SCIENTIFIC REPORTS

OPEN

Structural decomposition of decadal climate prediction errors: A Bayesian approach

Received: 6 July 2017
Accepted: 18 September 2017
Published online: 09 October 2017

Davide Zanchettin[1], Carlo Gaetan[1], Maeregu Woldeyes Arisido[1], Kameswarrao Modali[2], Thomas Toniazzo[3], Noel Keenlyside[4] & Angelo Rubino[1]

Decadal climate predictions use initialized coupled model simulations that are typically affected by a drift toward a biased climatology determined by systematic model errors. Model drifts thus reflect a fundamental source of uncertainty in decadal climate predictions. However, their analysis has so far relied on ad-hoc assessments of empirical and subjective character. Here, we define the climate model drift as a dynamical process rather than a descriptive diagnostic. A unified statistical Bayesian framework is proposed where a state-space model is used to decompose systematic decadal climate prediction errors into an initial drift, seasonally varying climatological biases and additional effects of co-varying climate processes. An application to tropical and south Atlantic sea-surface temperatures illustrates how the method allows to evaluate and elucidate dynamic interdependencies between drift, biases, hindcast residuals and background climate. Our approach thus offers a methodology for objective, quantitative and explanatory error estimation in climate predictions.

The multiannual forecast horizon inherent to decadal climate prediction requires that the complexity and uncertainties arising from the interaction between the climate response to external forcing and the evolution of internal modes of climate variability are accounted for[1,2]. To generate an ensemble of decadal forecasts, various procedures are applied in climate models to obtain sets of initial conditions based on the observed state of the climate system at a certain time[2]. The observed evolution of many geophysical parameters however resides outside the climatological envelope simulated by coupled climate models used in contemporary climate prediction systems, due to the presence of large systematic model biases with respect to observations[3–5]. These affect all of mean state, seasonal cycle and interannual internal variability. Decadal climate forecasts based on full-field initialization therefore unavoidably include a growing systematic error, which corresponds with the adjustment of the simulations from the assimilated state drawn from the observed climatology towards a state consistent with the biased climatology of the model[2]. This signal is commonly referred to as climate model drift[6].

Model drifts and biases can result from the erroneous representation of oceanic and atmospheric processes in climate models[1,4,7–11], but more generally they reflect our limited understanding of many of the interactions and feedbacks in the climate system and approximations and simplifications inherent to the numerical representation of climate processes (so-called parameterizations). Sea-surface temperature (SST) biases are often central in the characterization of climate model biases[3]. Typical examples of systematic model bias are the warm and weak upwelling systems at the eastern boundaries of the tropical oceans[12–14]. Among these, the severe warm SST bias in the southeastern tropical Atlantic has been especially studied[5,12,15–17].

A multi-model analysis of SST drifts in this region highlighted that similar biases arise there in different models through a variety of locally as well as remotely growing error mechanisms[12]. The dynamical processes governing model drifts (or their numerical representation) are identical to those of the simulated climate itself, so that climate and error signals are indistinguishable by first principles. The fact that model biases and drifts often have a prominent seasonal character[11,16] further increases the difficulty of evaluating, quantifying and understanding their impacts on the overall quality of simulated climates. Various techniques for drift mitigation have been proposed, such as anomaly initialization or flux correction[18]. Nonetheless, in decadal climate predictions model drifts

[1]University Ca'Foscari of Venice, Dept. of Environmental Sciences, Informatics and Statistics, Via Torino 155, 30170, Mestre Venezia, Italy. [2]Max Planck Institute for Meteorology, Bundesstrasse 53, 20146, Hamburg, Germany. [3]Uni Research, Bjerknes Centre for Climate Research, Bergen, Norway. [4]Geophysical Institute, University of Bergen, Postboks 7803, 5020, Bergen, Norway. Correspondence and requests for materials should be addressed to D.Z. (email: davide.zanchettin@unive.it)

remain most often regarded as mere biases, which are often corrected for by means of empirical techniques with varying complexity[19–22], but rarely investigated in depth[23].

A substantial improvement in our understanding of mechanisms contributing to the generation and propagation of systematic decadal climate prediction errors may be achieved by means of a process-oriented statistical characterization of their temporal and spatial complexity and a robust quantification of associated uncertainties. Here, we propose a structural decomposition analysis[24] of systematic hindcast errors (i.e., systematic discrepancies between values from "retrospective predictions" and the corresponding observation) in an ensemble of decadal climate predictions. Structural decomposition refers to a process model in which temporal changes of an observable variable are attributed to the discriminant effects (or "driving forces") underlying these changes. Accordingly, each time series of hindcast errors $D_j(t)$ in the ensemble is represented as the sum of a systematic component $\Delta(t)$, which captures the (systematic) portion of hindcast errors common to all realizations, and a non-systematic irregular component $\varepsilon_j(t)$, which is specific to the individual considered hindcast realization j. Here, $\Delta(t)$ is defined as a linear combination of systematic hindcast error $\delta(t)$ and systematic seasonal bias $\sigma(t)$. In particular, $\delta(t)$ evolves as a stochastic trend with slope $\beta(t)$: initially, when $\beta(t) \neq 0$, $\delta(t)$ describes the drift; then, when $\beta(t) \approx 0$, $\delta(t)$ corresponds to the climatological bias. Moreover, $\Delta(t)$ can incorporate the explanatory effects of a covariate $X(t)$ representing local and/or remote processes and quantified by the associated stochastic regression coefficient $\gamma(t)$ of $\Delta(t)$ on $X(t)$. As discriminant effects are non-observable, a state-space model is built in which the state vector includes all unobserved elements and the transition matrix describes their individual dynamics (see the methods section for details).

The variances of the error terms in the state-space model are unknown parameters, and inference is performed following a Bayesian approach that specifies prior distributions for them. Therefore, our modeling strategy consists of three hierarchical levels (see the methods section for details): the data level, where observational errors, $\varepsilon_j(t)$, for all hindcasts are accounted for conditionally on the systematic component $\Delta(t)$; the process level, where the systematic components of the bias, $\delta(t)$ and $\sigma(t)$, and its constituting discriminant effects, $X(t)$, are represented by means of a stochastic model; and the parameter level, where prior knowledge about the distributions of model errors (via the parameters τ^2) is formalized.

In this contribution, we choose to illustrate our hierarchical statistical framework and its value for a robust quantification of systematic model prediction errors and their associated uncertainty by means of a specific example. We discuss the diagnosed distinctive characteristics of the different error components and use them for a reliable identification of connections between drift dynamics and the physics of the simulated climate. This study thus focuses on SST errors in the Tropical and South Atlantic Ocean estimated for an ensemble of decadal hindcasts initialized from observations, with the MiKlip prototype system for decadal climate prediction[25] (see Methods).

