On the link between mean state biases and prediction skill in the tropics – An atmospheric perspective

INGO RICHTER, TAKESHI DOI, AND SWADHIN K. BEHERA

Application Laboratory, JAMSTEC, Yokohama, Japan

NOEL KEENLYSIDE

University of Bergen, Bergen, Norway

Climate Dynamics

submitted, 26 October 2016

revised, 17 April 2017

accepted, 13 July 2017

Corresponding author address:

Ingo Richter

Application Laboratory, JAMSTEC, 3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan

E-mail: richter@jamstec.go.jp

ABSTRACT

The present study examines how mean state biases in sea-surface temperature (SST), surface wind and precipitation affect model skill in reproducing surface wind and precipitation anomalies in the tropics. This is done using theoretical arguments, atmosphere-only experiments in the Coupled Model Intercomparison Project Phase 5 (CMIP5), and customized sensitivity tests with the SINTEX-F general circulation model. Theoretical arguments suggest that under certain conditions the root mean square error (RMSE) of a variable can be related to its variance and its mean, which indicates a direct link between bias and skill. The anomaly correlation coefficient (ACC), on the other hand, is generally not related to either the mean state or its variance, as several examples document. Multi-model atmosphere-only experiments with prescribed SST warming suggest that both ACC and RMSE of surface wind and precipitation are rather insensitive to warming on the order of 4 K. When SST biases from a free-running control simulation are prescribed in SINTEX-F, the ACC of surface wind is almost unaffected in the equatorial Pacific and Atlantic, while that of precipitation decreases noticeably in some regions but also increases in others. The RMSE of both fields shows widespread deterioration. There is a tendency for warm SST biases to increase the signal-to-noise ratio and sometimes ACC as well. The results suggest that, in the context of atmosphere-only simulations, improving SST and precipitation biases does not necessarily improve the skill in reproducing anomalies of surface wind and precipitation.

1. Introduction

The numerical simulation of weather and climate has made substantial progress over the last several decades (Edwards 2000; Richter et al. 2016). Nevertheless, systematic errors continue to pose a challenge to general circulation models (GCMs; de Szoeke and Xie 2008; Bellenger et al. 2013; Nagura et al. 2015; Richter et al. 2014a). While computational power has increased tremendously over the last few decades most climate models still cannot resolve scales below 100 km and even numerical weather prediction typically cannot resolve scales below 10 km. Due to these limitations to model resolution, many processes that occur on small spatial scales have to be parameterized. Among these processes are cumulus convection, boundary layer turbulence, and cloud microphysics. While such parameterizations have been reasonably successful, as demonstrated by the success of numerical weather prediction (NWP) in predicting weather and that of climate models in reproducing past and current climates, they necessarily involve the use of approximations, simplifications and ad-hoc assumptions, and also suffer from the limited availability of observational data. Thus deficiencies in parameterizations are thought to be the main cause of some of the persistent biases in GCMs, which, in the tropics, include errors in the mean position of the intertropical convergence zone (ITCZ; Li and Xie 2014), underrepresentation of low-level stratocumulus clouds (Richter 2015), and inadequate representation of the intraseasonal oscillation (Hung et al. 2013).

With the continuing increase in computing power it is possible that cumulus convection, whose representation requires a resolution of 5 km or higher, will be explicitly resolved over the next one to two decades, thus eliminating the need for cumulus parameterization. Boundary layer turbulence and cloud microphysics, on the other hand, require a model resolution that is several orders of magnitude higher and therefore will need to be parameterized even in the long term. This is one of the reasons why systematic model errors (or biases) will likely continue to be an issue in GCMs.

Much work has been done to identify biases and alleviate them. By way of motivation, such studies often state that GCM biases deteriorate the skill of seasonal predictions (as well as undermine confidence in global change projections). The often implicit assumption is that alleviating biases will lead to a more realistic representation of variability and more skillful predictions. Few studies, however, have thoroughly investigated this link between mean state bias and prediction skill. The ones that have been performed do point to a link but results are sometimes ambiguous. Perhaps the clearest evidence for a link comes from a study by Manganello and Huang (2009), who used a heat flux correction scheme to reduce sea-surface temperature (SST) errors in the eastern tropical Pacific. The flux correction, by design, drastically reduced the SST errors in the model but also led to more realistic SST variability in the eastern Pacific, with a peak in boreal winter, as observed, whereas the control simulation produced spurious peaks in spring and summer. Moreover, the improvement in mean and variability were accompanied by improved El Niño/Southern Oscillation (ENSO) prediction skill from lead month 6 onward. The authors linked the relatively poor skill in their control model to the spurious variability peak in summer: predictions initialized in January managed to grow SST anomalies until July but could not maintain them afterward. Similar results were obtained by Ding et al. (2015b) who found that climatological surface heat flux correction in a model with prescribed surface momentum anomalies dramatically increased the model’s ability to reproduce SST anomalies. They attributed the increased simulation skill to the influence of SST bias reduction on the climatology of surface wind stress and subsurface temperature.

Gualdi et al. (2005) show that increased atmospheric resolution in their model leads to both a reduced easterly surface wind bias over the equatorial Pacific and generally improved ENSO prediction skill. Lee et al. (2010) used a pattern correlation metric to examine the relation between mean state and prediction skill at 1-month lead time in the global tropics for a multi-model ensemble of reforecasts. They find that models with higher pattern correlation for the mean state also tend to have higher pattern correlation for the anomalies. This intermodel relation varies considerably depending on the season, and is more pronounced for SST than for precipitation. Magnusson et al. (2013) study the impact of heat and momentum flux correction on the prediction skill of a version of the European Centre for Medium Range Weather Forecasts (ECMWF) model. They obtain a slight improvement in SST anomaly correlation coefficient (ACC) in the eastern tropical Pacific at lead months 6 and 7 for the reforecasts with flux correction.

DelSole and Shukla (2010) used the DEMETER multi-model reforecasts to examine the related issue of whether skill is affected by model drift (i.e. the model's transition from observation-based initial conditions toward its biased equilibrium state during the forecast period). Based on the intermodel correlation between skill and bias over the first three forecast months they concluded that there is a robust inverse relation, particularly for the tropical Pacific. However, due to the fact that the intermodel correlations were based on only 7 models, there is some uncertainty regarding the results. It is also not clear to what extent the drift during the first three forecast months resembles the equilibrium bias.

There is no shortage of studies on the link between mean state biases and variability errors (e.g. Sperber and Palmer 1996; Guilyardi 2006; Spencer et al. 2007; Jin et al. 2008; Richter et al. 2014a; Ding et al. 2015a; Deppenmeier et al. 2016) but these studies typically do not examine how variability errors affect prediction skill. Thus it appears that much work remains to be done to study the link between seasonal prediction skill and model performance in terms of mean state and variability. This is an important issue with practical implications because a deeper understanding of the link between bias and prediction skill can help the community understand which improvement efforts are likely to yield the highest return in terms of added prediction skill. It might also inform us that some regions are not likely to benefit much from further model improvement, at least as far as seasonal prediction skill is concerned. The tropical Atlantic may turn out to be such a region. On the one hand, SST and surface wind biases are severe there (Davey et al. 2002; Richter et al. 2008; Richter et al. 2014a) and many prediction models still struggle to beat persistence forecasts, as can be seen in Fig. 1 for a model ensemble from the Climate Historical Forecast Project (CHFP) intercomparison (Kirtman and Pirani 2009). On the other hand, it has been shown that, despite severe biases, some models produce relatively realistic variability patterns (Richter et al. 2014a) and that the theoretical predictability may be much lower than in the tropical Pacific (Richter et al. 2014b). This suggests that some models are able to capture the relevant atmosphere-ocean coupling and that the limiting factor for forecast skill in the region might not be model biases but predictability. In this context it is instructive to consider the results of Tompkins and Feudale (2010), who showed that an enhanced network of ocean observations could improve the ECMWF forecast skill for the West African monsoon in the absence of any major model improvement.

Considering the issues discussed above we argue that it is important to obtain a deeper understanding of the link between mean state biases and prediction skill. The present study aims to take a first step in this direction by focusing on the ability of models to reproduce surface wind and precipitation anomalies when forced with observed SSTs, i.e. in an Atmospheric Model Intercomparison Project (AMIP)-style setting. One could consider this as a forecast at lead time 0, and it may provide an upper limit of the prediction skill one would expect to achieve. This statement should be qualified, however, because it has been shown that atmosphere-only experiments may misrepresent surface heat fluxes (Wang et al. 2005; Wu and Kirtmann 2005) and therefore coupling may increase precipitation skill at short lead times (Kang et al. 2004; Lee et al. 2010; DelSole and Shukla 2012).

Before examining model simulations, we will introduce the models and experiments used in section 2 and discuss some general considerations in section 3. To understand the impact of mean state SST warming on prediction skill we will examine three experiments from the Coupled Model Intercomparison Project Phase 5 (CMIP5) in section 4. These experiments prescribe observed SST from 1979-2008 but one experiment adds a spatially uniform 4K warming, while another one adds a patterned warming typical of global warming simulations. The purpose is to test the mean state dependence of the atmospheric response to SST anomalies. The impact of mean state biases will be further examined in dedicated sensitivity experiments with one particular model in section 5, while summary and conclusions will be given in section 6.

2. Method and experiment description

We examine the influence of biases on skill using two strategies. In the first, we study how skill is affected by intrinsic atmospheric biases, i.e. how a model's ability to reproduce surface wind and precipitation anomalies is affected by the respective biases in those fields. The second method is to examine how skill (in surface wind and precipitation) is affected by errors in the climatology of the SST boundary forcing. Details are given in the following.

2.1. CMIP5 experiments

We use AMIP-type simulations performed by several modeling centers for the CMIP5 model intercomparison. The basic experiment, called AMIP, consists of an atmospheric GCM (AGCM) forced with observed SST for the period 1979-2008. Monthly mean SSTs are interpolated to daily values. This experiment will be used to pursue the first strategy, i.e. exploring the influence of intrinsic atmospheric biases on skill. The SST boundary forcing is essentially the same in all models (except for interpolation errors) and thus we expect bias and skill of our fields of interest to be determined mainly by the atmospheric model component. It is, of course, possible that extraneous influences, e.g. from the land surface model obscure the bias-skill relation.

Two additional AMIP-type experiments in the CMIP5 archive allow us to pursue the second strategy, i.e. examining the sensitivity of skill to errors in the SST boundary forcing. In experiment amip4K a constant value of 4 K is added to the AMIP SST everywhere over the ice-free oceans. Experiment amipFuture adds a warming pattern that is also constant in time, but varies in space. The pattern is derived from the ensemble average of several global warming simulations and thus includes, among others, warming that is enhanced at the equator. The original purpose of amip4K and amipFuture was to explore various aspects of climate change. In the present study, on the other hand, we are not concerned with climate change, but rather use these experiments as a convenient way to investigate how the surface wind and precipitation responses are affected by unrealistic SST values ("unrealistic" in the sense of being inconsistent with the present-day greenhouse gas forcing applied in the models), with the standard AMIP experiment serving as our control.

To compare experiments, we calculate the ensemble average over a set of 11 models (Table 1), which is the largest subset that performed all three experiments. The resulting time series spans the period 1979-2008 (360 monthly means). The multi-model climatology and anomalies are calculated based on this ensemble average. For experiment AMIP, we also analyze intermodel correlations, with climatology and anomalies calculated for each model separately.

2.2. Sensitivity experiments with SINTEX-F

While the SSTs in amip4K and amipFuture provide a good opportunity to explore surface wind and precipitation sensitivity to the mean state, the SST distributions are quite different from typical model biases. To test specifically how SST biases affect the ability of a model to reproduce surface wind and precipitation we perform experiments with one particular model, the SINTEX-F GCM. This model was developed under the European Union-Japan collaboration project (Luo et al. 2003) and is based on the European SINTEX model (Gualdi et al. 2003). The version used here consists of the ECHAM 4.6 AGCM (Roeckner et al. 1996), the OPA 8.2 oceanic GCM (OGCM; Madec et al. 1998), and the OASIS 2.4 coupler (Valcke et al. 2000). The atmospheric resolution is T106 (approximately 1.1°), with 19 vertical levels, 4-5 of which are inside the planetary boundary layer. The oceanic resolution is 2° x 2° with the meridional resolution increasing to 0.5° at the equator. The OGCM has 31 vertical levels, 19 of which lie within the top 400m.

The control experiment (CTRL hereafter) is a simulation in which SSTs are strongly restored to the Optimally Interpolated SST (OISST; Reynolds et al. 2002) observations from 1982 to 2014. The strong restoring results in SST boundary conditions that are very similar to an AMIP-type simulation but may differ from observations on the order of 0.1 K. CTRL comprises 9 ensemble members, which are generated by using three restoring time scales (1-day, 2-day, 3-day) and three settings for the surface momentum flux formulation (Luo et al. 2005).

We perform two sensitivity tests, both of which use SSTs from a free-running 500-year control simulation (FR-CTRL). In the first experiment (Atl_bias), SST biases from FR-CTRL are imposed on the tropical Atlantic between 30°S and 30°N. This is achieved by subtracting OISST climatology (stratified by calendar month) from the original SST boundary conditions and adding the corresponding FR-CTRL climatology values. Thus the SST anomalies are the same as in OISST but the mean state in the tropical Atlantic is that of FR-CTRL and therefore features all the biases of the latter. In the second experiment (Pac_bias), an analogous procedure is applied to the tropical Pacific between 30°S and 30°N. Experiments with SST restoring in various regions using SINTEX-F were also performed by other authors, e.g. Sasaki et al. (2015), but none of these restored the SST to a biased state. The sensitivity experiments also consist of 9 ensemble members, which were generated by perturbing the SST boundary conditions with random values of 0.01 K amplitude.

Our reference data set for surface wind is the European Centre for Medium Range Weather Forecasts (ECMWF) Interim reanalysis (Dee et al. 2011). For precipitation we use the Global Precipitation Climatology Project (GPCP) version 2.2, which is a blend of satellite and station data (Adler et al. 2003).

3. General considerations

3.1. Relation between MSE, ACC, and biases

Here we examine whether there is an explicit mathematical link between biases (in both mean and variability) on the one hand and prediction skill on the other. The measures of prediction skill examined here are the anomaly correlation coefficient (ACC) and the mean square error (MSE). ACC is defined as the Pearson correlation coefficient between the predicted (p) and observed anomalies (o):

$$r(p,o) = \frac{\sum_{i=1}^{n}(p_i - \bar{p})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^{n}(p_i - \bar{p})^2}\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2}}$$

where the overbar denotes the seasonally stratified climatological time average, p_i and o_i are the seasonally stratified total values, and i is the time index. Likewise, MSE is defined through the following equation:

$$MSE^*(p,o) = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2,$$

where the asterisk indicates the use of total fields in the calculation of MSE. ACC and MSE* are related through the following equation (e.g. Barnston 1992):

$$MSE^*(p,o) = std^2(p) + std^2(o) - 2std(p)std(o)r(p,o) + b^2(p) \quad (1)$$

where p and o denote the total fields, *std* is standard deviation, and b(p) is the mean bias of the prediction, i.e. $b = \bar{p} - \bar{o}$. In (1), MSE* and bias are explicitly related because the total fields are used. Once the seasonal mean is removed, as is routinely done in seasonal prediction, the bias term drops out and (1) becomes

$$MSE(p,o) = std^2(p) + std^2(o) - 2std(p)std(o)r(p,o) \quad (2)$$

Eq. 2 shows that MSE decreases with increasing ACC. Further, if the predicted variance is much larger than the observed one, it is easy to see that the first term on the right-hand side of (2) dominates (independent of ACC). In this case, MSE is essentially determined by the standard deviation of the prediction model. Likewise, if *std(o)* >> *std(p)* then *std(o)* dominates MSE.

