Noname manuscript No.
(will be inserted by the editor)

Spatio-temporal quantification of climate model errors in a Bayesian framework

Maeregu Woldeyes Arisido · Carlo Gaetan · Davide Zanchettin · Jorge Lopez Parages · Angelo Rubino

Received: date / Accepted: date

Abstract Numerical output from coupled atmosphere-ocean general circulation models is a key tool to investigate climate dynamics and the climatic response to external forcings, and to generate future climate projections. Coupled climate models are, however, affected by substantial systematic errors or biases compared to observations. Assessment of these systematic errors is vital for evaluating climate models and characterizing the uncertainties in projected future climates. In this paper, we develop a spatio-temporal model based on a Bayesian hierarchical framework that quantifies systematic climate model errors accounting for their underlying spatial coherence and temporal dynamics. The key feature of our approach is that, unlike previous studies that focused on empirical and purely spatial assessments, it simultaneously determines the spatial and temporal features of model errors and their associated uncertainties. This is achieved by representing the spatio-temporally referenced data using weighting kernels that capture the spatial variability efficiently while reducing the high dimensionality of the data, and allowing the coefficients linking the weighting kernels to temporally evolve according to a random walk. Further, the proposed method characterizes the bias in the mean state as the time-invariant average portion of the spatio-temporal climate model errors. To illustrate our method, we present an analysis based on the case of near-surface air temperature over the southeastern tropical Atlantic and bordering region from a multi-model ensemble mean of historical simulations from the fifth phase of the Coupled Model Intercomparison Project. The results demonstrate the improved characterization of climate model errors and identification of non-stationary temporal and spatial patterns.

Keywords Bayesian hierarchical method · Climate model · Climate model errors · CMIP5 · Spatial statistics · Spatio-temporal model

M.W. Arisido (✉)
University of Milano-Bicocca, via Cadore 48, 20900 Monza, Italy
E-mail: maeregu.arisido@unimib.it
Tel.: +393665067068

C. Gaetan · D. Zanchettin · J. L. Parages · A. Rubino
Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Via Torino 155, 30172 Venice, Italy

1 Introduction

Coupled climate models use mathematical approximations of physical and biogeochemical processes to simulate the transfer of energy and mass within and across the various compartments of the climate system (Flato et al. 2013). Numerical simulations performed with such models are used to investigate climate dynamics and the climatic response to external forcings, to predict climate evolution and to generate future climate projections, where climate changes as a result of natural as well as anthropogenic forcings can be investigated (Tebaldi et al. 2005; Flato et al. 2013). Despite their continued improvements in representing atmospheric and oceanic physical processes, simulations performed with the current generation of coupled climate models suffer from substantial deficiencies (e.g., Hooten et al. 2008). Among these, of special relevance are the systematic errors that affect the mean state, seasonality and interannual-to-decadal variability simulated by climate models compared to observations (Hawkins et al. 2014; Wang et al. 2014). These systematic errors are commonly referred to as climate model biases (e.g., Cannon 2017).

Systematic climate model errors develop due to inadequate representation of relevant oceanic and atmospheric processes in climate models (e.g., Hawkins et al. 2014). These imperfections are largely attributed to either the limited understanding of many of the interactions and feedbacks in the climate system or to numerical oversimplifications of well-known processes, so-called parameterizations (Jun et al. 2008). One of the most serious errors shared by climate models is the strong warm sea-surface temperature bias in the south-eastern part of the tropical Atlantic (Flato et al. 2013; Zanchettin et al. 2017). Multiple causes have been identified at its origin, in different models, including local factors, such as the along-shore wind-stress and surface heat fluxes (e.g., Wahl et al. 2015; Milinski et al. 2016), and larger-scale or even remote phenomena, such as the propagation into the south eastern tropical Atlantic of downwelling anomalies generated at the equator (e.g. Toniazzo and Woolnough 2014).

Due to the severity of climate model biases, and their unavoidable impacts on the quality of the simulations, error identification, quantification and correction have become relevant topics of applied climate research (Cannon 2017). In general, current analytic approaches to evaluation and correction of coupled climate model errors determine how much the distributional properties of a climatically relevant quantity obtained from a climate simulation - or analogously from an ensemble of climate simulations - differ from those obtained from observational data for a certain time period and spatial domain (e.g., Jun et al. 2008; Liu et al. 2014). To this purpose, various statistical techniques have been proposed, including the empirical analyses of varying complexity (Richter and Xie 2008; García-Serrano et al. 2012) and bias estimation on a grid point by grid point basis (e.g., Boberg and Christensen 2012). Further, research interests on a Bayesian hierarchical assessment of climate model errors are increasing. The Bayesian paradigm allows quantifying systematic errors using full probabilistic inferences based on the posterior distributions derived from the proposed method. Recent studies focusing on the Bayesian estimation of climate model errors using spatially aggregated geophysical data includes Tebaldi et al. (2005), Buser et al. (2009) and Buser et al. (2010). More recently, Arisido et al. (2017) devised a purely spatial Bayesian hierarchical model using gridded data to determine the underlying spatial patterns in climate model biases, thus resolving the limitations of previous works that relied on spatial aggregation or grid-points separately.

In this paper, we develop a spatio-temporal model based on a Bayesian hierarchical approach in order to characterize and quantify climate model errors by explicitly accounting for their spatial and temporal dependencies within a single framework. Spatio-temporal characterization of climate model biases is motivated by the fact that such errors feature the same spatial and temporal complexity of the simulated climate itself, as both, climate and errors, stem and evolve based on the same numerical representation of physical processes (Zanchettin et al. 2017). To determine the spatial and temporal features of model errors, and their associated uncertainties, we represent the spatio-temporally referenced data using a set of weighting kernels (e.g., Higdon 1998) that capture the spatial variability efficiently while reducing the high dimensionality of the large-scale data. Our model specification is tailored to the well established state-space approach (Durbin and Koopman 2012), in which the spatio-temporal climate model error process is treated as a time series of non-stationary spatial fields, where space is assumed as continuous and time is discrete (Finley *et al.* 2012; Banerjee *et al.* 2014). We characterize the time-invariant bias in the mean state as the average portion of the spatio-temporally varying climate model error.

To illustrate our method, we present an analysis based on the case of annual-average near-surface air temperature for the period 1948-2005 over the southeastern tropical Atlantic and bordering area from a multi-model ensemble mean of historical simulations from the fifth phase of the Coupled Model Intercomparison Project (CMIP5, Taylor et al. (2012)). Focus on the ensemble mean allows reducing the complexity of the Bayesian treatment and attributing the temporal component of the error to the observed internal variability.

In the next section, we describe the data. In Section 3, we present the methodology, including the definition of climate model errors and our formulation of the Bayesian spatio-temporal model. Section 4 illustrates the results of the analysis. We provide a concluding discussion in Section 5.

