Statistical Structure of Intrinsic Climate Variability under Global Warming

XIUHUA ZHU

Centrum für Erdsystemforschung und Nachhaltigkeit, University of Hamburg, Hamburg, Germany

JOHN BYE

School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia

KLAUS FRAEDRICH

Max Planck Institute for Meteorology, Hamburg, Germany

ISABELLA BORDI

Department of Physics, Sapienza University of Rome, Rome, Italy

(Manuscript received 9 July 2015, in final form 21 April 2016)

ABSTRACT

Climate variability is often studied in terms of fluctuations with respect to the mean state, whereas the dependence between the mean and variability is rarely discussed. Here, a new climate metric is proposed to measure the relationship between means and standard deviations of annual surface temperature computed over nonoverlapping 100-yr segments. This metric is analyzed based on equilibrium simulations of the Max Planck Institute Earth System Model (MPI-ESM): the last-millennium climate (800–1799), the future climate projection following the A1B scenario (2100–99), and the 3100-yr unforced control simulation. A linear relationship is globally observed in the control simulation and is thus termed intrinsic climate variability, which is most pronounced in the tropical region with negative regression slopes over the Pacific warm pool and positive slopes in the eastern tropical Pacific. It relates to asymmetric changes in temperature extremes and associates fluctuating climate means with increase or decrease in intensity and occurrence of both El Niño and La Niña events. In the future scenario period, the linear regression slopes largely retain their spatial structure with appreciable changes in intensity and geographical locations. Since intrinsic climate variability describes the internal rhythm of the climate system, it may serve as guidance for interpreting climate variability and climate change signals in the past and the future.

1. Introduction

Climatic changes, whether anthropogenic or natural, can be described by climate modes or preferred "regimes" and changes in intensity and occupancy–occurrence frequency of these regimes (e.g., Cassou et al. 2004). These climate modes involve complex interactions between components of the climate system, which were already at work before human society could exert considerable impacts and form an essential part of the observed climate variability (e.g., Cobb et al. 2003). One scientific debate in recent decades concerns whether reported changes in these climate modes and climate extremes can be attributed to anthropogenic warming, and if so, to what extent (e.g., Tebaldi et al. 2006; Meehl et al. 2009). With the help of global climate models (GCMs), such a separation of anthropogenic contributions from natural signals is feasible from a technical point of view, but it has proven to be extremely difficult, if possible (e.g., Curry and Webster 2011; Huber and Knutti 2012).

Most climate analyses are based on statistical metrics in terms of means, covariances, and, more recently, higher moments (e.g., Goosse et al. 2005; Volodin and

Corresponding author address: Xiuhua Zhu, Centrum für Erdsystemforschung und Nachhaltigkeit, University of Hamburg, Grindelberg 5, Hamburg 20144, Germany.
E-mail: xiuhua.zhu@uni-hamburg.de

Yurova 2013; Huybers et al. 2014), which provide information of the underlying physical processes and help to understand climate extreme events. On shorter time scales (from days to seasons), relations between the first and second moments can be interpreted as baroclinic instability physics (e.g., Waterman and Jayne 2011). Such a relation has also been observed in sea ice thickness in the Arctic on a yearly scale (Blanchard-Wrigglesworth and Bitz 2014), suggesting that it may exist in different climate processes. Analyses on longer time scales, say year-to-year variability and century averages, have mostly focused on characterizing mean states and the related changes on the respective time scales (e.g., Watterson and Whetton 2011; Kosaka and Xie 2013). A possible link between the first and second moments on longer time scales, to our knowledge, has not been studied. On the other hand, considering the continuum of climate variability (Huybers and Curry 2006; Ault et al. 2013), such research is of particular interest.

In this study, we introduce a new climate metric characterizing the statistical relationship between means and standard deviations of annual surface temperatures computed over nonoverlapping segments to investigate internal climate variability. It reveals a linear relationship in an undisturbed GCM control simulation, which we term intrinsic climate variability (ICV). New features of the ICV in a warmer climate are presented. In addition, we find that the ICV metric, when large in magnitude, relates to asymmetric changes in temperature extremes, which may help interpret similar signals concerning anthropogenic warming.

This article is structured as follows: section 2 introduces GCM simulations and analysis methods; section 3 discusses ICV in relation to distribution symmetry and extremes; section 4 presents spatial features of ICV in the unforced control simulation; section 5 assesses ICV features in the last-millennium and A1B scenario simulations; and a discussion and conclusions are presented in section 6.

2. Model data and analysis method

a. Global climate model experiments

We analyze the millennium simulations of a state-of-the-art earth system model (MPI-ESM; Jungclaus et al. 2010), from which yearly surface temperatures are extracted. The MPI-ESM consists of the atmosphere model ECHAM5, the Max Planck Institute Ocean Model (MPIOM), the Jena Scheme for Biosphere–Atmosphere coupling in Hamburg (JSBACH), and the ocean biogeochemistry module, the Hamburg Model of the Ocean Carbon Cycle (HAMOCC5). The model setup enables the interactive simulation of the carbon cycle. The experiments are carried out with ECHAM5 in T31/L19 and MPIOM in GR3.0/L40 resolution, including a thermodynamic–dynamic sea ice model (for details see Jungclaus et al. 2010). Three experiments are analyzed: (i) The first is an unforced control run, where all external forcing is kept constant. (ii) The second is a last-millennium simulation with full forcing based on reconstructions of natural and anthropogenic forcing for the period of 800–2005 (five ensemble members). The full forcing comprises total solar irradiance (Vieira et al. 2011), volcanic aerosols (Crowley et al. 2008), land-cover changes (Pongratz et al. 2008), and greenhouse gas emissions, with CO_2 concentration being calculated by the model itself and the methane and nitrous oxide concentration being prescribed following MacFarling Meure et al. (2006). (iii) The third experiment is a future climate projection, following the IPCC A1B scenario (Nakicenovic et al. 2000). It covers the period of 2006–2200 with stabilized radiative forcing after 2100 and contains five ensemble members.

Since we are particularly interested in characterizing climate variability intrinsic to the climate system, our analysis is confined to the following three equilibrium periods: 1) the entire 3100 yr for the unforced control simulation, 2) the last-millennium simulation period (800–1799), and 3) the A1B equilibrium period (2100–99). The period of 1800–2005 is discarded because of its gentle shift related to modern warming, and so is the period of 2006–99 that is strongly influenced by transient changes (Fig. 1). In the following analysis, ensembles for the same simulation period are concatenated, which renders 5000 data points for the millennium simulation period (800–1799) and 500 yearly points for the A1B scenario (2100–99).

b. A new climate metric: Intrinsic climate variability

We define a new climate metric to measure the relation between means and standard deviations of annual surface temperatures (ST) computed over nonoverlapping segments. The segment length of 100 yr is chosen, because it is beyond the random walk length of around 30 yr (Bye et al. 2011, 2013, 2016) and provides reasonable statistics of all weather events, including extremes on decadal-to-interdecadal time scales.