In the case of *std(p)* = *std(o)*, i.e. predicted standard deviation is error-free, (2) can be rearranged to give $MSE(p,o) = 2std^2(o)[1 - r(p,o)]$, which states that MSE is a simple function of ACC and the observed variance (which equals the predicted variance).

Last, if ACC is close to 1, (2) can be approximated as $MSE(p,o) = std^2(p) + std^2(o) - 2std(p)std(o)$, or, after some manipulation,

$$\mathrm{RMSE}(p,o) = |std(p) - std(o)| \quad (3)$$

where RMSE is the root mean square error and the vertical bars denote the absolute value function. (3) suggests that, for ACC close to 1, RMSE is proportional to the absolute difference between the predicted and observed standard deviations. We test this relation for precipitation in the Niño 3.4 region (170º-120º W, 5º S-5ºN) using the AMIP multi-model ensemble. Figure 2a scatters RMSE versus the RHS of (3). The ACC is 0.9 or higher in most models so that the approximation used to derive (3) holds reasonably well. This is borne out by the high intermodel correlation coefficient of 0.96. RMSE is a little higher than what would be expected if (3) were to hold exactly, presumably due to the ACC being less than one (see Fig. 3a).

It is intuitively obvious that errors in the predicted variance will affect RMSE and so the high degree of intermodel correlation is essentially down to the ACC being consistently high across models. However, for an intermittent, positive definite variable like precipitation there is another aspect to Eq. (3), which may link skill to the mean state. Since precipitation is often zero but never less than zero, areas of high precipitation in the mean are also often areas of high precipitation variability. We thus suspect that regions with a wet precipitation bias also feature excessive variance and a high RMSE. Assuming an exact relation between mean and variance, Eq. (3) can be transformed into

$$RMSE(p,o) = c \cdot |\bar{p} - \bar{o}| = c \cdot |b(p)| \quad (4)$$

where c is a constant relating standard deviation of precipitation to its mean. To what extent this simple relation holds is examined in Fig. 2b, where, for each model, the RMSE of precipitation is scattered against the bias in the Niño 3.4 region. The intermodel correlation is 0.61, which is significant above the 99% level. As in Fig. 2a, there is an offset in the relation, which indicates the influence of the ACC.

3.2. Empirical link between ACC and bias

We have seen that, for the Niño 3.4 region, there is a moderately strong relation between mean state bias and RMSE of precipitation in regions where ACC is high. A more interesting question is whether there also is a relation between mean state and ACC. It can be seen from the definition of ACC that any linear transformation of the operands (i.e. observed and predicted time series) will leave its value unchanged. Intuitively one might expect severe biases (such as an ITCZ location bias) to affect ACC but such mean state biases do not figure into the mathematical definition of ACC (excepting the pathological case in which at least one of the time series is constant). Thus, based on the definition of ACC, there is no a-priori reason to expect that it should be influenced by errors in the mean state or variance. We can nevertheless examine whether there is empirical evidence for a link. To get a global view we show (seasonally unstratified) ACC and the annual mean bias of precipitation for the AMIP multi-model mean (Fig. 3a). There is a tendency for high ACC to be accompanied by small biases, particularly in the central and eastern equatorial Pacific and northern Indian Ocean. Other areas, like the equatorial Atlantic, tend to show the opposite behavior, i.e. high ACC accompanied by strong biases. The pattern correlation is 0.15, which is not significant at the 5% level. Surface zonal wind (Fig. 3b) also shows a variety of patterns, with high ACC in the eastern Indian Ocean almost collocated with the maximum easterly wind bias, while in the western equatorial Atlantic high ACC exists where the bias is very small. The pattern correlation between the two fields is -0.20, which is significant at the 5% level.

We further examine the relationship between bias and ACC in terms of intermodel spread. The Niño 3.4 June-July-August (JJA) precipitation is overestimated in most AMIP models and this anticorrelates with ACC at -0.68 (Fig. 4b). For precipitation over the equatorial Atlantic, on the other hand, the intermodel correlation is weakly positive, indicating a slight tendency for models with larger biases to have higher ACC (Fig. 4a, c). No significant relation exists for the equatorial Indian Ocean. Precipitation indices over monsoon regions (West Africa, South America, and India) show a similarly weak intermodel correlation (Fig. 4d-f). Furthermore, the general lack of useful prediction skill is striking. If the threshold is set at 0.5, a rather generous value, useful skill is only reached by one model each for West Africa and South America, while no model shows useful skill over India. As shown by previous studies, the absence of coupled feedbacks in AMIP experiments reduces the models' ability to reproduce precipitation patterns (Kang et al. 2004; Lee et al. 2010; DelSole and Shukla 2012) and thus skill may be somewhat higher in a seasonal prediction setting. Nevertheless, the generally low skill in these state-of-the-art models underscores the challenge of predicting rainfall in the monsoon regions, consistent with the study by Wang et al. (2009). As stated in their study, land surface initialization may be one way of improving skill for those regions.

A few examples of the ACC-bias relation for surface zonal wind are given in Fig. 5. For the western equatorial Atlantic (40-20ºW, 2ºS-2ºN; WEA hereafter) most models show the familiar westerly bias in March-April-May (MAM; Fig. 5a) as documented in Richter et al. (2008) and Richter et al. (2014a). Despite noticeable biases many models achieve an ACC of 0.8 or higher. A systematic relation between bias and ACC is not discernible. For the Niño 4 region (160ºE-150ºW, 5ºS-5ºN) in MAM, models are about evenly split into groups of westerly and easterly biases (Fig. 5b). There is a weak tendency for models with stronger easterly mean wind to have higher ACC (as indicated by the intermodel correlation of -0.44), even when the mean wind is more easterly than observed. The JJA surface zonal winds over the equatorial Indian Ocean (50-95ºE, 5ºS-5ºN) are too easterly in most models (Fig. 5c) and this is moderately correlated with a decrease in ACC (intermodel correlation 0.48). Overall, both precipitation and surface zonal wind tend to have higher ACC in those models with smaller biases but this is true only in some regions of the global tropics and many counter examples exist, most notably the equatorial Atlantic.

The tropical Pacific is known to have worldwide teleconnections (e.g. Horel and Wallace 1981) and thus one might expect to find that tropical Pacific precipitation biases adversely affect skill in other regions of the world. We examine this by repeating the analysis in Figs. 4 and 5 but with precipitation averaged over the Niño 3.4 region on the x-axis (not shown). Of the six regions shown in Fig. 4, only three show significant inter-model correlations. Apart from the tropical Pacific (see Fig. 4b) these are the tropical Indian Ocean and the Indian monsoon index (inter model correlations -0.42 and -0.37, respectively). For the latter, individual ACCs are low (see Fig. 4f), so that the tropical Indian Ocean relation appears to be the most interesting. According to this relation, models with pronounced positive precipitation bias in the Niño 3.4 region during boreal fall tend to have very low ACC for precipitation over the tropical Indian Ocean. Relating the surface wind ACCs shown in Fig. 5 to mean precipitation in the Niño 3.4 region does not produce any significant correlations. A more detailed investigation into the remote impact of tropical Pacific biases, particularly for the Indian Ocean, might produce interesting results but is out of scope for the present study.

4. Comparison of AMIP, amip4K, and amipFuture

In this section we examine multi-model ensemble means of three experiments in the CMIP5 archive (see Table 1 for a list of ensemble members). In amip4K, the prescribed SSTs are uniformly warmed by 4 K, relative to AMIP (Fig. 6b). This leads to a noticeable precipitation increase over the tropical Pacific and a more moderate increase over the tropical Atlantic and Indian Oceans. In amipFuture a typical global warming pattern is added to the AMIP SST, with enhanced warming in the deep tropics that is about 1 K warmer than in amip4K. Precipitation appears to respond in a non-linear way as values over the equatorial regions increase markedly compared to amip4K. Perhaps the most striking difference is seen over the eastern equatorial Pacific, where precipitation above 3 mm/day extends much farther east than in either AMIP or amip4K. The non-linear response in amipFuture is consistent with the result of previous studies that demonstrate the importance of the equatorial warming enhancement to rainfall patterns (Xie et al. 2010; Sobel and Camargo 2012; Huang et al. 2013).

4.1. Impact of SST warming on skill in the tropical Pacific

The ACC of Niño 4 surface zonal wind is almost unchanged in amip4K and amipFuture (Fig. 7a), despite pronounced precipitation changes (Fig. 6). In all experiments ACC is quite high and generally ranges between 0.8 and 0.9. This illustrates the strong influence of SSTs on surface winds in the tropical Pacific (consistent with Lindzen and Nigam 1987, and Zebiak and Cane 1987), an important part of the ENSO feedback loop (Bjerknes 1969; Neelin et al. 1998). ACC is even higher for precipitation in the Niño 3.4 region (Fig. 7b) with values up to 0.98 in AMIP and amip4K. Here amipFuture shows a noticeable decrease relative to the other two experiments, with ACC reduced by as much as 0.1 in April and May though this decrease is not statistically significant at the 95% level, based on Fisher's z transformation.

The higher SSTs in amip4K and amipFuture give rise to more intense precipitation over the tropics, which should increase variability and thus RMSE (see section 3). Moreover, since convection enhances the surface wind response to SST anomalies (Zebiak 1986; Richter et al. 2016) we also expect to see an increase in surface wind variability and, potentially, its RMSE. There is an opposing effect from the increase of atmospheric moisture and the stabilization of the atmospheric column (Held and Soden 2006). In the experiments under consideration, however, the variance of surface zonal wind does increase in the tropics (not shown). Consequently, the RMSE of surface zonal wind in the Niño 4 region (Fig. 7c) shows a more obvious skill deterioration as was the case for ACC, particularly for amipFuture. Nevertheless, the increase of RMSE relative to AMIP does not exceed 25% and none of the differences are significant at the 95% level, according to an F-test. For Niño 3.4 precipitation, the RMSE is more seriously affected (Fig. 7d), with values increasing by about a factor of 3 in amipFuture (statistically significant for all months). In amip4K the increase is only 10-20% and is only statistically significant in December.

The deterioration of precipitation RMSE in amipFuture points to a nonlinear response of precipitation and its variability in the equatorial Pacific region. Such nonlinearity is considered to play an important part in the skewness of ENSO (e.g. An and Jin 2004; Frauen and Dommenget 2010) and in the changes of ENSO variability under global warming (Power et al. 2013; Zheng et al. 2016). It is remarkable, however, that ACC is not much affected by this nonlinearity.

The area with the strongest decline in ACC is the central equatorial Pacific in MAM (not shown). Analysis of this area (160-120ºW, 5ºS-5ºN) reveals that observations, amip and amip4K all feature a highly non-linear relationship between precipitation and underlying SST (not shown): as SSTs exceed 28 ºC (32 ºC in the case of amip4K) the precipitation response becomes much more pronounced. In amipFuture, on the other hand, the relation is approximately linear, which leads to its decreased ACC.

4.2. Impact of SST warming on skill in the tropical Atlantic

In the WEA, the ACC of surface zonal wind anomalies is above 0.7 from April through June (Fig. 8a), with substantially lower values in other months. For the most part, ACC is very similar across experiments though AMIP tends to have higher values in boreal fall. These differences, however, are not statistically significant and, furthermore, skill in all three experiments is well below the usefulness level.

The ACC of precipitation over the equatorial Atlantic (50ºW-10ºE, 5ºS-5ºN; EQATL hereafter) exceeds 0.8 from May through July in all three experiments (Fig. 8b), which indicates a robust response to the pronounced SST anomalies that occur in that season (e.g. Carton and Huang 1994; Xie and Carton 1994; Richter et al. 2014a). In other months, ACC in the warming experiments both rises above and drops below the AMIP reference so that, on the whole, SST warming appears to have no systematic effect on the ACC of equatorial Atlantic precipitation.

RMSE of surface zonal wind over the western equatorial Atlantic tends to improve in the warming experiments from March through June (Fig. 8c), while in other months differences tend to be very small. None of these changes are statistically significant.

The RMSE of precipitation, on the other hand, increases significantly in the warming experiments, as expected from the increased mean (Fig. 6) and variability (not shown). This is particularly evident in amipFuture, where differences are statistically significant in several months.

5. Sensitivity tests with SINTEX-F

We first discuss the annual mean climatology of the SINTEX-F experiments. Biases generally have significant seasonal variability, particularly in the tropical Atlantic (Richter and Xie 2008). Nevertheless, the annual mean biases already feature many of the salient model errors. In the interest of brevity, we therefore discuss the annual mean biases only.

In experiment Atl_bias the tropical Atlantic SST of the SINTEX-F AMIP-like CTRL experiment is replaced with the biased climatology of the free running control simulation. This leads to annual mean SST in the eastern tropical Atlantic being up to 3 K warmer than in either CTRL or the observations (Fig. 9c). The Atlantic ITCZ responds by broadening meridionally, as can be seen from the increased precipitation south of the equator in Fig. 9c. Precipitation also changes in some other regions, with increase in the equatorial Indian Ocean and decrease in the South Pacific Convergence Zone (SPCZ).

In experiment Pac_bias (Fig. 9d) the equatorial Pacific SSTs are warmer than observed, particularly in the eastern basin. The SPCZ intensifies, extends further eastward, and becomes more zonally oriented, while the north-equatorial ITCZ weakens. This leads to a pronounced double ITCZ structure, a common bias in GCMs (de Szoeke and Xie 2008; Li and Xie 2014). Precipitation in other basins is not affected much.

Tropical precipitation is overestimated in all three experiments, a common problem in GCMs (Richter et al. 2016) that is most apparent over the Pacific warm pool. Note that for the latter region the precipitation differences across the experiments are relatively small compared to their difference from observations.

Surface zonal wind is biased westerly over the central and eastern equatorial Pacific in CTRL (Fig. 9b). A westerly bias is also seen over the equatorial Atlantic, which is typical of most GCMs (Richter et al. 2008). When SST biases are prescribed in the tropical Atlantic the westerly bias intensifies (Fig. 9c). This demonstrates the amplification of westerly wind biases in AGCMs by SST biases, as shown for CMIP3 (Richter et al. 2008) and CMIP5 (Richter et al. 2014a) models. In Pac_bias the westerly wind bias over the equatorial Pacific deteriorates noticeably (Fig. 9d). This is consistent with the surface winds responding to the reduced zonal SST gradient.