2 Data

The dataset comprises observational data and climate model outputs. The latter are obtained from deterministic numerical models, and it is a common practice to consider the model output as data. We use monthly-mean data obtained from the NCEP reanalysis (Kalnay et al. 1996) as our observational reference data. Reanalysis data are the output of a state-of-the-art analysis/forecast system with data assimilation using past data from 1948 to the present. The data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA. Reanalysis data are therefore not direct observations, yet they facilitate the purposes of this study by providing gridded records of absolute temperatures. This is an advantage compared to other observational products that provide anomalies as main gridded output, such as the temperature series produced by the Climatic Research Unit of the University of East Anglia (Brohan et al. 2006). The use of pseudo-observations as reference target to determine systematic climate errors is discussed in Zanchettin et al. (2017) and Arisido et al. (2017). Our climate model outputs are originally based on monthly-mean data from an ensemble of six *historical* full-forcing climate simulations contributing to CMIP5. The data covers the period 1948-2005, for which we derive yearly-mean time series of both observations and simulations over the southeastern tropical Atlantic and bordering area, which is defined geographically as the region covering the latitude range 40°S to 0°N and the longitude range 20°W to 30°E.

3 Methods

3.1 Definitions and notations

Climate model error (hereafter referred to as deviation) is determined by comparing output data simulated from the climate models against observations. We let $Y_t(\mathbf{s})$ and $X_t(\mathbf{s})$ to represent the observed and the simulated value of a certain geophysical quantity, respectively, at spatial location $\mathbf{s}$, $\mathbf{s} \in \{\mathbf{s}_1,\dots,\mathbf{s}_n\}$ in a region $\mathscr{D} \in R^2$ and time t, $t \in \{1,\dots,T\}$. We derive the spatio-temporal climate model deviation as

$$D_t(\mathbf{s}) = Y_t(\mathbf{s}) - X_t(\mathbf{s}), \quad t = 1\dots,T \tag{1}$$

where $D_t(\mathbf{s})$ denotes the deviation of the simulated value relative to the observations at spatial location $\mathbf{s}$ and time t. For n spatial locations in $\mathscr{D}$, we observe the deviations $D_t(\mathbf{s}_1),\dots,D_t(\mathbf{s}_n)$ for the time t. Generally, statistical analysis of climate model deviations can be affected by the spatial misalignment between observations and model output since the model output and the observations are on different grids. We tackle this issue by linearly interpolating the model output data on the regular observational grid to ensure that $Y_t(\mathbf{s})$ and $X_t(\mathbf{s})$ are aligned on the same grid (see, e.g., Jun et al. 2008; Banerjee *et al.* 2014). One reason for using the linear interpolation method is that both reanalysis and climate model outputs feature high spatial resolution over the investigated domain. We therefore expect that the uncertainty due to the interpolation to minimally affects the results. For each year of the period 1948-2005, we consider a 19×19 $(n = 361)$ grid points. From the spatio-temporal deviation $D_t(\mathbf{s})$, we calculate the empirical *bias* $B(\mathbf{s})$ as $B(\mathbf{s}) = \sum_{t=1}^{T} D_t(\mathbf{s})/T$. In Figure 1(a), we show this spatially distributed $B(\mathbf{s})$, which is calculated by averaging $D_t(\mathbf{s})$ over the whole period 1948-2005. The spatial pattern of $B(\mathbf{s})$ exhibits the typical features of the climate model bias in the mean state over this study region, including the strong warm bias up to 5 kelvin over the Angola-Benguela front region. Another notable feature is the cold bias over the southeastern sub-tropics. Figure 1(b) shows the time series, $D(t)$, of the empirical deviation averaged over the considered spatial domain. The time series reflects the evolution of the deviation over the years, in which both short-term and long-term components highlight the portion of observed variability that is not captured by the ensemble-mean evolution. This includes, therefore, observed internal (i.e., spontaneous) variability, which is smoothed out in the ensemble mean. The long-term temporal evolution of $D(t)$ traces that of the Atlantic Multidecadal Oscillation (AMO), specifically its phase transitions in the 1970s (warm to cold) (e.g. Zanchettin et al. 2016).

3.2 Bayesian spatio-temporal model for climate model errors

The aim here is to formulate a statistical model to quantify and characterize climate model errors accounting for their inherent spatial and temporal dependencies. We specify the model in the Bayesian hierarchical framework based on three levels: *data*, *process*, and *parameters* (see, Berliner 2003; Cressie and Wikle 2015, for a comprehensive review). In this setup, our model specification is structured with (1) a data model describing the information given in the form of the empirically observed deviation, conditional on unobserved spatio-temporal deviation process under investigation; (2) the unobserved process featuring spatio-temporal characters described using a set of parameters and (3) the parameters that appear in the first two levels, and specify their prior beliefs according to Bayesian reasoning.

3.2.1 Data model

The idea is that in the evaluation of the bias $B(\mathbf{s})$ the local spatio-temporal effects should be filtered out. To model the deviation, we assume that the observed deviation $D_t(\mathbf{s})$ can be decomposed into two components:

$$D_t(\mathbf{s}) = M_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}), \tag{2}$$

where $M_t(\mathbf{s})$ is a spatio-temporal Gaussian random field and $\varepsilon_t(\mathbf{s})$ is a temporally and spatially uncorrelated zero mean Gaussian noise with variance σ_t^2. Note that the model is allowed to take into account for the heterogeneity in time. We assume that the noise component $\varepsilon_t(\mathbf{s})$ is independent of the deviation process $M_t(\mathbf{s})$. In practice we convey into the process $M_t(\mathbf{s})$ all smoothed spatio-temporal components that actually are blurred by the noise term. We further assume that the observed deviation $D_t(\mathbf{s})$ is conditionally independent in time given $M_t(\mathbf{s})$. Such assumptions lead to the data model in the form

$$[D_1(\mathbf{s}),\dots,D_T(\mathbf{s})|M_1(\mathbf{s}),\dots,M_T(\mathbf{s}),\sigma_1^2,\dots,\sigma_T^2] = \prod_{t=1}^{T}[D_t(\mathbf{s})|M_t(\mathbf{s}),\sigma_t^2] \tag{3}$$

where $[A]$ denotes the generic notation for the probability distribution of the random quantity A. Accordingly $[A|B]$ is the conditional distribution given B.

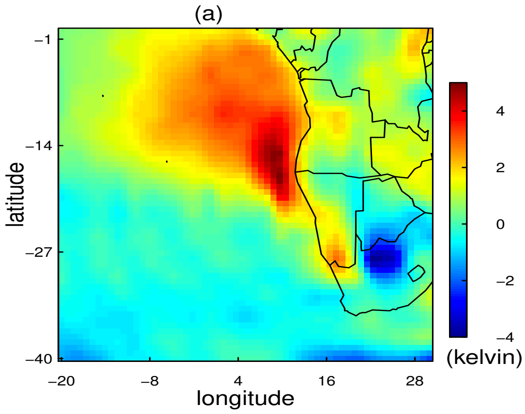

Fig. 1 (a) The empirical bias $B(\mathbf{s})$ in near-surface air temperature over the southeastern tropical Atlantic and bordering regions; (b) the temporal deviation $D(t)$, which is obtained by averaging $D_t(\mathbf{s})$ over the spatial domain.

