An overview of the performance of CMIP6 models in the tropical Atlantic: mean state, variability, and remote impacts

Richter, Ingo
Application Laboratory, Research Institute for Value-Added-Information, Japan Agency for Marine-Earth Science and Technology

Tokinaga, Hiroki
Research Institute for Applied Mechanics, Kyushu University

出版情報：Climate Dynamics. 55 (9-10), pp.2579-2601, 2020-08-14. Springer
バージョン：

An overview of the performance of CMIP6 models in the tropical Atlantic: mean state, variability, and remote impacts

INGO RICHTER[1], HIROKI TOKINAGA[2]

[1]*Application Laboratory, JAMSTEC, Yokohama, Japan*

[2]*Research Institute for Applied Mechanics, Kyushu University, Kasuga, Japan*

Climate Dynamics

submitted, 29 December 2019

revised, 7 May 2020

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

General circulation models of the Coupled Model Intercomparison Project Phase 6 (CMIP6) are examined with respect to their ability to simulate the mean state and variability of the tropical Atlantic and its linkage to the tropical Pacific. While, on average, mean state biases have improved little, relative to the previous intercomparison (CMIP5), there are now a few models with very small biases. In particular the equatorial Atlantic warm SST and westerly wind biases are mostly eliminated in these models. Furthermore, interannual variability in the equatorial and subtropical Atlantic is quite realistic in a number of CMIP6 models, which suggests that they should be useful tools for understanding and predicting variability patterns. The evolution of equatorial Atlantic biases follows the same pattern as in previous model generations, with westerly wind biases during boreal spring preceding warm sea-surface temperature (SST) biases in the east during boreal summer. A substantial portion of the westerly wind bias exists already in atmosphere-only simulations forced with observed SST, suggesting an atmospheric origin. While variability is relatively realistic in many models, SSTs seem less responsive to wind forcing than observed, both on the equator and in the subtropics, possibly due to an excessively deep mixed layer originating in the oceanic component. Thus models with realistic SST amplitude tend to have excessive wind amplitude. The models with the smallest mean state biases all have relatively high resolution but there are also a few low-resolution models that perform similarly well, indicating that resolution is not the only way toward reducing tropical Atlantic biases. The results also show a relatively weak link between mean state biases and the quality of the simulated variability. The linkage to the tropical Pacific shows a wide range of behaviors across models, indicating the need for further model improvement.

1. Introduction

The eastern equatorial Atlantic is marked by a pronounced seasonal cycle, with warm sea-surface temperature (SST) during boreal spring (MAM; Fig. 1a) giving way to a cold tongue in boreal summer (JJA; Fig. 1b). During the transition period, the intertropical convergence zone (ITCZ) shifts from a position just on the equator in MAM to a north-equatorial position in JJA, which is associated with profound precipitation changes over both South America and Africa. In addition to precipitation changes, the northward shift of the ITCZ leads to a strengthening of the equatorial trades that drives upwelling and is crucial to the development of the cold tongue.

This annual cycle is subject to variations in amplitude and phasing, which gives rise to a pattern of variability that has been termed the Atlantic Niño or Atlantic zonal mode (AZM). In addition, the SST gradient between the northern and southern tropical Atlantic undergoes fluctuations on interannual and longer time scales, which leads to a phenomenon known as the Atlantic meridional mode (AMM).

The representation of the tropical Atlantic mean state and its variability has been a challenge for general circulation models (GCMs) since the early years of climate modeling. Particularly mean state biases in the equatorial Atlantic have received much attention over the last 15-20 years (e.g. Davey et al. 2002; Richter and Xie 2008; Richter et al. 2014a; Richter 2015; Voldoire et al. 2019). The development of the equatorial Atlantic cold tongue in late spring and early summer is usually underrepresented and delayed in GCMs, leading to a warm SST bias in the eastern equatorial Atlantic. At the same time, models typically produce a cold SST bias in the western equatorial Atlantic, and thus SSTs in JJA increase from west to east along the equator, which is opposite to the observed gradient (Davey et

al. 2002; Richter and Xie 2008; Richter et al. 2014a). Since equatorial SST and surface zonal winds are tightly coupled through the Bjerknes feedback (Bjerknes 1969; Keenlyside and Latif 2007), such SST biases are accompanied by weaker than observed equatorial trade winds, i.e. a westerly surface wind bias. In terms of model biases, the Bjerknes feedback can briefly be described as follows. An initial warm SST bias in the eastern equatorial Atlantic lowers the sea-level pressure (SLP) there and thus leads to a westerly surface wind bias to the west. The westerly bias, in turn, deepens the thermocline in the east and renders cooling through upwelling less efficient, which reinforces the initial SST bias. Due to the coupled nature of these processes one might think that the bias problem originates in small initial SST biases that are subsequently amplified. Such SST biases, in turn, may originate from relatively small deficiencies or systematic errors in the oceanic component of GCMs. Studies by Chang et al. (2007) and Richter and Xie (2008), however, pointed out that westerly wind biases are actually most pronounced in MAM, when SST biases along the equator are small, and relatively weak in JJA. This suggests an atmospheric origin of the westerly wind bias, which is further supported by the fact that MAM westerly wind biases also occur in atmospheric GCMs (AGCMs) forced with observed SSTs. Richter and Xie (2008) show that the westerly wind bias is associated with an erroneous deepening of the thermocline that, presumably, weakens upwelling-related cooling in the following months and thus leads to the prominent cold tongue bias in JJA. This mechanism has since been supported by a number of GCM sensitivity studies (e.g. Wahl et al. 2011; Richter al. 2012; Voldoire et al. 2019).

As the westerly wind bias in AGCMs seems to be a major factor in equatorial Atlantic SST biases, there have been attempts to further trace back its origins. Richter and Xie

(2008) suggest that it may be due to a dry bias over the Amazon region that is pervasive in AGCMs. Such a dry bias may lead to a shift in the Atlantic branch of the Walker circulation, with erroneously weak convection (or even subsidence) over equatorial South America leading to a westerly bias over the Atlantic Ocean to the east. This hypothesis has been supported by a few studies (Richter et al. 2012; Zermeno-Diaz and Zhang 2013). In addition, Richter et al. (2014ab) pointed out the close correspondence between the amount of precipitation south of the equator and the strength of the equatorial trade winds. If further corroborated, this would link the equatorial Atlantic westerly wind bias to the wider problem of excessive precipitation south of the equator during MAM (Dai 2006; Richter et al. 2016) that manifests as a southward shift of the ITCZ in the tropical Atlantic (Richter et al. 2014a) and as a double ITCZ in the tropical Pacific (Mechoso et al. 1995; de Szoeke and Xie 2007; Lin 2007; Li and Xie 2014). Other hypotheses for the westerly wind bias include deficiencies in the presentation of the atmospheric boundary layer (Pauluis et al. 2004; Thomas Toniazzo, personal communication), and errors in the meridional cross-equatorial SST gradient (Wang et al. 2014; Song et al. 2015).

While the atmospheric components of GCMs appear to be responsible for a substantial portion of equatorial Atlantic biases, the oceanic components likely contribute as well. The most likely candidate for intrinsic OGCM biases is the too diffuse equatorial thermocline (Xu et al. 2014b), which is particularly problematic in the equatorial and coastal upwelling regions.

Tropical Atlantic biases are not limited to the equator. There are also pervasive cold biases in the west of the basin that are particularly pronounced in the subtropics (Fig. 2). Conversely, there are warm SST biases in the eastern tropical Atlantic that are most

pronounced in the subtropical upwelling regions at the eastern boundaries. In particular the severe biases in the Angola-Benguela upwelling area (ABA) have received much attention (e.g. Xu et al. 2014ab; Richter 2015; Small et al. 2015; Milinski et al. 2016; Patricola and Chang 2017; Kurian et al. 2020), with a consensus emerging that the detailed structure of the along-shore wind is crucial to a realistic representation of the coastal upwelling that drives much of the cooling in the region (see a recent review by Oettli et al. 2020). To simulate these along-shore winds, the horizontal resolution of AGCMs appears to be an important component (Harlaß et al. 2018; Kurian et al. 2020).

SST biases in the equatorial Atlantic have wider implications. The warmer than observed SSTs in the eastern equatorial Atlantic trigger deep convection in the region (Fig. 2cd; Richter et al. 2014a), which is typically not observed there (Fig. 1). This generally leads to reduced precipitation over West Africa in JJA (Fig. 2cd) and a delayed onset of the Africa monsoon (Steinig et al. 2018). An erroneous shift in deep convection over the tropical Atlantic has also been implicated in remote impacts on the Pacific, with McGregor et al. (2018) suggesting that tropical Atlantic biases may partly be responsible for the underestimated SST variability in the eastern tropical Pacific at decadal time scales.

An important question is whether tropical Atlantic mean state biases also affect the simulated variability and the skill of seasonal predictions for the region. Few studies have addressed this in detail but some results suggest that even in the presence of substantial mean state biases models are able to capture some aspects of the observed variability (Richter et al. 2014; Richter et al. 2018), and that biases are not a major detriment to prediction skill (Richter et al. 2018; Richter et al. 2020). There is even more uncertainty regarding the impact of model biases on climate change projections for the region, though a recent study

indicates that models with smaller equatorial Atlantic SST biases project a larger response to climate forcing in the region (Park et al. 2020).

In the present study we would like to use model output from the most recent coupled model intercomparison project (CMIP6; introduced in section 2) to examine whether the ability of state-of-the-art GCMs to reproduce the mean state of the tropical Atlantic has improved since the previous CMIP5 intercomparison (section 3.1), whether the behavior of current models confirms previous findings regarding the origin of biases (section 3.2), and whether the performance of models is related to their resolution (section 3.3). In section 4 we will examine the link between mean state biases and variability errors in the tropical Atlantic, while in section 5 we examine how biases may affect linkages to the tropical Pacific. Summary and discussion are given in section 6.

2. Model description and methods

We focus on the CMIP6 experiment piControl, in which fully coupled GCMs are run with greenhouse gas concentrations held at 1850 levels. Using this experiment has the advantage that relatively long simulations are available and that there are usually no systematic long-term trends, which facilitates the analysis of interannual variability (though we do remove the linear trend from all data sets before analysis). The 33 models analyzed in the present study are listed in Table 1.

The disadvantage of using piControl is that the radiative forcing is somewhat different from that of the 1979-2018 reference data used to evaluate model performance. Using the same period from experiment "historical" would obviate this problem but the shorter datasets would add uncertainty in the estimation of the model mean state and variability. Since the mean state biases in the tropical Atlantic are typically substantially larger than

160 the difference between the two experiments, we choose to analyze experiment piControl.
161 It should be noted however, that the weaker greenhouse gas forcing in piControl tends to
162 result in a slight underestimation (overestimation) of warm (cold) SST biases.

163 Another issue regarding model evaluation is that there can be a tropics-wide offset in
164 SST due to misrepresentation of cloud cover (Li and Xie 2012) and differences in radiative
165 tuning (Hourdin et al. 2017). Such biases are not specific to the tropical Atlantic and one
166 could argue that they are extraneous to our analysis. One could remove the offset by sub-
167 tracting the tropical mean SST from each model and the observations. For the present study,
168 however, we decided to look at the straightforward rather than the relative bias because
169 there are nonlinear processes, like tropical convection, that do depend on absolute SST to
170 some extent. We have, however, repeated most analyses for the relative biases as well and
171 provide the results in the Supplementary Material.

172 To examine the atmospheric origin of GCM biases we analyze simulations from the
173 experiment "amip", in which atmospheric GCMs are run in atmosphere-only mode with
174 observed SST forcing for the period 1979-2014. This experiment will serve to examine the
175 extent to which biases in atmospheric model fields are already present when forced with
176 realistic SST, which has implications for error sources.

177 To examine the potential benefits of increasing resolution in reducing GCM biases we
178 examine models from experiment control-1950 of the high-resolution model intercompar-
179 ison project (HighResMIP). These models form the ensemble ens-hires and are listed in
180 Table 2. Note that several of the models in control-1950 are also part of piControl (CNRM-
181 CM6-1, CNRM-CM6-1-HR, HadGEM3-GC31-LL, HadGEM3-GC31-MM, and MPI-
182 ESM1-2-HR). These models have been removed from ens-hires, while the model CESM-

183 H, which is not formally part of control-1950 but uses a similar protocol, has been added.
184 As another way of examining the influence of model resolution on biases we compare high-
185 resolution (HR) models with their low-resolution (LR) counterparts. The individual model
186 pairs are shown in the Supplementary Material, while the mean difference of HR versus
187 LR is presented in the main text. The 5 model pairs are listed in Table 3.

188 To keep the intercomparison manageable we will mostly focus on model ensemble
189 means rather than individual models. We do, however, provide figures with all the individ-
190 ual models in the Supplementary Material. The basic ensemble comprises all 33 models
191 listed in Table 1 and is called ens-cmip6. To compare piControl and amip biases we use
192 those 24 models in ens-cmip6 that have corresponding amip simulations available to form
193 ensemble ens-amip (ens-amip-c for the coupled models from piControl and ens-amip-a for
194 the atmosphere-only models from amip). These models are marked with the superscript "A"
195 in Table 1. The 10 high-resolution models listed in Table 2 form ensemble ens-hires. Since
196 model resolution varies considerably across ens-hires models, we select a subset of 5 mod-
197 els with particularly high (low) resolution to form ensemble ens-HR (LR), and these mod-
198 els are marked with the superscript "HR" ("LR") in Table 2. 8 models in ens-hires have
199 corresponding atmosphere-only simulations (highresSST-present) available, and these are
200 used to form ens-hires-a. Additionally, we also use an ensemble of CMIP5 piControl mod-
201 els to get a rough sense of the changes in model biases between these two model genera-
202 tions. This model ensemble is called ens-cmip5 and its 36 members are listed in Table S1
203 of the Supplementary Material. Note that the mix of modeling centers in ens-cmip5 and
204 ens-cmip6 is not the same so that there is no one-to-one correspondence between the two
205 ensembles.

206 Our reference data for SST, 10m wind, latent heat flux (LHF), and sea-level pressure
207 (SLP) is the European Centre for Medium Range Weather Forecasting (ECMWF) reanal-
208 ysis 5 (ERA5; Hersbach et al. 2018) for the period 1979-2018. Precipitation data is from
209 the Global Precipitation Climatology Project (GPCP) version 2.3 (Adler et al. 2018).

210 # 3. Mean state

211 ## 3.1. piControl simulations

212 The ensemble mean of the preindustrial control simulations (ens-cmip6) shows a west-
213 erly wind bias on the equator in MAM, accompanied by a southward shift of the ITCZ (Fig.
214 2a; see Fig. S2 in the Supplementary Material for plots of individual models). In addition,
215 the observed intense precipitation over equatorial South America during MAM (Fig. 1a) is
216 severely underestimated in the models (Fig. 2a). There are cold SST biases in the west and
217 warm SST biases on the equator and in the northeastern and southern part of the basin,
218 particularly in the coastal upwelling regions. In JJA (Fig. 2c), warm SST biases in the east-
219 ern equatorial Atlantic are more pronounced, and so are the cold biases in the western
220 equatorial and subtropical Atlantic. The westerly wind bias on the equator is weaker in JJA
221 than in MAM but is still very noticeable, consistent with the excessive precipitation over
222 the eastern equatorial Atlantic (Fig. 2c) and the attendant low SLP (not shown). While
223 precipitation is excessive over the equatorial Atlantic it is deficient over West Africa.

224 Comparison with the CMIP5 GCM ensemble (Fig. 2bd) suggests that, on average,
225 there are only small changes from CMIP5 to CMIP6, not all of them toward improve-
226 ment. The warm SST bias in the eastern equatorial Atlantic is slightly exacerbated, as
227 is the wet precipitation bias over the same region. On the positive side, both the warm
228 bias in the Benguela upwelling region and the cold SST bias in the subtropics are

reduced in CMIP6. The overall visual impression is that the pattern of SST biases has changed very little, with a basin-wide positive offset explaining the reduction of subtropical cold biases and deterioration of equatorial warm biases. This is supported by high pattern correlations between CMIP6 and CMIP5, which are 0.91 and 0.93 for the MAM and JJA SST biases, respectively. The differences between CMIP6 and CMIP5 in JJA are highlighted in Fig. 3a, which reveals that the coastal warm biases are improved in CMIP6. This is consistent with the more intense along-shore winds in CMIP6 promoting upwelling and cooling. The bias reduction in the Angola-Benguela upwelling region, however, is only small and limited to JJA (see Fig. 4c).

We examine the annual climatological cycle in regions of particular interest. Here we focus on ensemble means and a few representative models. Plots for all models can be found in the Supplementary Material (Fig. S4). SST in the ATL3 region (20°W-0°, 3°S-3°N) represents the strength of the cold tongue and is commonly used as an indicator of the state of the AZM. The ERA-5 reanalysis (thick black line in Fig. 4a) shows the intensification of the cold tongue from April through August. The ens-cmip6 ensemble captures the seasonal evolution of ATL3 SST rather well but is offset by a positive value that ranges from 0.5 K in October to 2 K in July. The ens-cmip5 ensemble is consistently cooler than ens-cmip6 but that difference is small in JJA when the bias peaks. We quantify the deviation of CMIP6 GCMs from the ERA-5 reference using the root-mean-square error (RMSE) averaged over the entire annual cycle. The best 3 models (HadGEM3-GC31-MM, CNRM-CM6-1-HR, and UKESM1-0-LL), according to this metric, only have a weak warm bias. In particular the HadGEM3-GC31-MM is almost indistinguishable from the reference, except during boreal winter, when SST are warmer by up to 1 K. The models with the most

252 severe biases (CAMS-CSM1-0, GISS-E2-1-G, and GISS-E2-1-G-CC) feature warm biases
253 of up to 5 K during the cold tongue season but still reproduce the basic seasonality.

