Response to Referee #1

García-Franco, JL., Gray LJ, Osprey S.

Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford. Parks Road Oxford, United Kingdom OX1 3PU email: jorge.garcia-franco@physics.ox.ac.uk

---

We thank Reviewer 1 for the positive and constructive comments on the manuscript. We agree that this analysis is valuable for understanding the model performance of the UKMO models in a relatively less studied monsoon region. We now address each of your comments (in blue), highlighting the changes done to the manuscript as a result of your comment (in purple). In short, we have identified two major concerns that you presented and addressed them by including the AMIP simulation in all the analyses, to better understand the role of SSTs biases, as well as further analysing ENSO characteristics and the impact of ENSO diversity on the teleconnections to these monsoon systems by including a figure in the manuscript showing the different simulated and observed responses to the different types of ENSO events.

1. Specific comments

Line 8: The abstract describes that "[the model has] a stronger intraseasonal variation than observed". Note that intraseasonal variability (in the way that most readers will understand the term, i.e. the Madden-Julian Oscillation or Boreal Summer IntraseasonalOscillation) is not at all examined in this study. The authors are really describing aspects of the annual cycle (e.g. the mid-season drying). Thus, the wording here needs to be changed to avoid the language of intraseasonal variability. (See also later similar comment.)

The revised manuscript has changed the language to specify that these models have a stronger difference between the two peaks of precipitation and the mid-season dry period.

Line 9: While the Atlantic ITCZ is assessed, what of the SACZ? Is it relevant for sucha study of the South American monsoon system?

The SACZ is very relevant as a major driver of variability in precipitation and circulation in the SAMS. We have addressed your comment by adding a supplementary figure comparing the modelled and simulated SACZ spatial patterns and seasonal cycles, although the abstract makes no mention of this. This figure is now mentioned in the manuscript in section 3.2 as follows:

The South Atlantic Convergence Zone (SACZ) is a nortwest-southeast oriented band of convection and is a prominent influence on the South American Monsoon mean and extreme rainfall (Carvalho et al., 2004; Marengo et al., 2012). UKESM1 and GC3 appear to reasonably simulate the spatial pattern of active SACZ days and the seasonal cycle of SACZ activity (Figure S2).

Line 12: I think it is fair to say that ENSO characteristics (amplitude, frequency, longitu-dinal position, meridional spread, pattern, skewness...) are not at all assessed in this study. Thus, a more accurate form of words is needed here in order to avoid giving thereader such a misconception, e.g. revised wording should focus on the AMS responseto ENSO.

The wording has been changed to focus on the teleconnections and the response of the AMS to ENSO. However, the characteristics of ENSO in the models, such as those mentioned by the reviewer are in good agreement with observations. For example, Menary et al. (2018) showed that the power spectrum of ENSO agrees better with the observed HadSST than most CMIP5 and CMIP3 models. We discuss this further in a reply to another comment below.

Line 15/16: Instead of “between the two model configurations”, in the abstract the sentence should be worded to emphasize the scientific (rather than technical) meaning of this, namely that Earth System processes appear to make no difference to the monsoon simulation.

Done.

Line 16: At this stage the various resolutions involved have not been described so the use of the term “medium resolution” may confuse the reader, since it naturally implies there has been a comparison made with both lower and higher resolutions. Being familiar with the model framework used, I know that there is a higher resolutionversion of the HadGEM3 model although it is not studied here. The authors need torethink their terminology for this description. (See also later comments.) In addition,the resolutions used need to be explained in the abstract since they mean very different things to different readers (and their definition also changes with as this article ages).

We now use the terminology used in previous papers (e.g Menary et al., 2018) using the HadGEM3 model at N96 and N216 resolution, referring to these runs as low and medium-resolution. The term ”high-resolution” is now erased from the manuscript as to avoid possible confusion with higher resolution versions of the MOHC models.

Line 25/26: In an analogous fashion, what about the parts of South America north of the equator and their annual cycle? Can they be aligned to the NAMS?