5.1. Impact of SST biases on skill in the Tropical Pacific

The ACC of surface zonal wind in the Niño 4 region (Fig. 10a) shows that the skill of CTRL is comparable to that of the AMIP ensemble (Fig. 7a). Overall, skill scores in the equatorial Pacific are relatively robust to the presence of SST biases (Fig. 10). The ACC of Niño 4 surface zonal wind is essentially the same in all three experiments, with differences less than 0.05 that are not statistically significant at the 95% level (Fig. 10a). The ACC of Niño 3.4 precipitation (Fig. 10b) decreases significantly in January, May, and August but never by more than 0.1.

A horizontal map of ACC for precipitation in MAM is presented in Fig. 11. CTRL shows maximum ACC in the eastern equatorial Pacific. The difference plot (Fig. 11b) reveals that the Niño 3.4 region chosen for Fig. 10 includes areas of both significant increase and decrease in ACC. The most evident decrease is in the eastern tropical Pacific, where CTRL had high skill (Fig. 11a). Significantly increased ACC is found in the western equatorial Pacific, and in the far eastern Pacific centered at 10ºS and 10ºN. The changes in ACC very roughly correspond to those in mean precipitation in that both tend to decrease on the equator and increase away from it (Fig. 11b).

We examine the region of the largest ACC decrease (140-105ºW, 5ºS-5ºN; EEP hereafter) by scattering simulated vs. observed MAM precipitation (Fig. 12a). It is evident that both mean and variability of precipitation are reduced in Pac_bias (see also Fig. 11b). The two points with the highest precipitation in the GPCP observations correspond to the years 1983 and 1998, both of which followed exceptionally strong El Niño events. CTRL reproduces these high precipitation events fairly well while Pac_bias does not, which contributes to the drop of ACC from 0.95 in CTRL to 0.55 in Pac_bias.

Convection in the tropics is thought to be sensitive to the absolute value of the underlying SST (Graham and Barnett 1987) though there may be no critical threshold (Zhang 1993). We examine to what extent background SST changes in the EEP contribute to the drop in ACC by scattering precipitation versus SST, both averaged over the EEP (Fig. 12b). While the EEP has a warm bias in the annual mean (Fig. 9d), in MAM it is cooler than observed by about 0.3 K (Fig. 12b). Convection in the tropics can be sensitive to small SST changes, but closer inspection of Fig. 12b shows that, even for the same SST values, Pac_bias has much lower precipitation than CTRL or the observations. Furthermore, in Pac_bias, precipitation for some SST values below 28ºC turns out to be higher than that for SST above 28ºC. Thus the local SST change does not seem sufficient to explain the drastic reduction of mean precipitation in the region (Fig. 11b).

A meridional section averaged from 140-105ºW (Fig. 13) reveals an SST decrease of almost 1 K just north of the equator, which is partially offset by an increase south of the equator and therefore not apparent in the area average. Additionally, there is a warm bias of almost 2 K further poleward in both hemispheres for Pac_bias (Fig. 13). This is accompanied by anomalous subsidence and lower tropospheric divergence over the equatorial region and anomalous rising motion off the equator in both hemispheres. The analysis suggests that SST biases both in the EEP and in the subtropics create an environment in the EEP that is less conducive to convection. This makes it more susceptible to factors other than the underlying SST.

The RMSE of surface zonal wind in the Niño 4 region is almost unchanged in Pac_bias (Fig. 10c), with no significant differences in any month. The RMSE of precipitation in the Niño 3.4 region (Fig. 10d) significantly decreases in April and October. This improvement in RMSE is explained by the decreased variability in Pac_bias (not shown, but inferable from the precipitation decrease in Fig. 11b; see Eq. 3). There also is a pronounced and significant increase of RMSE in December.

It is interesting to note that the RMSE of precipitation consistently increases in the Atl_bias experiment (Fig. 10d) though this is only statistically significant in January and August. The result suggests remote influences on the Pacific from the severe tropical Atlantic SST bias.

5.2. Impact of SST biases on skill in the Tropical Atlantic

CTRL reproduces the surface zonal wind anomalies in the WEA with an ACC of approximately 0.8 from April through June (Fig. 14a), comparable to the performance of the AMIP multi-model ensemble (Fig. 8). Atl_bias features slightly lower ACC in those months but also higher ACC in other months. None of the differences are statistically significant. The ACC of precipitation for the EQATL index in CTRL is highest from May through July (Fig. 14b). The ACC in Atl_bias reduces in most months (significantly so in April, May and December). An increase of ACC occurs in August and September but is not statistically significant.

RMSE deteriorates more markedly than ACC (Fig. 14cd), with significant differences in many months. This is a consequence of the warm SST bias in Atl_bias, which increases both mean (Fig. 15) and variability (not shown) of precipitation, exacerbating the biases in CTRL. Excessive variability in precipitation directly leads to a high RMSE (Eq. 3) and, through its impact on wind variability, indirectly contributes to the high RMSE of that quantity.

The horizontal map of climatological precipitation in CTRL, averaged over April through June (AMJ; Fig. 15a), shows maximum precipitation off the West African coast at about 5ºN. ACC, on the other hand, is highest over the central equatorial Atlantic. In response to the warm bias in Atl_bias, the ITCZ shifts southward, resulting in precipitation decrease north of the equator and increase south of it (Fig. 15b). This behavior is roughly mirrored by the ACC, similarly to the Pacific case. Only some areas feature statistically significant changes. For the ACC of surface zonal wind, the horizontal map shows the highest skill in the western equatorial Atlantic off the coast of Northeast Brazil (Fig 15c). In Atl_bias, this area only shows a slight decrease (Fig. 15d and Fig. 14a) but further east the impact is more visible, though still not statistically significant.

We examine how the SST biases affect the precipitation response to equatorial warm events by compositing SST and precipitation anomalies in July for Atlantic Niño years (1984, 1988, 1991, 1995, 1996, 1999, 2008) and plot horizontal maps for the observations and the two experiments (Fig. 16). The observations show wet precipitation anomalies between the equator and 10ºN extending over northeast Brazil to the west and Africa to the east (Fig. 16a). The most pronounced rain anomalies occur off Northwest Africa and over the central equatorial Atlantic, while maximum SST anomalies occur in the eastern equatorial Atlantic (ATL3 region). Large areas of warm SST anomalies in the tropical southeast Atlantic are not accompanied by increased precipitation. CTRL, which is forced with essentially the same SST field, reproduces the precipitation response fairly well although precipitation anomalies are too intense (by a factor 3 approximately) and too narrow in the meridional direction (Fig. 16b). Precipitation anomalies are unrealistic in Atl_bias because, in addition to being excessive, they are shifted southeastward (Fig. 17c). To a first approximation, the precipitation response in Atl_bias just follows the underlying SST anomaly pattern, which is not the case in the observations and CTRL. Pattern correlation with observations for the area 50ºW-10ºE, 10ºS-10ºN yields 0.72 and -0.02 for CTRL and Atl_bias, respectively, thus confirming the visual impression. The drastic deterioration of the precipitation response in Atl_bias thus appears to be due to the unrealistic sensitivity to the underlying SST anomalies. This in turn, directly relates to the mean state SST bias, which is most severe in the southeastern tropical Atlantic (Fig. 9c) and thus creates an environment that is unrealistically conducive to deep convection.

The increased sensitivity to SST anomalies in Atl_bias, however, can also lead to increased skill by allowing a robust precipitation response to emerge that would otherwise be drowned out by atmospheric internal noise and remote influences. This is suggested by the increased ACC of precipitation over the tropical southeast Atlantic (Fig. 15b). We quantify this by calculating the signal-to-noise ratio (SNR) for the tropical Atlantic. An easy way to estimate SNR is to calculate ensemble mean variance divided by intra-ensemble variance. The result for AMJ confirms that SNR is indeed increased over the southeastern tropical Atlantic (shading in Fig. 17). This roughly corresponds with the SST bias during the season (contours in Fig. 17), though this relation is certainly complicated by other factors. Compositing events with anomalously high precipitation over the southeastern tropical Atlantic (10ºW-10ºE, 10ºS-0) confirms higher skill in Atl_bias for this particular region (not shown).

6. Summary

6.1. Summary

We have investigated the link between GCM biases and prediction skill in the tropics through theoretical considerations and AMIP-style sensitivity tests. Our metrics for model skill have been RMSE and ACC, and we have applied these to the variability of surface winds and precipitation.

Taking the well-known relation among RMSE, standard deviation and ACC as our starting point (Eq. 2), we have shown that, if ACC is close to 1, there is a simple relation between RMSE and the observed and simulated standard deviations that holds for any field. Thus, RMSE becomes a simple function of the simulated standard deviations. For areas in which models have consistently high skill, such as the equatorial Pacific, this relation clearly emerges in a multi-model scatter plot of AMIP models (Fig. 2a). For a positive definite field like precipitation, there is a relatively close relation between mean and variability. This establishes a link between the mean and RMSE of precipitation or, in other words, bias and skill. A multi-model scatter plot suggests that this relation holds reasonably well for the equatorial Pacific.

By definition, ACC is not explicitly related to the mean state and, consistently, multi-model plots scattering the ACC against bias of precipitation do not reveal a systematic link, except for the equatorial Pacific (Figs. 3, 4b). For the three monsoon regions examined, the scatter plots also suggest a general absence of useful prediction skill, although the lack of coupled feedbacks in AMIP may contribute to this. Equatorial surface zonal winds do not show a strong relation between mean and ACC either, though there is some suggestion of a link for the equatorial Indian Ocean (Fig. 5c). Biases over the tropical Pacific appear to have some negative remote impact on skill over the Indian Ocean but more analysis will be needed to substantiate this relation.

Multi-model AMIP-style simulations with prescribed warming patterns over the global oceans indicate that ACC and RMSE are rather insensitive to SST changes on the order of 4 K. Only for precipitation the RMSE deteriorates noticeably due to the excessive variability that results from the warming.

In two sensitivity experiments with the SINTEX-F GCM, SST biases from a free-running control simulation were prescribed over either the tropical Atlantic or the tropical

Pacific, while leaving the anomalies as in the control simulation. The ACC of surface zonal wind is mostly unaffected in some key equatorial regions. Precipitation shows some more obvious decrease, particularly over the equatorial Atlantic. The RMSE of precipitation and surface zonal wind deteriorates noticeably over the equatorial Atlantic because the variability of these fields increases. Conversely, RMSE is not affected significantly over the equatorial Pacific, where mean precipitation and its variability tend to decrease.

Composite analysis of equatorial Atlantic warm events (Atlantic Niños) reveals that the warm SST biases in the eastern tropical Atlantic are associated with excessive sensitivity of precipitation anomalies to the underlying SST. This suggests that the unrealistically warm SST produce an environment conducive to deep convection that reacts very sensitively to warm SST anomalies, even when observations show no such sensitivity.

While the excessive sensitivity to local SST anomalies often deteriorates the skill of precipitation, it can also increase it under certain circumstances. This appears to be the case for the southeastern tropical Atlantic where the signal-to-noise ratio is increased over warm SST biases. The spatial pattern of ACC (15ab) supports this notion because it shows increased skill in the southeastern tropical Atlantic. In the eastern equatorial Pacific, on the other hand, cold SST biases are accompanied by a reduction in mean precipitation during MAM, and ACC decreases (Fig. 11). Recent studies indicate that the signal-to-noise ratio is underestimated in climate models (e.g. Eade et al. 2014; Scaife et al. 2014) and thus warm SST biases may be able to compensate for this deficiency in some scenarios.

6.2. Discussion

Our results indicate that there is generally no straightforward linear correspondence between mean state biases in SST, precipitation and surface wind on the one hand, and the ability of a model to reproduce surface wind and precipitation anomalies on the other. Particularly the skill for surface wind seems largely unaffected by the mean state SST. For precipitation, there is some indication that cool SST biases reduce the signal-to-noise ratio and skill. This relation, however, can also work in the opposite direction, i.e. skill increases when the mean SSTs are positively biased. To summarize this, our results indicate that reducing the amplitude of SST biases does not necessarily lead to increased skill.

That SST biases affect precipitation more than surface wind can be explained by the non-linearity of precipitation. Changes in the SST distribution have a strong influence on whether a region permits or does not permit deep convection. This was evident in the Pacific bias experiment, where local cold biases and off-equatorial warm biases conspired to effectively suppress the precipitation response to warm SST anomalies in eastern equatorial Pacific. Conversely, the warm SST bias in the southeast Atlantic produced an environment that was unrealistically conducive to convection, leading precipitation to respond to SST anomalies where it would not in nature.

Our sensitivity tests assess the impact of mean state SST biases only and thus assume that variability patterns remain unaffected. This will generally not be the case for coupled seasonal predictions because mean state wind biases will change, among others, the simulated temperature stratification of the oceans and therefore the areas of strong air-sea coupling. Thus SST variability patterns and their timing may change significantly in coupled prediction experiments and, for the case of free running coupled simulations, such changes are well documented (e.g. Richter et al. 2014a for the tropical Atlantic, and Bellenger et al. 2013 for the tropical Pacific). Therefore, the impact of model biases on coupled prediction runs cannot be addressed here.

We stress that our results do not suggest that reducing surface wind and precipitation biases is futile. Rather we have shown that, in the narrow context of AMIP-like experiments, one cannot necessarily expect increased skill from improving mean state biases.

Acknowledgments

The authors would like to thank Prof. Shang-Ping Xie for helpful comments on the manuscript. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison which provides coordinating support and led development of software infrastructure for CMIP, and the climate modeling groups for making available their model output. The authors thank the two anonymous reviewers for their constructive comments, which helped to improve the manuscript.

References

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P (2003) The version 2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-Present). J Hydrometeor 4:1147–1167

An, SI, Jin F-F (2004) Nonlinearity and Asymmetry of ENSO. J Clim 17:2399–2412.

Barnston AG (1992) Correspondence among the correlation, RMSE, and Heidke forecast verification measures: refinement of the Heidke score. Weather Forecast 7:699–709

Bellenger H, Guilyardi E, Leloup J, Lengaigne M, Vialard J (2013) ENSO representation in climate models: from CMIP3 to CMIP5. Clim Dyn, doi:10.1007/s00382-013-1783-z.