Climate state can now be projected onto a new state space spanned by the first two moments of segmental statistics: namely, the segmental mean ST and the related standard deviation (STD) in each segment. This projection is depicted in Fig. 2. Given a time series, N nonoverlapping segments with equal length are constructed, $S_j, j = 1, \ldots, N$; the segmental mean STs
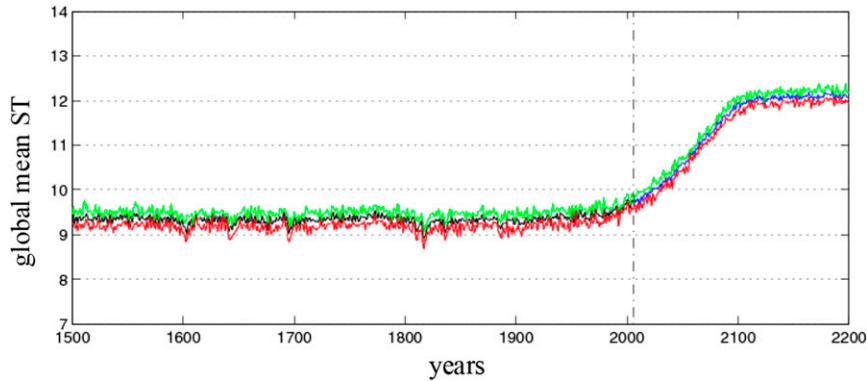
FIG. 1. Global mean surface temperature (°C) of last-millennium (showing only the period of 1500–2005) and A1B (2006–2200) simulations, separated by the vertical dashed–dotted line. Both experiments have five ensemble members; ensemble means are shown in black (to 2005) blue (2005 on); green (red) lines denote the upper (lower) boundaries given by the ensemble members.

and the related STDs form (x_j, σ_j) pairs (Fig. 2a) that are embedded in the new state space (Fig. 2b). Accordingly, the relation between the two moments, x_j and σ_j is assessed using the outlier-resistant nonparametric Theil–Sen regression estimator (Theil 1950a,b,c; Sen 1968) and characterizes the variability of the climate:

$$\sigma_j = \alpha x_j + b \quad j = 1 \ldots N.$$

The quantities (x_j, σ_j) refer to the mean ST and STD of segment S_j (Fig. 2a), α the regression slope, and b the error term or the noise; N equals the total number of segments in the time series. For a time series from the control simulation that comprises 3100 yearly data points, as is the case with Fig. 2a, given the segment length of 100 yr, we obtain $N = 31$ nonoverlapping segments. The resulting (x_j, σ_j) pairs with $j = 1, \ldots, 31$, correspond to 31 points in the mean ST–STD plane (black dots, Fig. 2b) with α equal to -0.68. Positive slopes ($\alpha > 0$) suggest STD increasing with increasing mean ST, while negative ones ($\alpha < 0$) indicate decreasing STD versus increasing mean ST. The significance of the slope is tested using a two-tailed t test ($p = 0.05$). It is shown in further sections that, in an unforced control simulation, such linear relations are globally observed (Figs. 5c,d); therefore, they reflect the internal rhythm
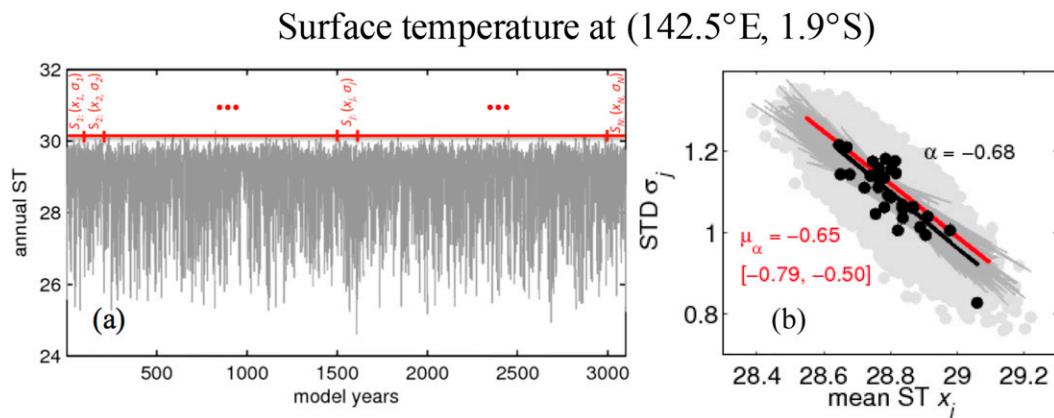

FIG. 2. Surface temperature at (1.9°S, 142.5°E) (control simulation) and its regression slope α: (a) 3100-yr annual ST time series and (b) mean ST–STD (standard deviation) slope. With a segment length of 100 yr, annual ST in (a) is split into N nonoverlapping segments S_j with mean ST and STD denoted as (x_j, σ_j), $j = 1, \ldots, N$, $N = 31$. These (x_j, σ_j) pairs determine the Theil–Sen regression slope $\alpha = -0.68$ [black in (b)]. In (b), bootstrap results are shown in gray, with the median slope $\mu_\alpha = -0.65$ constrained by the two-sided 95% Studentized bootstrap pivotal interval $[-0.79, -0.50]$; the red line denotes the linear fit with the median slope.

of the climate system and are termed intrinsic climate variability.

c. Bootstrap resampling

Since any GCM experiment represents only one realization of the true climate state, bootstrap resampling is introduced to improve the estimates of the ICV metric. The underlying assumption is that the statistical properties of the population can be inferred from a large number of samples by resampling the observed data (Efron 1981). The number of the replicas should be large enough to cover the whole range of the population; each replica has the same length as the observed data. These replicas produce a distribution of the statistic of interest, the regression slope in the context here. If the replicas are independent of each other, the central limit theorem holds: that is, the location parameter of the distribution can approximate the statistic (regression slope) of the population, regardless of the underlying distribution; its spread, often known as the standard error and described by confidence intervals, decreases with increasing sample size and thus corresponds to increasing precision of the estimate (Krzywinski and Altman 2013).

The MPI-ESM simulation is treated as the observed data. Take one time series in the control simulation as an example (Fig. 2a). The MPI-ESM time series is regarded as a record that harbors the information of the population. Then 1000 replicas are produced, each with the same data length of 3100 yr. Subsequently, the Theil–Sen regression analysis (following section 2c) is applied to each of the replicas, generating 1000 slope estimates, of which the median μ_a is to approximate the regression slope α of the MPI-ESM time series; the two-sided 95% studentized bootstrap pivotal interval following Carpenter and Bithell (2000) serves as its confidence interval. As expected, the median slope $\mu_a = -0.65$ is reasonably close to that of the MPI-ESM time series ($\alpha = -0.68$) (Fig. 2b); the 95% confidence interval [$-0.79, -0.50$] measures the related precision.

In this study, the intentionally biased bootstrap method (Hall and Presnell 1999), or block bootstrapping, is employed to generate the replica time series, in which blocks of observations, rather than individual data points, are resampled to preserve the data dependence or memory (e.g., Hall et al. 1995; Kitamura 1997). The block length is essentially determined by memory properties inherent to the system, which result from physical and dynamical processes active at different time scales, such as the El Niño–Southern Oscillation (ENSO) that is dominant on inter-annual time scales (Tziperman et al. 1994), the Pacific decadal oscillation (Mantua et al. 1997), and the across-scale long-term memory (Zhu et al. 2006, 2010).

As a rough estimate of the memory, we calculate the decorrelation time scale, approximated by the time lag when the autocorrelation function first becomes negative. It is between 1 and 3 yr in most of the tropical region, below 10 yr over continents, and between 10 and 20 yr in mid-to-high-latitude oceans; small patches of long decorrelation time scales (30–35 yr) are found in the Okhotsk Sea, in the Greenland Sea, and along the eastern coast of Antarctica (not shown). Hence, to set the block length to a value beyond 20 yr does not lose generality. In our analysis, the block length is set to 31 (20) yr for the control (last-millennium and A1B) simulation, such that 100 randomly selected blocks of 31 yearly values form one replica for the control simulation, and 250 (25) blocks of 20 yearly values for the last-millennium (A1B) equilibrium period.

d. Sample size and precision of slope estimation

The central limit theorem indicates that the bigger the sample size is, the more precise the estimate is (Krzywinski and Altman 2013), and the narrower the related confidence interval (as a measure of the precision) becomes. Therefore, estimates should always be discussed together with the related precision. This aspect is especially important for the A1B equilibrium period (2100–99) owing to its rather small sample size for determining regression slopes: 5 segments of 100-yr length (out of 5 ensemble members), in comparison to 31 for the control simulation and 50 for the last-millennium simulation.