Results

We consider the spatially averaged SST in a region of the south-eastern tropical Atlantic containing the Angola-Benguela frontal zone (SST_{ABF}, the region is defined in the methods section). Empirical hindcast errors consist in large seasonal errors superposed on a warm climatological mean bias (grey lines in Fig. 1a). The posterior marginal distributions, i.e., the posterior distributions of the individual systematic error components obtained by the structural decomposition, characterize three major drift phases: an initial strong warming ($\beta(t) > 0$) in the first two hindcast years, which peaks at ~4 °C; a subsequent progressive weak cooling ($\beta(t) < 0$), which extends into the 7th hindcast year; and finally a transition into climatological bias conditions ($\beta(t) \approx 0$) quantified as a bias of 3.12 [3.00, 3.23] °C – estimated as median and 5th–95th percentile range of $\delta(t = 90 \ldots 120)$. Annual and semiannual seasonal biases – $\sigma^A(t)$ and $\sigma^{SA}(t)$ – with amplitudes of ~1.3 °C and ~1.05 °C, respectively, interfere constructively in July/August and destructively in January/February. Further, the semiannual bias component is substantially damped during the first hindcast year, possibly highlighting the effects of the initial coupling shock. The associated Bayesian analysis (Fig. 1b) demonstrates that the data strongly modify our (weakly informative) prior assumption about the model's parameters, leading to well constrained posterior estimates. Variance of errors at the data level (τ^2_D, see equation (2) of the methods section) is substantially larger than those at the process level, and therefore more strongly contribute to the overall uncertainty in the hindcast errors. Posterior distributions of error variances are noticeably skewed at the process level, particularly for the seasonal bias component (τ^2_σ).

The residuals $r_j(t)$, i.e., the differences between $D_j(t)$ and the Bayesian estimates of $\Delta(t)$ (see Fig. 1a, bottom panel), become roughly stationary around zero and homoskedastic (i.e., consistent with being drawn from the same distribution at all time steps t) from roughly the third integration year onward, indicating that our decomposition captures the bulk evolution of systematic hindcast errors. However, systematic departures from zero of the residuals are diagnosed in the first two hindcast years. The positive residuals in the first few months and the negative residuals around January of the second integration year possibly reflect unresolved issues with the seasonal error (i.e., the initial shock). Figure 2 expands the residual analysis to the case of individual hindcasts: stripes of similar colors in Fig. 2a that are parallel to the black dashed lines characterize specific events that the decadal climate prediction system systematically fails to predict. Among these events are the El Niños of 1972/73, 1982 (concomitant with the El Chichón eruption), 1991/92 (concomitant with the Pinatubo eruption) and 1997/98. For long hindcast times (>2 years), failure to predict these events is not unexpected since their occurrence is determined by the chaotic nature of the climate system. It also reflects the difficulties inherent to simulating the Atlantic response to the Pacific[26].

The hindcasts also fail to predict multiannual SST_{ABF} anomalies. The residual error signal displays an interdecadal fluctuation with predominant negative errors before the mid-1970s and predominant warm errors afterwards. This signal robustly emerges in different seasons (Fig. 2a, right panel) and for different hindcast times (Fig. 2b). Accounting for the delay of the error signal with respect to the initialization year, its interdecadal

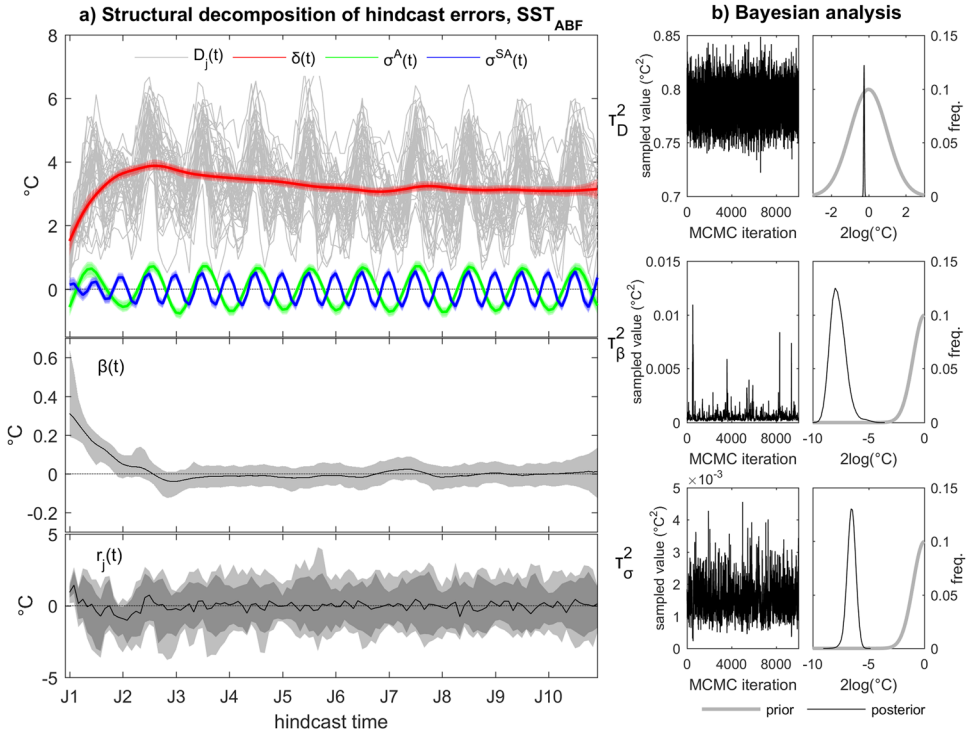

Figure 1. Structural decomposition and Bayesian analysis of decadal climate prediction errors. (**a**) (Top panel) grey lines: empirical hindcast errors in monthly mean spatially-averaged SST in the Angola-Benguela front region; colored lines/shading: associated posterior marginal estimates of systematic error components including drift/bias (red), annual seasonal bias (green) and semiannual seasonal bias (blue); (mid panel) posterior estimate of the stochastic trend; (bottom panel) posterior residual errors. Lines are posterior marginal medians, dark/light shading are 5th–95th and 1st–99th percentile ranges. (**b**) (Left) trace plots of the Monte Carlo Markov chains for the three model parameters; (right) posterior (black) versus prior (grey) distributions, shown as logarithms for better visualization. The probability distributions are plotted using 25 bins.

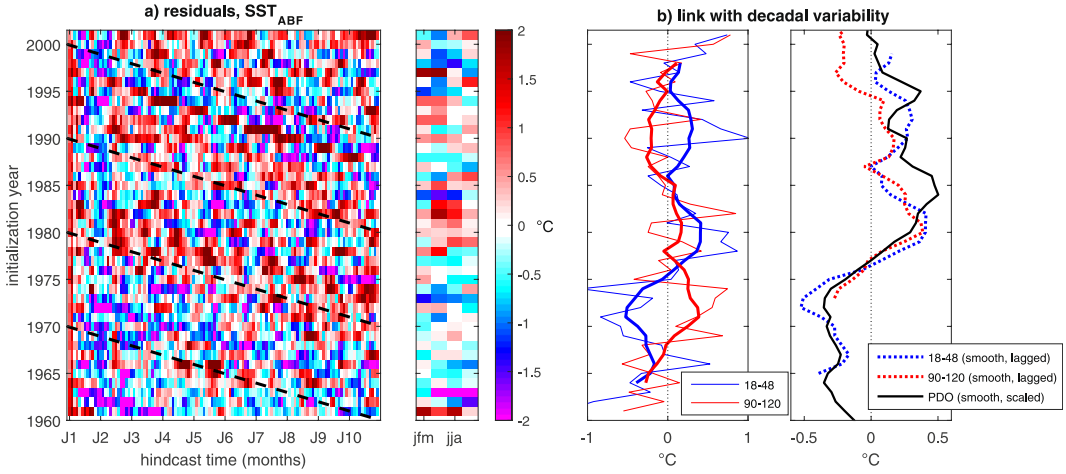

Figure 2. Analysis of the residuals ($r_j(t)$) of the Bayesian decomposition of monthly mean error signals in spatially-averaged SST in the Angola-Benguela front region. (**a**) Left: posterior median of monthly residuals for each hindcast; black dashed lines indicate the position of selected years along the forecast time axis, and provide a visual guideline to detect systematic drift-corrected error components. Right: seasonal residuals averaged for each simulated year (years refer to different integration years in the different hindcasts). (**b**) Left: temporal evolution of residuals calculated for two different hindcast times (blue: average during hindcast time 18–48 months; red: average during hindcast time 90–120 months); thick lines are 7-year smoothed data (see methods). Right: connection between residuals and the interdecadal internal climate variability expressed by the (7-year smoothed) observed annual-average Pacific Decadal Oscillation (PDO) index. Residuals are 7-year smoothed and lagged by the corresponding average hindcast time (lags are 3 and 9 years for the blue and red curve, respectively).