3.2.2 Process model

The process model characterizes the spatio-temporal deviation process $M_t(\mathbf{s})$. Once we determine $M_t(\mathbf{s})$, an important interest will be to estimate the more appropriate time-invariant bias $\tilde{B}(\mathbf{s})$ for the study period as an average of $M_t(\mathbf{s})$, i.e., $\tilde{B}(\mathbf{s}) = \sum_{t=1}^{T} M_t(\mathbf{s})/T$. The spatio-temporal process $M_t(\mathbf{s})$ is driven by a large scale spatial component that changes stochastically but smoothly in time and a site specific component. The spatial large scale component at time t is represented by a linear combination of p spatial kernel functions $\{\psi_k(\mathbf{s}) : k = 1,\ldots,p\}$ as in Higdon (1998), i.e., $\sum_{k=1}^{p} \psi_k(\mathbf{s})\beta_{t,k}$, where $\beta_{t,k}$ is the coefficient parameter for kernel k. The whole formulation is given by

$$M_t(\mathbf{s}) = \psi(\mathbf{s})'\beta_t + \nu_t(\mathbf{s}) \tag{4}$$

$$\beta_t = \beta_{t-1} + \omega_t \tag{5}$$

$$\nu_t(\mathbf{s}) = \nu_{t-1}(\mathbf{s}) + \delta_t(\mathbf{s}) \tag{6}$$

where $\psi(\mathbf{s}) = \{\psi_1(\mathbf{s}),\ldots,\psi_p(\mathbf{s})\}'$ and $\beta_t = \left(\beta_{t,1},\ldots,\beta_{t,p}\right)'$. The number of kernels p is chosen to be much less than the number of spatial data points n. The choice of the kernels is further discussed in section 3.3. Equation (5) states that the $p \times 1$ vector of the linear coefficients β_t change according to a random walk process, where the evolution error ω_t is assumed as an independently and identically distributed zero mean Gaussian process with variance-covariance matrix Σ_ω. Then, equation (6) defines the site specific component $\nu_t(\mathbf{s})$ in order to account for the underlying spatial correlation, capturing it's Markovian dependence in time. More specifically, $\delta_t(\mathbf{s})$ follows a zero mean spatial Gaussian process with covariance function C_t, which is specified as $C_t(\mathbf{s},\mathbf{s}';\theta_t) = \tau_t^2\rho(\mathbf{s},\mathbf{s}';\phi_t)$, where $\theta_t = \{\tau_t^2,\phi_t\}$ and $\rho(.;\phi_t)$ is a correlation function with ϕ_t controlling the correlation decay and τ_t^2 representing the spatial variance. Any valid spatial correlation function can be used to define $\rho(.;\phi)$ (e.g., see Cressie 1993). Here we use the exponential function, i.e., $C_t(\mathbf{s},\mathbf{s}';\theta_t) = \tau_t^2\exp(-\phi_t||\mathbf{s}-\mathbf{s}'||)$, where $||\mathbf{s}-\mathbf{s}'||$ is the Euclidean distance between locations $\mathbf{s}$ and $\mathbf{s}'$. Further, for each time point t, ω_t is uncorrelated with $\varepsilon_t(\mathbf{s})$.

Of note, the variance of $M_t(\mathbf{s})$ at any time t is a function of the site $\mathbf{s}$, as can be shown by calculating it from equation (4). Similarly, the covariance of the deviation between any two sites is also a function of the sites. It follows that the deviation $M_t(\mathbf{s})$ at time t is a non-stationary spatial process. The different levels of the Bayesian hierarchical approach discussed above can be formulated within a state-space form (Gelfand et al. 2005; Durbin and Koopman 2012). That is, combining the data model (2) and the process models (4)-(6) yields

$$D_t(\mathbf{s}) = \psi(\mathbf{s})'\beta_t + \nu_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}) \tag{7}$$

$$\beta_t = \beta_{t-1} + \omega_t \tag{8}$$

$$\nu_t(\mathbf{s}) = \nu_{t-1}(\mathbf{s}) + \delta_t(\mathbf{s}) \tag{9}$$

where (7) is the measurement equation, and (8,9) are the transition equations. While (7) is similar to the measurement equation of the standard state space model, we recognize that assuming a random walk process in transition equations is a simplification from the more general specification (as provided in, e.g., West and Harrison 1997). The random walk is chosen to provide adequate flexibility for computation and eases the interpretation (e.g., Finley *et al.* 2012). Nontheless, the model can be extended to a more general specification, including higher order autoregressive structure.

3.2.3 Parameter model

We complete the model specification by assigning prior probability distributions for the initial conditions $\{\beta_0, v_0(\mathbf{s})\}$ and the model parameters $\{\Sigma_\omega, (\sigma_1^2, \theta_1), \ldots, (\sigma_T^2, \theta_T)\}$. Prior distributions for these parameters are generally taken to be non-informative. For the initial conditions, we specify a Guassian process prior in the form $\beta_0 \sim N(\mu_{\beta_0}, \Sigma_{\beta_0})$ where μ_{β_0} is a vector of length p and Σ_{β_0} is a $p \times p$ covariance matrix, and $v_0(\mathbf{s}) = 0$. Recalling $\theta_t = \{\tau_t^2, \phi_t\}$, for the measurement error variance σ_t^2 and the spatial variance τ_t^2 we assign the Inverse-Gamma priors $\sigma_t^2 \sim \mathrm{IG}(a_1, b_2)$ and $\tau_t^2 \sim \mathrm{IG}(a_2, b_2)$ for each t, where $\mathrm{IG}(a,b)$ denotes the inverse gamma distribution with shape parameter a and scale parameter b.

Here $\{\mu_{\beta_0}, \Sigma_{\beta_0}, a_1, b_1, a_2, b_2\}$ are called hyper-parameters in the Bayesian context, and their values could either be chosen or could be assigned another priors (see, e.g., Gelman 2006). Some physical intuition is used for these the parameters where relevant information is available. It should be noted that the space-time data is large, so the results are believed to be dominated by the data used, rather than the choice of the hyperparameters (Vanem et al. 2012). Indeed, a sensitivity analyses on some of the hyperparameters indicated that the results are not substantially sensitive to the choice of exact values.

We choose $\mu_{\beta_0} = 0, \Sigma_{\beta_0} = \mathbf{I}_p, a_1 = a_2 = 3$ and $b_1 = b_2 = 100$. For the spatial decay parameter ϕ_t of the exponential spatial correlation function, we assign the uniform prior in the form $\phi_t \sim \mathrm{U}(0.001, 0.03)$, which corresponds to the support ranges from 100 to 3000 km. Since the maximum distance between any two locations in the study region is 1030 km, the specified support well covers the full extent of the spatial domain. For the $p \times p$ evolution matrix Σ_ω, we assume the inverse-Wishart prior probability distribution, $\Sigma_\omega \sim \mathrm{IW}(p+1, \mathbf{I}_p)$, with the degrees of freedom parameter taking the value $p+1$ and the scale parameter being the $p \times p$ identity matrix $\mathbf{I}_p$, as we assume independence between the elements of the coefficient vector β_t.