While parts of southern Central America, e.g., Panama and northern South America, e.g., Colombia, may fit the definition of global monsoon, characterised by a strong seasonal contrast in precipitation, the regions experience a slightly different seasonal cycle, as they are transition zones between the North and South American monsoon with significant rainfall rates in the fall and spring season. Perhaps for this, or other historic reasons, these regions are excluded from North and South American Monsoon literature (Adams and Comrie, 1997, Arritt et al., 2000, Vera et al., 2006, Jones and Carvalho, 2013, Geil et al., 2013), and they are rarely discussed in this manuscript. However, the manuscript does clarify this by stating that the manuscript uses the standard definitions for the North and South American Monsoon, and additionally the MSD region. The manuscript now reades:

This study uses the definitions for the North and South American Monsoons as previous studies (Vera et al., 2006; Marengo et al., 2012), with additional analysis on the MSD of southern Mexico and Central America (Magaña et al., 1999; Perdigón-Morales et al., 2018).

Line 63: How have CMIP5 models misrepresented the magnitude of the seasonalcycle? Are they systematic under- or over-estimations, or a mixture depending on themodel?

The manuscript now clarifies this point. CMIP5 models showed an earlier, but clearly observed, onset date, but the retreat date was unclear from precipitation time-series, due to problems in simulating the large and regional-scale processes associated with the retreat, i.e., the displacement of the subtropical jet. We have thus changed this paragraph as follows:

The CMIP5 simulations of the North American Monsoon misrepresented aspects of the seasonal cycle of precipitation and overestimated the peak monsoon rainfall (Geil et al., 2013; Sheffield et al., 2013a). Most CMIP5 models simulated an earlier onset date, but improved from CMIP3 since the onset date showed a clear separation of rainy and dry seasons in daily precipitation time-series. In contrast, the simulated retreat date was unclear in most models which highlighted problems for these models to simulate the regional changes during retreat stage (Geil et al., 2013; Sheffield et al., 2013a).

Line 67: Be specific as to what the CMIP5 models have improved upon. Presumably itis the CMIP3 models. A more specific and detailed description of biases in the CMIP5 cohort for the 3 regions is given.

Table 1: Which version of the TRMM algorithm is used? V7 for example is knownto perform better over orography such as the Andes (see Zulkafli et al., 2014,https://doi.org/10.1175/JHM-D-13-094.1).

We use the product from satellite 3B42, algorithm V7, the table now clarifies the version used.

Line 96/97: What is the evidence that TRMM provides the most reliable source of information on rainfall for this region? Are there citable studies intercomparing satellite,gauge and merged datasets for the NAMS and/or SAMS?

There are scarce studies that compare precipitation datasets in these regions and in-situ ground based stations managed by local governments have a short record or calibration problems, therefore there is hardly evidence to support the original statement of the manuscript. As the original sentence then lacked support from the evidence, the new manuscript now simply argues that TRMM is used as it has been considered by literature (cited in the manuscript) as a reliable source of information about the spatial and temporal characteristics of precipitation and therefore typically used in GCM evaluation in this region. Furthermore, we do point the reader to relevant comparison studies over several regions of the AMS as follows:

The TRMM dataset has a high horizontal and temporal resolution and was used in several CMIP assessments (Geil et al., 2013; Jones and Carvalho, 2013) as a reliable source of precipitation (Carvalho et al., 2012). Therefore, we use TRMM as our best estimate for the spatial and temporal characteristics of the AMS rainfall. However, the period covered by TRMM (1998-2018) is too short to analyse statistically robust teleconnections or variability, so we use GPCP, GPCC and CHIRPS for their longer period. Although a thorough validation and comparison of these datasets across the AMS domain is missing, several studies have analysed one or more of these datasets in regions of the AMS (e.g. Franchito et al., 2009; Dinku et al., 2010; Trejo et al., 2016).

Lines 100-124: Note that the ocean model horizontal resolutions have not been listed.

The corresponding ocean model horizontal resolutions are now listed in these lines.

Line 129: Clarify whether this is surface temperature or surface-air (i.e. 1.5 or 2 metre)temperature that is being considered.

We are using 2-m near surface air temperature. Table 1 and this section now makes clarifies this point.

Line 132: The Welch t-test should be defined in the methods or referenced here. How does this differ from a student's t-test? Plus, how are the different ensemble members dealt with relative to this?