254 The average of 10m zonal wind over the ATL4 region (45°-20°W, 3°S-3°N) is often
255 used to measure wind variability associated with the AZM, and is also a good indicator of
256 the westerly wind biases. The annual cycle of ATL4 10m zonal wind in the reanalysis
257 shows a relaxation of the equatorial trades in MAM that is about 2 m/s weaker than the
258 November peak. Most models overestimate the amplitude of this annual cycle and some of
259 them produce the weakest winds one month later than observed, in May (Fig. 4b). The
260 MCM-UA-1-0, one of the top 3 models in terms of its RMSE relative to ERA-5, has ap-
261 proximately the right amplitude of the annual cycle and realistic values of ATL4 10m zonal
262 wind in MAM but produces a secondary minimum in July, resulting in unrealistic seasonal
263 evolution. Among the 3 worst models in terms of RMSE all still capture the basic season-
264 ality of the reanalysis, albeit with a greatly exaggerated amplitude.

265 SST in the ABA (ocean points in 8°E-coast, 20-10°S) is a measure of upwelling inten-
266 sity and often used to characterize the intensity of Benguela Niños (Shannon et al. 1986;
267 review by Oettli et al. 2020), which are warm events along the southwest African coast that
268 are typically accompanied by reduced upwelling and biological productivity. The reanaly-
269 sis shows the warmest values in March and the coldest in August (Fig. 4c), which resembles
270 the seasonality of the ATL3 region. The CMIP6 ensemble mean overestimates SST
271 throughout the year, with a bias of almost 4 K in August. The differences with the CMIP5
272 ensemble are very small. The best 3 CMIP6 models are relatively close to the reanalysis
273 but still overestimate SST by about 2 K except the MCM-UA-1-0, which has a warm bias

of only about 0.5 K and an excellent seasonality. On the upper end of the scale, the bias can be as large as 7 K.

Since we will also examine teleconnections with the tropical Pacific in section 5, we show the Niño 3.4 index (SST averaged over 170-120°W, 5°S-5°N; Fig. 4d), which is typically used to measure the state of El Niño-Southern Oscillation (ENSO). The CMIP6 ensemble mean only has a weak cold bias of about 0.5 K and captures the annual cycle rather well. The CMIP5 ensemble has a very similar annual cycle but is consistently cooler (more biased) by about 0.4 K). The 3 worst models have more severe cold biases but still have a rather realistic annual cycle.

The relation of the MAM westerly wind bias and JJA SST biases is examined in a multi-model scatter plot (Fig. 5). This reveals that the models with stronger westerly bias in the ATL4 tend to have a more severe warm bias in the ATL3, which is consistent with previous model intercomparisons (Richter and Xie 2008; Richter et al. 2014a). The intermodel correlation is 0.48, which is comparable to the one obtained from CMIP5 models (0.51). This suggests that in CMIP6 too, biases in MAM westerly winds and JJA SST are related, though the correlation is only moderate. One possible explanation is that SST offsets across models partly obscure the link (Li and Xie 2012; Hourdin et al. 2017). Indeed, when the tropical mean SST is subtracted, the intermodel correlation increases to 0.74.

3.2. piControl vs. amip

In this subsection we examine to what extent the prominent westerly wind biases during MAM are already present in amip simulations with observed SST forcing. This has implications for the origins of the westerly bias.

297 A longitude-time section along the equator shows that the SST bias in the CGCM en-
298 semble has a distinct seasonality (Fig. 6a). Up until May, biases are between 0.5 to 1 K in
299 the eastern equatorial Atlantic, and close to zero in the west. This is followed by rapid bias
300 growth in June and July, with peak values at the eastern boundary exceeding 2.5 K. The
301 westerly wind bias follows a different seasonality, with maximum values of more than 3
302 m/s in May, when the SST biases are still relatively weak. The relatively weak response of
303 the 10 m zonal wind to the severe SST biases in June and July may seem surprising at first
304 but consideration of the total precipitation over the equator (not shown, but inferable from
305 the wet precipitation bias shown in Fig. 6c) reveals that the wind response is modulated by
306 the presence of deep convection. Since the ITCZ moves to the north in June and July, this
307 leads to a weakening of the wind response on the equator. This amplification of the surface
308 wind response to deep convection and diabatic heating has long been recognized in the
309 context of ENSO (Webster 1981; Zebiak 1986; Harrison and Vecchi 1999), while Richter
310 et al. (2017) showed that this mechanism is a crucial element in the variability of the equa-
311 torial Atlantic as well. In addition to being subject to the modulating effect of deep con-
312 vection, the pronounced westerly bias may also be influenced by off-equatorial SST biases,
313 either directly through their impact on the SLP distribution, or indirectly through their in-
314 fluence on ITCZ latitude. This is supported by the intermodel correlation of SST biases
315 with the 10m zonal wind bias in the ATL4 region (not shown), in which the westerly bias
316 is positively correlated with the SST bias in the ABA, and negatively with the SST bias in
317 the northern tropical Atlantic. Such linkages have also been found in previous studies (e.g.
318 Xu et al. 2014a).

Comparison with the seasonal evolution of biases in the amip ensemble (Fig. 6b) shows that the westerly bias is 2-3 times weaker in the absence of SST biases but that it still follows the same basic seasonality. In difference to the CGCM ensemble, however, the westerly bias is most pronounced one month earlier in April. This indicates strong amplification of the westerly bias by equatorial SST biases during May in the CGCM ensemble.

Latitude-time sections of precipitation biases, averaged between 40-20°W, show how, relative to observations, the Atlantic ITCZ is shifted south throughout the year in both the CGCM and AGCM ensembles (Figs. 6c and 6d). Note how the 10m wind biases collocate with the region of the strongest wet precipitation biases from January through June (Fig. 6c), regardless of latitude. As the precipitation biases roughly mark the position of the ITCZ, the suggestion is that wind biases are linked to the erroneous position of the ITCZ, as pointed out by Richter et al. (2014a). Note also, that, since we are looking at the central basin, the local SST biases are small and cannot explain the excessive precipitation. The north-south dipole of SST biases (Fig. 2a) may explain part of the southward ITCZ shift but, as evidenced by the AGCM ensemble with prescribed SST (Fig. 6d), other factors must play a role as well.

The influence of SST biases on surface wind biases is highlighted by a horizontal map of the difference between the CGCM and AGCM ensembles in JJA (Fig. 7). There is erroneous convergence of the surface winds into the area of pronounced deep convection bias over the eastern equatorial Atlantic, that is located just north of the maximum warm SST bias. As the amip ensemble does not have any westerly bias in JJA (as suggested by Fig. 6b), this clearly shows the influence of the SST biases.

3.3. control-1950 simulation of high-resolution models

The ensemble mean of the CMIP6/HighResMIP control-1950 simulations significantly improves the equatorial biases of SST, surface wind, and precipitation both in MAM and JJA (Figs. 3b, 3c, 8a, 8d). While weak westerly wind biases are still found over the equatorial Atlantic in MAM, the warm SST biases drop below 0.5 K, a value 50% less than that of the CMIP6 piControl ensemble. However, large SST biases are still present over the ABA region in MAM and over the coastal regions of the tropical eastern Atlantic, suggesting that even HighResMIP models tend to underestimate the strength of observed coastal upwelling. The three models with the highest resolution, CESM-H, HadGEM3-GC31-HM, and ECMWF-IFS-HR, successfully reproduce observed patterns of the equatorial cold tongue and trade winds (Fig. S8c and S8d). Their ensemble mean shows neither the severe warm SST bias (Fig. S8d) nor the surface westerly wind bias over the equatorial Atlantic (Figs. S8c). The coastal upwelling is also well reproduced both in MAM and JJA, a significant improvement compared with low-resolution models of the CMIP5/6 piControl simulations. The only large bias in the three models is a northward displacement of the ITCZ. This probably strengthens the cross-equatorial southeasterly trade winds that act to sustain the east-west tilt of the thermocline and enhance evaporative cooling at the sea surface.

The impact of high resolution is further examined by the comparison of high-resolution models from control-1950 with their corresponding low-resolution models in piControl (Figs. 8b, 8c, 8e, and 8f). The southward shift of the ITCZ during MAM is much weaker in the high-resolution models and the westerly wind bias on the equator greatly reduced (Figs. 8b and 8c). Equatorial SST biases are much smaller in the high-resolution than in the low-resolution ensemble (Figs. 8e and 8f). SST biases along the southwest

366 African coast also show substantial improvement, even though the along-shore winds,
367 thought to be crucial for upwelling, are actually weaker. This may be due to remote influ-
368 ences from the equator (Xu et al. 2014b), the delayed effect of the MAM wind biases, or a
369 combination of both. The delayed effect of MAM equatorial wind biases is evident in the
370 equatorial SST bias during JJA, when westerly biases are comparable in both ensembles
371 but SST biases are much more severe in the low-resolution models. This can be seen as a
372 consequence of the strong westerly bias during MAM in the low-resolution models. Note
373 that wet precipitation biases over the equatorial Atlantic and dry biases over West Africa
374 remain problematic even in the high-resolution models.

375 We also compare the annual cycles of SST and surface zonal wind biases along the
376 equator (Fig. 9). While their seasonal evolutions are similar to CMIP6 piControl simula-
377 tions (Fig. 6a), the amplitudes are much smaller in HighResMIP control-1950 simulations
378 (Fig. 9a). Interestingly, the coupled HighResMIP models exhibit roughly the same magni-
379 tude of MAM zonal wind biases as their corresponding atmosphere-only simulations (Figs.
380 9a and 9d), while in the low-resolution models the westerly bias becomes more severe upon
381 coupling (Figs. 9c and 9f). The slight improvement of the MAM zonal SST gradient in the
382 equatorial Atlantic (Fig. 9a vs. Fig. 9c) likely contributes to this but off-equatorial SST
383 biases probably have a role as well. In particular the improved meridional SST gradient in
384 the high-resolution models (Fig. 8b vs. Fig 8c) is likely to be important here.

385 # 4. Variability

386 ## 4.1. Equatorial variability

387 The seasonally stratified standard deviation of ATL3 SST in the ERA-5 reanalysis (Fig.
388 10a) peaks in June, with a small secondary maximum in December associated with the

389 Atlantic Niño II (Okumura and Xie 2006). In the CMIP6 ensemble mean (ens-cmip6), the
390 standard deviation of ATL3 SST has a similar seasonal cycle with comparable amplitude,
391 though the peak occurs one month later, in July. As in section 3, we calculate for each
392 model the average RMSE relative to ERA-5 and plot the top and bottom three models
393 according to this metric in Fig. 10a (see Fig. S10a for plots of all the models). The three
394 top models (MPI-ESM1-2-HR, CanESM5, and IPSL-CM6A-LR) have a remarkably real-
395 istic cycle with the peak of variability occurring in the right month. Note that none of these
396 models is among the top three in terms of the annual cycle of ATL3 SST (Fig. 4a), though
397 they are all in the upper half according to the RMSE criterion. The bottom three models
398 (NESM3, MIROC-ES2L, and NorESM2-LM) display a wide range of behaviors, with
399 NESM3 peaking in February, MIROC-ES2L having a double peak in May and August,
400 and NorESM2-LM having the peak in the right month but severely overestimating it. Of
401 these models, the NorESM2-LM ranks relatively high (#8) in terms of the annual cycle of
402 SST itself, suggesting again that the link between mean state biases and variability is not
403 straightforward. Note that the ATL3 SST of NorESM2-LM ranks even higher (#2) when
404 the tropical SSTs are subtracted first (Figs. S4aa and S4aaa), indicating that the good per-
405 formance of the model in terms of mean bias is robust.

406 ens-cmip5 has the same annual cycle as ens-cmip6 but with a weaker, and therefore
407 less realistic, amplitude. The general impression of superior performance of the CMIP6
408 models is confirmed by more detailed statistics (not shown). In terms of the RMSE metric,
409 none of the CMIP6 models has values above 0.2, while 5 CMIP5 models do. An additional
410 metric, the correlation between model and ERA-5, reveals that 6 CMIP6 models have val-
411 ues above 0.8, while only one CMIP5 model achieves the same.

412 The standard deviation of ATL4 10m zonal wind peaks in April in the ERA-5 reanal-
413 ysis (Fig. 10b). The ensemble mean overestimates variability and has the peak occurring
414 one month later than the reanalysis. Even the top 3 models (in terms of the RMSE with
415 ERA-5) suffer from a one-month delay in peak variability, though all of them get the am-
416 plitude about right. The three bottom models severely overestimate variability and in two
417 of them variability peaks as late as June. The #1 model (CanESM5) is also the model with
418 the smallest RMSE for ATL4 10m zonal wind (Fig. 4b), but the other two models (MPI-
419 ESM1-2-HR and BCC-CSM2-MR) are actually among the worst-performing in terms of
420 the mean state bias. Likewise, the bottom three models in term of their annual cycle of
421 standard deviation have very different mean state biases. We note that the standard devia-
422 tion of ATL4 10m zonal wind is smaller in ens-cmip5 than in ens-cmip6 and therefore
423 closer to the reanalysis.
424 The standard deviation of ABA SST in the reanalysis has a prominent peak in April
425 (Fig. 10c), associated with the Benguela Niño. This peak is underestimated in ens-cmip6
426 and occurs three months later, in July. Even the top three models (HadGEM3-GC31-MM,
427 CNRM-CM6-1-HR, and MPI-ESM1-2-HR) feature a delay by 1-2 months and underesti-
428 mate variability. Atmospheric model resolution may play a role in the SST biases of the
429 upwelling region as indicated by previous studies (Harlaß et al. 2018; Kurian et al. 2020).
430 All three of the top models have relatively high resolution in their atmospheric components
431 (< 1° horizontally, >= 85 vertical levels). Two of the bottom three models (BCC-ESM1
432 and GISS-E2-1-H) have a weak and delayed annual cycle, while the third (AWI-CM-1-1-
433 MR) severely overestimates variability and peaks in August. Two of these three models

434 employ relatively coarse resolution in their atmospheric components (> 2.5° horizontally,
435 <= 40 vertical levels).

436 The Niño 3.4 index (Fig. 10d) in the reanalysis displays the well-known peak variabil-
437 ity in December-January-February (DJF). In comparison, ens-cmip6 has less variability in
438 DJF and more variability in other months, though the peak still occurs in DJF, while ens-
439 cmip5 has less variability than observed in almost all months. The best three CMIP6 mod-
440 els in terms of the RMSE criterion (SAM0-UNICON, HadGEM3-GC31-MM, and
441 FGOALS-g3), while displaying stronger seasonality, suffer from a similar problem as ens-
442 cmip6 in that they tend to overestimate variability in most months, except during DJF.
443 Overall, however, the seasonal cycle of variability is quite realistic in these models. The
444 bottom three models (GISS-E2-G-1-CC, NorESM2-LM, and FGOALS-f3-L) all exagger-
445 ate variability and show little seasonal preference.

446 To examine more closely the evolution of warm AZM events we composite fields on
447 the ATL3 index exceeding one standard deviation in JJA. ERA-5 shows the typical evolu-
448 tion of AZM events, with westerly 10m wind anomalies in the ATL4 region preceding
449 positive SST anomalies in the ATL3 (Fig. 11a). The wind anomalies peak in May, one
450 month before the peak of the SST anomalies. Also of note are the southward shift of the
451 equatorial Atlantic ITCZ (indicated by precipitation in the SEQ area: 40°W-10°E, 10°S-
452 4°S), and a weakening of SLP in the southern tropical Atlantic (STA; 20°W-20°E, 25°S-
453 5°S), both of which become prominent in March. These two indicators have previously
454 been associated with the developing phase of the AZM (Richter et al. 2014a for the ITCZ
455 shift, Nnamchi et al. 2016 for the SLP anomalies). The ensemble mean of piControl simu-
456 lations, ens-cmip6, is obtained by first calculating the composites for each model and

457 subsequently averaging over all models. ens-cmip6 shows a similar evolution of AZM
458 events as the reanalysis, with the peak in ATL4 westerly wind anomalies preceding the
459 peak of the ATL3 SST anomalies by one month but, as already suggested by Fig. 10a,
460 events are delayed by one month. The southward shift of the ITCZ also occurs in ens-cmip6
461 but peaks two months later than in the ERA-5, whereas the STA SLP anomalies are weakly
462 negative throughout the period examined.