3.3 Implementation

First we discuss the choice of the spatial kernel vector $\psi(s)$. Several types of kernel functions have been suggested, including Gaussian kernels (Stroud et al. 2001), harmonic functions (e.g., Furrer et al. 2007) and bisquare functions (Kang et al. 2012). In this paper we have considered a Gaussian kernel specified as

$$\psi_k(s) = \exp\{-(s - c_k)'\Sigma^{-1}(s - c_k)/2\}, \quad k = 1, \ldots, p \quad (10)$$

where c_k denotes the center of the kernel and Σ determines the shape. The number of kernels p, their locations and shapes must be chosen. These choices are often based on the presence of prior information such as smoothness and spatial dependence related to the spatial process under study (Stroud et al. 2001). If we choose spherically shaped kernels, i.e., $\Sigma = \kappa I_2$ on R^2 and $\kappa > 0$, and the centers belong to a regular grid over an unbounded domain, the resulting spatial process approximates a covariance function of a stationary isotropic process when the number of kernels p is very large. Alternatively, a geometrically anisotropic process may be obtained if we choose non-spherical Gaussian kernels. One way to assess the shape of Σ is to perform variogram analyses for different directions (see, e.g., Cressie 1993). Our preliminary analysis using variograms at several time points suggests that isotropy is a plausible assumption for $M_t(\mathbf{s})$. An example of the variogram plot for $t = 1970$ is shown in Figure 2(a) for the directions: $0°, 45°, 90°, 135°$ (i.e. North, Northeast, East and Southeast direction, respectively). The variogram does not reveal strong anisotropy in the four directions at small distances since the patterns are quite similar to each other. Figure 2(b) shows the $p = 25$ equally-spaced and spherically shaped Gaussian kernels with scale $\Sigma = 0.5 I_2$ on R^2 that are used in the main analysis. At the end of section 4 we further investigate the sensitivity of results for the different choices of p.

Once a reasonable choice of the kernels $\psi(s)$ is made, the model can be implemented in the Bayesian framework. For parameter estimation and associated inference, we seek to obtain the posterior distribution of the unknown parameters $\{\beta_0, \Sigma_\omega, (\beta_1, \sigma_1^2, \theta_1), \ldots, (\beta_T, \sigma_T^2, \theta_T)\}$. For a particular location $\mathbf{s}$, the posterior distribution can be given in the form

$$\begin{aligned}[\beta_0, \beta_{1:T}, \Sigma_\omega, \sigma_1^2, \theta, \ldots, \sigma_T^2, \theta_T | D_{1:T}(\mathbf{s})] \propto \\ \prod_{t=1}^{T} [D_t(\mathbf{s}_i) i = 1, \ldots, n | \beta_t, \sigma_t^2] \times [\beta_0] \times \\ \prod_{t=1}^{T} [\beta_t | \beta_{t-1}, \Sigma_\omega] \times \prod_{t=1}^{T} [\sigma_t^2] \times \prod_{t=1}^{T} [\theta_t] \times [\Sigma_\omega]\end{aligned} \quad (11)$$

with notations as in Cressie and Wikle (2015). Clearly, the normalizing constant for (11) cannot be found analytically. So, we use the Markov Chain Monte Carlo (MCMC) method (Gilks et al. 1996) with Gibbs sampler and random walk Metropolis steps (Robert and Casella 2013). For the random walk Metropolis step, a multivariate normal (same

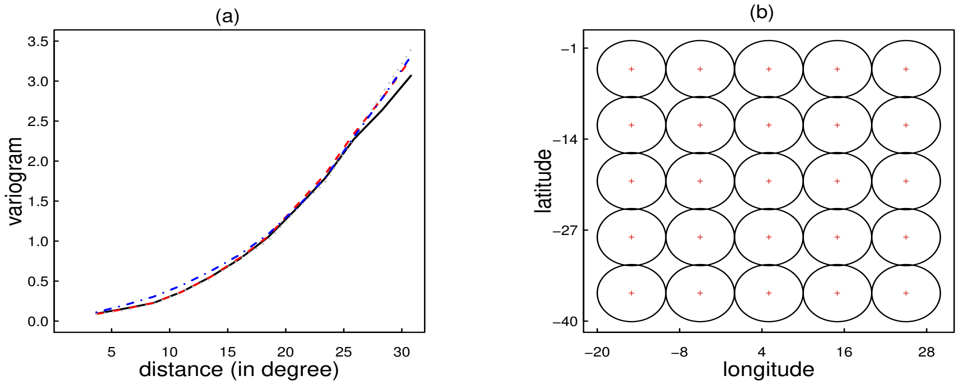

Fig. 2 (a) Empirical variogram for the time $t = 1970$ for the four different directions (black solid 0°, red dashed 45°, gray dotted 90°, blue dashed 135°). The variogram was analyzed using the robust estimator by Cressie (1993); (b) The spherically-shaped 25 equally-spaced Gaussian kernels used for the main analysis. Red crosses indicate the centers of the kernels.

dimension as the number of model parameters) proposal distribution is used. Based on inspection of graphical tools of the simulation history to assess convergence, we run the Gibbs sampler for 10,000 simulation steps and discarding the first 5,000 as the burn-in period. We performed the analysis using the *spBayes* package (Finley *et al.* 2015) in the freely available R computing environment. The computation time depends mainly on the size of the kernel vectors, the spatial coverage and the number of time points. For our main analysis, a Gaussian kernel vector with length 25, a regular grid of $19 \times 19 = 361$ sites and $T = 58$ years, the computations take about 15 hours on a 64-bit linux workstation version Linux Mint 18.2. We then summarized draws from the posterior MCMC in terms of mean, median and standard deviation to perform posterior inference about the unknowns.

4 Results

Figure 3 shows the posterior means of the deviation process $M_t(\mathbf{s})$ for the years $t \in \{1950, 1960, 1970, 1980, 1990, 2000\}$. The posterior means are estimated using the 25 Gaussian kernels that are shown in Figure 2(b). These results corroborate the purely spatial results of Arisido et al. (2017) where a broader tropical Atlantic region was considered. The most prominent feature is that the warm error over the Angola-Benguela front region persists throughout the simulated period, with the maximum value exceeding 4 kelvin and extending westward beyond 10°W. However, the severity of the climate model error estimates is noticeably different across the years, with differences in the local deviation of more than 1 kelvin (e.g., between 1980 and 1950). The shown exemplary posterior spatial fields reflect an (inter)decadal modulation of the warm error over the southeastern tropical Atlantic, with alternating decades of strong (roughly 1955-1965, 1980s and 1990s) and moderate (late 1940s-early 1950s, 1970s and early 2000s) errors. Further, substantial variations through time are observed in the severity of the warm error extending southeastern over the Namibia desert.

The corresponding uncertainty estimates of the posterior $M_t(\mathbf{s})$ are shown in Figure 4, which indicate that the posterior estimates of $M_t(\mathbf{s})$ are most uncertain in the regions affected by cold errors and, more generally, they are more uncertain over ocean than over the land. Particularly, uncertainty is largest in the West-tropics with maximum standard error reaching 0.8, which is more pronounced in the period 1960, 1970 and 1990. The posterior estimates of $M_t(\mathbf{s})$ are, conversely, more certain in regions affected by warm errors particularly over the Angola-Benguela front region, where the minimum standard error is estimated to be about 0.2. This ocean-land contrast reflects topographic effects and the different spatio-temporal scales of characteristic ocean and land processes.