A reference and explanation about the use of the Welch t-test is now given in this line. The p-value is estimated from each ensemble member and these p-values are then combined into a single p-value using Fisher's method (Fisher, 1992). We have thus included the following the lines suggested by the reviewer:

"Only statistically significant differences are shown, according to a Welch t-test (Wilks, 2011), which accounts for the difference in sample size and variance between model and observations/reanalysis data. The significance for simulations with multiple ensemble members is estimated first for each ensemble member and then combined into a single probability or p-value using Fisher's method (Fisher, 1992)."

Line 137/138: Does the stronger Bolivian LLJ support a stronger seasonal cycle/monsoon in the region north of the equator in South America during boreal summer? (I.e. in the South America component of the NAMS.)

Although the regions north of the equator do experience a monsoonal climate as defined by Wang et al. (2017) they are not comprised in North American Monsoon Literature. The stronger Bolivian LLJ is suggestive of stronger moisture transport to the south-eastern part of the South American Monsoon and would also suggest more rainfall than observed in this region while also drying the Amazon. The manuscript now reads:

"The South America Low-Level Jet, the low-level northwesterly flow in Bolivia, observed in Figure 1a, is stronger in the simulations. This stronger than observed jet is suggestive of a stronger moisture transport to the La Plata Basin, with has been associated with a drying of the Amazon and positive precipitation anomalies at the exit region of the jet (Marengo et al., 2012; Jones and Carvalho, 2018)."

Line 147/148: The physical outcome of this needs to be made explicitly clear to the reader, namely it appears that the inclusion of Earth System processes makes no difference to the SAMS.C3

Done. The manuscript now reads in these lines:

"The inclusion of Earth System processes appears to make no improvement on the low-level circulation biases."

Line 149/150: A better summary of the changes in historical forcing (compared to the pre-industrial) needs to be described in lines 119-124 in order for the reader to be able to understand possible changes. Clearly the reader will know that global GHG emissions have increase, but what are the relevant/local patterns of aerosol emissions,land-use change etc. between the two experiments?

A more detailed description of the differences between the historical and piControl experiments is given in section 2.2:

In contrast to the pre-industrial control experiments, the historical experiments use time-varying aerosol and greenhouse gas emissions and land-use change (Eyring et al., 2016). In Latin-America, land-use change for agricultural purposes have dramatically decreased tree cover in Central America and south-eastern Brazil since the 1950s (Lawrence et al., 2012), thereby affecting the surface energy balance. The regional emissions of carbonaceous aerosols, nitrogen oxides and volatile organic compound in Latin American megacities are also considered in the historical experiments. These emissions are noteworthy, e.g., due to the impact of black carbon emissions by increased biomass burning in the Amazon and northern Central America (Chuvieco et al., 2008).

Line 150-152: Given the length of the pre-industrial control integrations that are available (and given the small size of the forcing when compared to the historical experiment), the internal variability of quantities such as those listed here (and elsewhere through the results) within the pre-industrial should be considered as a means to understanding the significance of any change.

Both in the case of Figure 2e,f and Figure 7,8 h, the differences between historical and piControl have been shown only where the historical experiments shows a statistically significant difference from the piControl variability, as defined by a Welch t-test between the two experiments.

Line 175: I understand the logic, but the chosen model comparison mixing UKESM with the GC3 model appears rather unclean.

The choice of model comparison was based on the fact that the four low resolution simulations (GC3 N96-pi, GC3-hist, UKESM1-pi and UKESM-hist) were virtually indistinguishable for precipitation, ITCZ and low-level circulation biases. Comparing the low resolution coupled model with the medium resolution coupled or the AMIP simulation provides a better. For brevity, we now compare one coupled low resolution simulation with the coupled medium-resolution and the AMIP simulation. The manuscript now clarifies that this choice of simulations shown was based on showing the main biases and the differences.

Line 193: In what way is the low-level wind structure biased?

The manuscript now describes the wind biases:

The modelled low-level wind in the coupled model structure shows significant biases near the ITCZ. These wind biases are observed as stronger wind vectors converging toward the ITCZ during boreal summer and spring and stronger wind vectors diverging away from the equator during boreal winter.

Lines 171-222 and onwards: All of the comparisons whether maps or seasonal cycles would benefit from a table of quantitative comparisons between the various datasets, such as pattern correlations (or just correlations for the seasonal cycles) and RMSE.This is standard practice in multi-model evaluation studies.