To examine how the (un)certainty in the slope estimate evolves with the sample size, we choose a time series (1.9°S, 142.5°E) from the last-millennium simulation period (800–1799), which has a regression slope $\alpha = -0.70$ (Fig. 3a; the thick dotted–dashed line in Fig. 3b). Its bootstrap median slope μ_a equals -0.63 and is bounded by the 95% studentized confidence interval [$-0.78, -0.47$] (block length 20 yr), which says that with a 95% probability the true regression slope will lie within this range. Bootstrapping is also applied for bootstrap lengths of 3000, 1500, and 500, corresponding to the first 600, 300, and 100 yr of the millennium period, respectively (Fig. 3b). As expected, the shorter the bootstrap length is (the smaller the sample size), the wider the confidence interval becomes (thus, the worse the precision of the estimate is).

If we take the 95% confidence interval for the full millennium period as criteria (given by edges of the first box in Fig. 3b), the above test can be used to detect changes in ICV. If an ICV slope (from a different time period) lies within this interval, the system can be regarded to be statistically unchanged. This is the case in Fig. 3b, where median slopes estimated for different

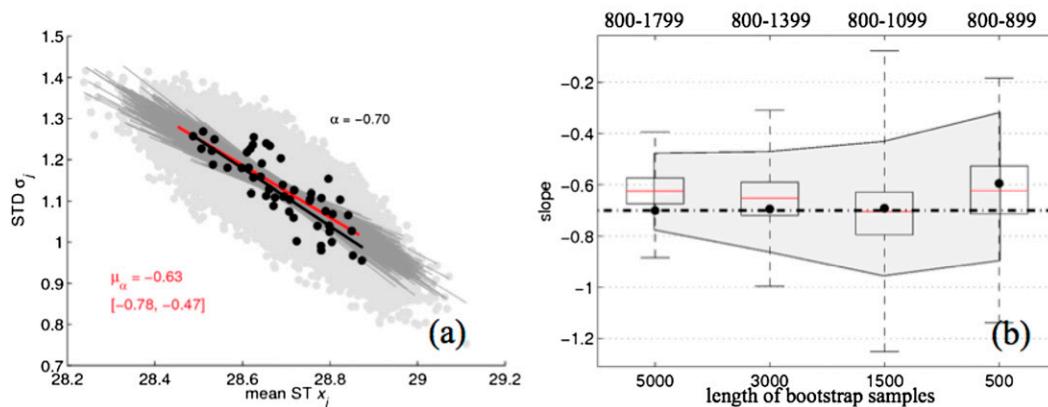## Surface temperature at (142.5°E, 1.9°S) (*last millennium*)

FIG. 3. Precision of slope estimation in relation to sample size: (a) mean ST–STD slope, as in Fig. 2b, but for the time series from the last-millennium simulation at [1.9°S, 142.5°E]; $\alpha = -0.70$, denoted in (b) by the thick dashed line; and (b) boxplot of bootstrap slope estimates for the periods of 800–1799, 800–1399, 800–1099, and 800–99; the respective 95% confidence intervals are gray shaded; red lines (black dots) denote the respective median bootstrap (MPI-ESM) slope; box edges mark the 25th and 75th quantiles of the bootstrap slopes.

bootstrap lengths all lie within these criteria (and in Fig. 6b brown hatched areas). On the contrary, as soon as the ICV slope exceeds this interval, it indicates that significant changes have emerged above the internal variability of the current climate (see areas hatched in green in Fig. 6c).

We notice that median slopes (red lines in Fig. 3b) do not always lie closer to the true value (thick dotted–dashed line) than the respective MPI-ESM slopes (dots). In fact, the benefit of using bootstrap median slopes, instead of the estimates solely from the MPI-ESM simulation, becomes preferable when the sample size is limited. We have compared the deviations of the median slopes and the MPI-ESM slopes of the corresponding periods with respect to the true slopes (the MPI-ESM slope of the full millennium period) on a global scale (not shown). The former outperforms the latter on 32% of the global grid points when the first 600 yr (800–1399) are considered for bootstrapping, and this outperformance ratio increases to 75% when only the first 100 yr (800–99) are considered. Therefore, the overall increase of the outperformance ratio globally is at the expense of precision at each grid point.

3. Distribution symmetry and extremes in relation to ICV

The linear correlation between the mean STs and STDs suggests that the yearly ST is not normally distributed. In fact, non-Gaussian distributions have been reported in various climate variables (e.g., White 1980; Fraedrich and Kietzig 1983; Sardeshmukh and Sura

2009; Ogata et al. 2013; Ng et al. 2014). Moreover, strong skewness may account for the covariance between the mean and the extremes (Huybers et al. 2014). Therefore, the following question naturally arises: Is the ICV metric solely a reflection of the underlying skewed distribution? And what are the implications of ICV for climate extremes? They are first discussed with two example time series (Fig. 4); spatial perspectives are presented in later sections.

a. ICV in relation to skewness

Two time series from the control simulation are first analyzed. One is located in the western (1.86°S, 142.5°E, $\alpha = -0.68$) and the other in the eastern (1.9°S, 108.8°W, $\alpha = 0.71$) equatorial Pacific (Figs. 4b,e). The Jarque–Bera test that considers both skewness and kurtosis is employed to test whether the sample data match a normal distribution (Jarque and Bera 1987). Both example time series failed the test at the 5% significance level and are thus non-normally distributed. In spite of the large magnitudes of both regression slopes, one time series is strongly skewed ($S = -1.2$), and the other is almost symmetric ($S = 0.1$) (Figs. 4a,d). Therefore, the strength of the linear mean ST–STD relation does not reveal whether the distribution is skewed or not.

Next, we try to examine the extent to which the skewness, or more precisely, the statistical property of the sample data, contributes to the reported linear mean ST–STD relation. For each MPI-ESM time series, we generate a surrogate dataset (10^6 data points) that shares the same statistical parameters as listed in Figs. 4a,d; bootstrapping and regression analysis are applied to the
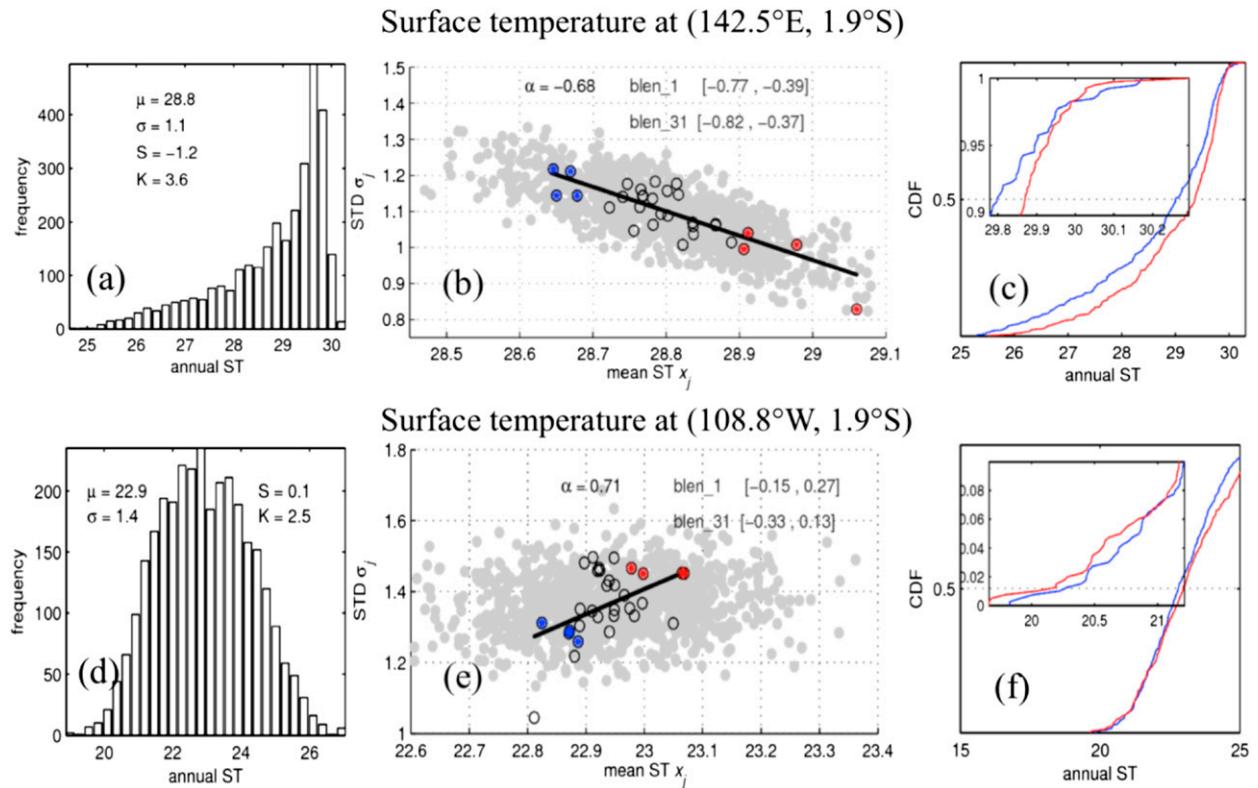
FIG. 4. Example ST time series at (a)–(c) 1.9°S, 142.5°E and (d)–(f) 1.9°S, 108.8°W (control simulation): (a),(d) histograms (first four moments listed); (b),(e) mean ST–STD slopes; (c),(f) cumulative distribution functions of yearly data in segments denoted as red–blue stars in (b),(e). Gray dots in (b),(e) show segmental mean STs and STDs of bootstrap samples sharing the same distribution as in (a),(d); the 95% confidence intervals for block length of 1 and 31 are listed. Distribution differences between blue and red lines in (c),(f) passed the Kolmogorov–Smirnov test at the 1% significance level; insets show the highest and lowest 10%.