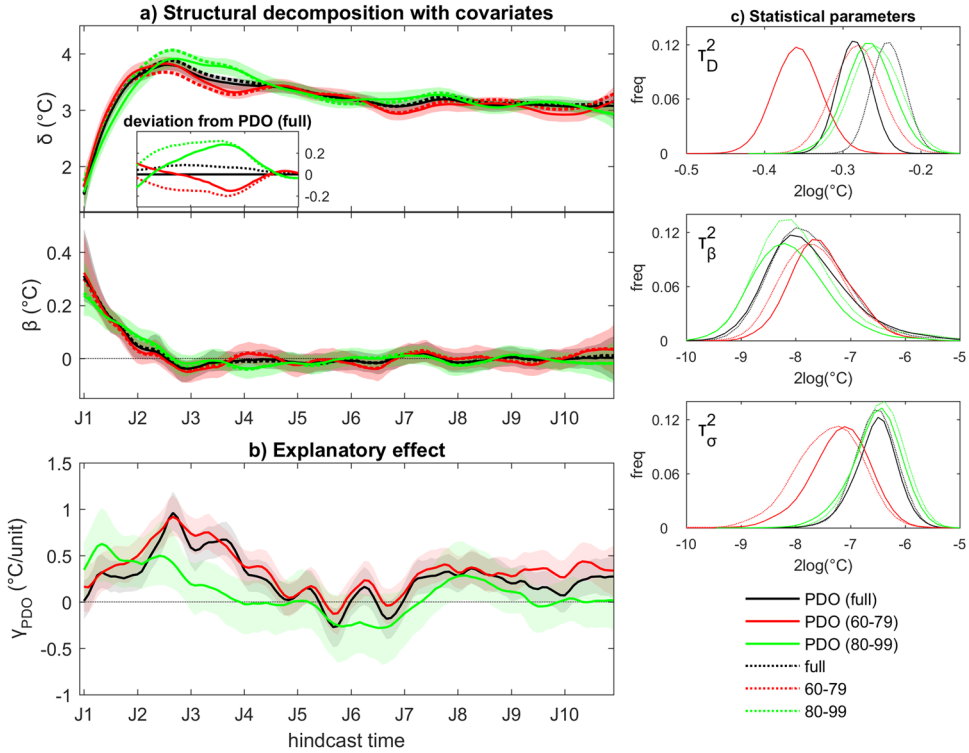

Figure 3. Explanatory effect of unpredicted inter-decadal climate variability on the structural decomposition of decadal climate prediction errors. (**a**) Posterior marginal estimates of drift/bias (δ) and stochastic trend (β) from hindcast errors in monthly mean spatially-averaged SST in the Angola-Benguela front region for three ensembles (full ensemble, sub-ensemble including initialization years from 1960 to 1979, sub-ensemble including initialization years from 1970 to 1999). Results including the observed Pacific Decadal Oscillation (PDO) index as covariate (continuous lines: median; shading: 5th–95th percentile range) are compared with corresponding results without any covariate (dashed lines: median). The inset in the top panel illustrates the differences in the posterior medians relative to the estimate for the full-ensemble with PDO for the 13–48 month hindcast period. (**b**) Posterior marginal estimates of regression coefficients γ for the PDO index in the different setups. (**c**) Posterior empirical distributions of the main parameters in the different model setups, shown as logarithm for better visualization. Representation as in Fig. 1.

modulation can be shown to match the shift of the Pacific Decadal Oscillation (PDO) from a cold to a warm phase in the mid-1970s (Fig. 2b, right panel).

Based on this empirical analysis, the statistical model for the estimation of systematic SST_{ABF} errors is expanded to assess the explanatory effect of the PDO. In practice, the observed PDO index is included as an additional covariate in the state-space model, and the posterior marginal distributions of all unknowns obtained from the expanded model are compared with those obtained in the original analysis illustrated in Fig. 1. Consistent with the analysis of residuals without co-variates, the PDO index is lagged by 24 months, i.e., $X(t) = PDO(t\text{-}24)$. The approach is applied to the full hindcast ensemble and for two equally-sized sub-ensembles defined by the period of the hindcast initialization year, i.e., 1960–1979 and 1980–1999. In the full-ensemble analysis, the stochastic regression coefficient $\gamma_{|PDO}(t)$ indicates that the PDO index is informative about SST_{ABF} hindcast errors, with strongest impact in the second and third hindcast years (black line/shading in Fig. 3b). Inclusion of the PDO index does not appreciably affect the posterior estimation of the parameters of the systematic components (posterior estimates as black lines for τ^2_β and τ^2_σ in Fig. 3c), but it substantially reduces uncertainty in the non-systematic component (see τ^2_D in Fig. 3c). Similar damping of τ^2_D is inferred also for the two sub-ensembles. The PDO index is relevant for the estimation of SST_{ABF} hindcast errors for both sub-ensembles, but with a smaller impact in the 1980–1999 sub-ensemble (Fig. 3b), in line with the weaker relationship seen in the empirical analysis (Fig. 2b). The PDO-corrected posterior estimates of $\delta(t)$ agree with the full-period estimate better than the corresponding estimates from a model without covariates, particularly around the peak warm error phase (compare dashed and continuous lines in Fig. 3a). A plausible interpretation may be given as follows. A teleconnection exists between the PDO and the southeastern tropical Atlantic, which is observed as a negative PDO-SST_{ABF} correlation; the positive explanatory term $\gamma_{|PDO}$ (in this context, a sort of correction term) suggests that this teleconnection is active also in the simulated climate; removing the effects of remote forcing from the erroneous Pacific state leads to improved estimation of the local drift component. In general, it can be concluded that the estimation of the drift and of its uncertainty depends, to some extent, on the systematic portion of hindcast errors that depends on (inter)decadal internal climate variability. Our method could therefore be used in practice to improve drift estimation by accounting for known explanatory factors of such internal climate variability.