As pointed out in section 3.2.2, the posterior estimate of the bias $\tilde{B}(\mathbf{s})$ is obtained as an average of the posterior $M_t(\mathbf{s})$. Figure 5 presents the posterior $\tilde{B}(\mathbf{s})$ (panel a) and its associated uncertainty estimate (panel b). Overall, the posterior estimate shows the obvious warm bias along the Angola-Benguela front region. We notice that the posterior

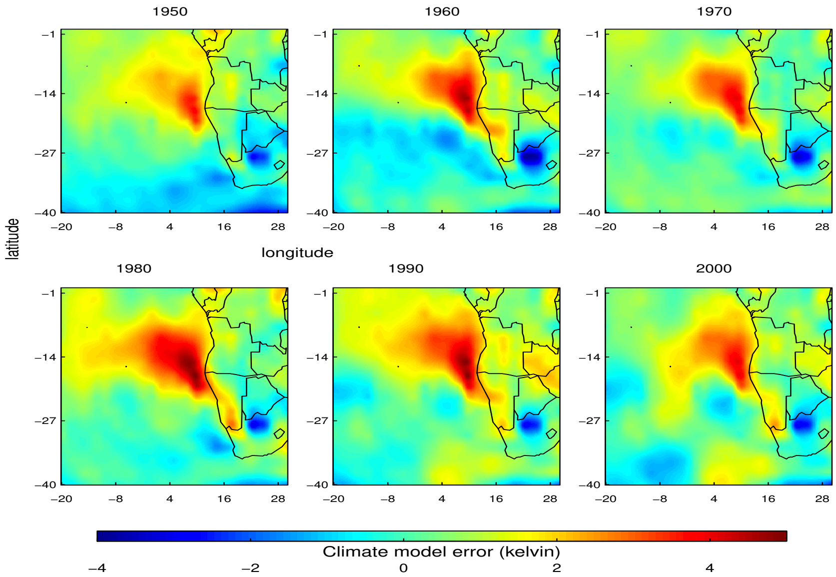

Fig. 3 The posterior means of the spatio-temporal deviation process $M_t(\mathbf{s})$ over the southeastern tropical Atlantic and bordering regions for the years $t \in \{1950, 1960, 1970, 1980, 1990, 2000\}$.

bias $\tilde{B}(\mathbf{s})$ agrees well in its general features, with the empirical bias estimate $B(\mathbf{s})$ (Figure 1a), which implies that our model captures the most prominent features in the data. In particular, both $B(\mathbf{s})$ and $\tilde{B}(\mathbf{s})$ capture the warm error over the Angola-Benguela front region, particularly including its meridionally elongated core around 17°S off the African coast and its elongated shape protruding west along 14°S latitude. Nonetheless, the Bayesian spatio-temporal approach allows to gain deeper insights about the climate model error, in particular concerning the spatial dependency of the diagnosed features, and the associated posterior uncertainty estimation. The fact that physically plausible features emerge in Figure 5(a), including sharp coastal effects and the signature of oceanic waves, manifests about the detail and quality of the spatial bias estimation allowed by the proposed statistical model. Furthermore, the posterior estimates of uncertainty (Figure 5b) highlight regions where the quantification of the bias is less certain. In particular, the small bias in the more equatorial regions of the south Atlantic Ocean are affected by large uncertainty, whereas the large bias over the Angola-Benguela front region is associated to small uncertainty.

To determine the overall temporal character of the spatio-temporal process $M_t(\mathbf{s})$, Figure 5(c) shows the posterior temporal deviation $M(t)$ as spatially averaged $M_t(\mathbf{s})$ overlay on the corresponding observed deviation $D(t)$. The posterior deviation $M(t)$ appears to be smaller than the corresponding observed deviation $D(t)$. Furthermore, the posterior estimate has a smoother evolution compared to that of the empirical deviation, which in turn suggests that the variability in the empirical deviation is greater than the variability in the estimate of the posterior deviation. To further assess the posterior estimate of the spatio-temporal process $M_t(\mathbf{s})$ for more localized features within the study domain, the upper panel in Figure 6 depicts the time series trends of the posterior averages of $M_t(\mathbf{s})$ for four subregions, whose locations are indicated in Figure 6 lower panel. The four subregions were selected for illustrating the evolution of $M_t(\mathbf{s})$ over: the Angola-Benguela front region, the Namibia desert and two open ocean regions at the northern and southern edges of the south Atlantic gyre, respectively. A trend in a subregion is calculated by averaging the posterior information over the spatial domain of the subregion, where the spatial domain of the subregions are not necessarily equal. The time series of the local deviations indicate that a sharp transition from a cold bias to a neutral bias is diagnosed over the southern edge of the south Atlantic gyre, where the cold bias is most severe in the decade 1950-1960 of the analysis (Trend 3). Over the northern edge of the gyre, the temporal evolution shows a slow transition, from the late 1960s to the early 1990s, from a warm bias to a neutral bias situation (Trend 4). As expected, the Angola-Benguela front region features the highest warm bias over the whole period of study (Trend 1), consisting of a long-term warming trend with superimposed noticeable decadal variability. Similarly, the Namibia desert experiences a relatively warm bias (Trend 2), indi-

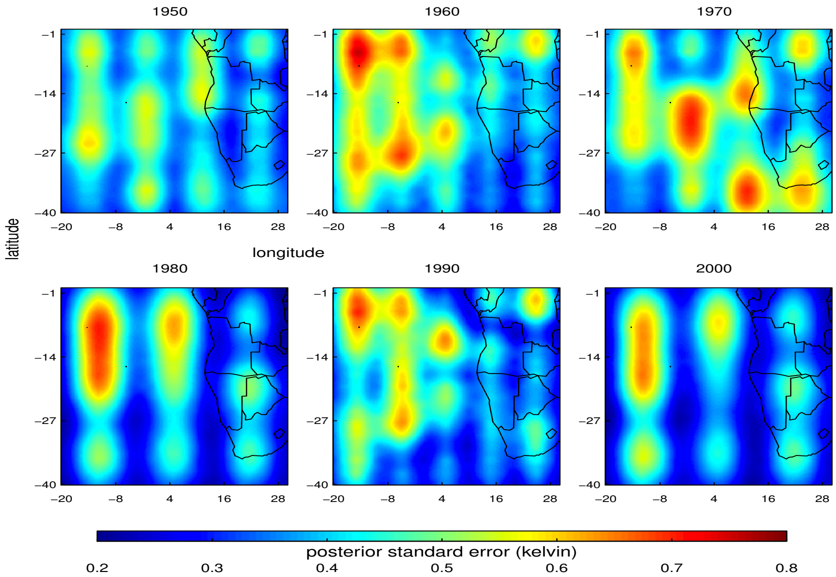

Fig. 4 Posterior standard errors associated to the posterior mean fields of $M_t(\mathbf{s})$, which are shown in Figure 3.

cating a long-term warming with superimposed noticeable decadal variability as well.

In Figure 7 we show the posterior medians and the associated 95% credible end points for the variance components $\{\sigma_t^2, \tau_t^2\}$. The evolutions of the variance of the observed deviation σ_t^2 (i.e., the nugget in the geostatistical term) and the spatial variance τ_t^2 exhibit temporal variability, validating our assumption to define time dependent variance parameters to take into account the heterogeneity in time of these parameters. Additionally, we can see that the spatial variance is greater than the nugget. In fact, the signal-to-noise ratio, which is computed as τ_t^2/σ_t^2 for comparing the strength of the two variance components, is substantially greater than one (not shown), coherent with the hypothesis that the nugget effect is often smaller than the spatial variance (e.g., Bakar and Sahu 2015). We note that the variance parameters here are allowed to change in time, but not according to a specific model structure. An interesting idea for further study is to explore the possibility of specifying the temporal dependence in terms of a variance model, such as the generalized autoregressive conditional heteroskedasticity (ARCH) oriented approach.