463 As we did for the standard deviation of indices (Fig. 10), we select the top and bottom
464 three models to show in Fig. 11. Here our metric is based on the RMSE of ATL3 SST and
465 the RMSE of ATL4 10m zonal wind, which are averaged from March through August (the
466 active phase of the AZM) and then added. To account for the different units of SST and
467 wind, we normalize the fields by their respective standard deviations before adding them.
468 In all the top three models, AZM events develop similarly to the ones in ERA-5 though
469 there are some differences. While CanESM5 mimics the observed one-month lag between
470 the ATL4 10m zonal wind and ATL3 SST (Fig. 11c), the two fields peak simultaneously
471 in CESM2 and IPSL-CM6A-LR (Figs. 11e and 11g). All three of these models feature a
472 southward shift of the Atlantic ITCZ but it is delayed relative to the reanalysis. The weak-
473 ening of STA SLP is present throughout the composite period but shows little development.
474 The bottom three models show rather disparate behavior. In the two GISS models (GISS-
475 E2-1-H and GISS-E2-1-G-CC), AZM events develop gradually from January through June
476 and, notably, in the presence of easterly wind anomalies (Figs. 11d and 11f). This evolution
477 suggests that the warm events in these two models are caused not by a Bjerknes type mech-
478 anism but rather by thermodynamic processes acting on their ocean mixed layers, which

479 are excessively deep (not shown). NorESM2-LM, on the other hand, shows behavior that
480 is qualitatively similar to the reanalysis but with an excessive amplitude (Fig. 11h).

481 Interannual variability in the ATL3 region has shown some decline over the last sev-
482 eral decades (Tokinaga and Xie 2011), which is particularly evident in the weak interannual
483 variability of the last two decades (Richter and Tokinaga 2020; Prigent et al. 2020). This
484 behavior could be due to the general slow-down of the Walker circulation under global
485 warming (Vecchi et al. 2006; Tokinaga et al. 2012) or merely multi-decadal modulation of
486 interannual variability associated with the Atlantic meridional overturning circulation
487 (AMOC; Haarsma et al. 2008; Polo et al. 2013; Martin-Rey et al. 2018). It is therefore of
488 interest to examine the low-frequency modulation of ATL3 interannual variability in the
489 piControl simulations with steady radiative forcing and compare it with the difference in
490 ATL3 standard deviation between the periods 1979-1998 and 1999-2018. We use the rank-
491 ing of the composites above to pick the top and bottom three models in terms of AZM
492 evolution. Here, however, we also require that the simulations be at least 500 years long,
493 so that some of the selected models differ from those shown in Fig. 11. The running stand-
494 ard deviation of ATL3 SST in the top three models suggests that the observed variability
495 decline of the last 40 years is still within the range of naturally occurring variability, though
496 just barely (Fig. 12). The bottom three models, on the other hand, all show modulations
497 that well exceed the observed shift.

498 **4.2. Subtropical variability (AMM)**

499 The AMM is commonly defined via a maximum covariance analysis (MCA) of 10m
500 wind and SST (e.g. Chiang and Vimont 2004; Amaya et al. 2017), though the differences
501 between NTA and STA SST indices has also been used in the literature (e.g. Servain et al.
502 1999). The NTA has larger variability than the STA (Amaya et al. 2017) and thus

503 dominates the NTA-STA index. This is confirmed by an analysis of the ERA-5, which
504 shows that the MCA based index and the NTA index are correlated at 0.95.

505 In the ERA-5, the standard deviation of NTA SST is highest from February through
506 May (Fig. 13a). The 10m zonal wind averaged over the same area shows highest variability
507 one month earlier, in January (Fig. 13b), which is consistent with the forcing of AMM
508 variability by trade wind variations and associated latent heat flux changes. ens-cmip6 has
509 a similar seasonal evolution but generally underestimates SST variability, while overesti-
510 mating 10m zonal wind variability, which is similar to the results Amaya et al. (2017) found
511 for CMIP5 and the behavior of our CMIP5 ensemble (Fig. 13). This suggests that the oce-
512 anic mixed layer is too deep in these models, thus requiring stronger wind forcing to pro-
513 duce SST anomalies of the observed magnitude. Such mean state errors can be inferred
514 from the excessively strong northeasterly trade winds in Fig. 2a, which deepen the mixed
515 layer through stirring and latent heat flux-induced surface cooling. The top three models
516 (in terms of NTA SST variability relative to ERA-5) have an annual cycle that is quite
517 similar to the reanalysis, both in terms of amplitude and phasing. The bottom three models,
518 on the other hand, severely underestimate variability and display no particular seasonality.
519 Interestingly, one of these models (GISS-E2-1-H) is in the top three in terms of NTA wind
520 variability (Fig. 13b). Conversely, one of the top three models in terms of SST variability
521 (E3SM-1-0) is in the bottom three of wind variability, because it overestimates the ampli-
522 tude. This again points to an excessively deep layer in the models, which means that a
523 model with realistic wind variability will underestimate SST variability.

524 We form AMM composites based on our AMM index (NTA SST minus STA SST)
525 exceeding 1 standard deviation for the MAM average (Fig. 14). The developing phase of

526 the AMM in ERA-5 is marked by westerly 10m wind anomalies over the NTA (indicating
527 a weakening of the northeasterly trades), accompanied by the expected negative (down-
528 ward) latent heat flux anomalies. In response to the surface heat flux forcing, NTA SSTs
529 grow in January, peak in February and persist until June. The STA SST show a similar
530 evolution but of opposite sign and weaker amplitude. LHF anomalies switch sign in May
531 and thereafter contribute to the decay of the AMM. The sign switch of LHF anomalies
532 results from the competing effects of growing SST anomalies (which act to increase LHF
533 anomalies) and decreasing surface wind anomalies (which act to decrease LHF anomalies).
534 The evolution of AMM events is relatively similar in ens-cmip6, although the wind and
535 LHF anomalies peak one month later in February. Similarly to the AZM composites, we
536 pick the top and bottom three models based on the sum of the RMSEs of NTA SST and
537 NTA 10m zonal wind, where both RMSEs are first averaged over the period January
538 through July. The top three models all show very similar behavior, with the AMM event
539 peaking in April, preceded by peaks in wind and LHF anomalies in January or February
540 (Figs. 14c, 14e, and 14g). The switch to positive LHF anomalies (damping of the AMM)
541 occurs in April or May, which is similar to the reanalysis composite. The wind anomalies
542 in the top three models have a realistic amplitude, even though they overestimate the stand-
543 ard deviation of 10m wind in the area (Figs. 13b and S13b). A possible reason for this
544 discrepancy is the constraint of covariability with the STA SST in the selection of events.
545 The bottom three models show very disparate behavior. The AMM events in
546 NorCPM1 are about 50% weaker than in the ERA-5 (Fig. 14d), while those in NESM3 and
547 E3SM-1-1 are stronger by about 30% and 50%, respectively (Figs. 14f and 14h). The

seasonal evolution of the composite fields, however, is still rather similar to that in the ERA-5.

5. Linkage to the tropical Pacific

We explore the linkage between the tropical Atlantic and Pacific using the AZM composites introduced in section 4 but showing a longer period and including anomalies of Niño 3.4 SST and Niño 4 (160°W-150°E, 5°S-5°N) 10m zonal wind (Fig. 15). In the ERA-5 reanalysis, fall and winter before the peak of the AZM event show approximately neutral conditions in terms of Niño 3.4 SST anomalies (Fig. 15a). This is consistent with previous studies pointing out the inconsistent influence of ENSO on the AZM (Chang et al. 2006; Lübbecke and McPhaden 2012), which may be partly due to low-frequency modulation (Joly and Voldoire 2010). Starting from spring there are easterly anomalies in the Niño 4 and cold anomalies in the Niño 3.4, and these continue to grow toward the end of the year. The AZM event, which peaks in June, appears to lead developments in the Pacific, giving the impression that the Atlantic is driving the Pacific, as suggested by some authors (Rodriguez-Fonseca et al. 2009; Ding et al. 2012; see Kucharski et al. 2016 for a review). It should be noted, however, that there are only a limited number of AZM events in the study period, and that none of the Pacific anomalies is statistically significant at the 95% level. ens-cmip6 shows very weak anomalies in the Pacific, though the evolution from positive to negative anomalies is somewhat similar to the ERA-5 reanalysis (Fig. 15b).

In the following we discuss a few individual models that exemplify distinct behavior in the tropical Pacific even though they all feature (by construction) positive AZM events in JJA. The IPSL-CM6A-LR and MIROC6 show negative wind and SST anomalies in the tropical Pacific that become statistically significant from boreal winter (Figs. 15c and 15d).

571 The behavior of these models is similar to that in the ERA-5 reanalysis. The HadGEM3-
572 GC31-MM and CNRM-CM6-1-HR have neutral conditions in the Pacific during the whole
573 period shown (Figs. 15e and 15f). Finally, the GISS-E2-1-G and the CAMS-CSM1-0 show
574 pronounced positive anomalies in the tropical Pacific that are already prominent in the first
575 month shown, October of year -1 (Figs. 15g and 15h). At the same time, these models show
576 weakly easterly (negative) wind anomalies in the ATL4 region from Oct year -1 through
577 May year 0. This kind of evolution is untypical for positive AZM events, which are thought
578 to rely on downwelling caused by westerly (positive) zonal wind anomalies. The sugges-
579 tion is that AZM events in the two models are due primarily to the widespread tropospheric
580 warming that accompanies El Niño events (Chang et al. 2006). It is not clear though why
581 these models do not develop the typical easterly ATL4 anomalies found in observations
582 during El Niño events.

583 We examine horizontal maps of the composites in MAM in the ERA-5 and three se-
584 lected models (Fig. 16; note that linear regression of MAM fields on the JJA ATL3 index
585 yields very similar results). The ERA-5 shows weakly cold SST anomalies in the eastern
586 equatorial Pacific accompanied by weak westerly anomalies that are in contrast to the east-
587 erly Niño 4 anomalies (Fig. 15). The competing effects of these wind anomalies may partly
588 explain the weak SST signal. In IPSL-CM6A-LR the SST and 10m zonal wind anomalies
589 are of the same sign in the eastern equatorial Pacific, and of opposite sign to the wind
590 anomalies in the equatorial Atlantic. This is consistent with the dynamic coupling of the
591 two basins through shifts in the Walker circulation (Chang et al. 2006). In HadGEM3-
592 GC31-MM, both SST and 10m wind anomalies are very small, not only over the equator
593 but over the entire eastern tropical Pacific. Since ENSO has realistic strength in this model,

the lack of anomalies in the composites suggests that these are averaged out because AZMs occur independently of ENSO in the simulation. Finally, GISS-E2-1-G shows pronounced warming of SST in the eastern tropical Pacific that extends into the tropical Atlantic, presumably through the atmospheric bridge.

6. Summary and discussion

6.1. Summary

We have evaluated the performance of CMIP6 models with respect to representing the mean state and variability of the tropical Atlantic, and its linkage to the tropical Pacific. Our results indicate that biases in the ensemble mean have seen both slight improvement and deterioration compared to an ensemble of CMIP5 models. In the eastern equatorial Atlantic, the warm SST biases are slightly exacerbated in CMIP6 relative to CMIP5, as is the erroneous southward shift of the Atlantic ITCZ. On the positive side, the equatorial SST gradient is slightly improved, the cold SST bias in the subtropical Atlantic is reduced, and the warm SST biases in the eastern upwelling regions have become slightly less severe.

While these results may not seem particularly encouraging, there are a few models that have drastically reduced SST and 10m wind biases in the equatorial Atlantic. Moreover, interannual variability in both the equatorial and subtropical Atlantic is relatively well simulated in many of the CMIP6 models, though this was already true for a number of CMIP5 models (Richter et al. 2014a).

Comparison of fully coupled models with their atmosphere-only counterparts indicates that westerly wind biases and the southward ITCZ shift in MAM continue to be a problem in CMIP6, and that these biases are amplified by coupled process in JJA, as suggested by previous studies (Richter and Xie 2008). There is, however, also an indication that the

617 southward ITCZ shift is quite sensitive to subtropical SST biases in coupled simulations,
618 particularly to those in the southeastern tropical Atlantic, which is consistent with Xu et al.
619 (2014a).

620 High-resolution models tend to perform better than the general model ensemble but do
621 not represent, on average, a breakthrough. The three models with the highest resolution,
622 however, do show very small biases, indicating that increasing resolution may be a way of
623 tackling the Atlantic bias problem. On the other hand, some models with relatively coarse
624 resolution, achieve a quite realistic mean state, which indicates that improvements in pa-
625 rameterizations may be another avenue of successfully reducing tropical Atlantic biases.

626 Tropical Atlantic variability is quite realistic in many CMIP6 models. These models
627 correctly capture the amplitude and seasonality of both equatorial Atlantic (AZM) and sub-
628 tropical Atlantic (AMM) variability. Regarding positive AZM events, many models cor-
629 rectly simulate the peak of ATL3 SST anomalies in JJA and the westerly wind anomalies
630 that drive development in the preceding months. Most models, however, continue to de-
631 velop the peak 1-2 months later than observed. Furthermore, model SSTs in the ATL3
632 appear to be less sensitive to equatorial surface wind forcing than the ERA-5 reference.
633 This hints at a thermocline that is too deep or too diffuse, which may be due to problems
634 inherent to the oceanic components of GCMs. Similar problems are apparent in the repre-
635 sentation of the AMM, where most models capture the amplitude of either wind anomalies
636 or SST anomalies, but not both. While some of the models with realistic variability also
637 have small mean state biases, there are a number of counter examples.

638 The linkage of the AZM to the tropical Pacific varies widely across models, with the
639 developing and mature phases of the AZM being accompanied by SSTs in the tropical

Pacific that can be opposite-signed, neutral, or same-signed, depending on the model. The same-signed SST anomalies in the tropical Pacific are not typical of observed AZMs and the models producing this pattern usually suffer from severe mean state biases. Neutral ENSO conditions throughout the AZM cycle are also relatively uncommon in observations, but the models exhibiting this behavior tend to have realistic equatorial variability in both the Atlantic and the Pacific.

6.2. Discussion

While the lack of substantial improvement in ensemble mean CMIP6 biases in the tropical Atlantic may seem discouraging at first, the fact that a few models were able to substantially reduce tropical Atlantic biases should be quite encouraging. Some examples for the equatorial Atlantic are the HadGEM family, the IPSL family, and the CanESM family. Closer examination of how successive generations of these models gradually improved, and how these improvements were achieved, may provide valuable guidance to other modeling centers.

It is also encouraging that variability patterns in the equatorial and subtropical Atlantic are now very well captured in several models. While already in CMIP5 some models had relatively small variability errors, this number has increased in CMIP6, with some models having very small errors. This should give confidence that some current models are useful tools to understand and predict variability in the tropical Atlantic. On the other hand, the wide range of behaviors across models in terms of linkages between the equatorial Atlantic and Pacific variability patterns indicates that there remains much room for improvement, though interdecadal modulation of the linkage may complicate identifying the "correct" model behavior. Clearly, simulating the intricacies of tropical basin interaction is a harder task than capturing the variability of one particular basin. Likewise, the remote impact of

tropical Atlantic variability on the surrounding continents will likely continue to pose challenges.

Finally, the current wide spread of tropical Atlantic model performance offers an opportunity to examine the impact of model biases on regional projections. This could be achieved by comparing the projections of models with small and large biases. While there are some indications that model biases do not have a very strong impact on prediction skill in the tropical Atlantic (Richter et al. 2018; Richter al. 2020), understanding their impact on long-term projections remains a challenge. Exploring this issue should help to increase confidence in climate change projections for the tropical Atlantic region.

Acknowledgments

We thank the three anonymous reviewers for their helpful comments. 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. This work was supported by the Japan Society for the Promotion Science KAKENHI, Grant Numbers 18H01281, 18H03726, and 19H05704.

References

Adler, R. F., and Coauthors, 2018: The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138, https://doi.org/10.3390

Amaya, D.J., DeFlorio, M.J., Miller, A.J. *et al.* WES feedback and the Atlantic Meridional Mode: observations and CMIP5 comparisons. *Clim Dyn* **49,** 1665–1679 (2017) doi:10.1007/s00382-016-3411-1

Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.,* **97,** 163–172.

Chang, P., Y. Fang, R. Saravanan, L. Ji, and H. Seidel, 2006: The cause of the fragile relationship between the Pacific El Niño and the Atlantic Niño. *Nature*, **443,** 324–328.

Chang CY, Carton JA, Grodsky SA, Nigam S (2007) Seasonal climate of the tropical Atlantic sector in the NCAR community climate system model 3: error structure and probable causes of errors. J Clim 20:1053–1070

Chiang JC, Vimont DJ (2004) Analogous Pacific and Atlantic meridional modes of tropical atmosphere-ocean variability. J Clim 17:4143–4158

Dai AG (2006) Precipitation characteristics in eighteen coupled climate models. J Clim 9:4605–4630

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

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

Ding, H., Keenlyside, N. S., & Latif, M. (2012). Impact of the equatorial Atlantic on the El Niño Southern Oscillation. Climate Dynamics, 38, 1965–1972. https://doi.org/10.1007/s00382-011-1097-y

Haarsma, R. J., E. Campos, W. Hazeleger, and C. Severijns, 2008: Influence of the meridional overturning circulation on tropical Atlantic climate and variability. *J. Climate*, **21**, 1403–1416, https://doi.org/10.1175/2007JCLI1930.1.

Harlaß, J., Latif, M. & Park, W. Alleviating tropical Atlantic sector biases in the Kiel climate model by enhancing horizontal and vertical atmosphere model resolution: climatology and interannual variability. *Clim Dyn* **50,** 2605–2635 (2018) doi:10.1007/s00382-017-3760-4

Harrison DE, Vecchi GA (1999) On the termination of El Niño. Geophys Res Lett 26:1593–1596

Hersbach, H., and Coauthors, 2018: Operational global reanalysis: progress, future directions and synergies with NWP, ECMWF ERA Report Series 27.