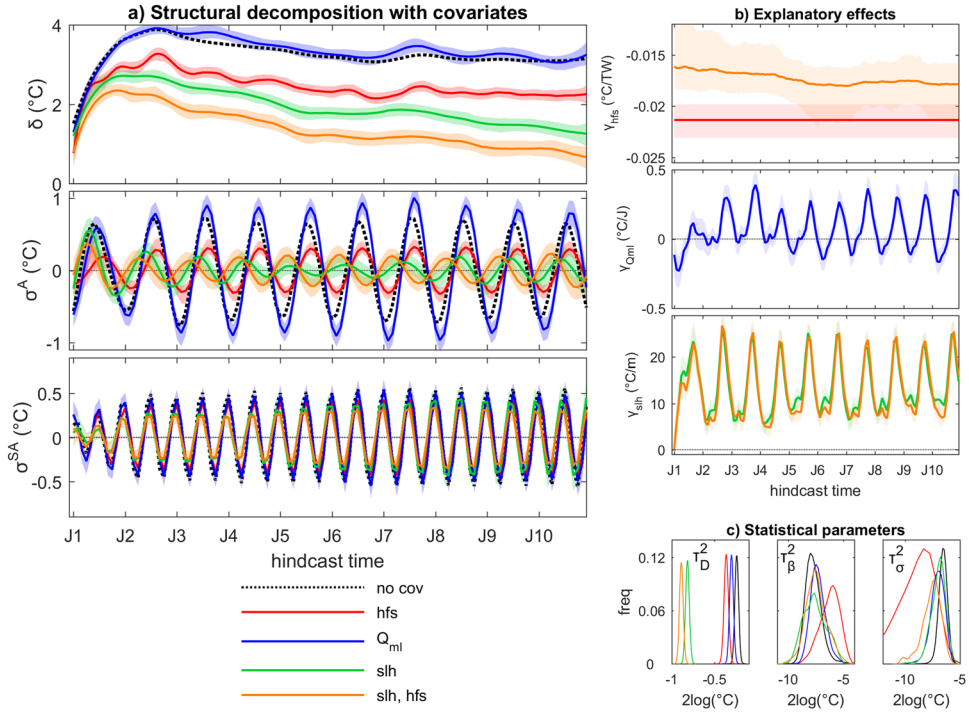

Figure 4. Explanatory effects of covariates on the structural decomposition of decadal climate prediction errors. (**a**) Posterior marginal estimates of systematic error components from hindcast errors in monthly mean spatially-averaged SST in the Angola-Benguela front region, including (from top to bottom) drift/bias δ, stochastic trend β, annual seasonal bias σ^A and semiannual seasonal bias σ^{SA}. Different choices of covariates are shown: adjusted regional surface heat flux into the ocean (shf, red), mixed layer heat content (Q_{ml}, blue), 1-month lagged sea-level height (slh, green), hfs and slh in conjunction (orange). Solid lines: median; shading: 5th–95th percentile range. The median estimate without covariate is shown for comparison (dashed black, from Fig. 1). (**b**) Posterior marginal estimates of regression coefficients γ for the two covariates in the different setups. (**c**) Posterior empirical distributions of the main parameters in the different model setups. Representation as in Fig. 1.

As a further example for the possibility of including explanatory covariates in the statistical model, Fig. 4a shows the structural decomposition of systematic SST_{ABF} hindcast errors when the effect of hindcast errors in selected terms of the heat budget for the local mixed layer are accounted for. The blue curves in Fig. 4a correspond to results of a model setup accounting for the explanatory effect of the mixed layer heat content (Q_{ml}). Whereas empirical estimates of systematic errors (see methods) in Q_{ml} and SST_{ABF} are significantly correlated ($r = -0.66$, $p < 0.001$), the Bayesian analysis indicates that systematic SST_{ABF} errors are largely unexplained by Q_{ml} alone, reflective of the weaker average correlation between errors of both variables in the individual hindcasts ($r = -0.32$) and suggesting a more complex causal association beyond the thermal state of the mixed layer.

Hindcast errors in both, adjusted surface heat flux (hfs) and 1-month lagged sea-level height (slh) reduce the error component $\delta(t)$ and damp the estimated annual and semiannual biases (red and green lines in Fig. 4a, respectively). Changes in slh tend to reflect ocean currents. Horizontal circulation changes tend to be associated with meridional shifts of the Angola-Benguela SST front, while changes in upwelling result in local shoaling or sinking of the thermocline. With slh as covariate, the annual seasonal bias $\sigma^A(t)$ nearly vanishes from the second hindcast year onward, and the semiannual seasonal bias $\sigma^{SA}(t)$ is also appreciably reduced. The larger amplitudes of $\sigma^A(t)$ during the first years suggest decoupling of surface from subsurface errors during the initial shock phase. The posterior regression coefficient $\gamma(t)$ (Fig. 4b) corresponding to this parameter illustrates a seasonally-varying connection between errors in SST that is strongest during austral spring, possibly reflecting the seasonality of wind-dependent errors. Analysis of the posterior marginal distributions of the statistical model parameters indicates that inclusion of slh as explanatory covariate largely damps the variance of the observational error (Fig. 4c), which is the major source of uncertainty in the original estimation (Fig. 1b).

An even larger reduction of $\delta(t)$ is obtained when the explanatory effects of hindcast errors in hfs and slh are accounted for in conjunction (orange lines in Fig. 4a). With this choice, $\delta(t)$ has values below 1 °C for long hindcast times, indicating that the climatological bias is almost completely damped when accounting for the explanatory effects of hfs and slh. The posterior estimate of $\sigma^A(t)$ obtained with this model configuration is phase-shifted compared to the original analysis and has larger amplitudes compared to the configuration with slh as the only covariate, which suggests concomitant effects on SST biases from misrepresented seasonal variations in hfs and slh that mutually compensate or reinforce each other depending on their phase.

The SST_{ABF} drift thus largely develops due to erroneous heat gain by the ocean through surface fluxes coupled with an erroneous redistribution of that heat in the water mass. In contrast, an only marginal connection is found

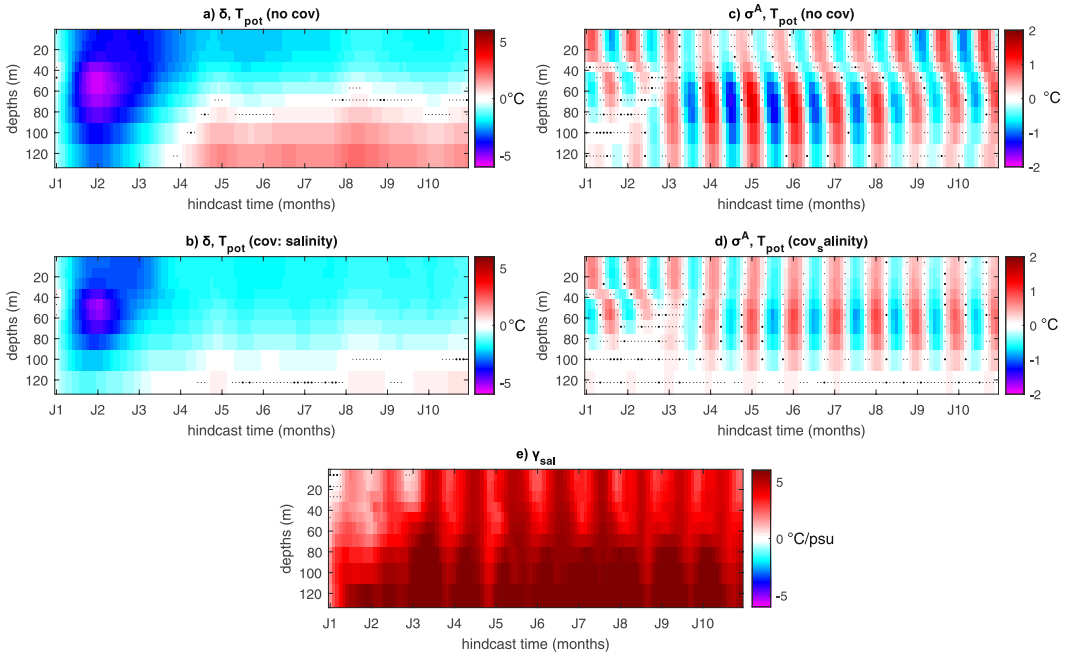

Figure 5. Evolution of systematic errors (level error δ and annual bias component σ^A) in seawater potential temperature at 44°S latitude and 50°W longitude for the upper ten model levels. (**a**,**c**) Standalone analysis; (**b**,**d**,**e**) analysis including the explanatory effect of local salinity errors. Posterior estimates were interpolated on a regular longitude-latitude grid before the analysis. Shading: median of posterior marginal errors. Large (small) dots indicate grid points where the estimated value is not significantly different from zero with high (low) confidence, i.e., where the distribution of values crosses the value of zero in between the 40th–60th (5th–95th) percentile range. Posterior estimates were interpolated on a regular longitude-latitude grid before the analysis.

with associated systematic errors in the mixed layer heat content, which are in turn dominated by the erroneous volume loss simulated in the mixed layer. Overall, the analysis highlights the complexity of errors in a region featuring a great variety of coupled dynamics and teleconnections[27]. Fully disentangling such complexity exceeds the scope of the illustrative analysis presented here.