Our model specification depends on the use of Gaussian kernel functions to describe the spatial features of the deviation process $M_t(\mathbf{s})$. We therefore investigate the adequacy of our model to the choice of Gaussian weighting kernels. The parameters $\{p, \Sigma, \kappa\}$ associated to the kernels may impact the model fit and the prediction. In particular, the choice of p largely determines the level of spatial detail in the context of dimension reduction techniques (e.g., Finley *et al.* 2012; Arisido et al. 2017). Hence, we perform a sensitivity analysis on the p parameter using three different sets of kernels, that is $p \in \{9, 18, 36\}$ fixing $\Sigma = 0.5\mathbf{I}_2$, to investigate the sensitivity of the results to these choices of p. Figure 8 shows the three sets of kernels, along with the corresponding posterior fields of the deviation process $M_t(\mathbf{s})$. The three different sets of kernels are shown in column (a). Noticeable differences emerge in the shape of $M_t(\mathbf{s})$ (column b) including the location and magnitude of the deviation. With $p = 9$, the larger separation between the kernels results in a strongly smoothed posterior estimate. Clearly, the pattern also misses detailed spatial features and misrepresents the deviation along the Angola-Benguela front, a region known to be affected by a strong warm bias. This suggests that a too small number of kernels insufficiently represents the spatial processes. With larger numbers of kernels, $p = 18$ and $p = 36$, the deviation process $M_t(\mathbf{s})$ captures well know features and produce detailed patterns with clearly apparent topographic characteristics. The fact that both $p = 18$ and $p = 36$ choices lead comparable posterior estimates of $M_t(\mathbf{s})$ indicates that the optimal choice of number of kernel would be between these two choices. Specifically, a choice closer to $p = 36$ allows a better approximation of the deviation process by capturing fine-scale local features, but the benefit being gained has to be balanced with computational feasibility and the applicability of the model with large number of kernels.

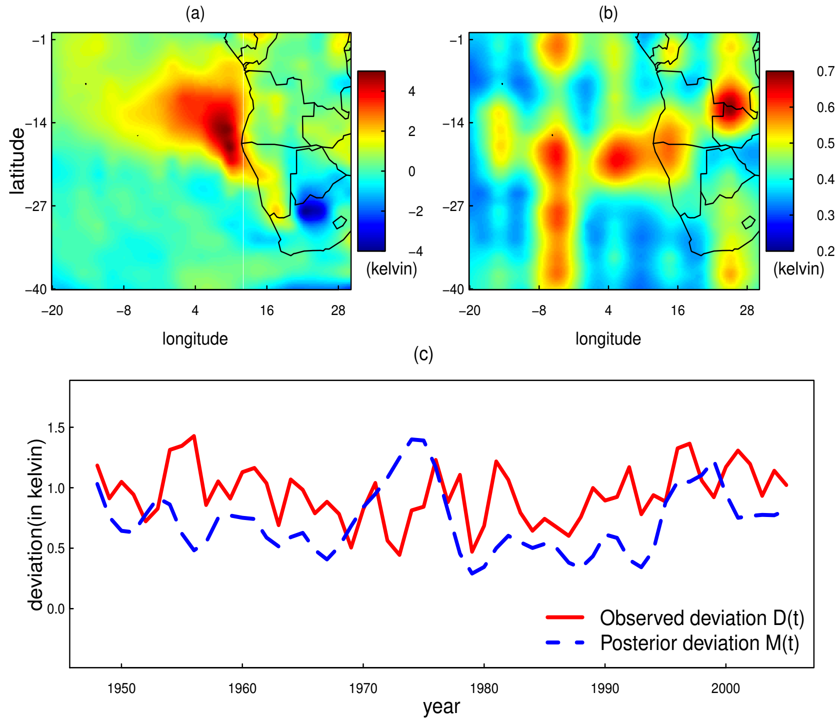

Fig. 5 (a) The posterior bias $\tilde{B}(\mathbf{s})$; (b) associated posterior standard error estimate; (c) time series plots of the posterior deviation $M(t)$ obtained as spatially averaged $M_t(\mathbf{s})$, overlay on the corresponding empirical deviation $D(t)$.

We also analyzed the predictive performance of the model to assess the goodness of the fit. In Figure 9(a) we show a residual, $D_t(\mathbf{s}) - M_t(\mathbf{s})$, surface plot at one randomly selected year. We observe that values of the residual surface plot varies from -0.3 to 0.3 kelvin and in most places the fitted values are close to the observations, particularly over the ocean. The largest discrepancies occur over land, in regions of strong climatic heterogeneity, where the residuals take the form of warm-cold dipoles. Figure 9(b) shows the observation against the posterior median of the time-varying fitted values together with the 95% credible intervals for 10 randomly selected locations. Again the fitted values are close to the observations. In fact both the observations and the fitted values lie within the 95% credible bands.

Finally, we assessed the sensitivity of the results to the assumption of isotropy for the spatial covariance structure of the deviation process $M_t(\mathbf{s})$. The empirical variogram in Figure 2 showed that treating $M_t(\mathbf{s})$ as an isotropic process was valid as no anisotropy was revealed. Nonetheless, when there is concern of strong anisotropy, it may be desirable to build a model that is able to handle such feature directly. There are various ways to address anisotropy (e.g., Higdon 1998; Banerjee *et al.* 2014). In the context of the current method, we can account for anisotropy by defining the 2×2 kernel covariance matrix Σ as diagonal where the first and the second diagonal elements are 0.8 and 0.3, respectively. Figure 9(c) and (d) show the $p = 25$ Gaussian kernels with this modified Σ and the corresponding posterior bias $\tilde{B}(\mathbf{s})$, respectively. The impact of the modified Σ is evident as the shape of the kernels is elliptical rather than spherical. Yet, the effect of these spherical kernels on the posterior estimate of $\tilde{B}(\mathbf{s})$ is less clear, since this posterior estimate is practically undistinguishable from the bias estimate in Figure 5. This supports the variogram analysis that considering $M_t(\mathbf{s})$ as an anisotropic process does not provide relevant additional advantages over the asumption of isotropy.

5 Discussion

We have proposed a Bayesian spatio-temporal model for assessing systematic errors in coupled climate models. A key feature of the work presented here is that the statistical model does not only quantify the errors by accounting for their spatial and temporal dependencies, but also determines the associated uncertainties using the posterior distributions. Spatio-temporal errors are characterized as non-stationary spatial fields over a discrete period of time, and the time-invariant bias in the mean state is estimated as the temporal average portion of the spatio-temporally varying climate model error.

The model was illustrated using the case of near-surface air temperature over the southeastern tropical Atlantic and

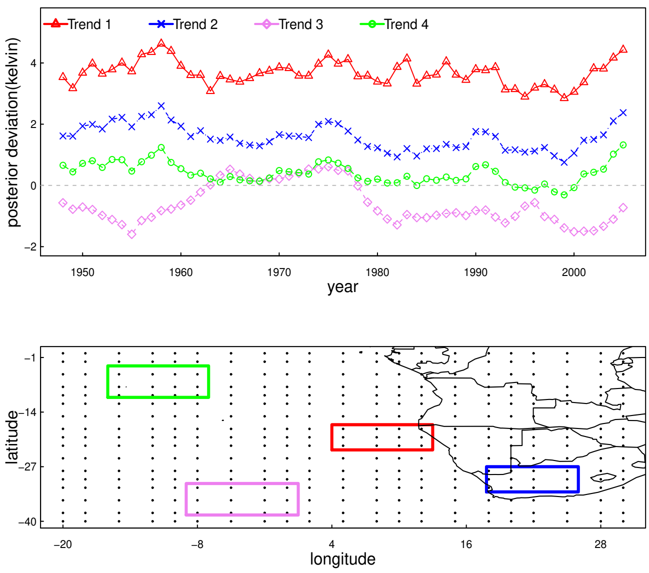

Fig. 6 Time series plots of the posterior average of $M_t(\mathbf{s})$ for selected four different subregions (upper panel) and the locations of the four different subregions (lower panel).