Hourdin, F., T. Mauritsen, A. Gettelman, J. Golaz, V. Balaji, Q. Duan, D. Folini, D. Ji, D. Klocke, Y. Qian, F. Rauser, C. Rio, L. Tomassini, M. Watanabe, and D. Williamson, 2017: The Art and Science of Climate Model Tuning. Bull. Amer. Meteor. Soc., 98, 589–602, https://doi.org/10.1175/BAMS-D-15-00135.1

726 Joly M. and A. Voldoire, 2010: Role of the Gulf of Guinea in the interannual variability of
727 the West African monsoon: what do we learn from CMIP3 coupled simulations? Int. J.
728 of Clim., 30(12), 1843-1856, DOI:10.1002/joc.2026.
729 Keenlyside, N. S., and Latif, M. (2007). Understanding equatorial Atlantic interannual var-
730 iability. *J. Clim.* 20, 131–142. doi: 10.1175/JCLI3992.1
731 Kucharski, F., Parvin, A., Rodriguez-Fonseca, B., Farneti, R., Martin-Rey, M., Polo, I.,
732 Mohino, E., Losada, T., Mechoso, C.R., 2016: The Teleconnection of the Tropical At-
733 lantic to Indo-Pacific Sea Surface Temperatures on Inter-Annual to Centennial Time
734 Scales: A Review of Recent Findings. Atmosphere, 7, 29.
735 Kurian, J., P. Li, P. Chang, C. M. Patricola, and J. Small, 2020: Impact of the Benguela
736 coastal low-level jet on the southeast tropical Atlantic SST bias in a regional ocean
737 model. *submitted*
738 Li, G., and Xie, S.-P., 2012: Origins of tropical-wide SST biases in CMIP multi-model
739 ensembles. *Geophys. Res. Lett.*, 39, L22703, doi:10.1029/2012GL053777.
740 Li G, Xie S-P (2014) Tropical biases in CMIP5 multimodel ensemble: The excessive equa-
741 torial Pacific cold tongue and double ITCZ problems. J Clim 27:1765-1780,
742 https://doi.org/10.1175/JCLI-D-13-00337.1
743 Lin, J.-L., 2007: The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean–atmos-
744 phere feedback analysis. *J. Climate*, **20**, 4497–4525.
745 Lübbecke, J. F., and M. J. McPhaden, 2012: On the inconsistent relationship between Pa-
746 cific and Atlantic Niños. *J. Climate*, **25**, 4294–4303.

747 Martin-Rey, M., I. Polo, B. Rodrıguez-Fonseca, T. Losada, and A. Lazar, 2018: Is There
748 Evidence of Changes in Tropical Atlantic Variability Modes under AMO Phases in the
749 Observational Record? J. Clim. 31, 515–536.
750 McGregor, S., M. F. Stuecker, J. B. Kajtar, M. H. England, and M. Collins, 2018: Model
751 tropical Atlantic biases underpin diminished Pacific decadal variability. *Nat. Climate*
752 *Change*, **8**, 493–498, https://doi.org/10.1038/s41558-018-0163-4.
753 Mechoso C, Robertson A, Barth N, Davey M, Delecluse P, Gent P, Tribbia J (1995) The
754 seasonal cycle over the tropical Pacific in coupled ocean-atmosphere general circulation
755 models. Mon Wea Rev 123: 2825– 2838
756 Milinski, S., J. Bader, H. Haak, A. C. Siongco, and J. H Jungclaus (2016), High atmos-
757 pheric horizontal resolution eliminates the wind-driven coastal warm bias in the south-
758 eastern tropical Atlantic, Geophys. Res. Lett., 43. 10.1002/2016GL070530
759 Nnamchi, H. C., Li, J., Kucharski, F., Kang, I., Keenlyside, N. S., Chang, P., et al. (2016).
760 An equatorial–extratropical dipole structure of the Atlantic Niño. J. Clim. 29, 7295–
761 7311. doi: 10.1175/JCLI-D-15-0894.1
762 Oettli, P., C. Yuan, and I. Richter, 2020: The Other Coastal Niño/Niña — The Benguela,
763 California and Dakar Niños/Niñas. In “Tropical and extra-tropical air-sea interactions”,
764 edited by S. K. Behera, Elsevier; *under review*
765 Okumura, Y., and S.-P. Xie, 2006: Some overlooked features of tropical Atlantic climate
766 leading to a new Nino-like phenomenon. *J. Climate*, 19, 5859-5874.
767 Park, W., and M. Latif, 2020: Sensitivity of CO_2-induced Tropical Atlantic Sector Climate
768 Changes to Model Bias. Climate and Atmospheric Science, *under review*

769 Patricola, C.M., Chang, P., 2017. Structure and dynamics of the Benguela low-level coastal
770 jet. Clim. Dyn. 49, 2765–2788.
771 Pauluis, O., 2004: Boundary layer dynamics and cross-equatorial Hadley circulation. *J.*
772 *Atmos. Sci.*, **61,** 1161–1173.
773 Polo, I., B. W. Dong, and R. T. Sutton, 2013: Changes in tropical Atlantic interannual
774 variability from a substantial weakening of the meridional overturning circulation. *Cli-*
775 *mate Dyn.*, **41**, 2765–2784, https://doi.org/10.1007/s00382-013-1716-x.
776 Prigent, A., J. Lübbecke, T. Bayr, M. Latif, C. Wengel, 2020: Weakened SST variability
777 in the tropical Atlantic Ocean since 2000. Clim. Dyn., *accepted*
778 Richter, I., and Xie, S.-P. (2008). On the origin of equatorial Atlantic biases in coupled
779 general circulation models. Clim. Dynam. 31, 587–598. doi: 10.1007/ s00382- 008-
780 0364- z
781 Richter, I., Behera, S. K., Masumoto, Y., Taguchi, B., Komori, N., and Yamagata, T.
782 (2010). On the triggering of Benguela Niños: remote equatorial versus local influences.
783 Geophys. Res. Lett. 37:L20604. doi: 10.1029/2010GL044461
784 Richter, I., Xie, S.-P., Wittenberg, A. T., and Masumoto, Y. (2012). Tropical Atlantic bi-
785 ases and their relation to surface wind stress and terrestrial precipitation. Clim. Dynam.
786 38, 985–1001. doi: 10.1007/s00382-011-1038-9
787 Richter, I., Behera, S. K., Masumoto, Y., Taguchi, B., Sasaki, H., and Yamagata, T. (2013).
788 Multiple causes of interannual sea surface temperature variability in the equatorial At-
789 lantic Ocean. Nat. Geosci. 643. doi: 10.1038/ngeo1660

790 Richter, I., Xie, S. P., Behera, S. K., Doi, T., and Masumoto, Y. (2014a). Equatorial Atlan-
791 tic variability and its relation to mean state biases in CMIP5. Clim. Dynam. 42, 171–
792 188. doi: 10.1007/s00382-012-1624-5
793 Richter, I., Behera, S. K., Doi, T., Taguchi, B., Masumoto, Y., and Xie, S. P. (2014b). What
794 controls equatorial Atlantic winds in boreal spring? Clim. Dyn. 43, 3091–3104. doi:
795 10.1007/s00382-014-2170-0
796 Richter, I. (2015). Climate model biases in the eastern tropical oceans: causes, impacts and
797 ways forward. WIREs Clim. Change 6, 345–358. doi: 10.1002/wcc.338
798 Richter, I., and Coauthors, 2016: An overview of coupled GCM biases in the tropics. Indo-
799 Pacific Climate Variability and Predictability, T. Yamagata and S. K. Behera, Eds.,
800 World Scientific Series on Asia-Pacific Weather and Climate, Vol. 8, World Scientific,
801 213–263, https://doi.org/10.1142/9789814696623_0008.
802 Richter, I., S.-P. Xie, Y. Morioka, T. Doi, B. Taguchi, and S. K. Behera, 2017: Phase lock-
803 ing of equatorial Atlantic variability through the seasonal migration of the ITCZ. *Cli-*
804 *mate Dyn.*, **48**, 3615–3629, https://doi.org/10.1007/s00382-016-3289-y.
805 Richter, I., Doi, T., Behera, S. K., and Keenlyside, N. (2018). On the link between mean
806 state biases and prediction skill in the tropics: an atmospheric perspective. Clim. Dynam.
807 50, 3355–3374. doi: 10.1007/s00382-017-3809-4
808 Richter, I. and T. Doi, 2019: Estimating the Role of SST in Atmospheric Surface Wind
809 Variability over the Tropical Atlantic and Pacific. J. Climate, 32, 3899–3915,
810 https://doi.org/10.1175/JCLI-D-18-0468.1

Richter, I., and H. Tokinaga, 2020: The Atlantic Niño: dynamics, thermodynamics, and teleconnections. In “Tropical and extra-tropical air-sea interactions”, edited by S. K. Behera, Elsevier; *under review*

Rodríguez-Fonseca, B., Polo, I., García-Serrano, J., Losada, T., Mohino, E., Mechoso, C. R., et al. (2009). Are Atlantic Niños enhancing Pacific ENSO events in recent decades? Geophys. Res. Lett. 36:20705.

Servain, J., I. Wainer, J. P. McCreary, and A. Dessier, 1999: Relationship between the Equatorial and meridional modes of climatic variability in the tropical Atlantic. *Geophys. Res. Lett.,* **26,** 485–488.

Shannon, L.V., Boyd, A.J., Brundrit, G.B., Taunton-Clark, J., 1986. On the existence of an El Niño-type phenomenon in the Benguela system. J. Mar. Res. 44, 495–520.

Small, R. J., E. Curchitser, K. Hedstrom, B. Kauffman, and W. Large, 2015: The Benguela upwelling system: Quantifying the sensitivity to resolution and coastal wind represen-tation in a global climate model. *J. Climate*, **28**, 9409–9432, doi:https://doi.org/10.1175/JCLI-D-15-0192.1.

Song, Z. S.-K. Lee, C. Wang, B. Kirtman and F. Qiao, 2015: Contributions of the atmos-phere-land and ocean-sea ice model components to the tropical Atlantic SST bias in CESM1. Ocean Modelling, 96, 280-290, https://doi.org/10.1016/j.ocemod.2015.09.008.

Steinig, S., Harlaß, J., Park, W. & Latif, M. Sahel rainfall strength and onset improvements due to more realistic Atlantic cold tongue development in a climate model. *Scientific Reports* **8**, 1–9 (2018).

Tokinaga, H., and S.-P. Xie, 2011: Weakening of the equatorial Atlantic cold tongue over the past six decades. *Nat. Geosci.*, **4**, 222–226.

834 Tokinaga, H., S.-P. Xie, A. Timmermann, S. McGregor, T. Ogata, H. Kubota, and Y. M.
835 Okumura, 2012: Regional patterns of tropical Indo-Pacific climate change: Evidence of
836 the Walker circulation weakening. *J. Climate*, **25**, 1689–1710.
837 Vecchi, G., Soden, B., Wittenberg, A. *et al.* Weakening of tropical Pacific atmospheric
838 circulation due to anthropogenic forcing. *Nature* **441,** 73–76 (2006) doi:10.1038/na-
839 ture04744
840 Voldoire, A., and Coauthors, 2019: Role of wind stress in driving SST biases in the Trop-
841 ical Atlantic, Clim. Dyn., doi:10.1007/s00382-019-04717-0, 53(5), 3481-3504.
842 Wahl S, Latif M, Park W, Keenlyside N (2011) On the tropical Atlantic SST warm bias in
843 the Kiel climate model. Clim Dyn 36:891–906
844 Wang, C., L. Zhang, S.-K. Lee, L. Wu, and C. R. Mechoso, 2014: A global perspective on
845 CMIP5 climate model biases. *Nat. Climate Change*, **4**, 201–205,
846 doi:https://doi.org/10.1038/nclimate2118.
847 Webster PJ (1981) Mechanisms determining the atmospheric response to large-scale sea
848 surface temperature anomalies. J Atmos Sci 38:554–571
849 Xu Z, Chang P, Richter I, Kim W, Tang G, 2014a: Diagnosing southeast tropical Atlantic
850 SST and ocean circulation biases in the CMIP5 ensemble. Clim Dyn 43(11):3123–3145
851 Xu, Z., M. Li, C. M. Patricola, and P. Chang, 2014b: Oceanic origin of southeast tropical
852 Atlantic biases. *Climate Dyn.*, **43**, 2915–2930, doi:https://doi.org/10.1007/s00382-013-
853 1901-y.
854 Zebiak SE (1986) Atmospheric convergence feedback in a simple model for El Niño. Mon
855 Weather Rev 114:1263–1271

856 Zermeno-Diaz, D., and C. Zhang, 2013: Possible root causes of surface westerly biases
857 over the equatorial Atlantic in global climate models. J. Climate, 26, 8154–8168,
858 doi:https://doi.org/10.1175/JCLI-D-12-00226.1.
859
860
861

Captions

Table 1. Model information about the CMIP6 piControl simulations used in this study.

Table 2. As in Table 1 but for the CMIP6-HighResMIP model and high-resolution CESM control simulations. Models used for the ensemble ens-HR (ens-LR) are indicated by superscript "HR" ("LR"). Models marked by * are removed from the ensemble ens-hires. Models marked by @ form ensemble ens-hires-a.

Figure 1. Climatological SST (color shading; ˚C), 10-m wind vectors (m s-1) and precipitation (blue shading; mm d-1) for 1979-2017. (a) MAM and (b) JJA seasons. SST and near-surface winds are from ERA5, precipitation from GPCP.

Figure 2. Biases of SST (shading; K), 10m wind (vectors; reference 2 m/s) and precipitation (contours; interval 2 mm/day; 0-contour omitted; negative contours dashed) for MAM (top row) and JJA (bottom row). The left and right columns show the ensemble average of piControl models from CMIP6 and CMIP5, respectively.

Figure 3. Difference between model ensembles in JJA for SST (shading; K), 10m wind (vectors; reference 1 m/s) and precipitation (contours; mm/day; 0-contour omitted; negative contours dashed). The ensemble differences shown are (a) CMIP6 minus CMIP5, (b) HighResMIP-MIP minus – CMIP5, and (c) HighResMIP minus CMIP6.

885 **Figure 4.** Climatological annual cycle of (a) ATL3 SST (°C), (b) ATL4 10m zonal
886 wind, (c) ABA SST (°C), and (d) Niño 3.4 SST. The thick black line shows the ERA-5
887 reanalysis, the thick grey line the ens-cmip6 ensemble average, and the dashed thick grey
888 line the ens-cmip5 ensemble average. Each panel also shows 6 individual models, which
889 are chosen based on an RMSE criterion relative to the ERA-5 reference. In each panel, the
890 top 3 models are the ones with the lowest RMSE (smallest error), and the bottom 3 are the
891 ones with the highest RMSE (largest error).
892
893 **Figure 5.** Multi-model scatter plot showing the climatological ATL4 10m zonal wind
894 (m/s) in MAM on the x-axis and the ATL3 SST (°C) in JJA on the y-axis. The black line
895 shows the regression line. The intermodel correlation is indicated in the lower right. Each
896 letter corresponds to one model, as indicated in the legend at the bottom.
897
898 **Figure 6.** The top row shows longitude-time section of biases of SST (shading; K)
899 and 10m zonal wind (contours; interval 0.5; negative contours dashed) for (a) an ensemble
900 of coupled piControl models, and (b) an ensemble of the corresponding atmosphere-only
901 models forced with observed SST (experiment amip). The bottom row shows latitude-time
902 sections of precipitation biases (shading; mm/day), 10m wind (vectors; reference 2 m/s),
903 and SST (contours; K; negative contours dashed), averaged from 40W to 20W for (c) the
904 coupled model ensemble, and (d) the atmosphere-only ensemble.
905

Figure 7. Difference between the piControl and amip ensembles in JJA for SST (shading; K), 10m wind (vectors; reference 2 m/s) and precipitation (contours; interval 1 mm/day; 0-contour omitted; negative contours dashed).

Figure 8. Biases of SST (shading; K), 10m wind (vectors; reference 2 m/s) and precipitation (contours; mm/day; 0-contour omitted; negative contours in blue) for MAM (top row) and JJA (bottom row). The left column shows the ensemble average over all models in the control-1950 experiment (high-resolution models), the center column the average over 5 selected models with particularly high resolution, and the right column the average over corresponding model versions with relatively low resolution.

Figure 9. Longitude-time section of biases of SST (shading; K) and 10m zonal wind (contours; interval 0.5; negative contours dashed) for the ensemble average of (a) all control-1950 models, (b) 5 selected models with particularly high resolution, and (c) the 5 corresponding low-resolution versions of those models. The bottom row shows the same fields but for the corresponding AMIP simulations.

Figure 10. As in Fig. 4, but for the seasonally stratified variability of each index, calculated as the standard deviation of the detrended anomalies.

Figure 11. AZM composites showing anomalies of ATL4 10m zonal wind (green line; m/s), ATL3 SST (blue line; K), SEQ precipitation (orange line; mm/day), and STA SLP (dark red line; hPa). The panels show (a) ERA-5, (b) ens-cmip6, (c), (e), (g) the 3

models with the smallest error relative to the ERA-5 reference, and (d), (f), (h) the 3 models with the largest error relative to the ERA-5 reference. The error is calculated as a combination of the RMSE of ATL3 SST and ATL4 10m zonal wind. See text for details. Filled circles indicate anomalies that are statistically significant at the 95% level.