surrogate data to estimate the distribution-determined ICV slopes; regression slopes of the 1000 bootstrap samples provide the 95% confidence bounds (listed in Figs. 4b and 4e, for block lengths of 1 and 31 yr, respectively). As expected, since there is no correlation between any neighboring data points in the surrogate data, the block length does not greatly influence the confidence bounds.

Random samples that are strongly skewed tend to produce large slope coefficients (Fig. 4b): the two bounds of the 95% confidence interval are considerably below zero in the case of strong negative skewness (Fig. 4b), whereas they are more or less equally distributed around zero in the case of the symmetric distribution (Fig. 4e). In fact, in the skewed case, it is impossible to distinguish the true parameter $\alpha = -0.68$ from the slopes of random samples, whereas, in the symmetric case, the regression slope $\alpha = 0.71$ lies well beyond the distribution-determined ranges.

Nonetheless, as will be shown in section 5, in some regions, even in the presence of strong skewness, the ICV regression slopes may have large magnitudes well

beyond the distribution-determined confidence intervals. Therefore, the ICV metric is determined by physical and dynamical processes and should not be understood from a purely statistical point of view.

b. ICV and climate extremes

Implications of the ICV metric for climate extremes are demonstrated with the same time series shown in Fig. 4. Four points at each end of the regression line (blue and red circles in Figs. 4b,e) are selected to construct the cumulative distribution function (Figs. 4c,f). Note that each point on the mean ST–STD plane (Figs. 4b,e) represents statistics of a segment containing 100 yearly points; thus, 400 yearly values in total form the respective cumulative distribution function. At both locations, the distribution change between the warm (red) and cold (blue) cases passes the two-sample Kolmogorov–Smirnov test at the 1% significance level (Figs. 4c,f). We further note that changes of different quantiles are of different magnitudes. In the case of Fig. 4c, the cumulative distribution function exhibits a general rightward shift in a warmer climate (red), while

the warmest 3% of the samples shift slightly leftward (see inset). It is noteworthy that the horizontal displacement of warm quantiles is smaller than that of cold quantiles. The opposite is observed in Fig. 4f: warm (cold) extremes become warmer (colder), and the horizontal shift of warm extremes is bigger than that of cold extremes. In short, changes in temperature extremes appear to be asymmetric. Consider the example in Fig. 4f, which is located in the Niño index region. When defining El Niño–Southern Oscillation events with a quantile–fixed SST threshold, an increase in intensity and occurrence of both El Niño and La Niña is to be expected in a warmer climate, with stronger increase in El Niño events.

We wish to stress that the asymmetric structure of climate extremes in presence of strong regression slopes is also found in other parts of the globe (not shown) and is thus not limited to the equatorial Pacific region. Keeping in mind that the analysis above is carried out in the undisturbed control simulation, the mean ST–STD linear relation and the related asymmetric structure of temperature extremes are thus intrinsic features of the climate system. Assuming that the climates under new forcings (the last-millennium and A1B simulations) are not fundamentally different from the undisturbed climate state (the control simulation), we expect to observe a similar mean ST–STD linear relation under different equilibrium states, and our task is to identify possible changes in terms of the ICV metric in these climate states with respect to our current climate, which may enable us to infer changes in the internal dynamics. Note that the term ''warmer climate'' refers to locally (segmentally) warmer mean ST (with respect to the equilibrium mean temperature) that occurs, owing to intrinsic variability of the system (rather than transient responses to varying external forcing).

4. ICV in the unforced control simulation: Spatial features

The climatological distributions of mean STs and STDs, calculated as averages of all 31 segments, are shown in Figs. 5a,b. The STD intensity is linked to local dynamical processes. For example, the STD maximum of about 1.4 K in the equatorial central Pacific Ocean and the local maxima in Brazil, Paraguay, Uruguay, and Australia reflect typical ENSO-related patterns (Müller and Roeckner 2008); intense variability of similar origin also exists over Africa. In the Polar Regions, sea ice variability is mainly responsible for the STD maxima in the Barents Sea, the Sea of Okhotsk, and around the coastline of Greenland, as well as in the Ross Sea and Weddell Sea (Bye et al. 2013, 2016). In contrast, the open Southern Ocean basins are characterized by low

variability with STD between 0.3 and 0.6 K. The striking pattern of the annual mean STD indicates the balance between polar and equatorial influences that characterizes the present climate, through a series of teleconnections, including the North Atlantic Oscillation, southern annular mode, and others, as summarized in section 3.6 of IPCC (2007).

We note the following from the spatial features of the ICV (Figs. 5c,d): (i) A greater area of the tropical ocean has a negative slope tendency (Fig. 5c), with the biggest patch of negative slopes over the Pacific warm pool region and the eastern Indian Ocean as well as over Australia. (ii) Negative slopes are also visible in high-latitude oceans: over the east of Japan and the Bering Sea and over the open ocean south and east of Greenland, extending to the Norwegian Sea. In these areas, sea ice processes are presumably involved, because, for instance, sea ice thickness in the Norwegian Sea and the Barents Sea also exhibits a linear relation between its segmental means (100-yr averages) and the related STDs (not shown). (iii) Positive slopes occur in the Arctic and over a large portion of high-latitude land areas: the Canadian Arctic Archipelago, central Asia, and the Antarctic coastal regions. Positive slopes are also observed in low-latitude land areas, including central Africa, northern South America, and the Indian peninsula, which indicates strong influences of ENSO-related teleconnections (Zheng 2014).

We now examine on a global scale whether and to what extent the ICV features in the control simulation (Figs. 5c,d) are due to the underlying distribution (Figs. 5e,f). While nonnormal distribution is globally observed (shaded in Fig. 5e), significant regression slopes that are beyond the local 95% reference ranges determined by the local distributions (shaded in Fig. 5f) still retain most of the spatial features displayed in Figs. 5c and 5d. This is particularly true in the tropical and high-latitude oceans, where the ICV metric is beyond what to expect from the background distribution, independent of whether it is normally distributed or not. Therefore, it is too simplistic to perceive the ICV features (Figs. 5c,d) merely from a statistical point of view; its characteristics are closely linked to the status of the climate system and should be understood as expressions of the underlying dynamical and physical processes.