Figures 1, 2, 7, 8 and 9 now show correlation and pattern correlation coefficients and RMSE.

Line 239: That the AMIP models “removed the spatial patterns” is strange wording. Did any bias remain at all? Generally, I think that this study could be significantly strengthened if a fuller comparison could be made between AMIP runs of these two models (which will be available as contributions to the CMIP6 DECK) could be thor-ughly compared with the coupled historical runs. The absence of SST bias would make for improved understanding.

The manuscript now includes results from the HadGEM3 GC3.1 AMIP simulation run at low resolution N96. The discussion is now updated to highlight the biases that are removed when the SST biases are removed, mainly the dry Amazon bias. However, biases in Central American rainfall and in the North American Monsoon are not improved even with "the right" SSTs.

Line 253: Here the run is referred to as “high-resolution” yet in the abstract it was medium resolution. The consistency within the manuscript needs to be improved.Could the manuscript not also examine a higher resolution version of the GC3 experiment, e.g. at N512?

The manuscript is now consistent with the wording of Williams et al. (2018), Menary et al. (2018) as to refer to the simulations as low (N96) and medium (N216) resolutions. Indeed the higher resolution simulation could be examined and compared, although long runs, pre-industrial or historical, at that resolution are not available and can therefore not be directly compared to the experiments used in this study or with observations.

Line 262: Are there any published onset measures for the AMS that could be used tomeasure this? And how is the onset objectively defined from Figure 9b? E.g. 1mm/dthreshold, or the maximum rate etc.?

There are some studies that analyse onset and retreat timings in precipitation time series or using other metrics such as OLR in the AMS. However, most of the methods are not suited to address model output as the required fields are not all provided by the modelling centres or are tailored to be used in one specific dataset and using these methods in model output requires further statistical treatment. Ongoing research by the authors aims to cover this shortcoming in the literature by presenting a robust method that can use precipitation time series from different datasets (observations, model, reanalysis) and in different monsoon regions. To address the specific comment of the reviewer, the statement in the manuscript now clarifies that onset timing is merely qualitatively well represented by the models.

Lines 256-287: In the tropics, and especially for monsoons, I would expect the sea-sonal cycle of precipitation to be discussed in the context of the lower tropospheric circulation. This doesn’t necessarily need to be done in the same paragraphs (the lay-out here is fine), but at the very least I would expect the discussions of precipitation biases here to reference the circulation biases for consistency. This is because of the intimate connection between circulation and precipitation in the tropics: winds providing moisture to the monsoon and the monsoon heating feeding back on the circulationto bring more moisture. At present the discussion is kept very separate. This could be aided by adding wind vectors to Figures 7 and 8.

The wind vectors have now been added to these Figures. Furthermore, the description of these figures and the discussions now couples the circulation and the precipitation to highlight, for instance, the relationship between the moisture transport away from the Amazon into southeastern Brazil and a corresponding dry Amazon bias and a wet southeastern Brazil bias in these models.

Lines 256-287: It would be preferable to have some contextual comparison with other contemporary models (or at least CMIP5). How did CMIP5 perform for the NAMS and SAMS (cite references)? Do the UKMO models here fit within that envelope or arethey better/worse? This will help improve the level of interest in this study outside thesingle modelling group. Furthermore, can the authors state how the current UKMOmodel versions (especially GC3.1) have advanced upon earlier versions (HadGEM3,HadGEM2-ES, even HadCM3) with respect to the AMS? Are there any published works mentioning those models? It would be useful for the community to understand if the simulation is being improved or whether significant biases are persisting.

We now provide context to assess whether these models have improved, which biases have been removed and which have persisted, with references, in the discussion section and for each monsoon region. For example for the North American Monsoon:

These results suggest model improvement on the simulation of the North American Monsoon sfrom previous versions of the MOHC models (Arritt et al., 2000), and most of the model cohorts of CMIP3 and CMIP5 (Geil et al., 2013). For example, most of CMIP5 models showed a very wet bias during monsoon maturity whereas rainfall during monsoon maturity in all the experiments of this study within 1 mm day-1 of observations. However, these models continue to show biases during monsoon retreat as rainfall does not decrease as sharply as in observations after mid-September.