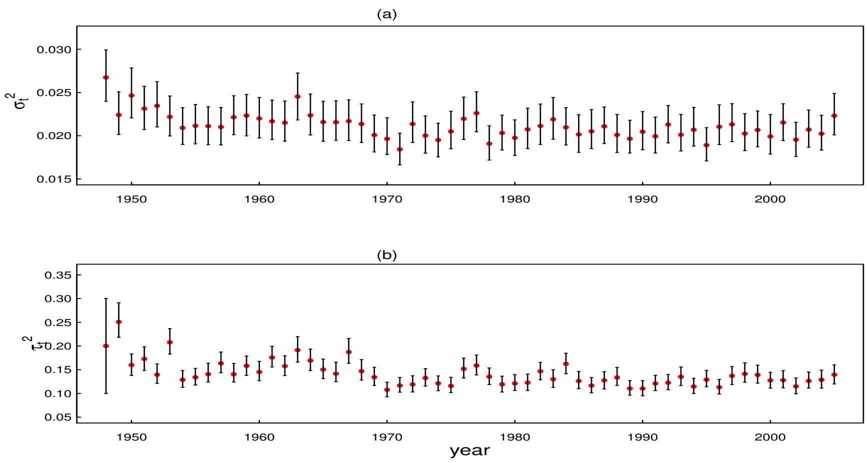

Fig. 7 Posterior median and 95% credible end points for time-varying variance parameters: (a) the variance of the observed data σ_t^2; (b) the spatial variance τ_t^2.

bordering regions from an ensemble average of six historical simulations contributing to CMIP5. Substantial warm error is estimated over the Angola-Benguela front region, persisting throughout the simulated period, but with noticeable decadal variations with amplitude of up to one kelvin. The posterior analysis showed that not only the estimate of the bias changes through time, but also the associated uncertainty. Another notable feature of the results is that the posterior overall temporal evolution in the investigated domain is smaller than the corresponding empirical estimate (see, Figure 5c). This is due to the fact that our statistical approach quantifies the error process by disentangling the noise component linked to the data, particularly those linked to the intrinsic interannual variability of the climate system (driven, for instance, by phenomena such as the El Nino-Southern Oscillation), and accounting for the underlying spatial correlation. The generality of the approach presented in this paper suits for the estimation of unknown quantities as well as their prediction for different spatial sites or forecast periods. The conditional dependency on the state at the previous time step allows for a straightforward extension of the model to the purpose of error forecasting. In particular, the use of long

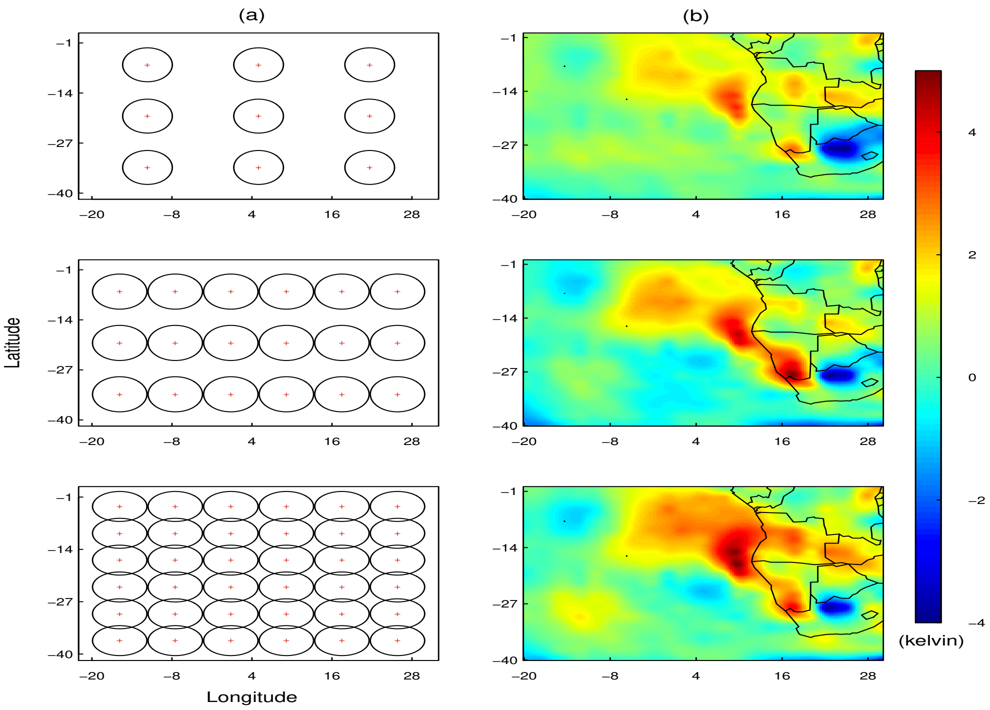

Fig. 8 Comparison of the posterior mean surfaces of the spatio-temporal deviation $M_t(\mathbf{s})$ at year $t = 1985$ for three different choices of the number of kernels p: (a) map of the employed sets of Gaussian weighting kernels; (b) the posterior mean surfaces of $M_t(\mathbf{s})$.

(spanning several decades) time series allows to obtain precise forecasts with an interannual-to-decadal horizon (West and Harrison 1997).

The proposed statistical model stimulates additional research, posing theoretical and computational challenges. We considered an ensemble average of climate simulations to be representative of climate simulation performances in the study region. A more comprehensive analysis can be envisaged in the form of a multivariate spatio-temporal oriented approach to allow assessments of spatio-temporal simulation errors from several climate models jointly. Further, we considered the exponential function based on Euclidean distance to specify the covariance function used to model the spatial dependence structure. Another future focus is a more flexible spatial covariance function and distance metric for a broader spatial region and a longer time period.

Acknowledgements The research leading to these results has received funding from the European Union, Seventh Framework Programme (FP7/2007-2013) under Grant agreement n 603521 - PREFACE.

Conflict of Interest: The authors declare that they have no conflict of interest.