5. ICV in the last-millennium and A1B simulations: Spatial features

In this section, we first introduce the ICV features in the last-millennium simulation. Based on this uniquely long experiment, the discussion regarding the sample size in relation to the precision of slope estimation, as
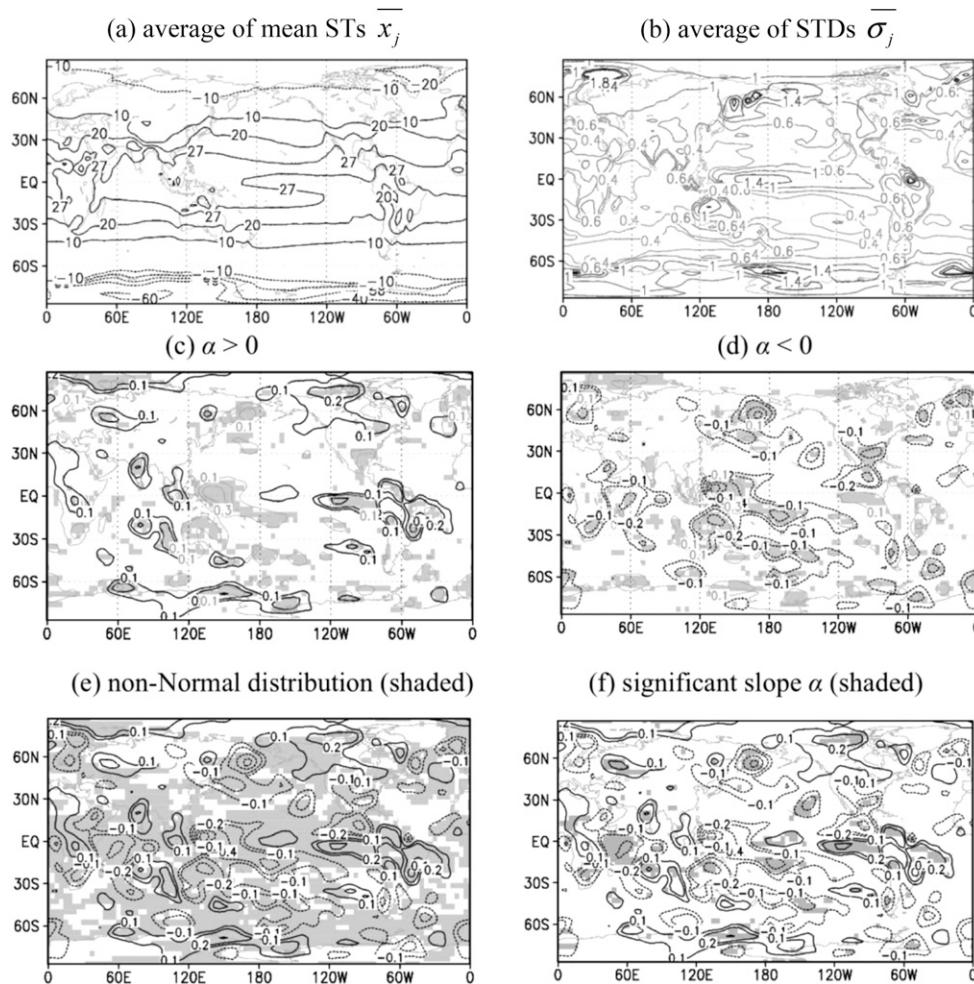
FIG. 5. Spatial patterns of ST and regression slope α (control simulation): (a) average of segmental mean STs (°C) and (b) segmental STDs (K), ($\overline{x}_j$, $\overline{\sigma}_j$); the overbar denotes averaging over all 31 segments of 100-yr length. (c)–(f) Regression slopes contoured at values of -0.8, -0.6, -0.4, -0.2, -0.1, 0.1, 0.2, 0.4, 0.6, and 0.8 for: (c) $\alpha > 0$ and (d) $\alpha < 0$; in (c),(d), gray lines denote the r-squared values of linear fitting; significant slopes ($p = 0.05$, a two-tailed t test) are shaded; shaded areas in (e) are areas where ST is non-normally distributed ($p = 0.05$, the Jarque–Bera test); in (f), shaded areas are significant slopes that are beyond distribution-determined signals ($p = 0.05$).

has been demonstrated in Fig. 3, is extended to a global scale. We demonstrate that bootstrapping with a limited sample size (5 segments of 100-yr length for the period of 800–99; five ensemble members are concatenated) reasonably reproduces the characteristics of the ICV metric in the full last-millennium period (section 5a); it is then applied to the A1B equilibrium simulation to estimate the ICV slope in a warmer climate; significant changes in the ICV metric are reported (section 5b).

a. ICV in the last-millennium simulation

The ICV metric shows noteworthy differences from that in the unforced control simulation (contour line in Fig. 6a vs Figs. 5c,d): (i) The area of negative slopes distinctly expands into the central tropical Pacific and

the entire Indian Ocean; the southern edge of this area extends toward 30°S in the last-millennium run (compared to about 15°S in the control simulation). (ii) Negative slopes now occur in the tropical Atlantic, extending southeastward and reaching the Africa continent. (iii) Positive slopes occur poleward of 60°S, in particular over Antarctica, while the coverage of positive slopes north of 60°N shrinks considerably (e.g., in the Barents Sea and Kara Sea, over central Eurasia and Canada).

Since, in contrast to the constant forcing in the control run, the last-millennium simulation is forced by time-varying reconstructions of natural and anthropogenic forcing, the above-noted differences can then be interpreted as new features of the ICV that reflect a new
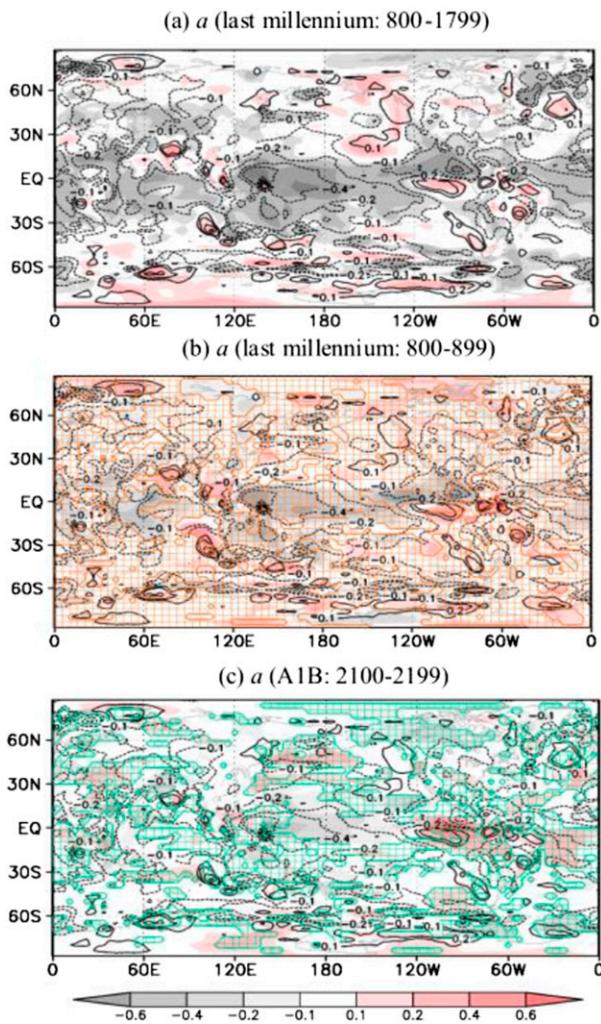
(a) a (last millennium: 800-1799)

(b) a (last millennium: 800-899)

(c) a (A1B: 2100-2199)

FIG. 6. Spatial patterns of regression slopes α for last-millennium [(a) 800–1799 and (b) 800–99] and A1B [(c) 2100–99] simulation. Median bootstrap slopes are shaded in gray and red; as reference, MPI-ESM slopes of 800–1799 are shown at color contour values of −0.6, −0.4, −0.2, −0.1, 0.1, 0.2, 0.4, and 0.6. Hatched areas in (b), (c) are median slopes lying inside (brown)–outside (green) of the 2.5th and 97.5th percentiles of bootstrap slopes for the full millennium period of 800–1799.

balance of feedbacks and interactions in the system as a response to the new forcing.