Line 288: In the deep tropics, OLR is not really going to tell us much more than wealready learn from precipitation, since much of the convection is deep. What is the nature of convection in the regions discussed? If any particular regions are dominated by shallow convection/warm rain, then this could be highlighted by references to relevant published works.

While we generally agree with this reviewer's comment that in the deep tropics during the wet season of each monsoon, OLR is highly correlated with precipitation and virtually indistinguishable, Fig. 10 does show interesting differences in OLR and ω between model and reanalysis that do not agree with the precipitation. Particularly in the MSD region, the first peak, MSD and second peak characteristics in precipitation do not agree with OLR. The OLR would suggest a relatively similar first peak magnitude in the simulations and a weaker second peak than observed, however, the simulated 'precipitation shows a significantly wetter (Fig. 9) first peak and very similar second peak compared to both TRMM and ERA5 precipitation. The analysis of OLR, q and ω may point to model biases in the treatment of convection and potential feedbacks. The height of convection influences the radiative balance, whereas characterising the strength of ascending and mid-level moisture aids to evaluate several aspects of the model's convection and microphysical schemes.

Line 297: How certain can we be about the tropospheric moisture in any case in areanalysis? What level of data is assimilated in some of these remote regions? Can any ground-truthing (really air-truthing!) be performed (even if not shown) using nearbyRS launches such as those publicly available from Wyoming?

According to the Wyoming website and the NOAA station archive, regular soundings have been made in Manaus and Leticia (in the Amazon region), at Empalme, Sonora, México (core North American Monsoon), at Guatemala City and San Cristobal de las Casas, Mexico (in the MSD region) and at Sao Paulo and Brazilia (southeastern Brazil) at least since 1979. These radiosonde observations are assimilated twice a day into ERA-5. Although scarce and not as widespread as in other regions these are valuable input into ERA-5 and while not ground truthing over the whole domain of each monsoon region, these are the best estimates of tropospheric moisture available and therefore, arguably, makes ERA-5, and other reanalyses, a good standard to compare against. Analysing reanalysed and modelled tropospheric moisture where no RS launches are assimilated into the reanalysis would in fact be subject mostly to the reanalysis model driving the variables, for example over the ocean. We hope that showing that there have been RS launches assimilated into ERA5 in all the regions analysed of this study would answer this reviewer's concern. A more thorough comparison between ERA-5 and RS would be most beneficial but outside the scope of our study.

Line 328: Unlike the implication in the abstract, there is no assessment made hereof general ENSO behaviour in these coupled models – and if the driving point of ateleconnection is faulty then resultant impacts over the AMS will hardly do well. Asummary of the behaviour of ENSO in these models with reference to a publishedassessment of their performance should be made.

This is an interesting and recurrent point by this reviewer. Menary et al. (2018) showed that the EN3.4 index has a similar power spectrum in the pre-industrial control experiments (see Figure 1 of this document) when compared to observations. The models also show a good representation of the perturbation to the Walker circulation by ENSO events (see Figure 2 of this document). The main patterns of variability (Figure 3) are also reasonably reproduced, particularly by the medium-resolution simulation. Of particular importance to the study at hand is ENSO diversity and the impact each different ENSO event has on the rainfall of the American Monsoon System. The characteristics of ENSO in these models are now summarized in section 5 in the manuscript and several of these figures have been added to a Supplementary information document that accompanies the manuscript.

Line 332: Is this in units of temperature (degC/K) or a normalised index in terms ofstandard deviations? Where is the index taken from or how have you calculated it?

The index has units of K, and was estimated from the HadSST dataset in the El Niño 3.4 region, as described in line 333 of the original manuscript. The figure now shows the units of the index.

Line 334: What are the years included in the observed composites of El Nino and LaNina? Has the impact of CP and EP El Nino events been considered and what doesthe published literature say about the different impacts of such events on the NAMSand SAMS?

Cai et al. (2020) and references therein show that ENSO teleconnections to SAMS depend on the type of ENSO, as shown by their indices (see another response and figures 3-6 below). For instance, Figure 4 of this document shows that GC3 N216 has ENSO events well represented in all the quadrants of the PC space. We now included a new figure (Figure 6 of this document and new Figure 13 of the manuscript) comparing observed and modelled responses to CP and EP ENSO events. This analysis would of course be improved by analysing ENSO diversity in future projections and providing a more thorough analysis of ENSO diversity, for instance, measuring the skewness in the PC space, but we considered this to be outside the scope of this study.