Figure 12. Running standard deviation of ATL3 SST anomalies (K) using a 20-year sliding window. Colored lines show 6 models from ens-cmip6 that were selected based on the same performance metric as in Fig. 11 but excluding those models with less than 500-years of simulation. The two horizonal lines show the standard deviation of ATL3 SST in the ERA-5 reanalysis for the periods 1979-1998 (top line) and 1999-2018 (bottom line).

Figure 13. As in Fig. 10 but for the standard deviation of (a) NTA SST (K), and (b) NTA 10m zonal wind (m/s).

Figure 14. AMM composites showing anomalies of NTA SST (green line; K), STA SST (blue line; K), NTA 10m zonal wind (orange line; m/s), and NTA surface latent heat flux (dark red line; W/m2*0.05). The panels show (a) ERA-5, (b) ens-cmip6, (c)-(e) the 3 models with the smallest error relative to the ERA-5 reference, and (f)-(h) the 3 models with the largest error relative to the ERA-5 reference. The error is calculated as a combination of the RMSE of NTA SST and NTA 10m zonal wind. See text for details.

Figure 15. AZM composites showing anomalies of ATL3 SST (green line; K), Niño 3.4 SST (blue line; K), ATL4 10m zonal wind (orange line; m/s), and Niño 4 10m zonal

wind (dark red line; m/s). The panels show (a) ERA-5, (b) ens-cmip6, (c)-(h) 6 models with distinct behavior in terms of the concomitant sign of Niño 3.4 SST anomalies: negative (2nd row), neutral (3rd row) and positive (4th row).

Figure 16. Horizontal maps of 4 of the composites shown in Fig. 15 (ERA-5, IPSL-CM6A-LR, HadGEM3-GC31-MM, and GISS-E2-1-G). The maps show the MAM average (one season before the peak of the AZM) for the following fields: SST (shading; K), 10 m wind (vectors; reference 1 m/s), and precipitation (contours; interval 1 mm/day; negative contours dashed).

A. Tables

Table 1. Model information about the CMIP6 piControl simulations used in this study.

Model name	Nation	Atmospheric model	Ocean model
		Atmosphere resolution	Ocean resolution
AWI-CM-1-1-MR	Germany	ECHAM6.3.04p1 T127 (~0.94˚× 0.94˚); 95 levels	FESOM 1.4 Unstructured grid in the horizontal with 830305 wet nodes; 46 levels
BCC-CSM2-MR[A]	China	BCC_AGCM3_MR T106 (~1.125˚× 1.125˚); 46 levels	MOM4 1/3˚ in 30˚S-30˚N, 1/3˚-1˚ in 30˚ - 60˚N/S, and 1˚ in high latitudes; 40 levels
BCC-ESM1[A]	China	BCC_AGCM3_LR T42 (~2.8125˚× 2.8125˚); 26 levels	MOM4 1/3˚ in 10˚S-10˚N, 1/3˚-1˚ in 10˚ - 30˚N/S, and 1˚ in high latitudes; 40 levels
CAMS-CSM1-0[A]	China	ECHAM5_CAMS T106 (~1.125˚× 1.125˚); 31 levels	MOM4 Primarily 1˚×1˚, and 1/3˚ in 30˚S-30˚N; 50 levels
CanESM5[A]	Canada	CanAM5 T63 linear gaussian grid (~2.8125˚× 2.8125˚); 49 levels	NEMO3.4.1 1˚ with refinement to 1/3˚ in 20˚S-20˚N; 45 levels
CESM2-WACCM[A]	USA	CAM6 Finite volume grid (0.9˚×1.25˚); 70 levels	POP2 320x384 longitude/latitude; 60 levels
CESM2[A]	USA	CAM6 Finite volume grid (0.9˚×1.25˚); 32 levels	POP2 320x384 longitude/latitude; 60 levels
CNRM-CM6-1-HR	France	Arpege 6.3 T359 (~35km); 91 levels	NEMO3.6 Primarily 0.25˚; 1442 x 1050 longitude/latitude; 75 levels
CNRM-CM6-1[A]	France	Arpege 6.3 T127 (~100km); 91 levels	NEMO3.6 Primarily 1˚; 362 x 294 longitude/latitude; 75 levels
CNRM-ESM2-1[A]	France	Arpege 6.3 T127 (~100km); 91 levels	NEMO3.6 Primarily 1˚; 362 x 294 longitude/latitude; 75 levels
E3SM-1-0[A]	USA	EAM v1.0 1˚ average grid spacing; 72 levels	MPAS-Ocean v6 Variable resolution 60 km to 30 km; 60 levels

Model name	Nation	Atmospheric model	Ocean model
		Atmosphere resolution	Ocean resolution
E3SM-1-1	USA	EAM v1.1 1˚ average grid spacing; 72 levels	MPAS-Ocean v6 Variable resolution 60 km to 30 km; 60 levels
EC-Earth3-Veg[A]	European countries	IFS cy36r4 TL255 (~70km); 91 levels	NEMO3.6 Primarily 1˚; 362 x 294 longitude/latitude; 75 levels
EC-Earth3[A]	European countries	IFS cy36r4 TL255 (~70km); 91 levels	NEMO3.6 Primarily 1˚; 362 x 294 longitude/latitude; 75 levels
FGOALS-f3-L[A]	China	FAMIL2.2 C96 (~1˚×1); 32 levels	LICOM3.0 Primarily 1˚; 360 x 218 longitude/latitude; 30 levels
FGOALS-g3[A]	China	GAMIL2 180x90 (~200km); 26 levels	LICOM3.0 Primarily 1˚; 360 x 218 longitude/latitude; 30 levels
GISS-E2-1-G-CC	USA	GISS-E2.1 2.5˚x2˚; 40 levels	GISS Ocean 1.25˚x1˚; 32 levels
GISS-E2-1-G[A]	USA	GISS-E2.1 2.5˚x2˚; 40 levels	GISS Ocean 1.25˚x1˚; 32 levels
GISS-E2-1-H	USA	GISS-E2.1 2.5˚x2˚; 40 levels	HYCOM Ocean ~1˚×1˚; 26 levels
HadGEM3-GC31-LL[A]	UK	MetUM-HadGEM3-GA7.1 N96 (1.875˚x1.25˚); 85 levels	NEMO-HadGEM3-GO6.0 Primarily 1˚ with meridional refinement down to 1/3 ˚ in the tropics; 75 levels
HadGEM3-GC31-MM[A]	UK	MetUM-HadGEM3-GA7.1 N216 (~0.83˚x0.55˚); 85 levels	NEMO-HadGEM3-GO6.0 Primarily 0.25˚; 1440 x 1205 longitude/latitude; 75 levels
IPSL-CM6A-LR[A]	France	LMDZ N96 (2.5˚x1.259˚); 79 levels	NEMO-OPA Primarily 1˚; 362 x 332 longitude/latitude; 75 levels
MCM-UA-1-0	USA	R30L14 3.75˚ x 2.5˚; 14 levels	MOM1.0 1.875 ˚x 2.5˚; 18 levels
MIROC-ES2L	Japan	CCSR AGCM T42 (~2.8125˚× 2.8125˚); 42 levels	COCO4.9 Primarily 1˚; 360 x 256 longitude/latitude; 63 levels

Model name	Nation	Atmospheric model	Ocean model
		Atmosphere resolution	Ocean resolution
MIROC6[A]	Japan	CCSR AGCM T85 (~1.4˚× 1.4˚); 81 levels	COCO4.9 Primarily 1˚; 360 x 256 longitude/latitude; 63 levels
MPI-ESM1-2-HR	Germany	ECHAM6.3 T127 (0.94˚x0.94˚); 95 levels	MPIOM1.63 Approximately 0.4˚; 802 x 404 longitude/latitude; 40 levels
MRI-ESM2-0[A]	Japan	MRI-AGCM3.5 TL159 (~120km); 80 levels	MRI.COM4.4 ~1˚x0.5˚; 360 x 364 longitude/latitude; 61 levels
NESM3[A]	China	ECHAM v6.3 T63 (1.9˚x1.9˚); 47 levels	NEMO3.4 Primarily 1˚; 384 x 362 longitude/latitude; 46 levels
NorCPM1[A]	Norway	CAM-OSLO4.1 ~2.5˚x2; 26 levels	MICOM1.1 1˚; 320 x 384 longitude/latitude; 53 levels
NorESM1-F	Norway	CAM4 ~2.5˚x2˚; 32 levels	MICOM 1˚; 360 x 384 longitude/latitude; 70 levels
NorESM2-LM[A]	Norway	CAM-OSLO ~2.5˚x2˚; 32 levels	MICOM 1˚; 360 x 384 longitude/latitude; 70 levels
SAM0-UNICON[A]	Republic of Korea	CAM5.3 with UNICON ~1˚x1˚; 30 levels	POP2 320 x 384 longitude/latitude; 60 levels
UKESM1-0-LL[A]	UK	MetUM-HadGEM3-GA7.1 N96 (1.875˚x1.25˚); 85 levels	NEMO-HadGEM3-GO6.0 Primarily 1˚ with meridional refinement down to 1/3 ˚ in the tropics; 75 levels

980 **Table 2.** As in Table 1 but for the CMIP6-HighResMIP model and high-resolution CESM
981 control simulations. Models used for the ensemble ens-HR (ens-LR) are indicated by su-
982 perscript "HR" ("LR"). Models marked by * are removed from the ensemble ens-hires.
983 Models marked by @ form ensemble ens-hires-a.

Model name	Nation	Atmospheric model	Ocean model
		Atmosphere resolution	Ocean resolution
CMCC-CM2-HR4$^{LR@}$	Italy	CAM4 ~1˚x1˚; 26 levels	NEMO3.6 1442 x 1051 longitude/latitude; 50 levels
CMCC-CM2-VHR4$^{HR@}$	Italy	CAM4 ~0.25˚x0.25˚; 26 levels	NEMO3.6 1442 x 1051 longitude/latitude; 50 levels
CNRM-CM6-1^{LR*}	France	Arpege 6.3 T127 (~100km); 91 levels	NEMO3.6 Primarily 1˚; 362 x 294 longitude/latitude; 75 levels
CNRM-CM6-1-HR^{HR*}	France	Arpege 6.3 T359 (~35km); 91 levels	NEMO3.6 Primarily 0.25˚; 1442 x 1050 longitude/latitude; 75 levels
EC-Earth3P-HR	Europe	IFS CY36R4 TL511 (1024 x 512 longitude/latitude); 91 levels	NEMO3.6 1442 x 1921 longitude/latitude; 75 levels
ECMWF-IFS-HR$^{HR@}$	UK	IFS CY43R1 Tco399 (1600 x 800 longitude/latitude); 91 levels	NEMO v3.4 1442 x 1021 longitude/latitude; 75 levels
ECMWF-IFS-LR$^{LR@}$	UK	IFS CY43R1 Tco199 (800 x 400 longitude/latitude); 91 levels	NEMO v3.4 362 x 292 longitude/latitude; 75 levels
HadGEM3-GC31-HM$^{HR@}$	UK	MetUM-HadGEM3-GA7.1 N512 (1024 x 768 longitude/latitude); 85 levels	NEMO-HadGEM3-GO6.0 1440 x 720 longitude/latitude; 75 levels
HadGEM3-GC31-LL^{LR*}	UK	MetUM-HadGEM3-GA7.1 N96 (1.875˚x1.25˚); 85 levels	NEMO-HadGEM3-GO6.0 Primarily 1˚ with meridional refinement down to 1/3 ˚ in the tropics; 75 levels
GFDL-CM4C192$^{@}$	USA	GFDL-AM4C192 C192 (720 x 360 longitude/latitude); 33 levels	GFDL-OM4p25 1440 x 1080 longitude/latitude; 75 levels
INM-CM5-H$^{@}$	Russia	INM-AM5-H 0.67˚x0.5˚ (540 x 360 longitude/latitude); 73 levels	INM-OM5-H 2160x1440 longitude/latitude; 40 levels
MPI-ESM1-2-HR^{LR*}	Germany	ECHAM6.3 T127 (384 x 192 longitude/latitude); 95 levels	MPIOM1.63 802 x 404 longitude/latitude; 40 levels
MPI-ESM1-2-XR$^{HR@}$	Germany	ECHAM6.3 T255 (768 x 384 longitude/latitude); 95 levels	MPIOM1.63 802 x 404 longitude/latitude; 40 levels
CESM-H	USA	CAM5 Approximately 0.25˚ (1152 x 768 longitude/latitude); 30 levels	POP2 Aproximately 0.1˚ (3600 x 2400 logitude/latitude); 62 levels

984

B. Figures

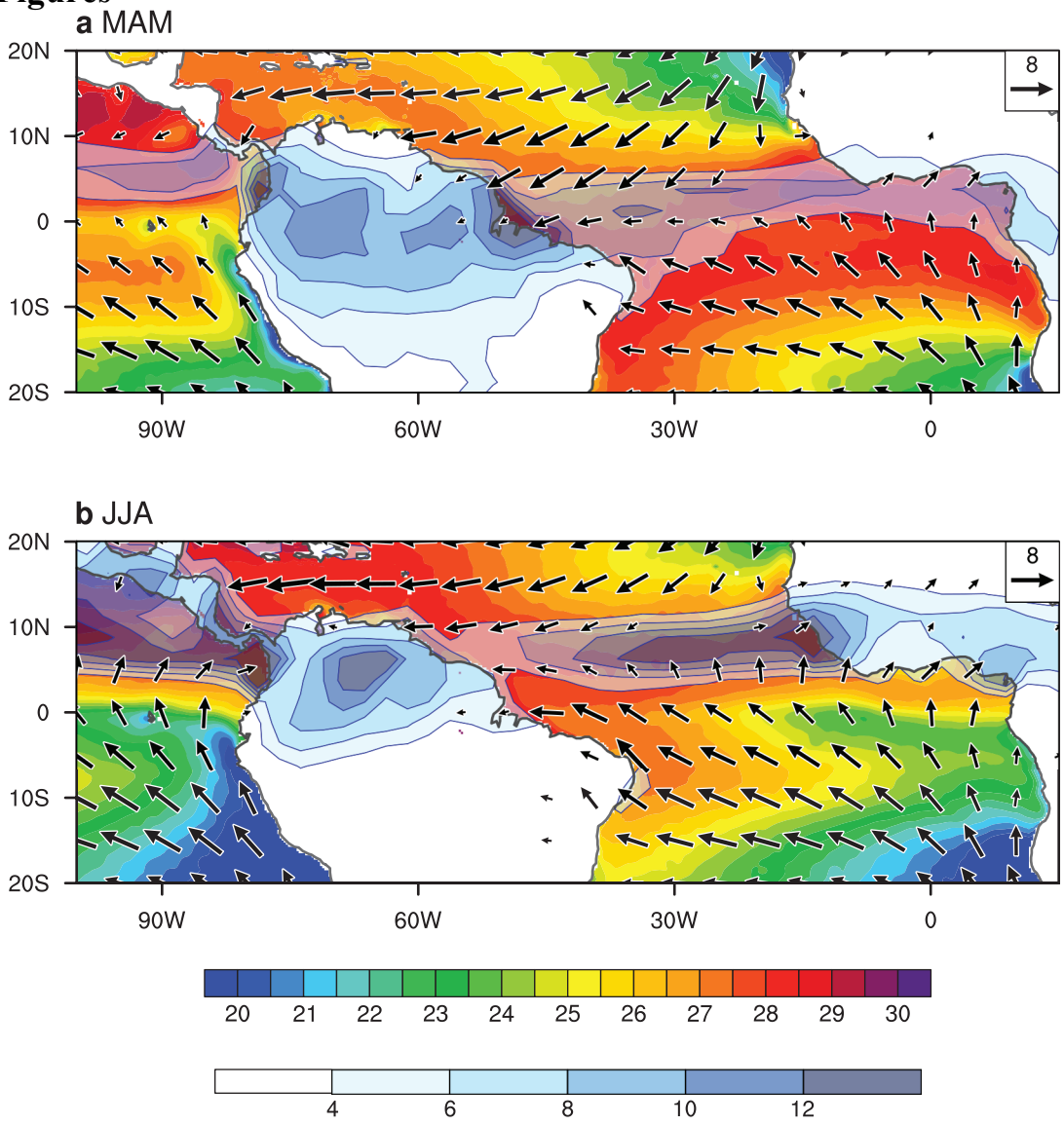


Figure 1. Climatological SST (color shading; ˚C), 10-m wind vectors (m s-1) and precipitation (blue shading; mm d-1) for 1979-2017. (a) MAM and (b) JJA seasons. SST and near-surface winds are from ERA5, precipitation from GPCP.