The bootstrapping method with the full bootstrap length (800–1799) rather faithfully reproduces the ICV metric of the MPI-ESM last-millennium simulation (shaded in Fig. 6a) in terms of magnitude and spatial coverage of both positive and negative slopes (Figs. 7a–c). In the following, we take the 95% confidence interval defined by the bootstrap samples (with the full bootstrap length) at each grid point as the local reference range: a regression slope being within or beyond this range is regarded to be statistically the same as (brown in Fig. 6b

and Figs. 7a–c) or significantly different from (green in Fig. 6c and Figs. 7d–f) the ICV metric of the MPI-ESM last-millennium simulation.

With a much shorter bootstrap length (800–99), the bootstrapping method can still reproduce the ICV metric to a large extent (areas where the median slopes are within the local 95% confidence intervals account for a spatial coverage of 75%, hatched in Fig. 6b). The good consistency is clearly seen in Figs. 7a–c. We also notice that bootstrapping with shorter bootstrap length tends to underestimate the coverage of negative slopes (Fig. 7b) and slightly overestimates the slope magnitude in the tropical region (Fig. 7c).

b. ICV in the A1B scenario period

The ICV metric of the A1B equilibrium period is shown in Fig. 6c (shaded). In comparison to the last-millennium simulation (contour line in Fig. 6c), the spatial feature of the regression slopes is largely retained: the strongest signals appear in the tropical region, and a distinct dipole structure occurs in the tropical Pacific, with negative slopes in the west and positive slopes in the east. On the other hand, considerable changes are observed globally (median slopes lying outside of the local 95% confidence intervals are hatched in green in Fig. 6c): (i) positive slopes expand to a bigger area in the tropics and subtropics; (ii) areas with negative slopes shrink accordingly (Figs. 7d,e), with reduced magnitude in the Indian Ocean and the western tropical Pacific (Fig. 6c); (iii) positive slopes intensify, in particular, in the tropical band and in the middle latitudes of the Southern Hemisphere (Fig. 7f); (iv) the expansion and enhancement of positive slopes is most obvious in the tropical region (including the eastern tropical Pacific, northern South America, and the southern tropical Atlantic), over Antarctica, and in the midlatitude oceans in the Northern Hemisphere. In particular, positive slopes prevail in the northwest Pacific region and in the central North Pacific, which are dominated by negative slopes in both the control and the last-millennium simulations (Fig. 6c vs Figs. 5c,d).

6. Conclusions and discussion

We propose a new climate metric to measure the linear relationship between mean surface temperature (ST) and its standard deviation (STD), which are estimated for segments of 100-yr length. This relationship is characterized by two opposing tendencies (slopes): that is, higher mean STs are related to enhanced or reduced STDs, and they are found on a global scale: (i) meridionally pronounced regression slopes cover the tropical and high-latitude oceans, with a prevailing coverage of
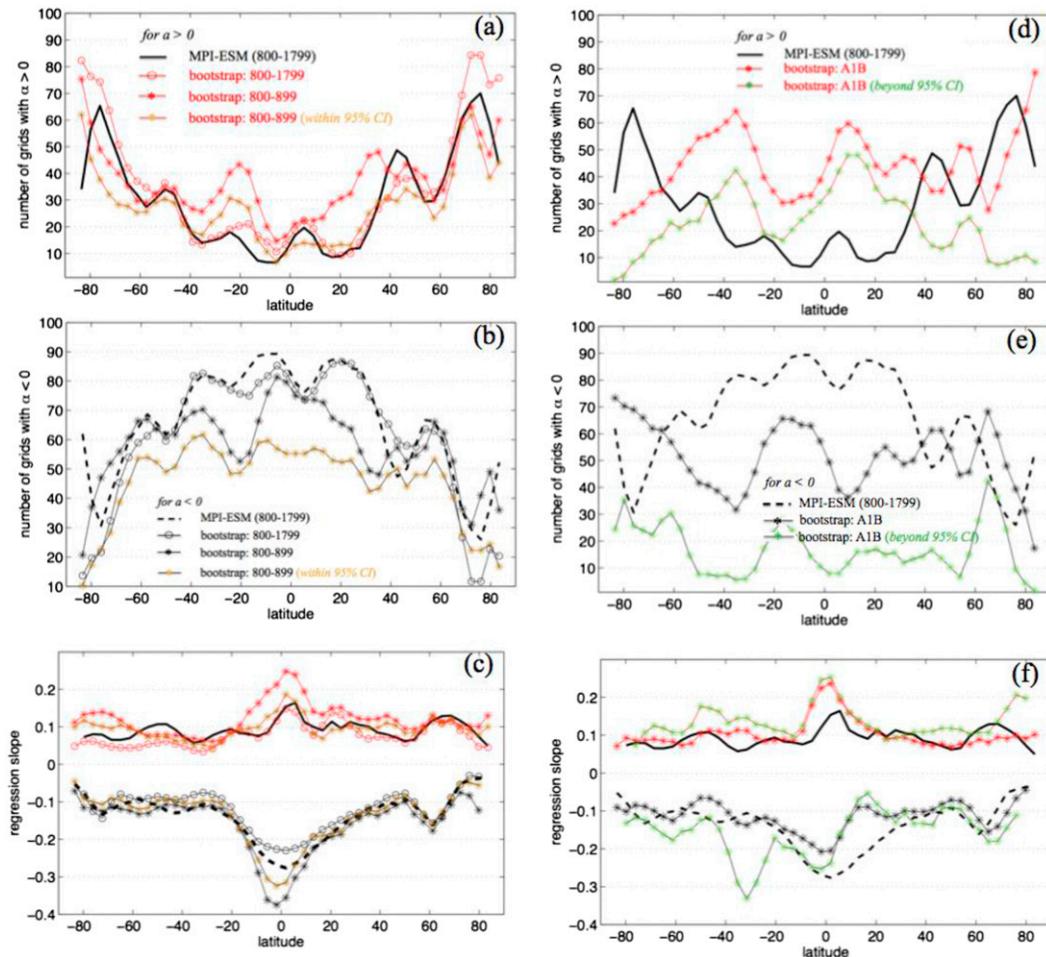
FIG. 7. Zonal statistics of regression slopes α from the (a)–(c) last-millennium and (d)–(f) A1B simulation. Zonal counts of (a),(d) positive and (b),(e) negative slopes; and (c),(f) zonal mean of slope magnitude. Thick black solid ($\alpha > 0$) and dashed ($\alpha < 0$) lines in all panels denote slopes of the MPI-ESM last-millennium experiment. Median slopes of bootstrapping samples are marked by circles for 800–799 and asterisks for a bootstrap length of 100 yr [(a)–(c) 800–99; (d)–(f) A1B] superimposed on thin red ($\alpha > 0$) and black ($\alpha < 0$) lines. The 2.5th and 97.5th percentiles, determined by bootstrap samples for the period of 800–1799, serve as the 95% confidence interval; slopes lying inside (outside) the interval are shown as brown (green) asterisks, corresponding to the hatched areas in Figs. 6b and 6c.

negative (positive) slopes in low-to-middle (high) latitudes (Figs. 7a–c); (ii) zonally, a distinct zonal dipole is found in the Indo-Pacific region, with negative slopes in the Indian Ocean and the western tropical Pacific and positive slopes in the eastern tropical Pacific. These spatial features indicate the balance between polar and equatorial influences and between the Indo-Pacific warm pool and the relatively colder eastern tropical Pacific. A series of teleconnections is likely involved, including the Hadley circulation, the Walker cells, and the ENSO mode (section 3.6 of IPCC 2007; Alexander et al. 2002; Tsonis et al. 2008). This linear relationship is also shaped by local processes, because such a relation is also observed in sea ice thickness in regions where sea

ice processes are active (e.g., in the Norwegian Sea, the Labrador Sea, and the Okhotsk Sea) (not shown). Since these features exist in the unforced control simulation, they are therefore intrinsic to the climate system, because of which they are termed intrinsic climate variability (ICV).