Line 351: It would be very instructive if wind vectors were added to Figure 12, enabling to reader to understand something of the mechanism by which ENSO controls rainfall anomalies in the AMS. The authors should then elaborate upon this in the text.

We agree with this reviewer's comment, we added the wind vectors to this Figure, but found that this would overcrowd the figure and decrease clarity. It is important to note that the mechanisms for the teleconnection to the subtropics are different from the teleconnection to the equatorial Amazon and therefore the most relevant wind anomalies for each teleconnection take place at different levels of the atmosphere. Figure 2 of this reply shows the Walker circulation anomalies in circulation and moisture. The Amazon region anomalies are closely related to this perturbation to the overturning circulation whereas the subtropical regions are affected by the perturbation of ENSO to Rossby wave-trains and the subtropical jets. We have added this figure to the supplementary material.

Line 348-350: It's not immediately obvious how the NAO links described are relevant to the study at hand. The authors should either make this clear or remove this text.

The boreal winter-time NAO has been shown to influence precipitation in Central American and the Caribbean (Giannini et al., 2000), as well as northern South America (Giannini et al., 2004). Therefore, capturing the response in the North Atlantic SLP field may be important to capture secondary aspects of ENSO teleconnections. The link is now explained in the manuscript as:

"While the models seem to be able to capture this response of the NAO, the simulated response is weaker than observed, which may be relevant to simulate a secondary effect of the NAO on Central American and northern South American rainfall (Giannini et al., 2000, 2004)."

Lines 365-370: The authors should consider whether the lack of nonlinearity in the modelled ENSO response reflects the lack of diversity of simulated ENSO in the model(e.g. the lack of distinct central Pacific or east Pacific events).

Figures 5 and 4 of this document show the spread of boreal winter SST patterns as measured by the principal component analysis, as shown by (Cai et al., 2020). The variability of these models appears to cover a range of ENSO diversity, except perhaps missing extreme events such as the 97-98 and 2015-2016. The patterns associated with Central and East Pacific positive ENSO events (Figure 5) agree with the patterns of HadSST. These figures have now been added to the Supplementary material and the main results are introduced in section 5 of the new manuscript, to validate the fact that ENSO diversity seems, to a first degree, well represented in these models. A more thorough analysis would better evaluate how different are EP and CP in the observations and in the simulations, using metrics such as skewness or perhaps the degree of coupling of the SSTs to the Walker circulation.

In any case, the teleconnections of the different types of ENSO events to South America (shown in Figure 6) appear to be independent of the ENSO type in the simulations. This may be because the model diversity is not representative of the observed ENSO diversity in other metrics, or perhaps because the simulated SST patterns do not couple to the atmospheric circulation in the same way, but this warrants further analysis, outside of the scope of this study.

Line 376: The authors could be more explicit on the likely kink between cloud cover and the warm bias in the SAMS domain. If precipitation is too weak, this should be stated explicitly. (Note there would also be a soil-moisture feedback as a result.)

The manuscript now makes the link between precipitation, cloud cover and temperature explicit:

In the Amazon, the simulations showed a warm bias (+2 K) during austral spring and summer, a typical feature of previous models (Jones and Carvalho, 2013), and a colder than observed southeastern Brazil. These biases were linked with decreased cloud cover and less rainfall over the Amazon and more high clouds and rainfall in southeastern Brazil (Figures 7 and 9). The low cloud cover, warm and dry Amazon biases are intertwined with the low-level circulation from the Atlantic into the South American continent. The biases in the circulation during austral summer were observed as a northerly flow anomaly over the central and southern Amazon, a feature that has been associated with a stronger moisture transport away from the Amazon (Marengo et al., 2012; Jones and Carvalho, 2018). During the period of maximum mean rainfall rates in February, the simulations can overestimate rainfall by 3 mm day -1 in southeastern Brazil and underestimate rainfall in the Amazon by a similar rate.