Figures 1c, 3c and 4c consistently show that the largest uncertainty that affects the estimation of the SST_{ABF} drift stems from the data model parameter τ^2_D included in the observation equation of our statistical model (see the methods section). We tested the impact of using different input data to the dynamic linear model by constructing different hindcast ensembles, from small sub-ensembles to that used in the main analysis to super-ensembles including multiple realizations for each hindcast (results not shown). Tighter constrains on τ^2_D are generally achieved by increasing the size of the ensemble, without significant impacts on its estimated mean. By contrast, reducing the ensemble size can lead to major discrepancies in the mean estimation. It is especially in this case that the proposed Bayesian hierarchical formulation could be valuable – through the implementation of prior knowledge about uncertainty parameters and/or explanatory factors, as shown in Fig. 3 – to partly overcome the dependency of drift uncertainty estimation on the quality and quantity of available empirical information about the drift.

So far, the state-space decomposition method has been illustrated for a single predictand, SST_{ABF}. Its analytic potential, however, is exploited more fully when it is applied simultaneously to spatially distributed data. Here, we illustrate how it may aid the characterization of the spatio-temporal complexity of decadal climate prediction errors, and assist in understanding the underlying dynamics with an application to gridded upper-ocean potential temperature data in the southern Atlantic mid-latitudes, near the Brazil–Malvinas confluence zone. The results indicate intricate variability in the sub-surface ocean for the $\delta(t)$ and $\sigma^A(t)$ components (see supplementary movie). For instance, at [44°S, 50°W] during the initial drift phase negative (i.e., cold) errors ($\delta(t) < 0$) are diagnosed throughout the water column down to 120 m depth, with peak negative values around 50 m depth, while errors of opposite sign characterize near-surface and subsurface evolution during the climatological bias phase (Fig. 5a). In the annual bias component, downwelling pulses of both warm and cold errors are observed (see Fig. 5c). Inclusion of hindcast errors in local seawater salinity as covariate in the state-space model indicates that a considerable part of potential temperature errors in the subsurface ocean is linked with concomitant errors in salinity. Following the initial drift phase, the warm climatological bias below 80 m depth is largely damped in the model including salinity errors as covariate (compare panels a and b of Fig. 5). Also, annual bias fluctuations are damped in the sub-surface ocean while they are phase-shifted in the near-surface layer (compare panels c and d of Fig. 5), as highlighted by the strong seasonal variations of the regression coefficient $\gamma_{|sal}(t)$ in the near-surface layers (Fig. 5e). Thus, coherent seasonal error evolutions are detected throughout the upper ocean column (Fig. 5d). Overall, the analysis highlights once more how our statistical model aids the identification and quantification of linkages between drift and biases affecting different covarying processes. Specifically, it links downwelling of

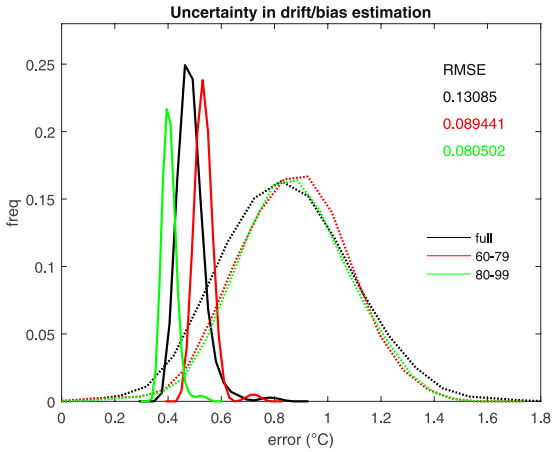

Figure 6. Comparison between the empirical distribution (dashed line) and Bayesian posterior distribution (solid line) of systematic decadal climate prediction errors. Results refer to the monthly mean spatially-averaged SST in the Angola-Benguela front region for the three ensembles referring to different initialization periods used in Fig. 3. Black: full ensemble; red: sub-ensemble including initialization years from 1960 to 1979; green: sub-ensemble including initialization years from 1980 to 1999). Analysis is for drift plus seasonal biases (i.e., $\delta + \sigma^{A} + \sigma^{SA}$, see equation (3) in the methods section). The shown error distributions are obtained from the monthly estimates calculated over all forecast lead times. These are the 1st–99th percentile ranges of the posterior samples for the BHM, and the unbiased sample standard deviation for the empirical method. The probability distributions are plotted using 13 bins. Numbers are the root mean squared errors (RMSEs) between the mean Bayesian and mean empirical estimates of the drift.

systematic warm (i.e., buoyancy positive) hindcast errors with overcompensation by corresponding local salinity errors.

Discussion and Conclusions

We propose a Bayesian hierarchical framework for the unified statistical assessment of systematic hindcast errors in decadal climate predictions. The major novelty of our approach is that the proposed state-space model allows for an explicit statistical estimation of the temporal evolution of major systematic error components, including drift, climatological bias and seasonal biases. It also allows to quantify the explanatory effect of co-varying processes, and to separately evaluate the associated uncertainties at the data, process and parameter levels.

We present an illustrative application of the model for spatially-averaged SST in the Angola-Benguela front region – where coupled climate models are typically affected by a strong warm bias[5] – simulated by an ensemble modelling system intended for decadal climate predictions. Different aspects of drift/bias quantification and interpretation are discussed, and we demonstrate the value of the proposed statistical approach as a diagnostic, exploratory and hypothesis-testing tool.

First, we show that by virtue of the separation between data and process levels, the hierarchical method allows for well-constrained estimates of statistical uncertainty for the different systematic error components (Fig. 1). Compared to currently adopted empirical approaches, the Bayesian estimates of drift uncertainty yield narrower distributions (Fig. 6). The more reliable estimation of drift uncertainty – at the process level – resulting from this rigorous methodology can provide more robust estimates of prediction skills compared to traditional simple averages. In particular, the posterior ensemble realizations of the drift (and of other systematic hindcast error components) obtained from the Bayesian hierarchical model gives a forecast drift correction analogous to the empirical estimate used in current approaches[28].

Second, we show that drift estimation is appreciably affected by the interdecadal climate state and evolution, including extant global teleconnections between hindcast errors (Fig. 2). The shorter the analysis period, the larger the risk of erroneously attributing to drift a portion of (inter)decadal climate variability which the system systematically fails to predict. We show that this issue can be in principle overcome within our statistical modeling framework by accounting for the explanatory effect of such unresolved (inter)decadal climate variability (Fig. 3).

Finally, we show that the proposed structural decomposition permits a robust quantification of non-stationary explanatory effects of the bias and their uncertainty (Figs 4 and 5), leading to the possibility of probing causal hypotheses for the origin of model biases with a statistically rigorous quantification that is inaccessible to commonly used approaches in traditional multivariate assessments of systematic decadal climate prediction errors. We stress that the proposed hierarchical framework uniquely allows for a fully traceable quantification of explanatory effects based on the interdependencies expressed in each individual hindcast.