Bjerknes J (1969) Atmospheric teleconnections from the equatorial Pacific. Mon Weather Rev 97:163-172

Carton JA, Huang B (1994) Warm events in the tropical Atlantic. J Phys Oceanogr 24:888–903

Davey MK et al (2002) STOIC: a study of coupled model climatology and variability in topical ocean regions. Clim Dyn 18:403–420

Dee DP et al (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137:553–597

DelSole T, Shukla J (2010) Model fidelity versus skill in seasonal forecasting. J Clim 23:4794-4806

DelSole T, Shukla J (2012) Climate models produce skillful predictions of Indian summer monsoon rainfall. Geophys Res Lett 39:L09703. doi:10.1029/2012GL051279

Deppenmeier A-L, Haarsma RJ, Hazeleger W (2016) The Bjerknes feedback in the tropical Atlantic in CMIP5 models. Clim Dyn 7:2691. doi:10.1007/s00382-016-2992-z

de Szoeke SP, Xie S-P (2008) The tropical eastern Pacific seasonal cycle: Assessment of errors and mechanisms in IPCC AR4 coupled ocean-atmosphere general circulation models. J Clim 21:2573-2590

Ding H, Keenlyside N, Latif M, Park W, Wahl S (2015a) The impact of mean state errors on equatorial Atlantic interannual variability in a climate model. J Geophys Res Oceans. doi:10.1002/201 4JC010384

Ding H, Greatbatch RJ, Latif M, Park W (2015b) The impact of sea surface temperature bias on equatorial Atlantic interannual variability in partially coupled model experiments. Geophys Res Lett 42:5540–5546.

Eade R, Smith D, Scaife A, Wallace E, Dunstone N, Hermanson L, Robinson N. Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?. Geophys Res Let 41:5620-5628, doi:10.1002/2014GL061146

Edwards PN (2000) A brief history of atmospheric general circulation modeling. General Circulation Model Development: Past, Present, and Future, D. A. Randall, Ed., Academic Press, 67–90

Frauen C, Dommenget D (2010) El Nino and La Nina amplitude asymmetry caused by atmospheric feedbacks. Geophys Res Lett 37:L18801

Graham NE, Barnett TP (1987) Sea surface temperature, surface wind divergence, and convection over tropical oceans. Science 238:657–659

Gualdi S, Navarra A, Guilyardi E, Delecluse P (2003) Assessment of the tropical Indo-Pacific climate in the SINTEX CGCM. Ann Geophys 46:1–26

Gualdi S, Alessandri A, Navarra A (2005) Impact of atmospheric horizontal resolution on El Niño Southern Oscillation forecasts. Tellus 57A:357–374

Guilyardi E (2006) El Nino-mean state-seasonal cycle interactions in a mutil-model ensemble. Clim Dyn 26:329–248

Held IM, Soden BJ (2006) Robust responses of the hydrological cycle to global warming. J Clim 19:5686–5699. doi:10.1175/JCLI3990.1

Horel JD, Wallace JM (1981) Planetary-scale atmospheric phenomena associated with the Southern Oscillation. Mon Wea Rev 109:813–829

Huang P, Xie SP, Hu K, Huang G, Huang R (2013) Patterns of the seasonal response of tropical rainfall to global warming. Nat Geosci 6:357–361

Hung MP, Lin JL, Wang W, Kim D, Shinoda T, Weaver SJ (2013) MJO and convectively coupled equatorial waves simulated by CMIP5 climate models. J Clim 26:6185–6214

Jin EK, et al (2008) Current status of ENSO prediction skill in coupled ocean-atmosphere models. Clim Dyn 31:647–664

Kang IS, Lee JY, Park CK (2004) Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. J Clim 17:834–844

Kirtman B, Pirani A (2009) The State of the Art of Seasonal Prediction Outcomes and Recommendations from the First World Climate Research Program (WCRP) Workshop on Seasonal Prediction. Bull Am Meteor Soc, doi: 10.1175/2008BAMS2707.1

Lee JY, Wang B, Kang IS, Shukla J et al (2010) How are seasonal prediction skills related to models' performance on mean state and annual cycle? Clim Dyn 35:267–283

Li G, Xie S-P (2014) Tropical biases in CMIP5 multimodel ensemble: the excessive equatorial Pacific cold tongue and double ITCZ problems. J Clim 27:1765–1780

Lindzen RS, Nigam S (1987) On the role of the sea surface temperature gradients in forcing the low-level winds and convergence in the tropics. J Atmos Sci 44:2418– 2436

Luo JJ, Masson S, Behera SK, Gualdi S, Navarra A, Yamagata T (2003) South Pacific origin of the decadal ENSO-like variation as simulated by a coupled GCM. Geophys Res Lett 30:2250. doi:10.1029/2003GL018649

Luo JJ, Masson S, Behera SK, Shingu S, Yamagata T (2005) Seasonal climate predictability in a coupled AOGCM using a different approach for ensemble forecast. J Clim 18:4474–4497

Madec G, Delecluse P, Imbard M, Levy C (1998) OPA 8.1 ocean general circulation model reference manual. Tech. Rep. Note 11, LODYC/IPSL, Paris, France

Magnusson L, Alonso-Balmaseda M, Corti S, Molteni F, Stockdale T (2013) Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. Clim Dyn 41(9–10):2393–2409. doi:10.1007/s00382-012-1599-2

Manganello JV, Huang B (2009) The influence of systematic errors in the Southeast Pacific on ENSO variability and prediction in a coupled GCM. Clim Dyn 32:1015–1034

Nagura M, Sasaki W, Tozuka T, Luo J-J, Behera SK, Yamagata T (2013) Longitudinal biases in the Seychelles Dome simulated by 35 ocean-atmosphere coupled general circulation models. J Geophys Res Oceans, 118:831-846, doi:10.1029/2012JC008352

Neelin, JD, Battisti DS, Hirst AC, Jin F-F, Wakata Y, Yamagata T, Zebiak S (1998) ENSO Theory. J Geophys Res 103:14261-14290

Power S, Delage F, Chung C, Kociuba G, Keay K (2013) Robust twenty-first-century projections of El Niño and related precipitation variability. Nature 502(7472):541–545

Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002) An improved in situ and satellite SST analysis for climate. J Clim 15:1609-1625

Richter I, Xie S-P (2008) On the origin of equatorial Atlantic biases in coupled general circulation models. Clim Dyn 31:587–598

Richter I, Xie S-P, Behera SK, Doi T, Masumoto Y (2014a) Equatorial Atlantic variability and its relation to mean state biases in CMIP5. Clim Dyn 42:171–188. doi:10.1007/s00382-012-1624-5

Richter I, Behera SK, Doi T, Taguchi B, Masumoto Y, Xie S-P (2014b) What controls equatorial Atlantic winds in boreal spring? Clim Dyn 43(11):3091–3104

Richter I (2015) Climate model biases in the eastern tropical oceans: causes, impacts and ways forward. WIREs Clim Change 6: 345-358

Richter I, Chang P, Xu Z, Doi T, Kataoka T, Nagura M, Oettli P, de Szoeke S, Tozuka T, (2016) An overview of coupled GCM performance in the tropics. in Indo-Pacific Climate Variability and Predictability (Vol. 8), T. Yamagata and S. K. Behera, eds.

Roeckner E, Arpe K, Bengtsson L, Christoph M, Claussen M, Dümenil L, Esch M, Giorgetta M, Schlese U, Schulzweida U (1996) The atmospheric general circulation model ECHAM-4: model description and simulation of present-day climate. Tech. Rep. No. 218, Max-Planck-Institut für Meteorologie, Hamburg, Germany

Sasaki W, Doi T, Richards KJ, Masumoto Y (2015) The influence of ENSO on the equatorial Atlantic precipitation through the Walker circulation in a CGCM. Clim Dyn 44:191-202

Scaife AA, Arribas A, Blockley E, Brookshaw A, Clark RT, Dunstone N, Eade R, Fereday D, Folland CK, Gordon M, Hermanson L, Knight JR, Lea DJ, MacLachlan C, Maidens A, Martin M, Peterson AK, Smith D, Vellinga M, Wallace E, Waters J, Williams A (2014) Skillful long-range prediction of European and North American winters. Geophys Res Lett 41:2014GL059, doi:10.1002/2014gl059637

Sobel AH, Camargo SJ (2012) Projected future seasonal changes in tropical summer climate. J Clim 24:473–487

Spencer H, Sutton R, Slingo JM (2007) El Nino in a coupled climate model: sensitivity to changes in mean state induced by heat flux and wind stress corrections. J Clim 20:2273–2298

Sperber KR, Palmer TN (1996) Interannual tropical rainfall variability in general circulation model simulations associated with the atmospheric model intercomparison project. J Clim 9:2727-2750

Tompkins AM, Feudale L (2010) Seasonal ensemble predictions of West African Monsoon precipitation in the ECMWF system 3 with a focus on the AMMA special observing period in 2006. Wea Forecasting 25:768–788

Valcke S, Terray L, Piacentini A (2000) The OASIS coupler user guide version 2.4. Tech. Rep. TR/CGMC/00-10, CERFACE, Toulouse, France

Wang B, Ding QH, Fu XH, Kang IS, Jin K, Shukla J, Doblas-Reyes F (2005a) Fundamental challenge in simulation and prediction of summer monsoon rainfall. Geophys Res Lett 32:L15711

Wang B, Lee J-Y, Kang I-S, Shukla J, Park C-K, Kumar A. Schemm J, Cocke S, Kug J.-S, Luo J-J, Zhou T, Wang B, Fu X, Yun W-T, Alves O, Jin EK, Kinter J. Kirtman B, Krishnamurti T, Lau NC, Lau W, Liu P, Pegion P, Rosati T, Schubert S, Stern W, Suarez M, Yamagata T (2009) Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). Clim Dyn. doi: 10.1007/s00382-008-0460-0

Wu R, Kirtman B (2005) Roles of Indian and Pacific Ocean air–sea coupling in tropical atmospheric variability. Clim Dyn 25:155–170. doi:10.1007/s00382-005-0003-x

Xie S-P, Carton JA (2004) Tropical Atlantic variability: Patterns, mechanisms, and impacts. In *Earth Climate: The Ocean-Atmosphere Interaction*, C. Wang, S.-P. Xie and J.A. Carton (eds.), Geophysical Monograph, 147, AGU, Washington D.C., 121-142.

Xie SP, Deser C, Vecchi G, Ma J, Teng H, Wittenberg A (2010) Global warming pattern formation: sea surface temperature and rainfall. J Clim 23:966–986

Zebiak SE (1986) Atmospheric convergence feedback in a simple model for El Niño. Mon Wea Rev 114:1263-1271

Zebiak SE, Cane A (1987) A model El Niño-Southern oscillation. Mon Weather Rev 115:2262–2278

Zhang C (1993) Large-scale variability of atmospheric deep convection in relation to sea surface temperature in the tropics. J Clim 6: 1898-1913

Zheng XT, Xie S-P, Lu LH, Zhou ZQ (2016) Inter-model uncertainty in ENSO amplitude change tied to Pacific ocean warming pattern. J Clim, in press, doi: 10.1175/JCLI-D-16-0039.1.

Captions

Table 1 AMIP models used in this study. The "symbol" column shows the symbol used in the multi-model scatter plots. The "ensemble" column shows which models were used for the ensemble mean. The SINTEX-F model (first row) is not part of AMIP but was run in an AMIP-like configuration.

Fig. 1 Anomaly correlation coefficient (ACC) for SST in the ATL3 region as function of lead time for seasonal predictions from the CHFP model intercomparison (solid lines) and the SINTEX-F prediction system (dashed green line). The skill of the persistence reference prediction is indicated by the black solid line. All predictions were initialized on 1 February so that lead time 1 is centered on the middle of February, 2 on the middle of March etc. For some models, not all lead times were available.

Fig. 2 AMIP multi-model scatter plots of quantities calculated from JJA mean precipitation in the Niño 3.4 region. (a) Absolute difference of predicted and observed standard deviation versus root mean square error (RMSE). (b) Absolute difference of predicted and observed mean versus RMSE. Each model is marked by a letter, with "a" in the origin denoting observations. The model names can be looked up in Table 1. All quantities are calculated for the period 1979-2008.

Fig. 3 ACC (shading) and bias (contour lines; interval 0.5) of the ensemble average of 11 AMIP models for (a) precipitation (mm/day), and (b) surface zonal wind (m/s). The reference data are GPCP for precipitation and ERA-Interim for surface zonal wind. Dashed lines indicate negative values. The zero contour line has been omitted. The ACC is calculated for the entire time series (1979-2008; no seasonal stratification).

Fig. 4 AMIP multi-model scatter plot of mean precipitation versus its ACC for several regions and seasons: (a) equatorial Atlantic (50°W-10°E, 5°S-5°N) in MAM, (b) Niño 3.4 (170-120°W, 5°S-5°N) in JJA, (c) equatorial Indian Ocean (50-95°E, 5°S-5°N) in SON, (d) Sahel (land points in 20°W-40°E, 5-15°N) in JJA, (e) South American monsoon region (land points in 90-30°W, 25-5°S) in JJA, and (f) Indian monsoon region (land points in 65-95°E, 5-25°N) in JJA. Each letter corresponds to one model, with "a" denoting observations.

Fig. 5 AMIP multi-model scatter plot of surface zonal wind ACC and mean for the following regions and seasons: (a) western equatorial Atlantic (40-20°W, 2°S-2°N) in MAM, (b) Niño 4 (160°E-150°W, 5°S-5°N) in MAM, and (c) equatorial Indian Ocean (50-95°E, 5°S-5°N). Each letter corresponds to one model, with "a" denoting observations.

Fig. 6 Climatological annual mean of SST (shading; °C) and precipitation (contour lines; interval 3 mm/day) for an 11-member model ensemble in three experiments: (a) AMIP, (b) amip4K, and (c) amipFuture.

Fig. 7 Skill metrics in the equatorial Pacific for three AMIP-style experiments (amip, amip4K, and amipFuture), stratified by month, for the following quantities, and regions: (a) ACC of Niño 4 surface zonal winds, (b) ACC of Niño 3.4 precipitation, (c) RMSE of Niño 4 surface zonal winds, and (d) RMSE of Niño 3.4 precipitation. The reference data is ERA-Interim for winds and GPCP for precipitation. The dots indicate values that are significantly different from experiment AMIP at the 95% confidence level based on a Fisher's z transformation for ACC and an F-test for RMSE.

Fig. 8 As in Fig. 7, but for WEA surface zonal winds and equatorial Atlantic precipitation (50°W-10°E, 5°S-5°N).

Fig. 9 Climatological annual mean of SST (shading; °C), precipitation (contour lines; contour interval 2 mm/day) and surface winds (vectors; reference 5 m/s) in observations and the three AMIP-style experiments conducted with SINTEX-F. (a) Total fields for OISST (SST), GPCP (precipitation) and ERA-Interim (surface winds), (b) biases in CTRL, (c) biases in Atl_bias, and (d) biases in Pac_bias. The biases in panels b-d are with reference to the observations in panel a. The reference period is 1982-2014.

Fig. 10 As in Fig. 7 but for the following SINTEX-F experiments: CTRL (green line), Atl_bias (blue line), and Pac_bias (orange line). Skill scores are calculated from the 9-ensemble mean of each experiment for the period 1982-2014.

Fig. 11 (a) ACC (shading) and climatological mean (contours; interval 3 mm/day) of MAM precipitation in CTRL. (b) The difference between Pac_bias and CTRL for ACC (shading) and climatological mean precipitation (contours; interval 2 mm/day; negative contours dashed). In panel b, values significant at the 95% level are stippled.