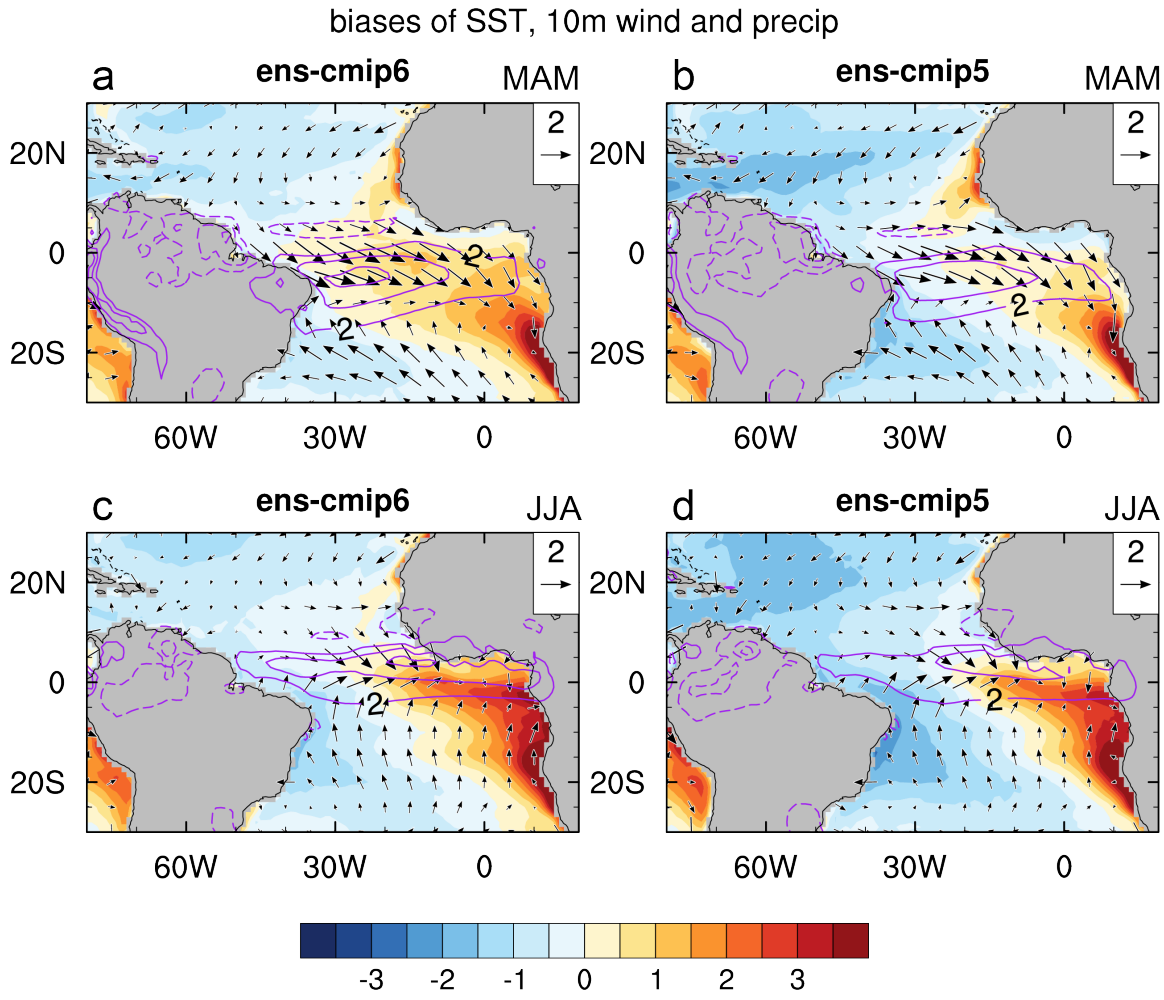


Figure 2. Biases of SST (shading; K), 10m wind (vectors; reference 2 m/s) and precipitation (contours; interval 2 mm/day; 0-contour omitted; negative contours dashed) for MAM (top row) and JJA (bottom row). The left and right columns show the ensemble average of piControl models from CMIP6 and CMIP5, respectively.

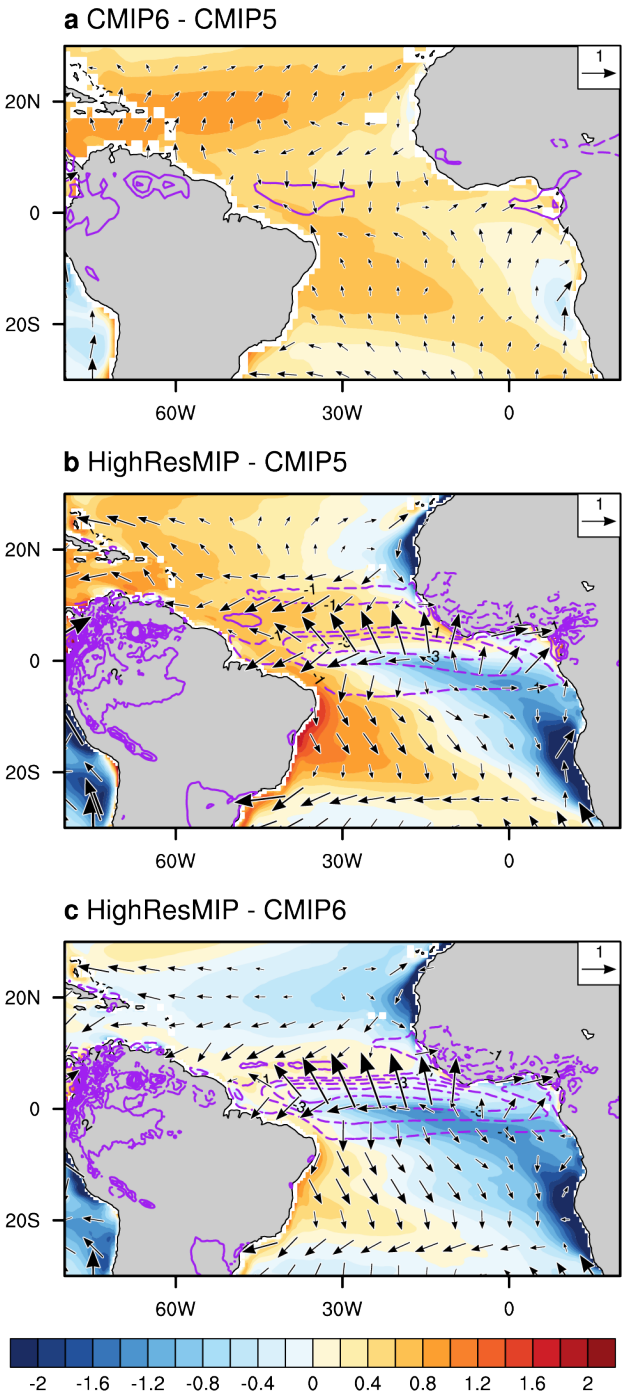


Figure 3. Difference between model ensembles in JJA for SST (shading; K), 10m wind (vectors; reference 1 m/s) and precipitation (contours; mm/day; 0-contour omitted; negative contours dashed). The ensemble differences shown are (a) CMIP6 minus CMIP5, (b) HighResMIP-MIP minus – CMIP5, and (c) HighResMIP minus CMIP6.

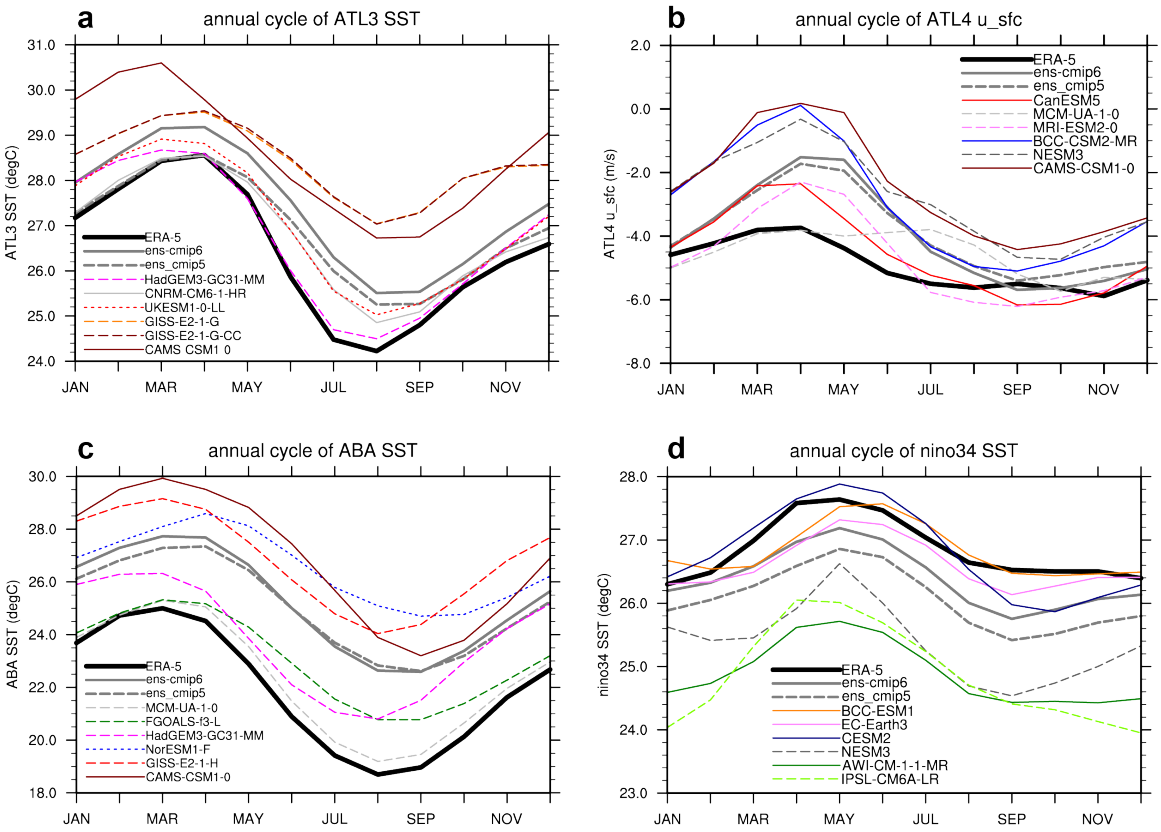
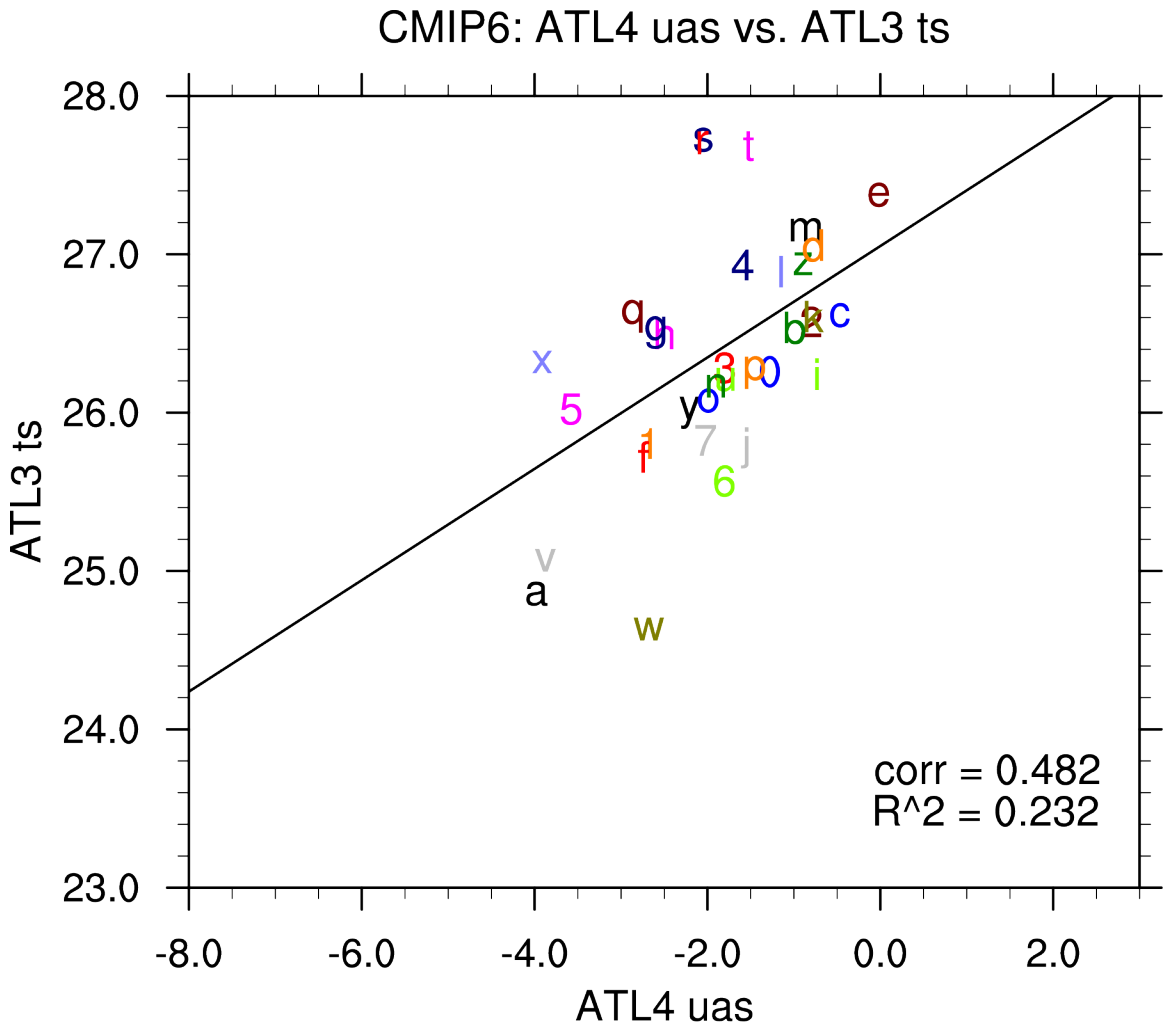


Figure 4. Climatological annual cycle of (a) ATL3 SST (°C), (b) ATL4 10m zonal wind, (c) ABA SST (°C), and (d) Niño 3.4 SST. The thick black line shows the ERA-5 reanalysis, the thick grey line the ens-cmip6 ensemble average, and the dashed thick grey line the ens-cmip5 ensemble average. Each panel also shows 6 individual models, which are chosen based on an RMSE criterion relative to the ERA-5 reference. In each panel, the top 3 models are the ones with the lowest RMSE (smallest error), and the bottom 3 are the ones with the highest RMSE (largest error).



a	era5	m	E3SM-1-1	y	MIROC6
b	AWI-CM-1-1-MR	n	EC-Earth3	z	MIROC-ES2L
c	BCC-CSM2-MR	o	EC-Earth3-Veg	0	MPI-ESM1-2-HR
d	BCC-ESM1	p	FGOALS-f3-L	1	MRI-ESM2-0
e	CAMS-CSM1-0	q	FGOALS-g3	2	NESM3
f	CanESM5	r	GISS-E2-1-G	3	NorCPM1
g	CESM2	s	GISS-E2-1-G-CC	4	NorESM1-F
h	CESM2-WACCM	t	GISS-E2-1-H	5	NorESM2-LM
i	CNRM-CM6-1	u	HadGEM3-GC31-LL	6	SAM0-UNICON
j	CNRM-CM6-1-HR	v	HadGEM3-GC31-MM	7	UKESM1-0-LL
k	CNRM-ESM2-1	w	IPSL-CM6A-LR		
l	E3SM-1-0	x	MCM-UA-1-0		

Figure 5. Multi-model scatter plot showing the climatological ATL4 10m zonal wind (m/s) in MAM on the x-axis and the ATL3 SST (°C) in JJA on the y-axis. The black line shows the regression line. The intermodel correlation is indicated in the lower right. Each letter corresponds to one model, as indicated in the legend at the bottom.

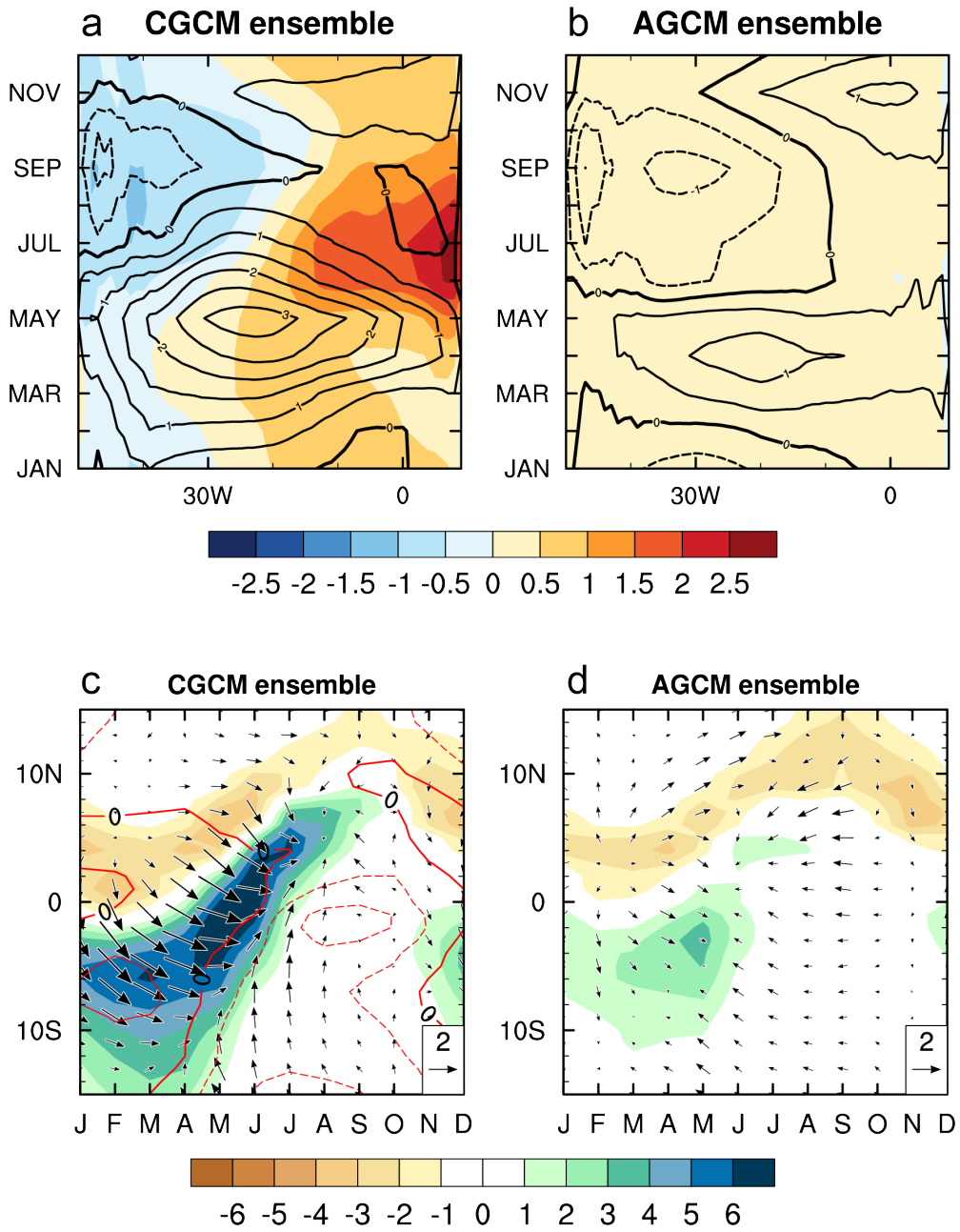


Figure 6. The top row shows longitude-time section of biases of SST (shading; K) and 10m zonal wind (contours; interval 0.5; negative contours dashed) for (a) an ensemble of coupled piControl models, and (b) an ensemble of the corresponding atmosphere-only models forced with observed SST (experiment amip). The bottom row shows latitude-time sections of precipitation biases (shading; mm/day), 10m wind (vectors; reference 2 m/s), and SST (contours; K; negative contours dashed), averaged from 40W to 20W for (c) the coupled model ensemble, and (d) the atmosphere-only ensemble.