We believe that the strength of these advantages, together with their modest computational requirements, justify the view that state-space models like the one proposed here, or variants of greater complexity (for instance accounting for spatial dependencies[29], can be used in future applications to robustly and quantitatively assess explanatory effects and uncertainties in climate forecasts. Ultimately, improved drift adjustment techniques such

as that outlined here will help to advance the quality of the estimation of the predictability and the predictive skills of a forecast system[20].

To conclude, the generality of our results make them relevant well beyond the specific decadal prediction system, the focus region or reference variable employed in this study. Compared with the existing empirical techniques for drift estimation, the structural decomposition and Bayesian hierarchical approach envisaged here represent an opportunity for progress in our understanding of systematic climate model errors: the method yields a reliable estimation of drift and bias within a unified framework; it permits an evaluation of causal relationships and teleconnections by embedding relevant information on associated dynamics at the process level; and it allows to make a quantitative separation of the process-related uncertainties from those associated to empirical data limitations.

Methods

Bayesian approach. Scientific literature contains numerous studies[29–34] describing the application of Bayesian hierarchical models (BHMs) in the field of climate research, especially to the assessment of climate model outputs. BHMs use Bayes' theorem to incorporate information from different sources, including observations, physical theories and experts' knowledge. They provide a flexible framework to developing consistent inference and prediction of unknown quantities under study, which overcomes single-value estimates of uncertainty.

The BHM general formulation is based on three building blocks: the data model, the process model and the parameter model. Suppose that $\boldsymbol{Z}$ represents the data, $\boldsymbol{Y}$ represents the process and $\boldsymbol{\theta}$ represents the parameters related to the data and the process model, we then have:

(a) Data model, $[\boldsymbol{Z}|\boldsymbol{Y}, \boldsymbol{\theta}]$
(b) Process model, $[\boldsymbol{Y}|\boldsymbol{\theta}]$
(c) Parameter model, $[\boldsymbol{\theta}]$

where [A] is the generic notation for the probability distribution of the random quantity A. In practice, (a) defines the statistical model representing the dependence of observations on the unknown process, (b) describes the conditional probability distribution of the process on the model parameters, and (c) describes the (prior) probability distribution of the parameters, which are treated as random quantities according to the Bayesian approach.

BHMs provide estimates of the unknowns ($\boldsymbol{Y}$ and $\boldsymbol{\theta}$) with associated uncertainty through the calculation of their posterior distribution conditioned to available observations, which is allowed by the Bayes' theorem:

$$[\boldsymbol{Y}, \boldsymbol{\theta}|\boldsymbol{Z}] = [\boldsymbol{Z}|\boldsymbol{Y}][\boldsymbol{Y}|\boldsymbol{\theta}][\boldsymbol{\theta}]/[\boldsymbol{Z}] \tag{1}$$

Direct evaluation of (1) to obtain the posterior distribution on the left term is often computationally intractable. This can be circumvented by generating samples from the posterior distribution through a Markov Chain Monte Carlo method[35].

Statistical model for the estimation of systematic decadal climate prediction errors. At the data level, the systematic error $\Delta(t)$ of the decadal climate prediction system at prediction time t is observed through differences between the predicted and the observed – or, analogously, assimilated – values ($D_j(t)$) from an ensemble of $j = 1,\ldots,p$ hindcasts initialized at different times, according to the following model:

$$D_j(t) = \Delta(t) + \varepsilon_j(t) \tag{2}$$

where $\varepsilon_j(t)$ is a Gaussian white noise random error with zero mean and variance τ^2_D. We assume, for simplicity, that the observation error has the same variance in all hindcasts.

At the process level, $\Delta(t)$ is decomposed into two components: the drift/bias $\delta(t)$ and the seasonal bias $\sigma(t)$, further split into annual ($\sigma^A(t)$) and semiannual ($\sigma^{SA}(t)$). Namely:

$$\Delta(t) = \delta(t) + \sigma^A(t) + \sigma^{SA}(t) \tag{3}$$

The drift/bias $\delta(t)$ changes through time according to a local linear trend:

$$\delta(t) = \delta(t-1) + \beta(t-1) + \varepsilon_\delta(t) \tag{4}$$

$$\beta(t) = \beta(t-1) + \varepsilon_\beta(t) \tag{5}$$

where $\varepsilon_\delta(t)$ and $\varepsilon_\beta(t)$ are uncorrelated Gaussian white noise random errors with zero mean and variance τ^2_δ and τ^2_β, respectively. The term $\beta(t)$ is a random walk. The effect of $\varepsilon_\delta(t)$ is to allow the level of the drift/bias to shift up and down, while $\varepsilon_\beta(t)$ allows the slope to change. The larger the variances, the greater the stochastic movements in the drift/bias. If $\tau^2_\delta = \tau^2_\beta = 0$, the local linear trend collapses to a linear deterministic trend.

Seasonal terms are modeled using harmonic functions[36]. Using monthly data with $k = 1$ for $\sigma^A(t)$ and $k = 2$ for $\sigma^{SA}(t)$, the kth harmonic function takes the general form

$$\sigma_k(t) = \zeta^1_k \cos(2\pi kt/12) + \zeta^2_k \sin(2\pi kt/12)$$

where ζ^1_k and ζ^2_k are constants. Like the local linear trend, the seasonal term can be built up recursively, leading to the (stochastic) model[36] for both components:

$$\begin{pmatrix}\sigma_k(t)\\ \sigma_k^*(t)\end{pmatrix} = \begin{pmatrix}\cos(k2\pi/12) & \sin(k2\pi/12)\\ -\sin(k2\pi/12) & \cos(k2\pi/12)\end{pmatrix}\begin{pmatrix}\sigma_k(t-1)\\ \sigma_k^*(t-1)\end{pmatrix} + \begin{pmatrix}\varepsilon_{\sigma,k}(t)\\ \varepsilon_{\sigma,k}^*(t)\end{pmatrix} k = 1, 2 \tag{6}$$

with $\varepsilon_{\sigma,k}(t)$ and $\varepsilon^*_{\sigma,k}(t)$ uncorrelated Gaussian noise random errors with zero mean and variance τ^2_σ, and * indicating the conjugate.

The process model (3) can be easily extended to include the effect of external factors in terms of additional explanatory variables. For one dimensional explanatory variable X(t), the model becomes:

$$\Delta(t) = \delta(t) + \sigma^A(t) + \sigma^{SA}(t) + \gamma(t)X(t) \tag{7}$$

In equation (7) we allow $\gamma(t)$ to vary according to a random walk:

$$\gamma(t) = \gamma(t-1) + \varepsilon_\gamma(t) \tag{8}$$

with $\varepsilon_\gamma(t)$ Gaussian white noises with zero mean and variance τ^2_γ. Time varying coeffcients $\gamma(t)$ allow us to take into account possible non stationary effects of the covariate. Again, if $\tau^2_\gamma = 0$, the effect collapses to be constant in time. Finally, the parameter level requires the specification of the prior distribution for the unknown parameters $\boldsymbol{\theta} = (\tau^2_D, \tau^2_\delta, \tau^2_\beta, \tau^2_\sigma, \tau^2_\gamma)$. In this study, we set $\tau^2_\delta = 0$ to obtain a smoothly varying error $\delta(t)$, which then better corresponds to the drift/bias. We define weakly informative lognormal (LN) priors in the form LN(0,1) for all parameters.