Fig. 12 MAM total precipitation (mm/day) averaged over the eastern equatorial Pacific (140-105°W, 5°S-5°N) scattered against (a) GPCP observations averaged in the same way, and (b) underlying SST (°C) averaged in the same way. Green indicates observations, blue CTRL, and orange Pac_bias. Regression lines are calculated for individual data sets and plotted in the corresponding colors. The correlation coefficient (r) and slope (m) are shown in the upper left.

Fig. 13 Difference between Pac_bias and CTRL. The upper panel shows a latitude-pressure section of geopotential height (shading; m), and meridional and vertical velocity (arrows; units: m/s for meridional velocity and hPa/hr (multiplied by -10) for pressure velocity; upward arrows indicate rising motion and vice versa), averaged over the eastern Pacific (140-105°W). The lower panel shows the SST difference averaged over the same longitude range.

Fig. 14 As in Fig. 10 but for the WEA (panels a and c) and EQATL (panels b and d) indices.

Fig. 15 (a) ACC (shading) and climatological mean (contours; interval 3 mm/day) of MAM precipitation in CTRL. (b) The difference between Atl_bias and CTRL for ACC (shading) and climatological mean precipitation (contours; interval 2 mm/day; negative contours dashed). (c) and (d) As in (a) and (b) but for surface zonal wind. In (b) and (d), values significant at the 95% level are stippled. The precipitation contour lines are repeated in (c) and (d) to facilitate assessing their collocation with the ACC of surface zonal wind.

Fig. 16 SST (shading; ºC) and precipitation anomalies (contours; mm/day) in July, composited on Atlantic Niño years (1984, 1988, 1991, 1995, 1996, 1999, 2008) for (a) GPCP observations, (b) CTRL, and (c) Atl_bias. The precipitation contour interval is 0.5 mm/day in (a), and 1 mm/day in (b) and (c). The zero-contour line has been omitted.

Fig. 17 April-May-June (AMJ) difference of Atl_bias and CTRL in terms of Signal-to-noise-ratio (SNR; shading), and SST (contours; ºC). SNR is estimated as the ensemble mean variance divided by the inter-ensemble variance. The SST difference between the two experiments is essentially identical to the bias in Atl_bias because SSTs in CTRL are strongly restored toward observations.

A. Tables

model	horizontal grid	# vertical levels	symbol	ensemble
SINTEX-F	T106 (1.1 º)	19	b	
ACCESS1-0	1.875º x 1.25º	38	c	
ACCESS1-3	1.875º x 1.25º	38	d	
bcc-csm1-1	T42 (2.8º)	26	e	yes
bcc-csm1-1-m	T42 (2.8º)	26	f	
BNU-ESM	T42 (2.8º)	26	g	
CanAM4	T63 (1.8º)	35	h	yes
CCSM4	1.25º x 0.9º	26	i	yes
CESM1-CAM5	1.25º x 0.9º	26	j	
CMCC-CM	T159 (0.75º)	31	k	
CNRM-CM5	T127 (1.5º)	31	l	yes
CSIRO-Mk3-6-0	T63 (1.9º)	18	m	
EC-EARTH	T159 (1.25º)	62	n	
FGOALS-g2	2.8125º x 2.8125º	26	o	
FGOALS-s2	R42 (2.8º x 1.7º)	26	p	
GFDL-CM3	200 km (2º)	48	q	
GFDL-HIRAM-	C180 (0.5º)	32	r	
GFDL-HIRAM-	C360 (0.25º)	32	s	
GISS-E2-R	2º x 2.5º	29	t	
HadGEM2-A	1.875º x 1.25º	60	u	yes
inmcm4	2º x 1.5º	21	v	
IPSL-CM5A-LR	3.75º x 1.9º	39	w	yes
IPSL-CM5A-	1.25° x 2.5°	39	x	
IPSL-CM5B-LR	3.75º x 1.9º	39	y	yes
MIROC5	T85 (1.4º)	40	z	yes
MIROC-ESM	T42 (2.8º)	80	0	
MPI-ESM-LR	T63 (1.8º)	47	1	yes
MPI-ESM-MR	T63 (1.8º)	95	2	yes
MRI-AGCM3-	T319 (60km)	64	3	
MRI-AGCM3-	T959 (20km)	64	4	
MRI-CGCM3	T159 (1.125º)	35	5	yes
NorESM1-M	2.5º x 2.9º	26	6	

Table 1 AMIP models used in this study. The “symbol” column shows the symbol used in the multi-model scatter plots. The “ensemble” column shows which models were used for the ensemble mean. The SINTEX-F model (first row) is not part of AMIP but was run in an AMIP-like configuration.

B. Figures

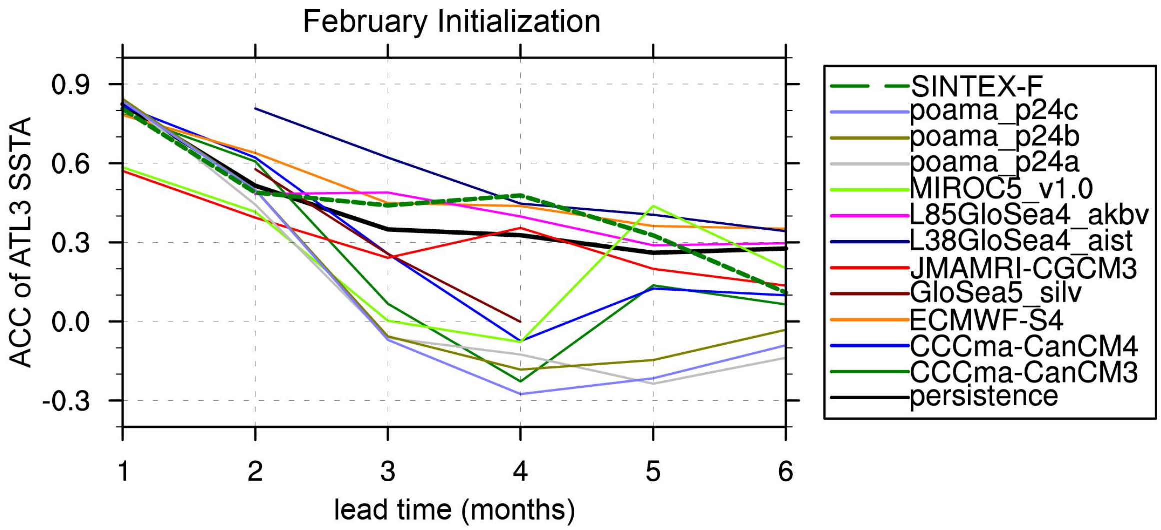

Fig. 1 Anomaly correlation coefficient (ACC) for SST in the ATL3 region as function of lead time for seasonal predictions from the CHFP model intercomparison (solid lines) and the SINTEX-F prediction system (dashed green line). The skill of the persistence reference prediction is indicated by the black solid line. All predictions were initialized on 1 February so that lead time 1 is centered on the middle of February, 2 on the middle of March etc. For some models, not all lead times were available.

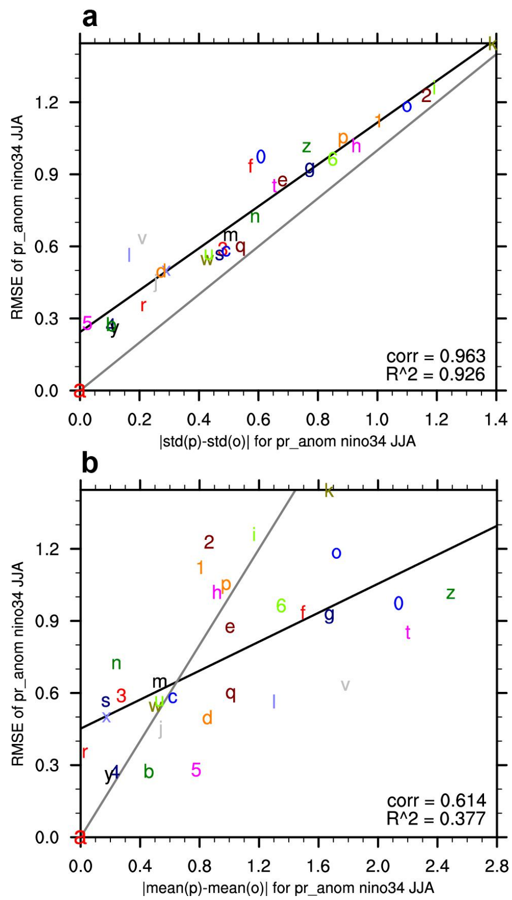

Fig. 2 AMIP multi-model scatter plots of quantities calculated from JJA mean precipitation in the Niño 3.4 region. (a) Absolute difference of predicted and observed standard deviation versus root mean square error (RMSE). (b) Absolute difference of predicted and observed mean versus RMSE. Each model is marked by a letter, with “a” in the origin denoting observations. The model names can be looked up in Table 1. All quantities are calculated for the period 1979-2008.

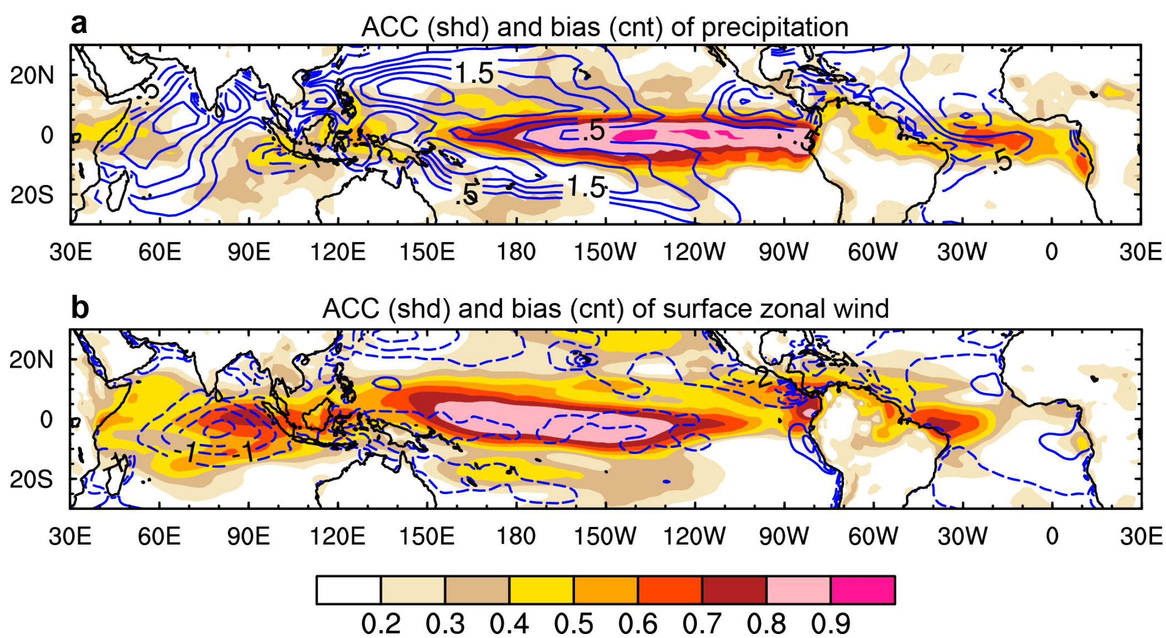

Fig. 3 ACC (shading) and bias (contour lines; interval 0.5) of the ensemble average of 11 AMIP models for (a) precipitation (mm/day), and (b) surface zonal wind (m/s). The reference data are GPCP for precipitation and ERA-Interim for surface zonal wind. Dashed lines indicate negative values. The zero contour line has been omitted. The ACC is calculated for the entire time series (1979-2008; no seasonal stratification).

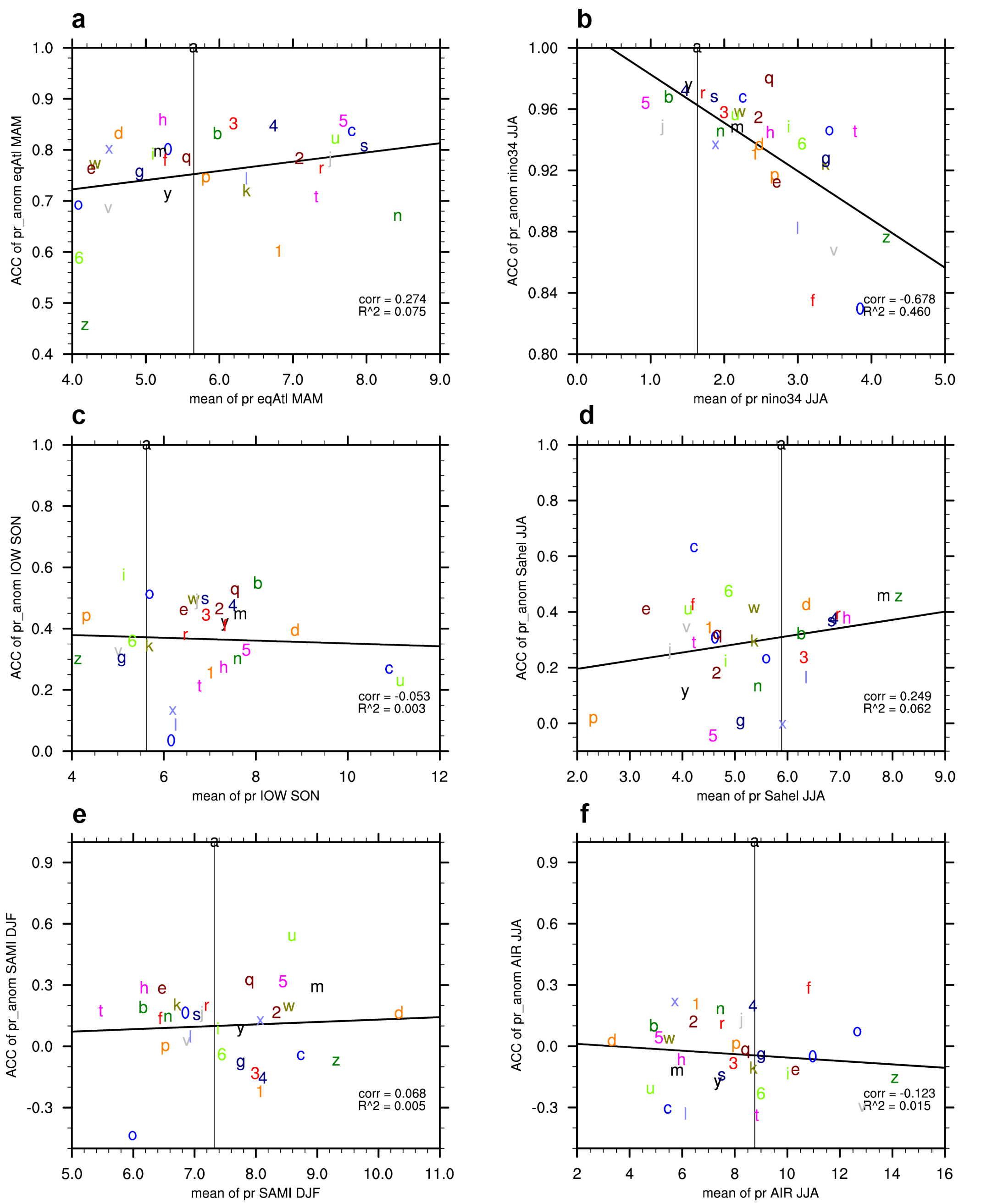

Fig. 4 AMIP multi-model scatter plot of mean precipitation versus its ACC for several regions and seasons: (a) equatorial Atlantic (50°W-10°E, 5°S-5°N) in MAM, (b) Niño 3.4 (170-120°W, 5°S-5°N) in JJA, (c) equatorial Indian Ocean (50-95°E, 5°S-5°N) in SON, (d) Sahel (land points in 20°W-40°E, 5-15°N) in JJA, (e) South American monsoon region (land points in 90-30°W, 25-5°S) in JJA, and (f) Indian monsoon region (land points in 65-95°E, 5-25°N) in JJA. Each letter corresponds to one model, with "a" denoting observations.