Lines 376-380: Finish the sentence by making explicit how the land-sea temperaturecontrast may feedback on the monsoon.

Addressed in the previous comment.

Line 391: Make explicit whether the Ryu and Hayhoe study was using CMIP5 models.

Yes, the study used CMIP3 and CMIP5 models, the manuscript now makes this explicit.

Line 393: With reference to the earlier comment on the abstract, the authors shouldavoid the terminology of intraseasonal variability here since the MJO/BSISO have not been assessed. Done.

Lines 371-404: In the conclusions I would want to see a more thorough synthesis ofthe results (e.g. how all the meteorological components fit together) than a summaryof each in turn. It would also be worth reflecting upon (if possible) how these modelssit in comparison to published literature on the AMS in CMIP3/5 models or on earlier versions of UKMO models.

The discussion section has been changed significantly to address this comment. The new manuscript now discusses each region of the AMS separately; for each region a summary of the biases in circulation, temperature and precipitation is given, indicating where they might be linked and finally whether these CMIP6 versions are an improvement from previous versions of the UKMO and CMIP5 models.

Line 413: See earlier comments on higher/medium resolution. Done.

Line 418: Need to see a summary of how the Earth System processes influence the response to forcing.

While we do not provide a thorough summary of the Earth System processes, as we did not investigate them explicitly and may be outside the scope of our study, the manuscript does state:

"A relevant difference between UKESM1 and GC3 is that warming over the historical period in Mexico and the Amazon is higher in UKESM1 than in GC3. This warming may be a consequence of the land-use change in these regions playing a role in the UKESM1 representation of soil-atmosphere feedbacks."

Figure 1: The domains used later in Figure 3 etc. need to be pictured somewhere, e.g.on this figure.

The domains are now shown in Figures 1, 7 and 8.

Trivia:

Lines 13/14: Perhaps replace "in subtropical America" [meaning USA?] with "inthe subtropical Americas".

To avoid confusion with native english speakers, we have opted to use the term "subtropical North and South America".

Line 21: Change "copuled" to "coupled". Done.

Line 42:"...and the dynamics the features largely characterise the MSD characteristics...". I don't understand what is meant here, something is wrong with the grammar.

Sentence has been reworded.

Line 43: Change "reproduce accurately" to "accurately reproduce".

Line 51: Remove hyphen from "South-America". Done.

Line 66: Space needed in "Met Office". Done.

Line 119: Replace "beginning for" with "covering"; replace "that include" with "of". Done.

Line 142: Change "temperature" to "temperatures". Done.

Line 171: Second "the" is not required. Done.

Line 184: brackets not needed around location point. Done.

Line 195: By "a minimum" do you mean "southernmost position"? This would be easier to understand.

Changed for "southernmost position".

Line 302: Replace "indicative" with "are indicative". Done.

Line 304: Clarify if the decreased omega is a reduction or increase in ascent.

Line 309: Mixture of singular and plural in this line. Done.

Line 331: By convention, "El" is not included when referring to the "Nino-3.4 index". Done.

Line 362: Change "opposite sign response" to "opposite signed response". Done.

References

Adams, D. K. and Comrie, A. C. (1997), 'The north American monsoon', *Bulletin of the American Meteorological Society* **78**(10), 2197–2214.

Arritt, R. W., Goering, D. C. and Anderson, C. J. (2000), 'The north american monsoon system in the hadley centre coupled ocean-atmosphere gcm', *Geophysical Research Letters* **27**(4), 565–568.

Cai, W., McPhaden, M. J., Grimm, A. M., Rodrigues, R. R., Taschetto, A. S., Garreaud, R. D., Dewitte, B., Poveda, G., Ham, Y.-G., Santoso, A. et al. (2020), 'Climate impacts of the el niño–southern oscillation on south america', *Nature Reviews Earth & Environment* **1**(4), 215–231.

Fisher, R. A. (1992), Statistical methods for research workers, *in* 'Breakthroughs in statistics', Springer, pp. 66–70.

Geil, K. L., Serra, Y. L. and Zeng, X. (2013), 'Assessment of cmip5 model simulations of the north american monsoon system', *Journal of Climate* **26**(22), 8787–8801.