Dynamic Linear Model implementation. The formulation of the statistical model above allows for a straightforward implementation within the dynamic linear model (DLM) framework. DLMs are based on a state-space approach, i.e., unobservable state variables are used that allow direct modeling of the process ($\boldsymbol{Y}$) generating the observed data ($\boldsymbol{Z}$)[36,37]. DLMs have the general form:

$$\boldsymbol{Z}(t) = \boldsymbol{F}\boldsymbol{Y}(t) + \mathbf{v}(t) \tag{9}$$

$$\boldsymbol{Y}(t) = \boldsymbol{G}\boldsymbol{Y}(t-1) + \mathbf{w}(t) \tag{10}$$

where t is the discrete time variable representing, in our case, monthly values, $\boldsymbol{Z}(t)$ is a vector of p observations at time t, $\boldsymbol{Y}(t)$ is the underlying state vector of dimension m, $\mathbf{G}$ is the mxm system matrix, and $\mathbf{F}$ is the mxp observation matrix. We suppose that $\mathbf{v}(t)$~N(0,$\mathbf{V}$) and $\mathbf{w}(t)$~N(0,$\mathbf{W}$) are the observation and model Gaussian errors, respectively, that are serially and mutually uncorrelated. In this formulation, the matrices $\mathbf{V}$ and $\mathbf{W}$ contain the model parameters $\boldsymbol{\theta}$. If we suppose that the unknown parameters are random the DLM formulation is a BHM where (9) and (10) are typically referred to as observation equation and system equation, respectively.

We have applied the DLM to spatially-averaged indices and, individually, to the model's grid-points over the investigated domain. Following (2) and (9), and having p values of D(t) (i.e., having p hindcasts) the observation vector is defined as $\boldsymbol{Z}(t) = \{D^1(t),\ldots, D^p(t)\}'$.

In the base model without explanatory covariates, following (3) and accounting for annual and semi-annual seasonalities in the decadal climate prediction errors, the state vector is defined as $\boldsymbol{Y}(t) = \{\delta(t), \beta(t), \sigma^A(t), \sigma^{A*}(t), \sigma^{SA}(t), \sigma^{SA*}(t)\}'$. Therefore, the dimension of the state vector is $m = 6$. We further assume $\mathbf{Y}(0)$~N(0,0.025), i.e., there is practically no systematic error during the assimilation. The observation matrix $\mathbf{F}$ is defined following (2) and (3).

The sequential definition of the process model (having a conditional dependency only on the previous time step) allows to use the Kalman filter formulas[36,37] for calculating the posterior distribution (1) inside a Monte Carlo Markov Chain algorithm. In particular, we use a slice-sampler algorithm[38] to iteratively sample from the full posterior distribution of the unknown parameters $\boldsymbol{\theta}$.

Empirical estimation of systematic decadal climate prediction errors. The BHM results are compared with empirical estimates obtained following the guidelines for drift/bias correction by the Decadal Climate Prediction Panel for the upcoming Coupled Model Intercomparison Project phase 6[28]. The associated uncertainty is calculated for each hindcast time t as the unbiased sample standard deviation of the $D_j(t)$ values from the $j = 1,\ldots,p$ hindcasts. Similar results are obtained using the 95% confidence interval of the standard error of the mean.

Decadal prediction system and dataset. The MiKlip prototype system for decadal climate predictions[25] is based on the low-resolution version of the Max Planck Institute - Earth System Model (MPI-ESM-LR). MPI-ESM-LR is a conglomerate of a coupled general circulation model and subsystem models for land and vegetation, and biogeochemistry. In MPI-ESM-LR, the atmospheric general circulation model ECHAM6[39] uses a T63/1.9° horizontal resolution and 47 hybrid sigma pressure levels that extend up to 0.01hPa; the ocean-sea ice model MPIOM[9] uses a 1.5° resolution bipolar grid with 40 z-levels. The time step for numerical integration is 600 s for ECHAM6 and 4320 s for MPIOM.

[9] describe the main oceanic biases, highlighting serious biases prevailing in the intermediate layers of the ocean, which reflect the inability of the model to maintain the correct water mass properties. In general, the ocean in MPI-ESM gets too warm and saline at intermediate levels and in the deep ocean whereas it is too cold and fresh in the upper layers. MPI-ESM-LR and similar configurations of MPI-ESM have been widely tested and used in studies of climate dynamics and variability[40,41].

We use the MiKlip decadal prediction experiments[25] based on full-field assimilation of the ORAS4 ocean reanalysis data[42]. The lagged initialization procedure is used to obtain the initial state from the historical assimilation run for each of the 15 ensemble members for each initialization year between 1960 and 2000. The external radiative forcing data are based on the recommendations for the Coupled Model Intercomparison Project phase 5 (CMIP5), an overview of which is outlined in[25]. The ensemble including the first ensemble members ("r1") was used in the main analysis. The associated historical assimilation run is used as analog for the observational targets.

Spatially-averaged indices for the Angola-Benguela front region used in the study refer to the domain spanning 10°S–20°S latitude and 10°E–15°E longitude (to the coastline). Empirical hindcast errors in monthly mean spatially-averaged SST in the Angola-Benguela front region are provided as supplementary information. The observed PDO index is the JISAO PDO index[43] available at: http://research.jisao.washington.edu/pdo/PDO.latest.txt. The original monthly index was low-pass filtered with an 85-month running average (corresponding to 7 years) before analysis to smooth out ENSO influences. Adjusted surface energy flux for the mixed layer and mixed layer heat content in the ABF region are estimated following[44]. All empirical distributions are smoothed with a 5-point moving average with symmetric Hanning weights.