If evaluated for global mean surface temperature, the ICV metric has a weak negative value ($\alpha = -0.07$ and -0.12 in the control and the last-millennium simulation, respectively); thus, relatively warmer (colder) global climates are associated with weaker (stronger) variability. Interestingly, such a relation has been observed on glacial–interglacial time scales (e.g., Müller et al. 2005), thereby demonstrating the encouraging property of the ICV metric.

We further demonstrate that ICV retains its statistical structure to a great extent in the warmer future but with reduced (increased) coverage and magnitude of negative (positive) slopes. The difference occurs mainly in the tropical and subtropical regions, which suggests a decreased equator-to-pole temperature gradient and thus a weaker and wider Hadley circulation (not shown). Meanwhile, negative slopes in the Indo-Pacific warm pool region weaken, while positive slopes in the eastern tropical Pacific expand and enhance, which corresponds to a weaker zonal temperature gradient in the warmer A1B future. These new ICV features are substantiated by the weakening of the tropical circulation and reflect changes in the internal dynamics in response to new external (e.g., A1B) forcing, which will complicate attribution of anthropogenic warming-related signals (e.g., Vecchi et al. 2006; Vecchi and Soden 2007).

Pronounced ICV regression slopes are related to asymmetric changes in temperature extremes. We find that, in a warmer climate, both extreme El Niño and La Niña events will exhibit increased occurrences and enhanced intensity; moreover, the increase in occurrence and intensity is more pronounced in El Niño than in La Niña. Similar changes in temperature extremes have been reported in relation to anthropogenic warming (Cai et al. 2014, 2015; Kodra and Ganguly 2014), whereas our results suggest that the asymmetric structure of changes in climate extremes may reflect an intrinsic rhythm of the climate system; thus, relating them to anthropogenic forcing requires careful underpinning.

We acknowledge the deficiencies of and discrepancies among GCMs in capturing physical processes, such as ENSO, sea ice–related processes, and others, and have performed the same analysis for some undisturbed control runs of CMIP5 models (about 1000-yr long). Similar ICV features are observed, but not in terms of the exact grid-wise location and intensity (not shown). In fact, the revealed divergence among GCMs leads us to believe that the ICV may indeed be a suitable climate metric for comparing GCM performances by providing long-term statistics of model-simulated internal variability and by enabling one to infer information of higher moments.

Finally, we like to stress that the phrase "a warmer climate/future" refers to the fluctuating climate means, estimated by 100-yr averages in the context here, within an equilibrium state. This should not be confused with temperature differences between two different equilibrium states: for example, between the last-millennium simulation and the scenario simulation (Fig. 1). It is also important to note that in all analyses presented here no transitional signals are included. Therefore, the statistical features of the ICV reported here characterize the natural rhythm of the internal dynamics, which itself, as a matter of fact, evolves dynamically, reflecting the active balance of feedbacks and interactions of all processes under a set of specific forcing. To study ICV in the transitional period requires a justified separation of the externally forced trend, whether linear, quadratic, or even higher order. This aspect is being tested and planned for a future report.

REFERENCES

Alexander, M. A., I. Bladé, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott, 2002: The atmospheric bridge: The influences of ENSO teleconnection on air–sea interaction over the global oceans. *J. Climate*, **15**, 2205–2231, doi:10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2.

Ault, T. R., J. E. Cole, J. T. Overpeck, G. T. Pederson, S. St. George, B. Otto-Bliesner, C. A. Woodhouse, and C. Deser, 2013: The continuum of hydroclimate variability in western North America during the last millennium. *J. Climate*, **26**, 5863–5878, doi:10.1175/JCLI-D-11-00732.1.

Blanchard-Wrigglesworth, E., and C. M. Bitz, 2014: Characteristics of arctic sea-ice thickness variability in GCMs. *J. Climate*, **27**, 8244–8258, doi:10.1175/JCLI-D-14-00345.1.

Bye, J. A. T., K. Fraedrich, E. Kirk, S. Schubert, and X. Zhu, 2011: Random walk lengths of about 30 years in global climate. *Geophys. Res. Lett.*, **38**, L05806, doi:10.1029/2010GL046333.

——, ——, S. Schubert, and X. Zhu, 2013: The changing length of the warming period of the annual temperature cycle in the high latitudes under global warming. *Atmos.–Ocean*, **51**, 309–318, doi:10.1080/07055900.2013.793594.

——, ——, ——, and ——, 2016: Future climate in the Far North. *Reflections on the Far North of Canada in the Twenty-First Century*: *Interdisciplinary Perspectives*, J. Nittel and S.-H. Ditze, Eds., Bochum, in press.

Cai, W., and Coauthors, 2014: Increasing frequency of extreme El Niño events under greenhouse warming. *Nat. Climate Change*, **4**, 111–116, doi:10.1038/nclimate2100.

——, and Coauthors, 2015: Increased frequency of extreme La Niña events due to greenhouse warming. *Nat. Climate Change*, **5**, 132–137, doi:10.1038/nclimate2492.

Carpenter, J., and J. Bithell, 2000: Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Stat. Med.*, **19**, 1141–1164, doi:10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F.

Cassou, C., L. Terray, J. W. Hurrell, and C. Deser, 2004: North Atlantic winter climate regimes: Spatial asymmetry, stationarity with time, and oceanic forcing. *J. Climate*, **17**, 1055–1068, doi:10.1175/1520-0442(2004)017<1055:NAWCRS>2.0.CO;2.

Cobb, K. M., C. D. Charles, H. Cheng, and R. L. Edwards, 2003: El Niño/Southern Oscillation and tropical Pacific climate

during the last millennium. *Nature*, **424**, 271–276, doi:10.1038/nature01779.

Crowley, T. J., G. Zielinski, B. Vinther, R. Udisti, K. Kreutz, J. Cole-Dai, and E. Castellano, 2008: Volcanism and the Little Ice Age. *PAGES News*, Vol. 16, No. 2, PAGES International Project Office, Bern, Switzerland, 22–23.

Curry, J. A., and P. J. Webster, 2011: Climate science and the uncertainty monster. *Bull. Amer. Meteor. Soc.*, **92**, 1667–1682, doi:10.1175/2011BAMS3139.1.

Efron, B., 1981: Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, **68**, 589–599, doi:10.1093/biomet/68.3.589.

Fraedrich, K., and E. Kietzig, 1983: Statistical analysis and wavenumber-frequency spectra of the 500 mb geopotential along 50°S. *J. Atmos. Sci.*, **40**, 1037–1045, doi:10.1175/1520-0469(1983)040<1037:SAAWFS>2.0.CO;2.

Goosse, H., H. Renssen, A. Timmermann, and R. S. Bradley, 2005: Internal and forced climate variability during the last millennium: A model-data comparison using ensemble simulations. *Quat. Sci. Rev.*, **24**, 1345–1360, doi:10.1016/j.quascirev.2004.12.009.

Hall, P., and B. Presnell, 1999: Intentionally biased bootstrap methods. *J. Roy. Stat. Soc.*, **B61**, 143–158, doi:10.1111/1467-9868.00168.