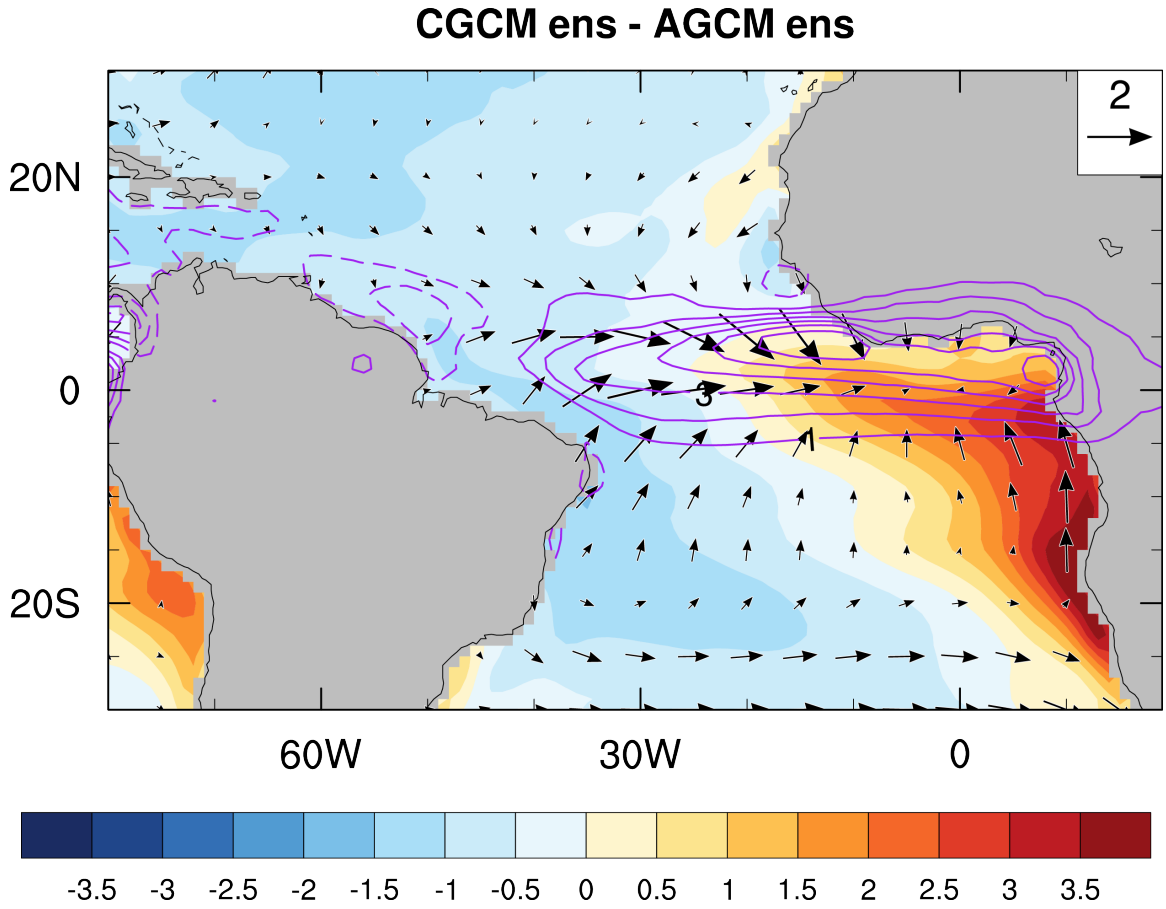


Figure 7. Difference between the piControl and amip ensembles in JJA for SST (shading; K), 10m wind (vectors; reference 2 m/s) and precipitation (contours; interval 1 mm/day; 0-contour omitted; negative contours dashed).

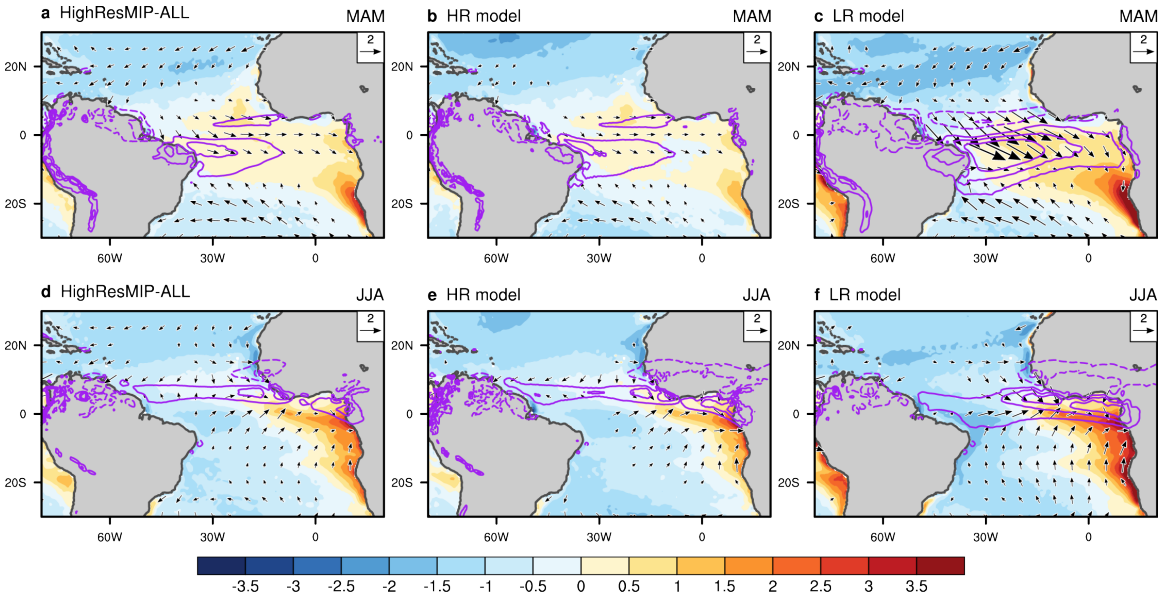


Figure 8. Biases of SST (shading; K), 10m wind (vectors; reference 2 m/s) and precipitation (contours; mm/day; 0-contour omitted; negative contours in blue) for MAM (top row) and JJA (bottom row). The left column shows the ensemble average over all models in the control-1950 experiment (high-resolution models), the center column the average over 5 selected models with particularly high resolution, and the right column the average over corresponding model versions with relatively low resolution.

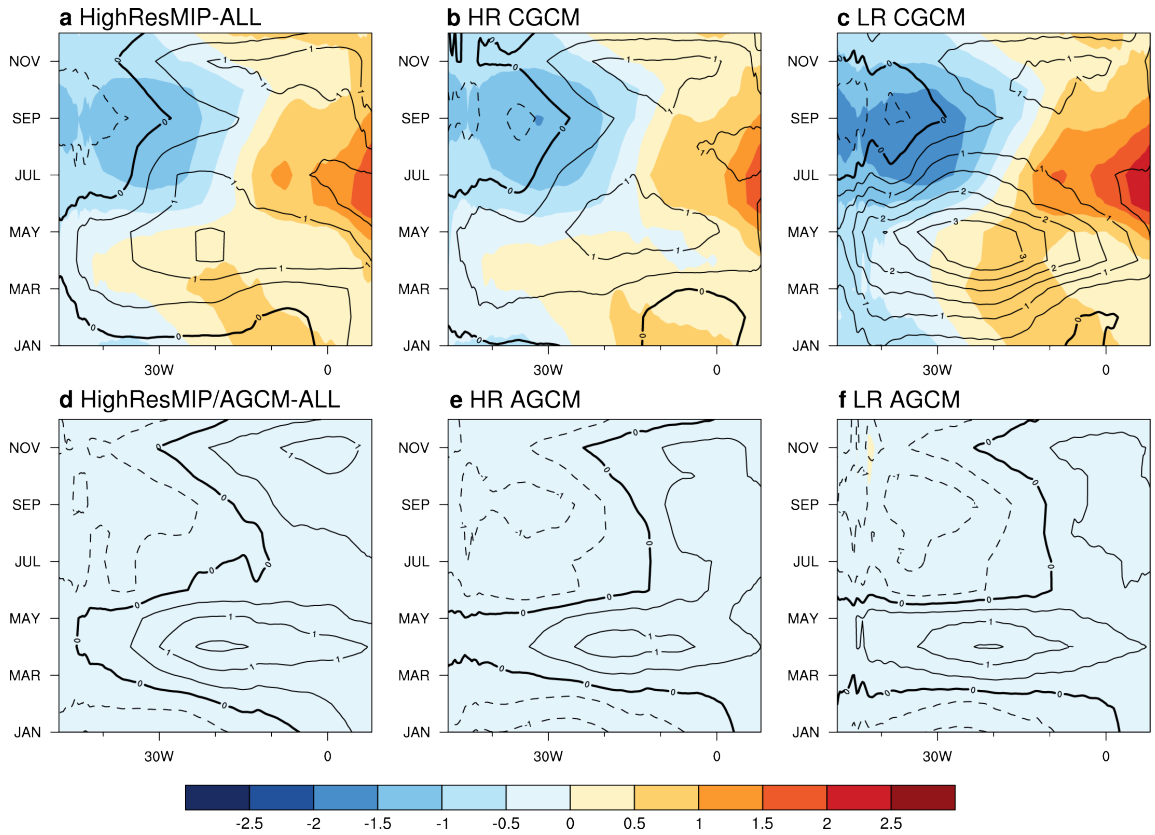


Figure 9. Longitude-time section of biases of SST (shading; K) and 10m zonal wind (contours; interval 0.5; negative contours dashed) for the ensemble average of (a) all control-1950 models, (b) 5 selected models with particularly high resolution, and (c) the 5 corresponding low-resolution versions of those models. The bottom row shows the same fields but for the corresponding AMIP simulations.

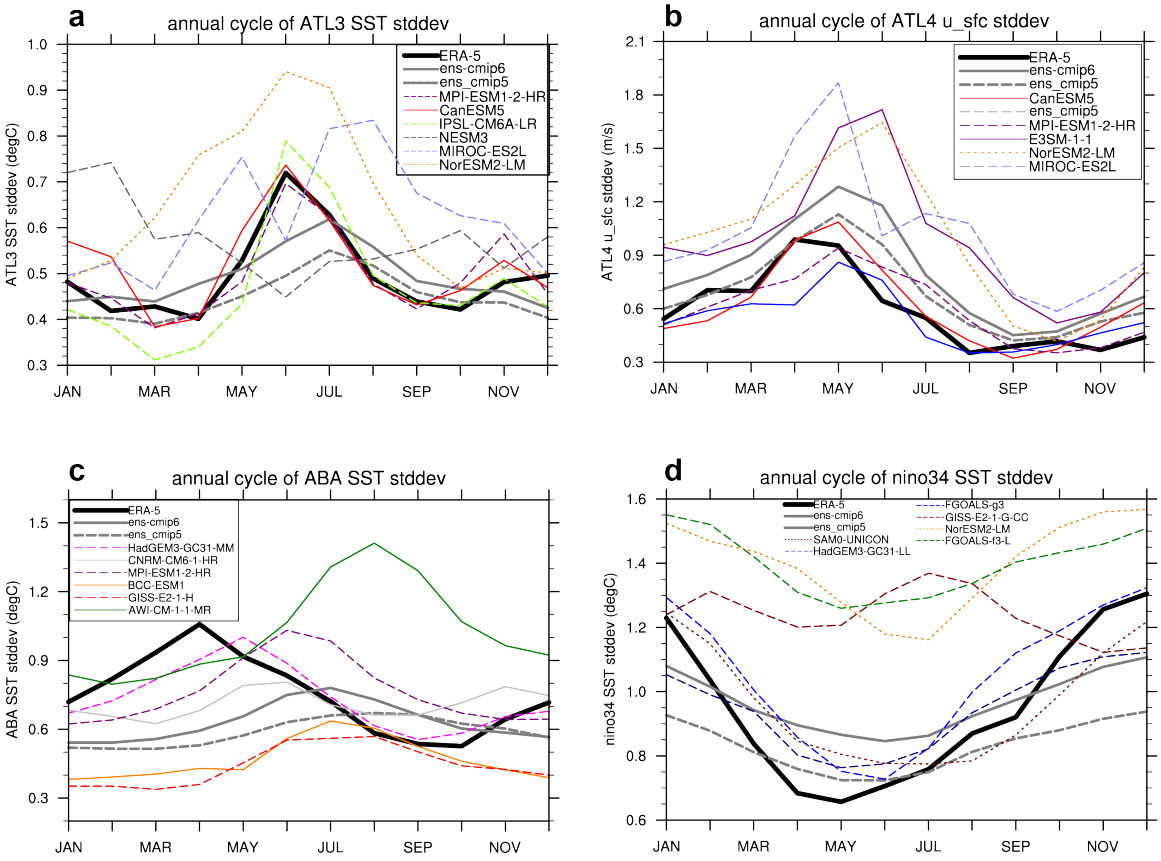


Figure 10. As in Fig. 4, but for the seasonally stratified variability of each index, calculated as the standard deviation of the detrended anomalies.

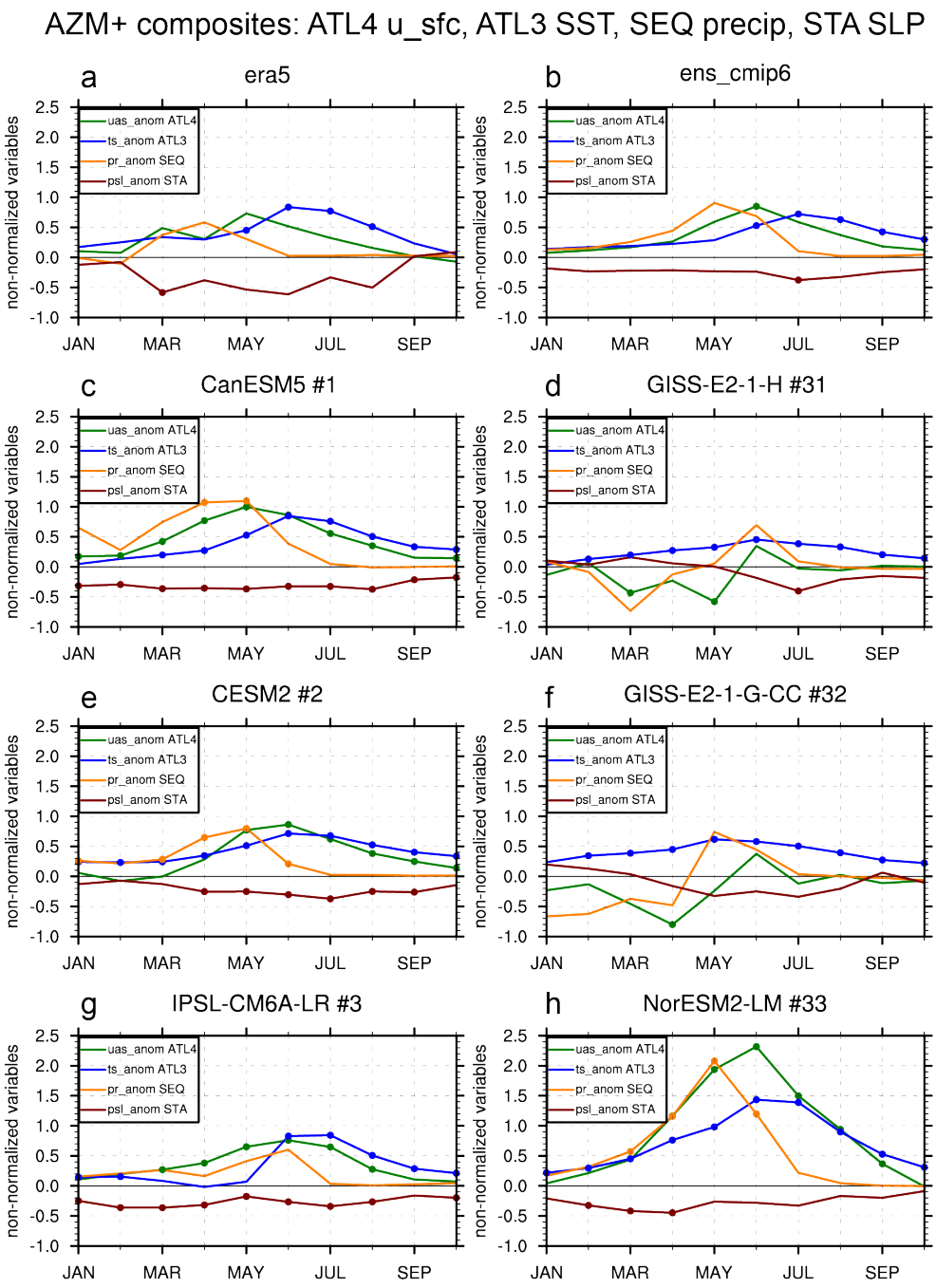


Figure 11. AZM composites showing anomalies of ATL4 10m zonal wind (green line; m/s), ATL3 SST (blue line; K), SEQ precipitation (orange line; mm/day), and STA SLP (dark red line; hPa). The panels show (a) ERA-5, (b) ens-cmip6, (c), (e), (g) the 3 models with the smallest error relative to the ERA-5 reference, and (d), (f), (h) the 3 models with the largest error relative to the ERA-5 reference. The error is calculated as a combination of the RMSE of ATL3 SST and ATL4 10m zonal wind. See text for details. Filled circles indicate anomalies that are statistically significant at the 95% level.