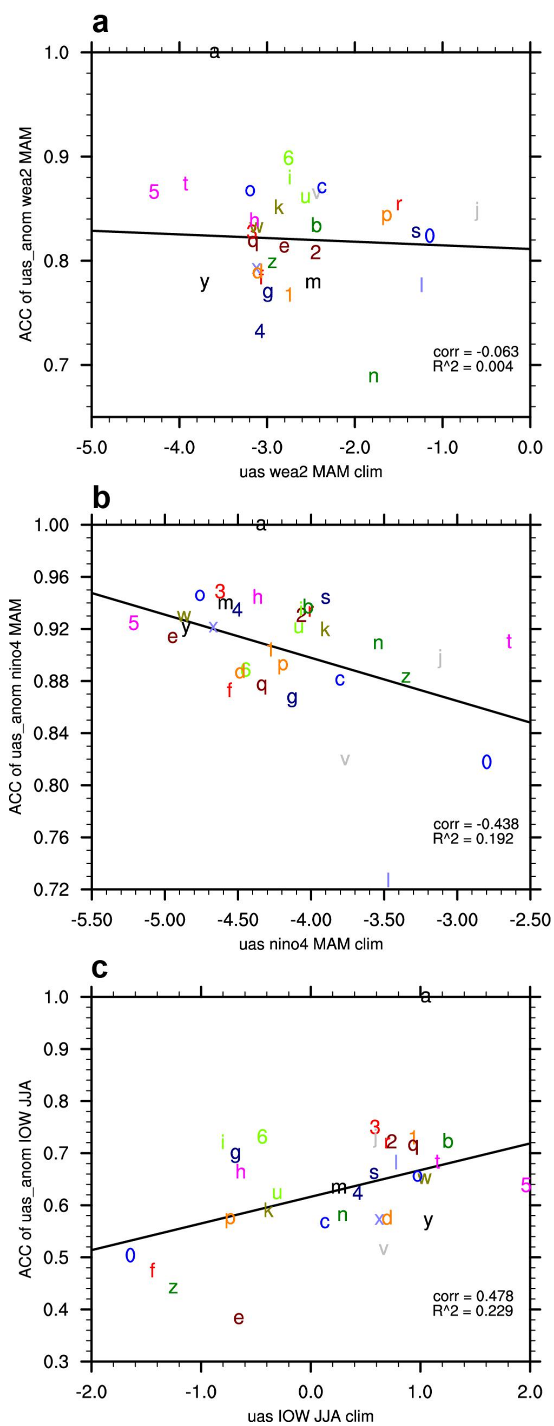

Fig. 5 AMIP multi-model scatter plot of surface zonal wind ACC and mean for the following regions and seasons: (a) western equatorial Atlantic (40-20°W, 2°S-2°N) in MAM, (b) Niño 4 (160°E-150°W, 5°S-5°N) in MAM, and (c) equatorial Indian Ocean (50-95°E, 5°S-5°N). Each letter corresponds to one model, with “a” denoting observations.

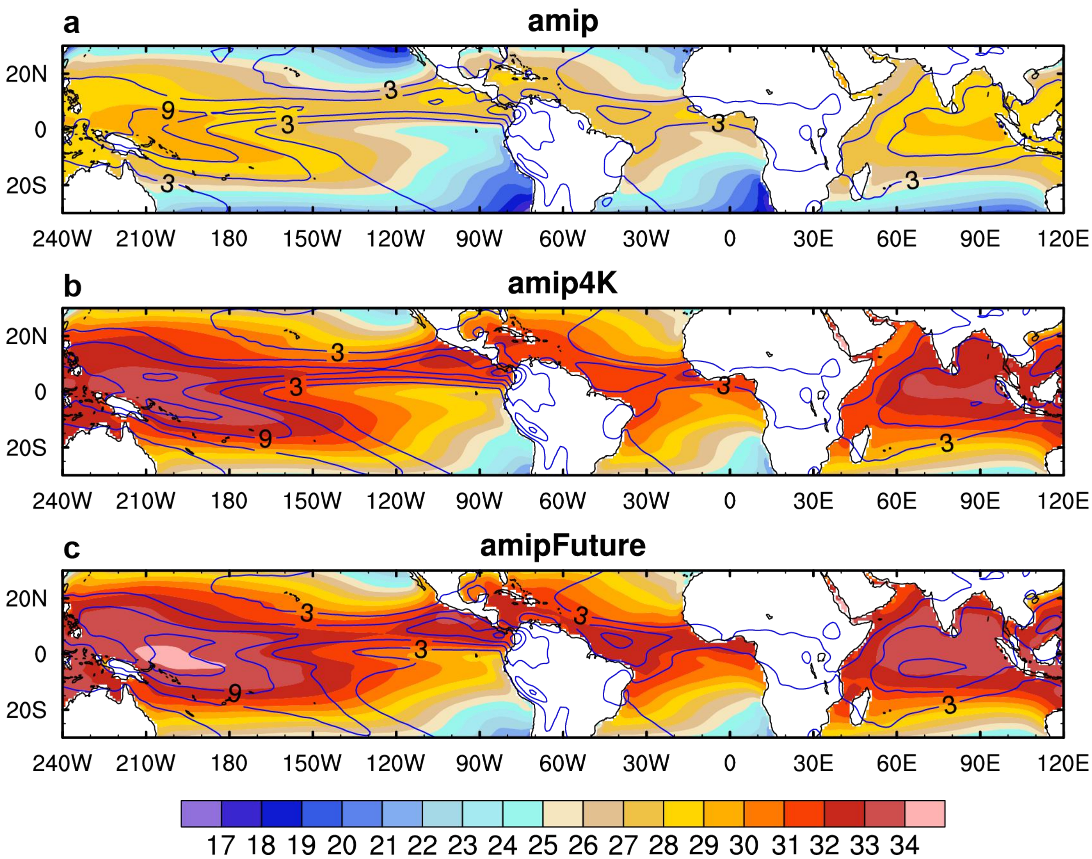

Fig. 6 Climatological annual mean of SST (shading; °C) and precipitation (contour lines; interval 3 mm/day) for an 11-member model ensemble in three experiments: (a) AMIP, (b) amip4K, and (c) amip-Future.

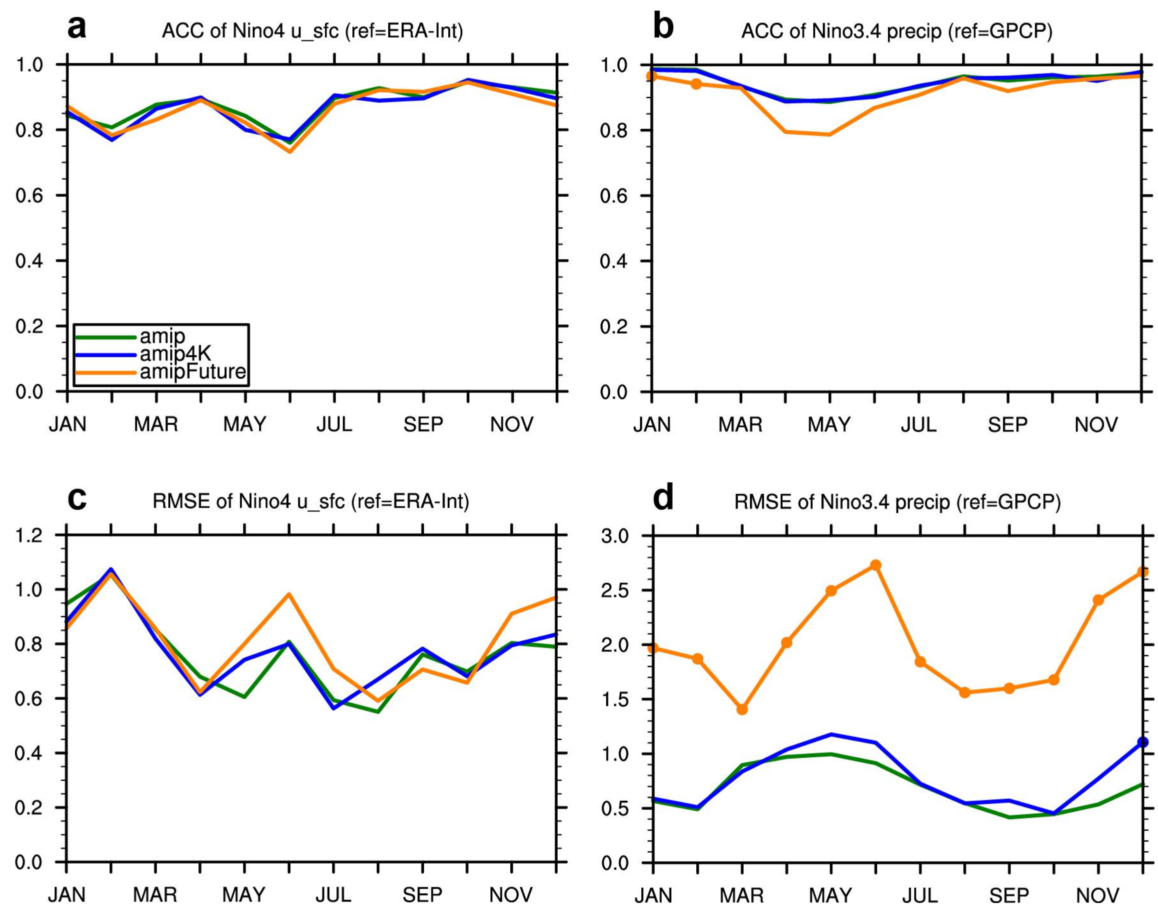

Fig. 7 Skill metrics in the equatorial Pacific for three AMIP-style experiments (amip, amip4K, and amipFuture), stratified by month, for the following quantities, and regions: (a) ACC of Niño 4 surface zonal winds, (b) ACC of Niño 3.4 precipitation, (c) RMSE of Niño 4 surface zonal winds, and (d) RMSE of Niño 3.4 precipitation. The reference data is ERA-Interim for winds and GPCP for precipitation. The dots indicate values that are significantly different from experiment AMIP at the 95% confidence level based on a Fisher's z transformation for ACC and an F-test for RMSE.

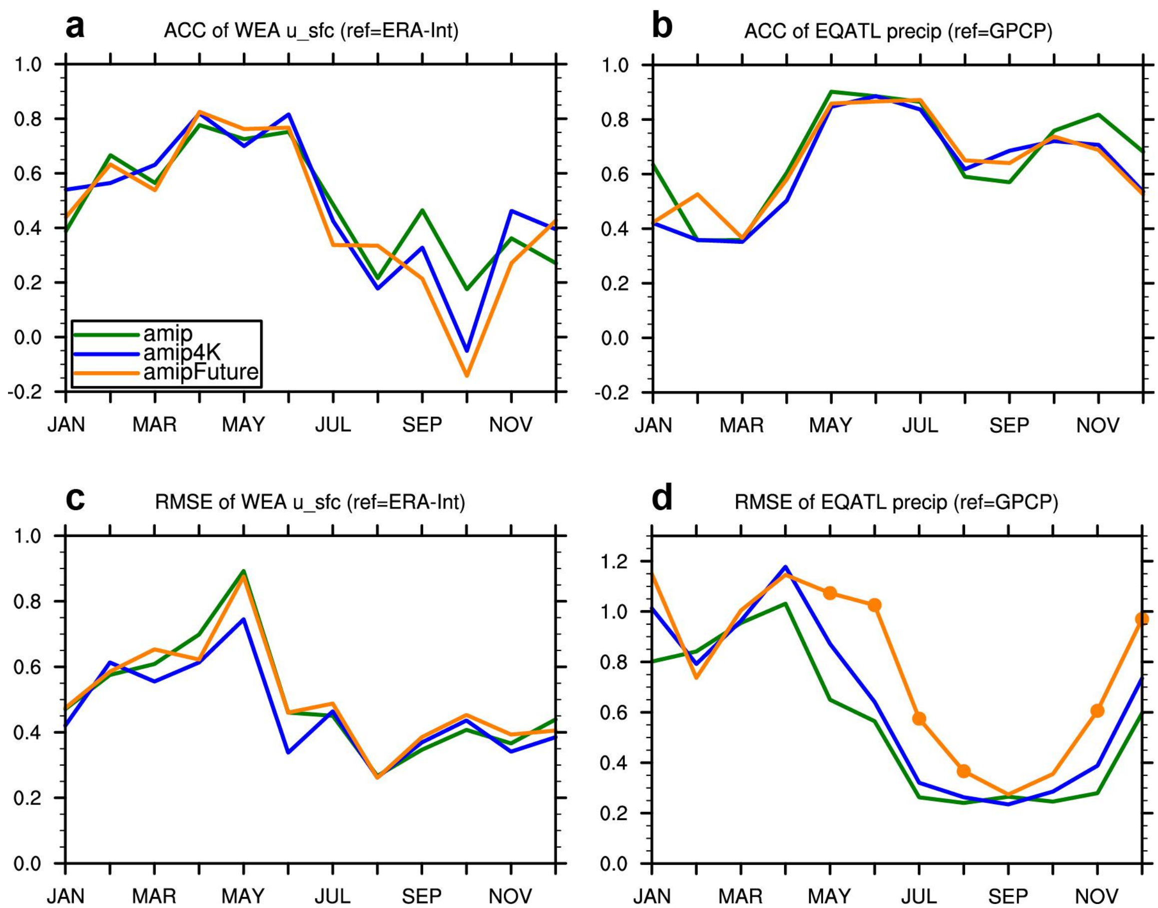

Fig. 8 As in Fig. 7, but for WEA surface zonal winds and equatorial Atlantic precipitation (50°W-10°E, 5°S-5°N).

