TRIBBIA, J., TSENG, Y.-H. and VERTENSTEIN, M. (2014). A new synoptic scale resolving global climate simulation using the Community Earth System Model. *Journal of Advances in Modeling Earth Systems* **6** 1065-1094.

STEIN, M. L. (2005). Statistical Methods for Regular Monitoring Data. *Journal of the Royal Statistical Society, Series B* **67** 667–687.

SUN, Y. and GENTON, M. G. (2011). Functional Boxplots. *Journal of Computational and Graphical Statistics* **20** 316-334.

TAYLOR, K., STOUFFER, R. and MEEHL, G. (2012). An Overview of CMIP5 and the Experiment Design. *Bullettin of the American Metereological Society* **93** 485–498.

TORN, R. D. (2016). Evaluation of Atmosphere and Ocean Initial Condition Uncertainty and Stochastic Exchange Coefficients on Ensemble Tropical Cyclone Intensity Forecasts. *Monthly Weather Review* **144** 3487-3506.

VAN VUUREN, D. P., EDMONDS, J., KAINUMA, M., RIAHI, K., THOMSON, A., HIBBARD, K., HURTT, G. C., KRAM, T., KREY, V., LAMARQUE, J.-F., MASUI, T., MEINSHAUSEN, M., NAKICENOVIC, N., SMITH, S. J. and ROSE, S. K. (2011). The Representative Concentration Pathways: An Overview. *Climatic Change* **109** 5–31.

WHITTLE, P. (1954). On Stationary Processes in the Plane. *Biometrika* **41** pp. 434-449.

YOSHIMORI, M., STOCKER, T. F., RAIBLE, C. C. and RENOLD, M. (2005). Externally Forced and Internal Variability in Ensemble Climate Simulations of the Maunder Minimum. *Journal of Climate* **18** 4253-4270.

DEPARTMENT OF MATHEMATICAL SCIENCES,
NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY,
TRONDHEIM, NORWAY
E-MAIL: geir-arne.fuglstad@ntnu.no

DEPARTMENT OF APPLIED AND
COMPUTATIONAL MATHEMATICS
AND STATISTICS,
UNIVERSITY OF NOTRE DAME, USA
E-MAIL: scastruc@nd.edu