——, J. Horowitz, and B.-Y. Jing, 1995: On blocking rules for the bootstrap and dependent data. *Biometrika*, **82**, 561–574, doi:10.1093/biomet/82.3.561.

Huber, M., and R. Knutti, 2012: Anthropogenic and natural warming inferred from changes in Earth's energy balance. *Nat. Geosci.*, **5**, 31–36, doi:10.1038/ngeo1327.

Huybers, P., and W. Curry, 2006: Links between annual, Milankovitch and continuum temperature variability. *Nature*, **441**, 329–332, doi:10.1038/nature04745.

——, K. A. McKinnon, A. Rhines, and M. Tingley, 2014: U.S. daily temperatures: The meaning of extremes in the context of nonnormality. *J. Climate*, **27**, 7368–7384, doi:10.1175/JCLI-D-14-00216.1.

IPCC, 2007: *Climate Change 2007: The Physical Science Basis.* Cambridge University Press, 996 pp.

Jarque, C. M., and A. K. Bera, 1987: A test for normality of observations and regression residuals. *Int. Stat. Rev.*, **55**, 163–172, doi:10.2307/1403192.

Jungclaus, J., and Coauthors, 2010: Climate and carbon-cycle variability over the last millennium. *Climate Past*, **6**, 723–737, doi:10.5194/cp-6-723-2010.

Kitamura, Y., 1997: Empirical likelihood methods with weakly dependent processes. *Ann. Stat.*, **25**, 2084–2102, doi:10.1214/aos/1069362388.

Kodra, E., and A. R. Ganguly, 2014: Asymmetry of projected increases in extreme temperature distributions. *Sci. Rep.*, **4**, 5884, doi:10.1038/srep05884.

Kosaka, Y., and S. Xie, 2013: Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, **501**, 403–407, doi:10.1038/nature12534.

Krzywinski, M., and N. Altman, 2013: Importance of being uncertain. *Nat. Methods*, **10**, 809–810, doi:10.1038/nmeth.2613.

MacFarling Meure, C., D. Etheridge, C. Trudinger, P. Steele, R. Langenfelds, T. van Ommen, A. Smith, and J. Elkins, 2006: Law Dome CO_2, CH_4 and N_2O ice core records extended to 2000 years BP. *Geophys. Res. Lett.*, **33**, L14810, doi:10.1029/2006GL026152.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on

salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1079, doi:10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

Meehl, G. A., A. Hu, and B. D. Santer, 2009: The mid-1970s climate shift in the Pacific and the relative roles of forced versus inherent decadal variability. *J. Climate*, **22**, 780–792, doi:10.1175/2008JCLI2552.1.

Müller, U. C., M. A. Geyh, J. Pross, and G. C. Bond, 2005: Cyclic climate fluctuations during the last interglacial in central Europe. *Geology*, **33**, 449–452, doi:10.1130/G21321.1.

Müller, W. A., and E. Roeckner, 2008: ENSO teleconnections in projections of future climate in ECHAM5/MIP-OM. *Climate Dyn.*, **31**, 533–549, doi:10.1007/s00382-007-0357-3.

Nakicenovic, N., and Coauthors, 2000: *Special Report on Emissions Scenarios.* Cambridge University Press, 599 pp.

Ng, B., W. Cai, and K. Walsh, 2014: Nonlinear feedbacks associated with the Indian Ocean dipole and their response to global warming in the GFDL-ESM2M coupled climate model. *J. Climate*, **27**, 3904–3919, doi:10.1175/JCLI-D-13-00527.1.

Ogata, T., S.-P. Xie, J. Lan, and X. Zheng, 2013: Importance of ocean dynamics for the skewness of the Indian Ocean dipole mode. *J. Climate*, **26**, 2145–2159, doi:10.1175/JCLI-D-11-00615.1.

Pongratz, J., C. H. Reick, T. Raddatz, and M. Claussen, 2008: A global land cover reconstruction AD 800 to 1992—Technical description. Max Planck Institute for Meteorology Rep. on Earth System Science 51, 72 pp.

Sardeshmukh, P. D., and P. Sura, 2009: Reconciling non-Gaussian climate statistics with linear dynamics. *J. Climate*, **22**, 1193–1207, doi:10.1175/2008JCLI2358.1.

Sen, P. K., 1968: Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Stat. Assoc.*, **63**, 1379–1389, doi:10.1080/01621459.1968.10480934.

Tebaldi, C., K. Hayhoe, J. M. Arblaster, and G. A. Meehl, 2006: Going to the extremes: An intercomparison of model-simulated historical and future changes in extreme events. *Climatic Change*, **79**, 185–211, doi:10.1007/s10584-006-9051-4.

Theil, H., 1950a: A rank-invariant method of linear and polynomial regression analysis I. *Nederl. Akad. Wet. Proc.*, **53**, 386–392.

——, 1950b: A rank-invariant method of linear and polynomial regression analysis II. *Nederl. Akad. Wet. Proc.*, **53**, 521–525.

——, 1950c: A rank-invariant method of linear and polynomial regression analysis III. *Nederl. Akad. Wet. Proc.*, **53**, 1397–1412.

Tsonis, A. A., K. L. Swanson, and G. Wang, 2008: On the role of atmospheric teleconnections in climate. *J. Climate*, **21**, 2990–3001, doi:10.1175/2007JCLI1907.1.

Tziperman, E., L. Stone, M. A. Cane, and H. Jarosh, 1994: El Niño chaos: Overlapping of resonances between the seasonal cycle and the Pacific Ocean–atmosphere oscillator. *Science*, **264**, 72–74, doi:10.1126/science.264.5155.72.

Vecchi, G. A., and B. J. Soden, 2007: Global warming and the weakening of the tropical circulation. *J. Climate*, **20**, 4316–4340, doi:10.1175/JCLI4258.1.

——, ——, A. T. Wittenberg, I. M. Held, A. Leetmaa, and M. Harrison, 2006: Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature*, **441**, 73–76, doi:10.1038/nature04744.

Vieira, L. E. A., S. K. Solanki, N. A. Krivova, and I. Usoskin, 2011: Evolution of the solar irradiance during the Holocene. *Astron. Astrophys.*, **531**, A6, doi:10.1051/0004-6361/201015843.

Volodin, E. M., and A. Y. Yurova, 2013: Summer temperature standard deviation, skewness and strong positive temperature anomalies in the present day climate and under global

warming conditions. *Climate Dyn.*, **40**, 1387–1398, doi:10.1007/
s00382-012-1447-4.

Waterman, S., and S. R. Jayne, 2011: Eddy-mean flow interactions in
the along-stream development of a western boundary current
jet: An idealized model study. *J. Phys. Oceanogr.*, **41**, 682–707,
doi:10.1175/2010JPO4477.1.

Watterson, I. G., and P. H. Whetton, 2011: Distributions of decadal
means of temperature and precipitation change under global
warming. *J. Geophys. Res.*, **116**, D07101, doi:10.1029/
2010JD014502.

White, G. H., 1980: Skewness, kurtosis, and extreme values
of Northern Hemisphere geopotential heights. *Mon. Wea.*

Rev., **108**, 1446–1455, doi:10.1175/1520-0493(1980)108<1446:
SKAEVO>2.0.CO;2.

Zheng, F., 2014: ENSO variability simulated by a coupled
general circulation model: ECHAM5/MPI-OM. *Atmos.
Oceanic Sci. Lett.*, **7**, 471–475, doi:10.1080/16742834.2014.
11447209.

Zhu, X., K. Fraedrich, and R. Blender, 2006: Variability regimes of
simulated Atlantic MOC. *Geophys. Res. Lett.*, **33**, L21603,
doi:10.1029/2006GL027291.

——, ——, Z. Liu, and R. Blender, 2010: A demonstration of long-
term memory and climate predictability. *J. Climate*, **23**, 5021–
5029, doi:10.1175/2010JCLI3370.1.