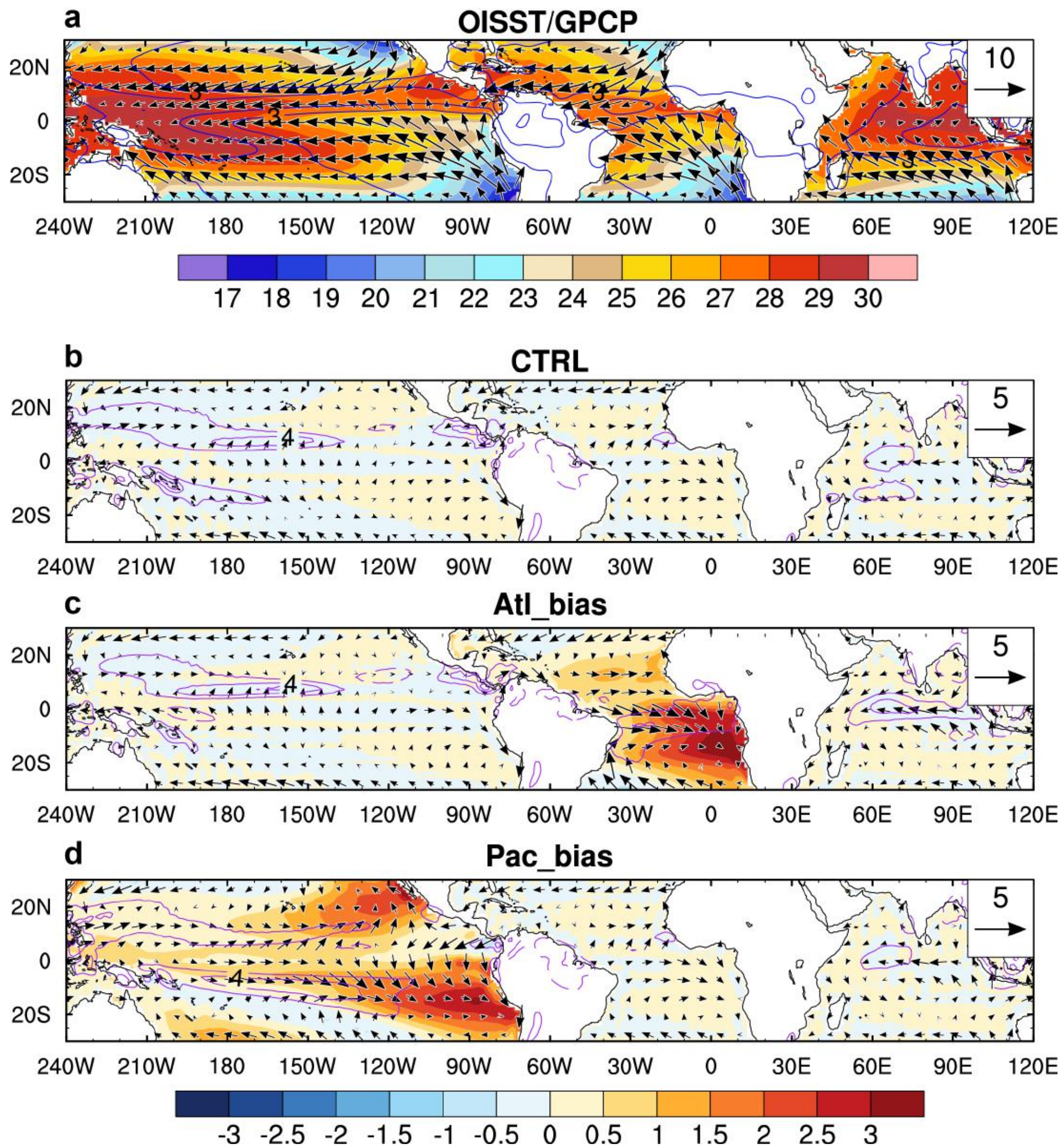

Fig. 9 Climatological annual mean of SST (shading; °C), precipitation (contour lines; contour interval 2 mm/day) and surface winds (vectors; reference 5 m/s) in observations and the three AMIP-style experiments conducted with SINTEX-F. (a) Total fields for OISST (SST), GPCP (precipitation) and ERA-Interim (surface winds), (b) biases in CTRL, (c) biases in Atl_bias, and (d) biases in Pac_bias. The biases in panels b-d are with reference to the observations in panel a. The reference period is 1982-2014.

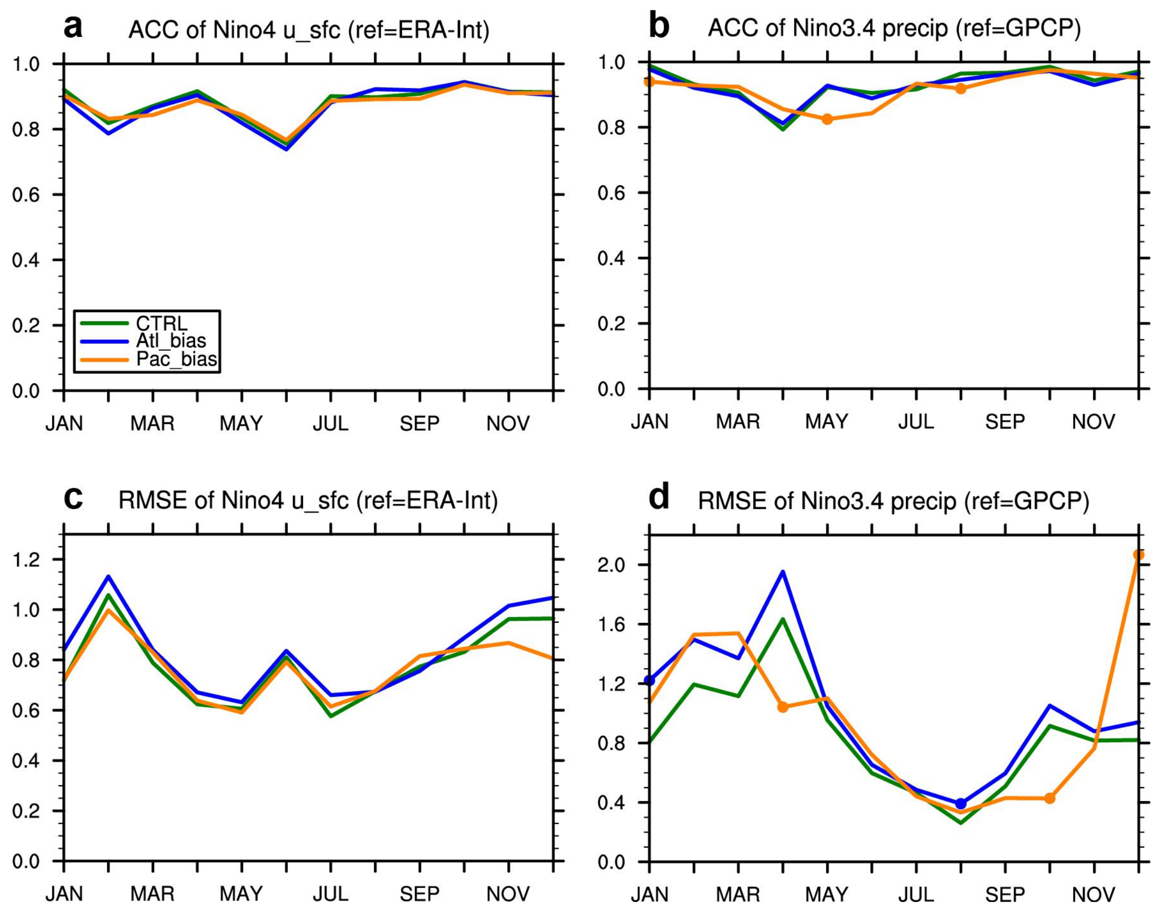

Fig. 10 As in Fig. 7 but for the following SINTEX-F experiments: CTRL (green line), Atl_bias (blue line), and Pac_bias (orange line). Skill scores are calculated from the 9-ensemble mean of each experiment for the period 1982-2014.

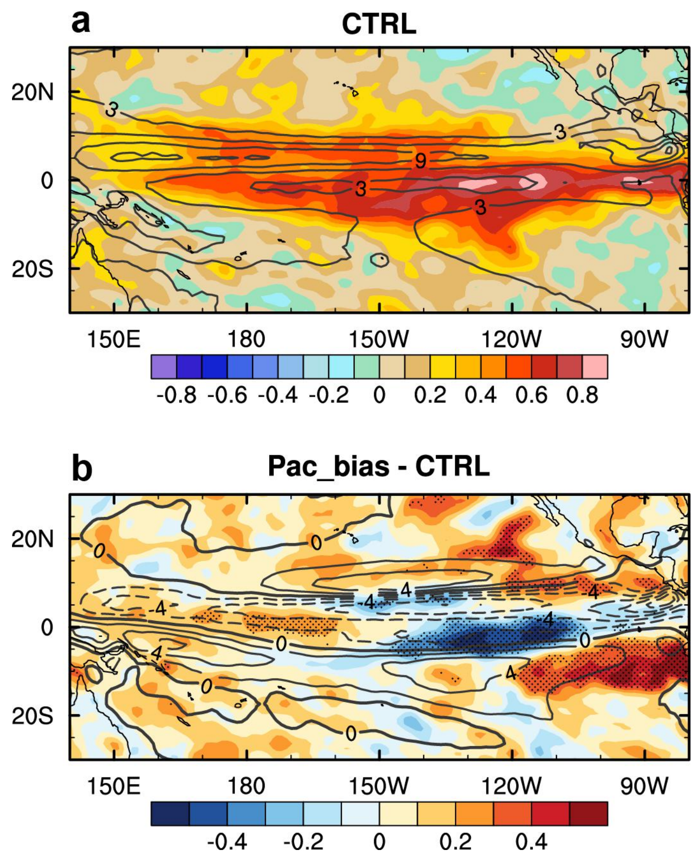

Fig. 11 (a) ACC (shading) and climatological mean (contours; interval 3 mm/day) of MAM precipitation in CTRL. (b) The difference between Pac_bias and CTRL for ACC (shading) and climatological mean precipitation (contours; interval 2 mm/day; negative contours dashed). In panel b, values significant at the 95% level are stippled.

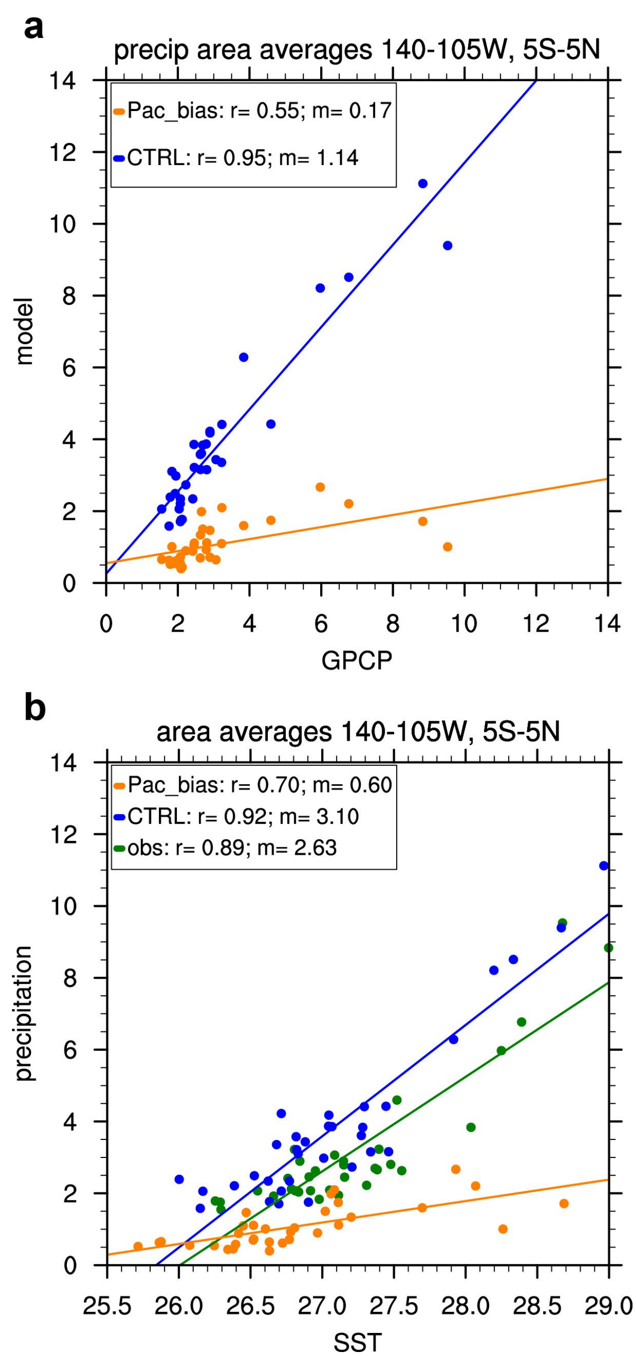

Fig. 12 MAM total precipitation (mm/day) averaged over the eastern equatorial Pacific (140-105°W, 5°S-5°N) scattered against (a) GPCP observations averaged in the same way, and (b) underlying SST (°C) averaged in the same way. Green indicates observations, blue CTRL, and orange Pac_bias. Regression lines are calculated for individual data sets and plotted in the corresponding colors. The correlation coefficient (r) and slope (m) are shown in the upper left.