References

1. Meehl, G. A. *et al*. Decadal climate prediction: an update from the trenches. *Bull. Am. Meteorol. Soc.* **95**, 243–267 (2014).
2. Kirtman, B. *et al*. Near-termClimate Change: Projections and Predictability in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al*.) (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013).
3. Wang, C., Zhang, L., Lee, S.-K., Wu, L. & Mechoso, C. R. A global perspective on CMIP5 climate model biases. *Nature Clim. Ch.* **4**, 201–205 (2014).
4. Hawkins, E., Dong, B., Robson, D. J., Sutton, R. & Smith, D. M. The interpretation and use of biases in decadal climate predictions. *J. Clim.* **27**, 2931 (2014).
5. Xu, Z. *et al*. Diagnosing southeast tropical Atlantic SST and ocean circulation biases in the CMIP5 ensemble. *Clim. Dyn.* **43**, 3123 (2014).
6. Sen Gupta, A., Jourdain, N., Brown, J. & Monselesan., D. Climate drift in the CMIP5 models. *J. Clim.* **26**, 8597–8615 (2013).
7. Voldoire, A. *et al*. The CNRM-CM5.1 global climate model: Description and basic evaluation. *Clim. Dyn.* **40**, 2091–2121 (2013).
8. Danabasoglu, G. *et al*. The CCSM4 Ocean Component. *J. Clim.* **25**, 1361–1389 (2012).
9. Jungclaus, J. H. *et al*. Characteristics of the ocean simulations in MPIOM, the ocean component of the Max Planck Institute Earth System Model. *J. Adv. Model. Earth Syst.* **5**, 422–446 (2013).
10. Sterl, A. *et al*. A look at the ocean in the EC-Earth climate model. *Clim. Dyn.* **39**, 2631–2657 (2012).
11. Ding, H., Keenlyside, N., Latif, M., Park, W. & Wahl, S. The impact of mean state errors on equatorial Atlantic interannual variability in a climate model. *J. Geophys. Res. Oceans* **120**, 1133–1151 (2015).
12. Toniazzo, T. & Woolnough, S. Development of warm SST errors in the southern tropical Atlantic in CMIP5 decadal hindcasts. *Clim. Dyn.* **43**, 2889–2913 (2014).
13. Small, R. J., Curchitser, E., Hedstrom, K., Kauffman, B. & Large, W. G. The Benguela upwelling system: quantifying the sensitivity to resolution and coastal wind representation in a global climate model. *J. Clim.* **28**, 9409–9432 (2015).
14. Ndoye, S. *et al*. SST patterns and dynamics of the southern Senegal-Gambia upwelling center. *J. Geophys. Res. Oceans* **119**, 8315–8335 (2014).
15. Richter, I. & Xie, S.-P. On the origin of equatorial atlantic biases in coupled general circulation models. *Clim. Dyn.* **31**, 587–598 (2008).
16. Wahl, S., Latif, M., Park, W. & Keenlyside, N. On the Tropical Atlantic SST warm bias in the Kiel Climate Model. *Clim. Dyn.* **36**, 891–906 (2011).
17. Milinski, S., Bader, J., Haak, H., Siongco, A. C. & Jungclaus, J. H. High atmospheric horizontal resolution eliminates the wind-driven coastal warm bias in the south eastern tropical Atlantic. *Geophys. Res. Lett.* **43**, 10455–10462 (2016).
18. Carrassi, A. *et al*. Full-field and anomaly initialization using a low-order climate model: a comparison and proposals for advanced formulations. *Nonlin. Proc. Geophys.* **21**, 521–537 (2014).
19. Garcia-Serrano, J. & Doblas-Reyes, F. J. On the assessment of near-surface global temperature and North Atlantic multi-decadal variability in the ENSEMBLES decadal hindcast. *Clim. Dyn.* **39**, 2025–2040 (2012).
20. Fyfe, J. C. *et al*. Skillful predictions of decadal trends in global mean surface temperature. *Geophys. Res. Lett.* **38**, L22801 (2011).
21. Kharin, V. V., Boer, G. J., Merryfield, W. J., Scinocca, J. F. & Lee, W. S. Statistical adjustment of decadal predictions in a changing climate. *Geophys. Res. Lett.* **39**, L19705 (2012).
22. Fučkar, N. S., Volpi, D., Guemas, V. & Doblas-Reyes, F. J. A posteriori adjustment of near-term climate predictions: Accounting for the drift dependence on the initial conditions. *Geophys. Res. Lett.* **41**, 5200–5207 (2014).
23. Sanchez-Gomez, E., Cassou, C., Ruprich-Robert, Y., Fernandez, E. & Terray, L. Drift dynamics in a coupled model initialized for decadal forecasts. *Clim. Dyn.* **46**, 1819–1840 (2016).
24. Hoekstra, R. & van der Bergh, J. J. C. J. M. Comparing structural decomposition analysis and index. *Energy economics* **25**, 39–64 (2003).
25. Marotzke, J. *et al*. MiKlip - a National Research Project on Decadal Climate Prediction. *Bull. Amer. Meteor. Soc.* **97**, 2379–2394 (2016).
26. Luebbecke, J. & McPhaden, M. On the inconsistent relationship between Pacific and Atlantic Niños. *J. Clim.* **25**, 4294–4303 (2012).
27. Lübbecke, J. F., Böning, C. W., Keenlyside, N. S. & Xie, S.-P. On the connection between Benguela and equatorial Atlantic Niños and the role of the South Atlantic Anticyclone. *J. Geophys. Res.* **115**, C09015 (2010).
28. Boer, G. J. *et al*. The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci. Model Dev.* **9**, 3751-3777 (2016).
29. Arisido, M. W., Gaetan, C., Zanchettin, D. & Rubino, A. A Bayesian hierarchical approach for spatial analysis of climate model bias in multi-model ensembles. *Stoch. Environ. Res. Risk. Assess.*, https://doi.org/10.1007/s00477-017-1383-2 (2017).
30. Duan, Q. & Phillips, T. J. Bayesian estimation of local signal and noise in multimodel simulations of climate change. *J. Geophys. Res. Atmos.* **115**, D18123 (2010).
31. Tebaldi, C., Smith, R. L., Nychka, D. & Mearns, L. O. Quantifying Uncertainty in Projections of Regional Climate Change: A Bayesian Approach to the Analysis of Multimodel Ensembles. *J. Clim.* **18**, 1524–1540 (2005).
32. Buser, C. M., Künsch, H. R., Lüthi, D., Wild, M. & Schär, C. Bayesian multi-model projection of climate: bias assumptions and interannual variability. *Clim. Dyn.* **33**, 849–868 (2009).
33. Kang, E. L., Cressie, N. & Sain, S. R. Combining outputs from the North American Regional Climate Change Assessment Program by using a Bayesian hierarchical model. *J. R. Stat. Soc. C* **61**, 291–313 (2012).
34. Reilly, J. *et al*. Uncertainty in climate change assessments. *Science* **293**, 430–433 (2001).
35. Robert, C. P., & Casella, G. Monte Carlo Statistical Methods. pp. 649 (Springer, 2004).

36. Laine, M., Latva-Pukkila, N. & Kyrölä, E. Analysing time-varying trends in stratospheric ozone time series using the state space approach. *Atmos. Chem. Phys.* **14**, 9707–9725 (2014).
37. Brogan, W. L. Modern Control Theory. pp. 736 (Prentice-Hall, 1974).
38. Radford, N. M. Slice Sampling. *Ann. Stat.* **31**, 705–767 (2003).
39. Giorgetta, M. A. *et al.* Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model Earth Syst.* **5**, 1–26 (2013).
40. Zanchettin, D., Bothe, O., Müller, W., Bader, J. & Jungclaus, J. H. Different flavors of the Atlantic Multidecadal Variability. *Clim. Dyn.* **42**, 381–399 (2014).
41. Moreno-Chamarro, E., Zanchettin, D., Lohman, K. & Jungclaus, J. H. An abrupt weakening of the subpolar gyre as trigger of Little Ice Age-type episodes. *Clim. Dyn.* **48**, 727–744 (2017).
42. Balmaseda, M. A., Mogensen, K. & Weaver, A. T. Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q. J. R. Meteorol. Soc.* **139**, 1132–1161 (2013).
43. Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. & Francis, R. C. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Am. Meteorol. Soc.* **78**, 1069–1079 (1997).
44. Zanchettin, D. *et al.* A decadally delayed response of the tropical Pacific to Atlantic multidecadal variability. *Geophys. Res. Lett.* **43**, 784–792 (2016).

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, and from the German Federal Ministry for Education and Research (BMBF) (MiKlip, project Nr. FKZ01LP1519A). The MPI-ESM simulations were carried out at the German Climate Computing Center, which also provided data services. We thank Wolfgang Mueller for having granted access to the MPI-ESM data. The statistical model is developed in matlab using the dlmsmo routine by Marko Laine (http://helios.fmi.fi/~lainema/dlm/). Part of the numerical calculations were done with the scientific computing system of Ca'Foscari (SCSCF). Primary data used in the analysis that may be useful in reproducing the author's work are archived by the Max Planck Institute for Meteorology and can be obtained by contacting miklip-mpi-esm@mpimet.mpg.de. We thank three anonymous reviewers, whose critical comments helped to improve the original study.

Author Contributions

C.G., D.Z. and A.R. conceived the study. D.Z. performed the DLM simulations and the analyses. K.M. calculated spatially-averaged indices. D.Z., T.T., K.M., C.G., M.A., N.K. and A.R. contributed to the discussion and writing the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-13144-2.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017