References

Arisido MW, Gaetan C, Zanchettin D, Rubino A (2017) A Bayesian hierarchical approach for spatial analysis of climate model bias in multi-model ensembles. Stochastic Environmental Research and Risk Assessment 31:2645-2657

Bakar KS, Sahu SK (2015) sptimer: Spatio-temporal Bayesian modelling using R. Journal of Statistical Software 63:1-32

Banerjee S, Carlin BP, Gelfand AE (2014) Hierarchical modeling and analysis for spatial data. CRC Press, New York

Berliner LM (2003) Physical-statistical modeling in geophysics. Journal of Geophysical Research (Atmospheres) 108:8776. doi:10.1029/2002JD002865

Boberg F, Christensen JH (2012) Overestimation of Mediterranean summer temperature projections due to model deficiencies. Nature Climate Change 2:433-436

Brohan P, Kennedy JJ, Harris I, Tett SF, Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. Journal of Geophysical Research: Atmospheres 111. doi:10.1029/2005JD006548

Buser CM, Knsch HR, Lthi D, Wild M, Schr C (2009) Bayesian multi-model projection of climate: bias as-

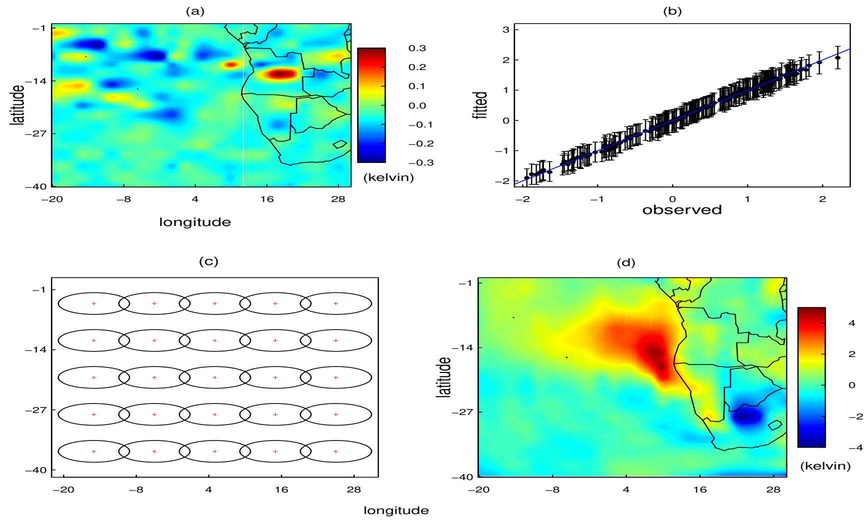

Fig. 9 (a) A residual surface plot at one randomly selected year, (b) observation against the posterior median of the fitted values together with the 95% credible intervals for 10 randomly selected locations, (c) Elliptically-shaped 25 Gaussian kernels with the red crosses indicating the centers of the kernels, and (d) The posterior bias $\tilde{B}(\mathbf{s})$ estimated using the kernels in (c).

sumptions and interannual variability. Climate Dynamics 33:849-868

Buser CM, Knsch HR, Weber A (2010) Biases and uncertainty in climate projections. Scandinavian Journal of Statistics 37:179-199

Cannon AJ (2017) Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. Climate Dynamics. doi:10.1007/s00382-017-3580-6

Cressie N (1993) Statistics for spatial data. Wiley, New York

Cressie N, Wikle CK (2015) Statistics for Spatio-temporal Data. John Wiley and Sons, New Jersey

Durbin J and Koopman, SJ (2012) Time Series Analysis by State Space Methods. Oxford University Press, Oxford

Finley AO, Banerjee S, Gelfand AE (2012) Bayesian dynamic modeling for large space-time datasets using Gaussian predictive processes. Journal of Geographical Systems 14:29-47

Finley AO, Banerjee S, Gelfand AE (2015) spBayes for large univariate and multivariate point-referenced spatio-temporal data models. Journal of Statistical Software 63:1-24

Flato G, Marotzke J, Abiodun B, Braconnot P, Chou SC, Collins W, Cox P, Driouech F, Emori S, Eyring V, Forest C, Gleckler P, Guilyardi E, Jakob C, Kattsov V, Reason C, Rummukainen M (2013) Evaluation of climate models. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Furrer R, Sain SR, Nychka D, Meehl GA (2007) Multivariate Bayesian analysis of atmosphereocean general circulation models. Environmental and Ecological Statistics 14:249-266

García-Serrano J, Doblas-Reyes FJ (2012) On the assessment of near-surface global temperature and North Atlantic multi-decadal variability in the ENSEMBLES decadal hindcast. Climate dynamics 39:2025-2040

Gelfand AE, Banerjee S, Gamerman D (2005) Spatial process modelling for univariate and multivariate dynamic spatial data. Environmetrics 16:465-479

Gelman A (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Brown and Draper). Bayesian Analysis 1:515-534

Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov chain Monte Carlo in Practice. Chapman and Hall, London

Hawkins E, Dong B, Robson J, Sutton R, Smith D (2014) The interpretation and use of biases in decadal climate predictions. Journal of Climate 27:2931-2947

Higdon D (1998) A process-convolution approach to modelling temperatures in the North Atlantic Ocean. Environmental and Ecological Statistics 5:173-190

Hooten MB, Wikle CK (2008) A hierarchical Bayesian nonlinear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove.

Environmental and Ecological Statistics 15:59-70
Jun M, Knutti R, Nychka DW (2008) Spatial analysis to quantify numerical model bias and dependence: how many climate models are there?. Journal of the American Statistical Association 103:934-947
Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y (1996) The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77:437-471
Kang EL, Cressie N, Sain SR (2012) Combining outputs from the North American regional climate change assessment program by using a Bayesian hierarchical model. Journal of the Royal Statistical Society: Series C (Applied Statistics) 61:291-313
Liu M, Rajagopalan K, Chung SH, Jiang X, Harrison J, Nergui T, ..., Choate J (2014) What is the importance of climate model bias when projecting the impacts of climate change on land surface processes?. Biogeosciences 11:2601-2622
Milinski SJ, Bader H, Haak AC, Siongco J, Jungclaus H (2016) High atmospheric horizontal resolution eliminates the wind-driven coastal warm bias in the southeastern tropical Atlantic. Geophysical Research Letter 43
Richter I, Xie SP (2008) On the origin of equatorial Atlantic biases in coupled general circulation models. Climate Dynamics 31:587-598
Robert C, Casella (2013) Monte Carlo Statistical Methods. Springer, New York
Stroud JR, Mller P, Sansó B (2001) Dynamic Models for Spatiotemporal Data. Journal of the Royal Statistical Society. Series B, Statistical Methodology 63:673-689
Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93:485-498
Tebaldi C, Smith RL, Nychka D, Mearns LO (2005) Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multi-model ensembles. Journal of Climate 18:1524-1540
Toniazzo T, Woolnough S (2014) Development of warm SST errors in the southern tropical Atlantic in CMIP5 decadal hindcasts. Climate Dynamics 43:2889-2913
Vanem, E., Huseby, A. B. and Natvig, B (2012) A Bayesian hierarchical spatio-temporal model for significant wave height in the North Atlantic. Stochastic environmental research and risk assessment 26:609-632
Wahl S, Latif M, Park W, Keenlyside N (2015) On the Tropical Atlantic SST warm bias in the Kiel Climate Model. Climate Dynamics 36:891-906
Wang C, Zhang L, Lee SK, Wu L, Mechoso CR (2014) A global perspective on CMIP5 climate model biases. Nature Climate Change 4:201-205
West M, Harrison PJ (1997) Bayesian Forecasting and Dynamic Models. Springer, New York
Zanchettin D, Bothe O, Rubino A, Jungclaus JH (2016) Multi-model ensemble analysis of Pacific and Atlantic SST variability in unperturbed climate simulations. Climate Dynamics 47:1073-1090
Zanchettin D, Gaetan C, Arisido MW, Modali K, Toniazzo T, Keenlyside N, Rubino A (2017) Structural decomposition of climate model uncertainties: A Bayesian approach. Scientific reports 7: 12862-12873