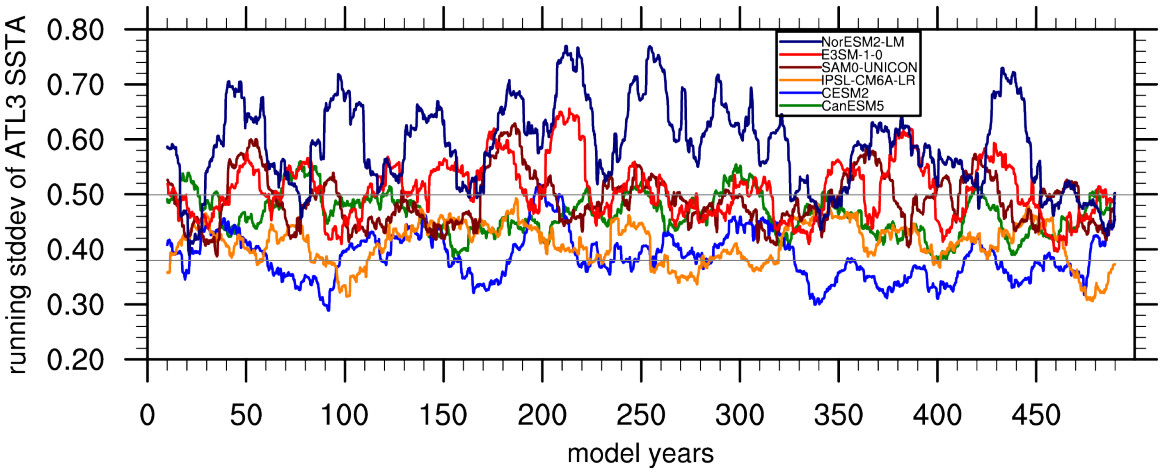


Figure 12. Running standard deviation of ATL3 SST anomalies (K) using a 20-year sliding window. Colored lines show 6 models from ens-cmip6 that were selected based on the same performance metric as in Fig. 11 but excluding those models with less than 500-years of simulation. The two horizonal lines show the standard deviation of ATL3 SST in the ERA-5 reanalysis for the periods 1979-1998 (top line) and 1999-2018 (bottom line).

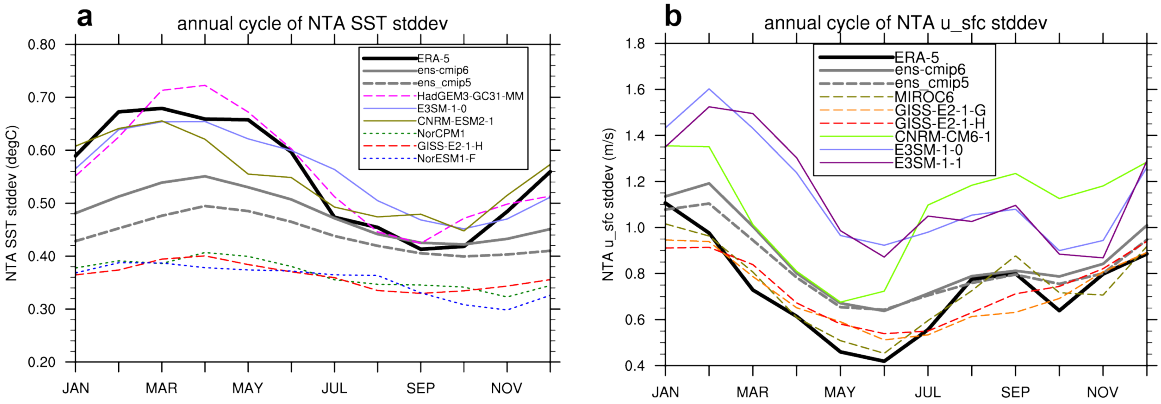


Figure 13. As in Fig. 10 but for the standard deviation of (a) NTA SST (K), and (b) NTA 10m zonal wind (m/s).

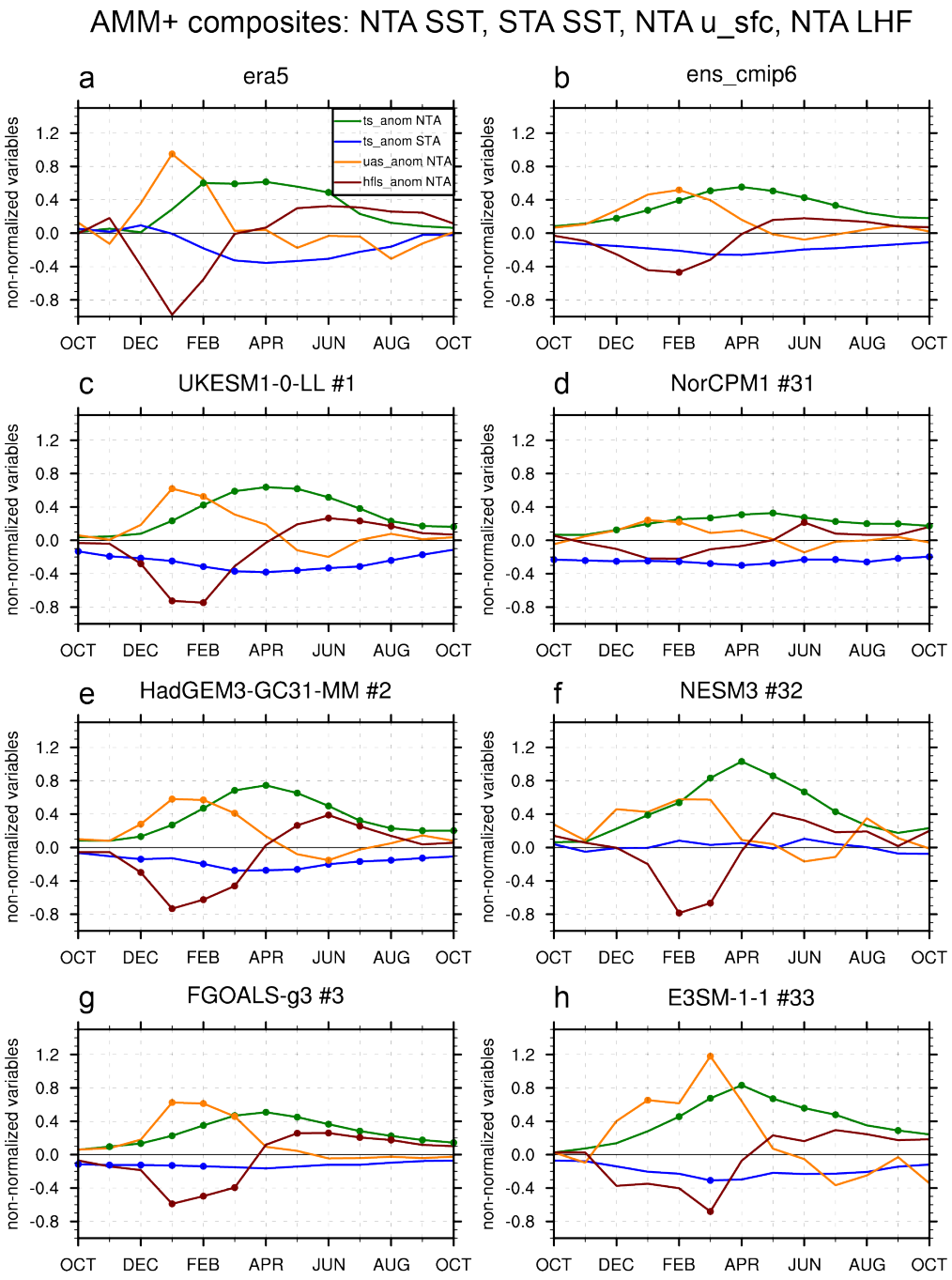


Figure 14. AMM composites showing anomalies of NTA SST (green line; K), STA SST (blue line; K), NTA 10m zonal wind (orange line; m/s), and NTA surface latent heat flux (dark red line; W/m2*0.05). The panels show (a) ERA-5, (b) ens-cmip6, (c)-(e) the 3 models with the smallest error relative to the ERA-5 reference, and (f)-(h) the 3 models with the largest error relative to the ERA-5 reference. The error is calculated as a combination of the RMSE of NTA SST and NTA 10m zonal wind. See text for details.

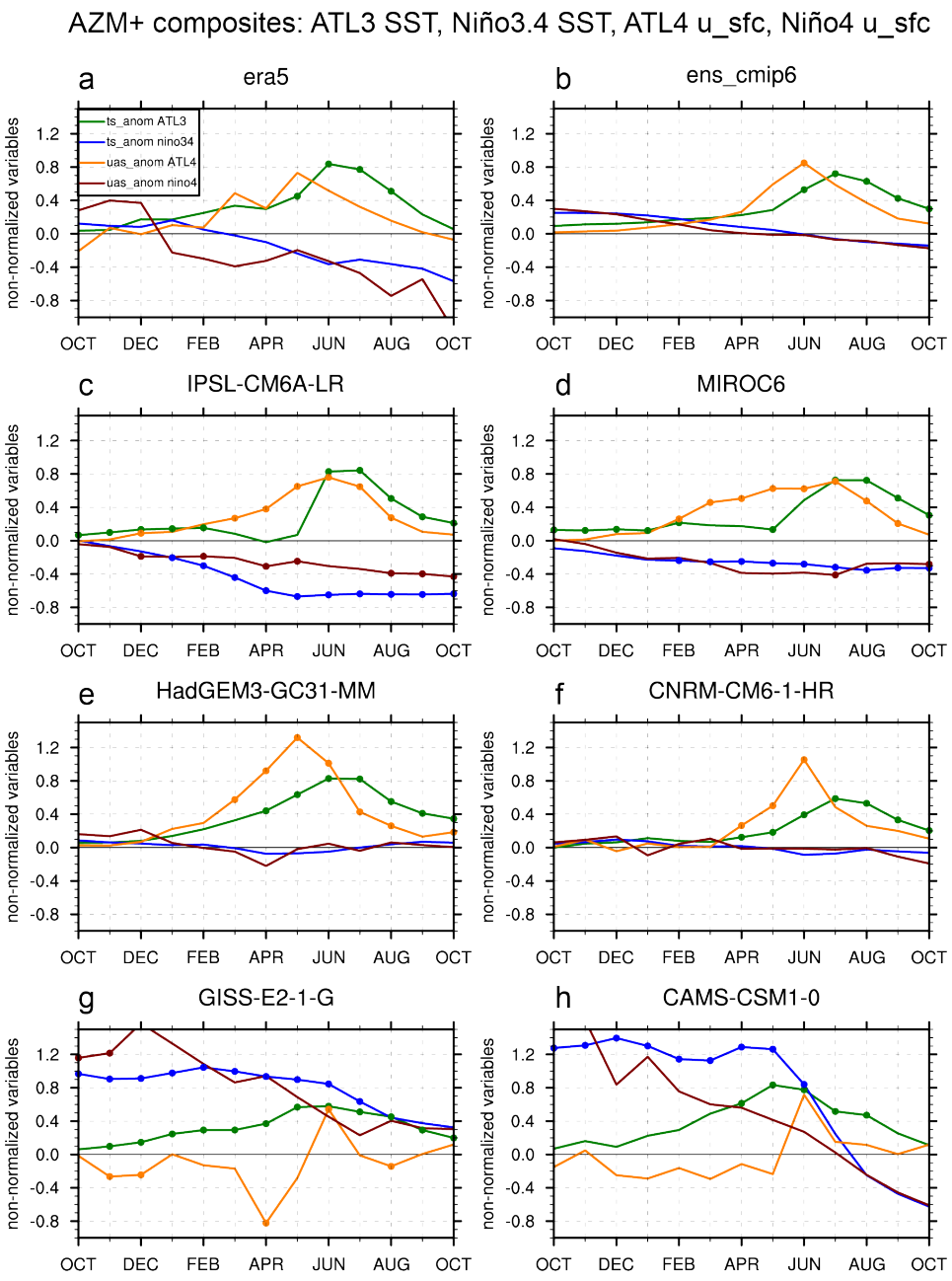


Figure 15. AZM composites showing anomalies of ATL3 SST (green line; K), Niño 3.4 SST (blue line; K), ATL4 10m zonal wind (orange line; m/s), and Niño 4 10m zonal wind (dark red line; m/s). The panels show (a) ERA-5, (b) ens-cmip6, (c)-(h) 6 models with distinct behavior in terms of the concomitant sign of Niño 3.4 SST anomalies: negative (2nd row), neutral (3rd row) and positive (4th row).

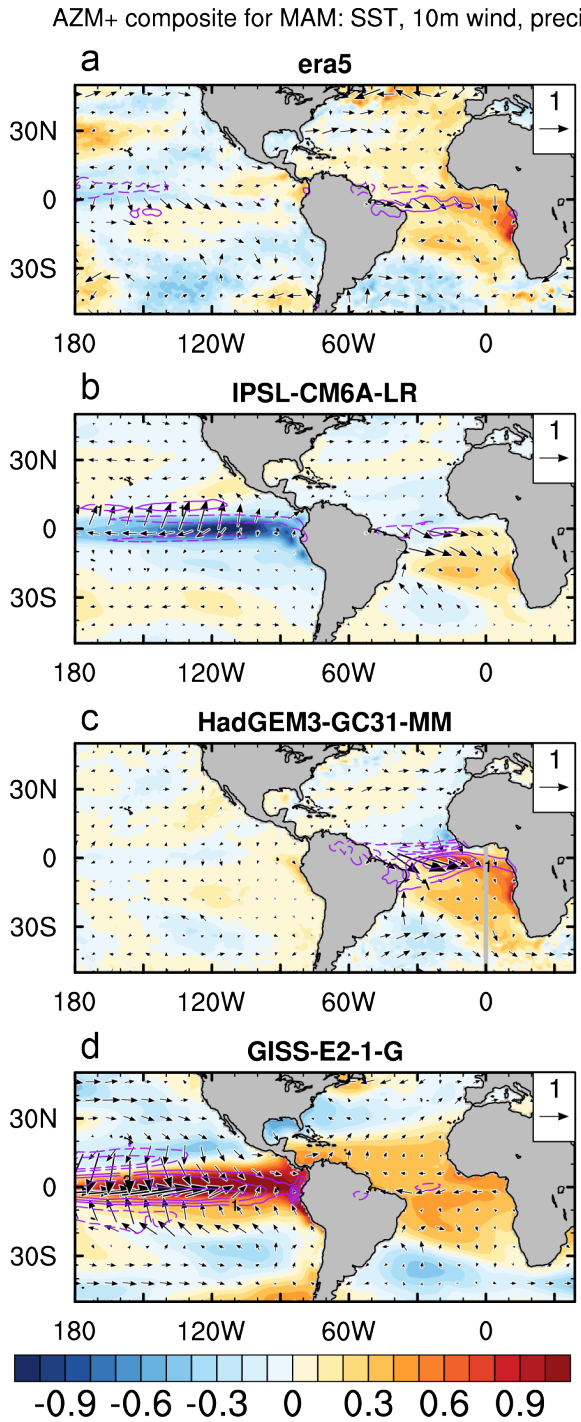


Figure 16. Horizontal maps of 4 of the composites shown in Fig. 15 (ERA-5, IPSL-CM6A-LR, HadGEM3-GC31-MM, and GISS-E2-1-G). The maps show the MAM average (one season before the peak of the AZM) for the following fields: SST (shading; K), 10 m wind (vectors; reference 1 m/s), and precipitation (contours; interval 1 mm/day; negative contours dashed).