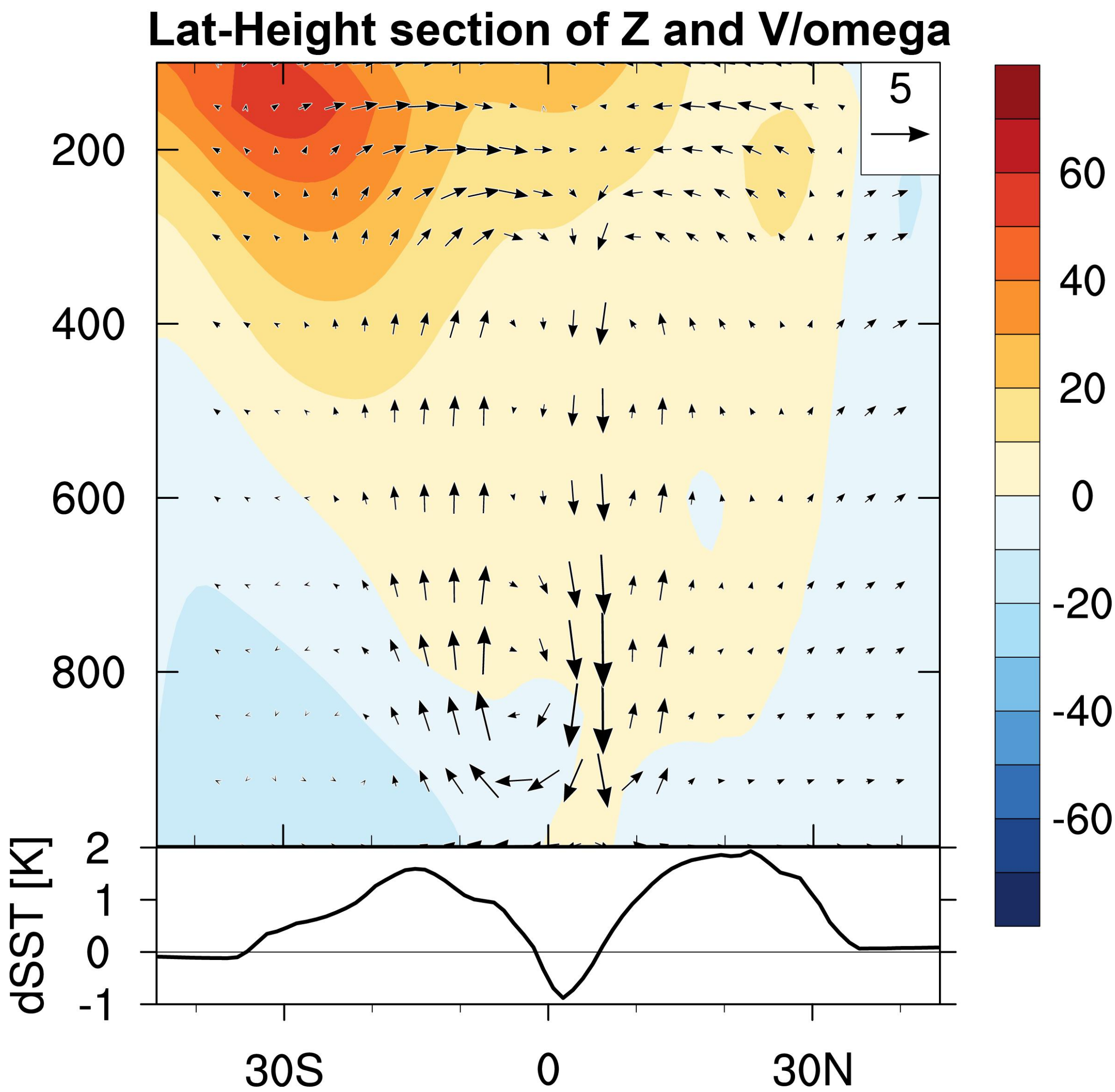

Fig. 13 Difference between Pac_bias and CTRL. The upper panel shows a latitude-pressure section of geopotential height (shading; m), and meridional and vertical velocity (arrows; units: m/s for meridional velocity and hPa/hr (multiplied by -10) for pressure velocity; upward arrows indicate rising motion and vice versa), averaged over the eastern Pacific (140-105°W). The lower panel shows the SST difference averaged over the same longitude range.

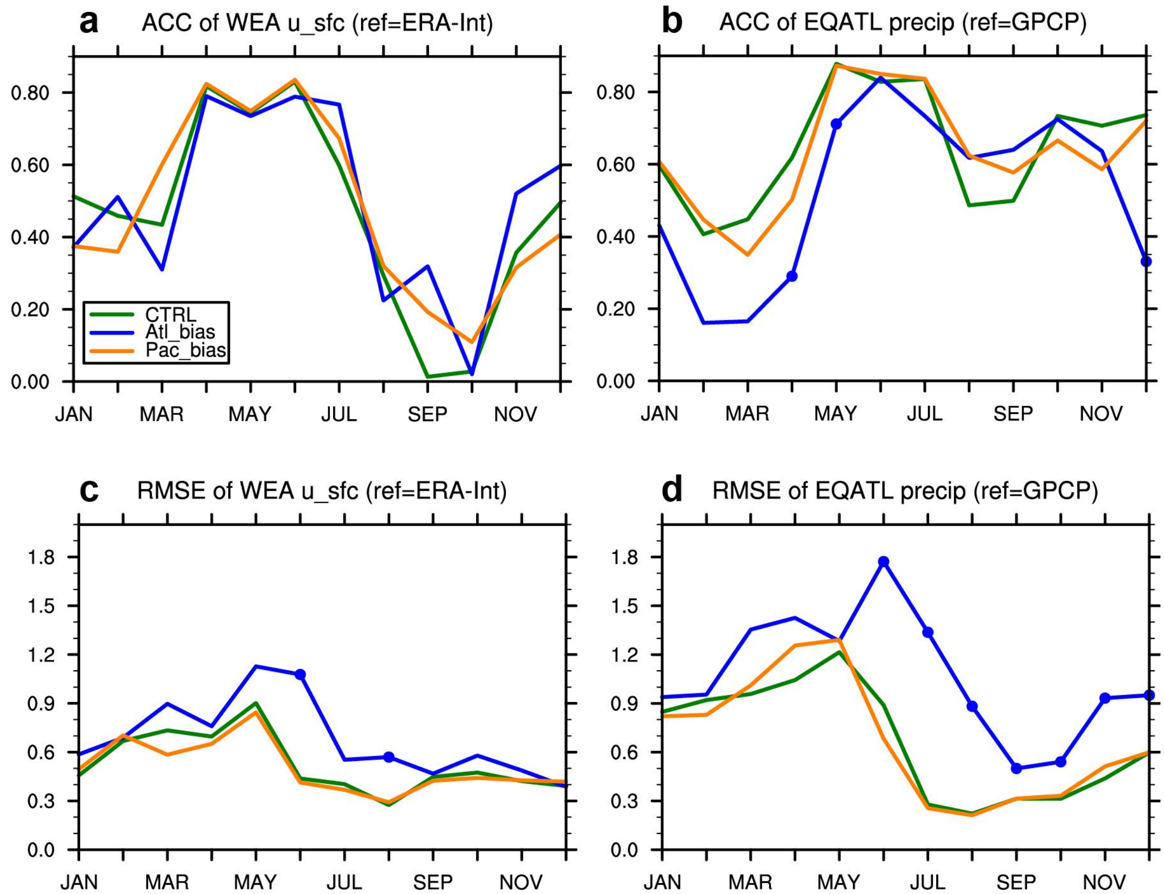

Fig. 14 As in Fig. 10 but for the WEA (panels a and c) and EQATL (panels b and d) indices.

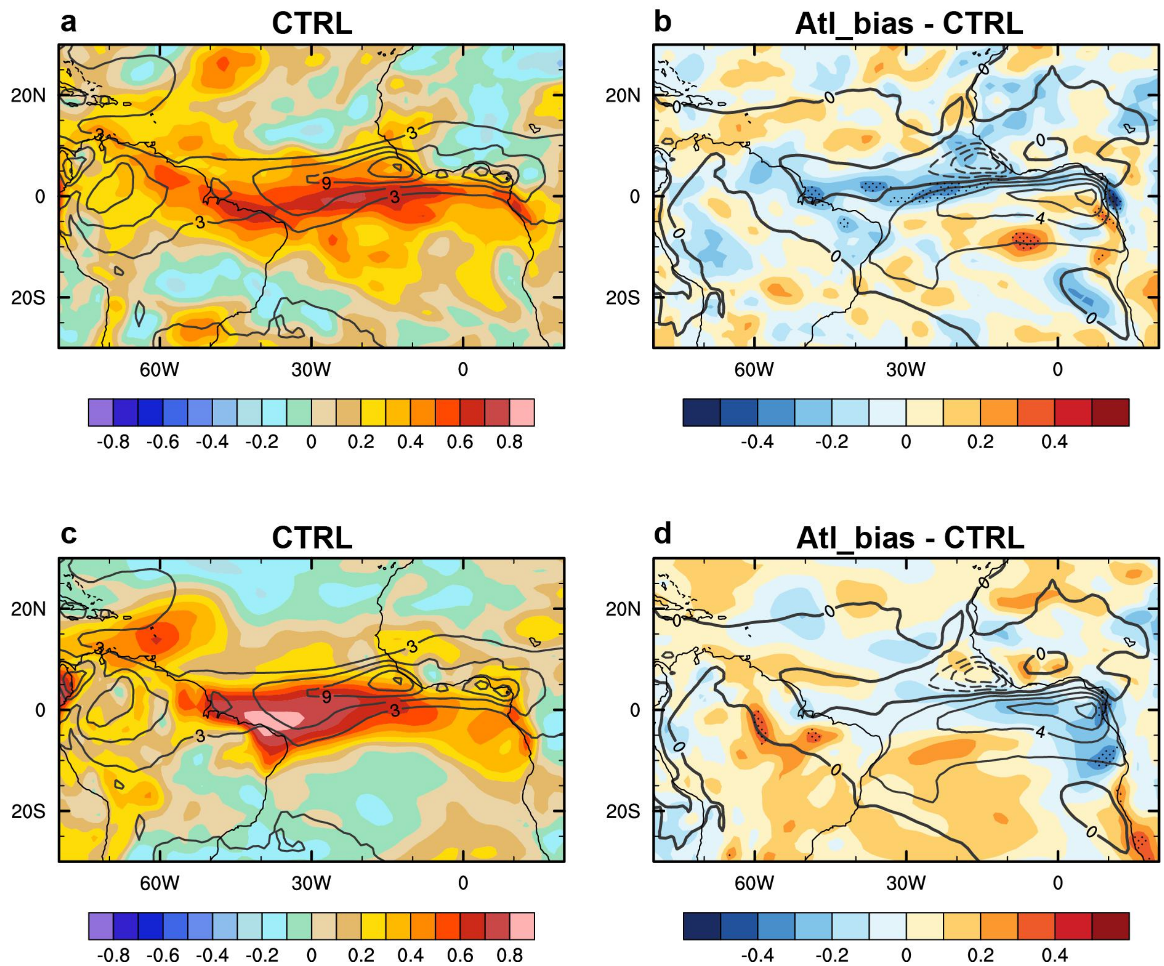

Fig. 15 (a) ACC (shading) and climatological mean (contours; interval 3 mm/day) of MAM precipitation in CTRL. (b) The difference between Atl_bias and CTRL for ACC (shading) and climatological mean precipitation (contours; interval 2 mm/day; negative contours dashed). (c) and (d) As in (a) and (b) but for surface zonal wind. In (b) and (d), values significant at the 95% level are stippled. The precipitation contour lines are repeated in (c) and (d) to facilitate assessing their collocation with the ACC of surface zonal wind.

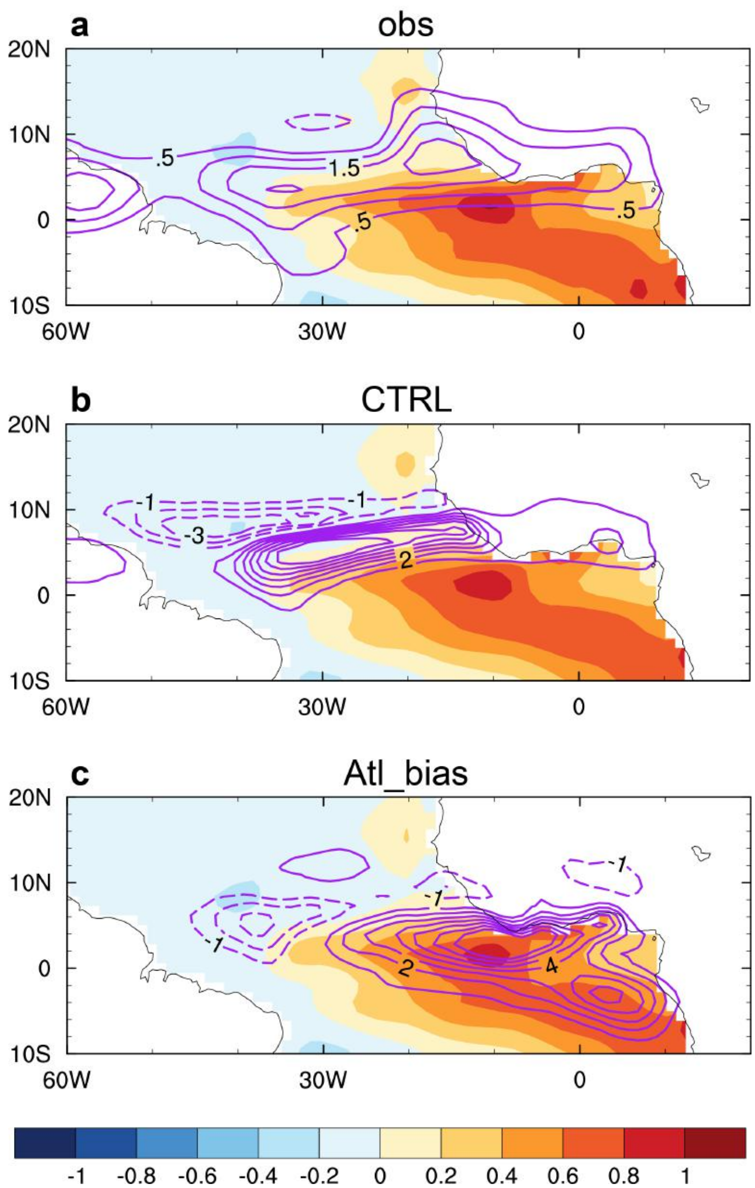

Fig. 16 SST (shading; ºC) and precipitation anomalies (contours; mm/day) in July, composited on Atlantic Niño years (1984, 1988, 1991, 1995, 1996, 1999, 2008) for (a) GPCP observations, (b) CTRL, and (c) Atl_bias. The precipitation contour interval is 0.5 mm/day in (a), and 1 mm/day in (b) and (c). The zero-contour line has been omitted.

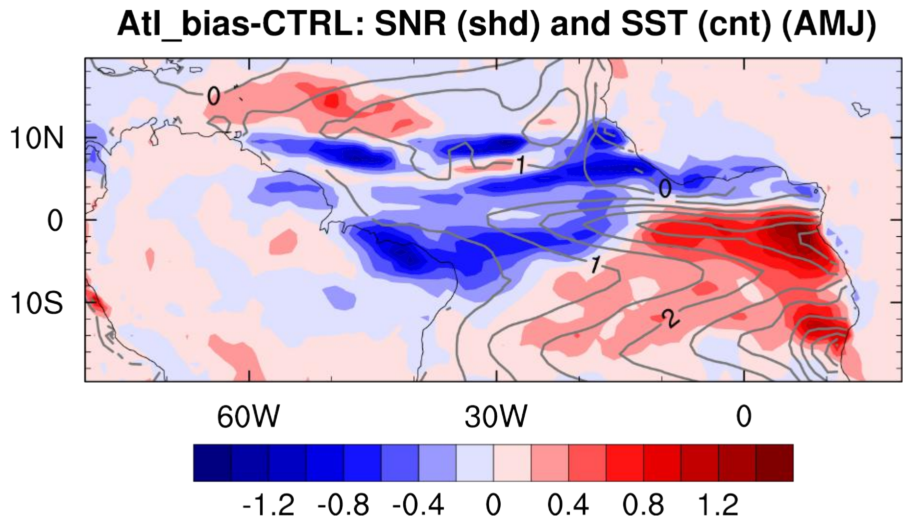

Fig. 17 April-May-June (AMJ) difference of Atl_bias and CTRL in terms of Signal-to-noise-ratio (SNR; shading), and SST (contours; ºC). SNR is estimated as the ensemble mean variance divided by the inter-ensemble variance. The SST difference between the two experiments is essentially identical to the bias in Atl_bias because SSTs in CTRL are strongly restored toward observations.