Giannini, A., Kushnir, Y. and Cane, M. A. (2000), 'Interannual variability of Caribbean rainfall, ENSO, and the Atlantic Ocean', *Journal of Climate* **13**, 297–311.

Giannini, A., Saravanan, R. and Chang, P. (2004), 'The preconditioning role of tropical Atlantic variability in the development of the ENSO teleconnection: Implications for the prediction of Nordeste rainfall', *Climate Dynamics* **22**, 839–855.

Jones, C. and Carvalho, L. M. (2013), 'Climate change in the South American monsoon system: present climate and CMIP5 projections', *Journal of Climate* **26**(17), 6660–6678.

Menary, M. B., Kuhlbrodt, T., Ridley, J., Andrews, M. B., Dimdore-Miles, O. B., Deshayes, J., Eade, R., Gray, L., Ineson, S., Mignot, J., Roberts, C. D., Robson, J., Wood, R. A. and Xavier, P. (2018), 'Preindustrial control simulations with hadgem3-gc3. 1 for cmip6', *Journal of Advances in Modeling Earth Systems* **10**(12), 3049–3075.

Vera, C., Higgins, W., Amador, J., Ambrizzi, T., Garreaud, R., Gochis, D., Gutzler, D., Lettenmaier, D., Marengo, J., Mechoso, C. R., Nogues-Paegle, J., Dias, P. L. S. and Zhang, C. (2006), 'Toward a unified view of the American monsoon systems', *Journal of Climate* **19**(20), 4977–5000.

Wang, P. X., Wang, B., Cheng, H., Fasullo, J., Guo, Z., Kiefer, T. and Liu, Z. (2017), 'The global monsoon across time scales: Mechanisms and outstanding issues', *Earth-Science Reviews* **174**, 84–121.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T. and Xavier, P. K. (2018), 'The met office global coupled model 3.0 and 3.1 (gc3. 0 and gc3. 1) configurations', *Journal of Advances in Modeling Earth Systems* **10**(2), 357–380.

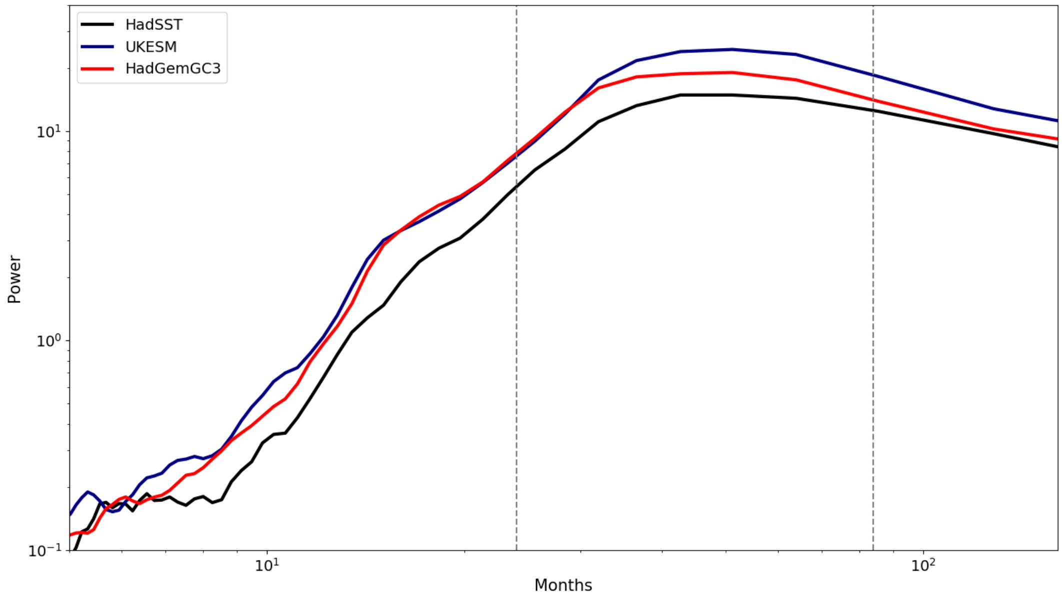

Figure 1: Power spectrum of the ENSO 3.4 index in pre-industrial control simulations of the HadGEM3 and UKESM1 models and HadSST data. The gray lines indicate the 2 and 7 yr period.

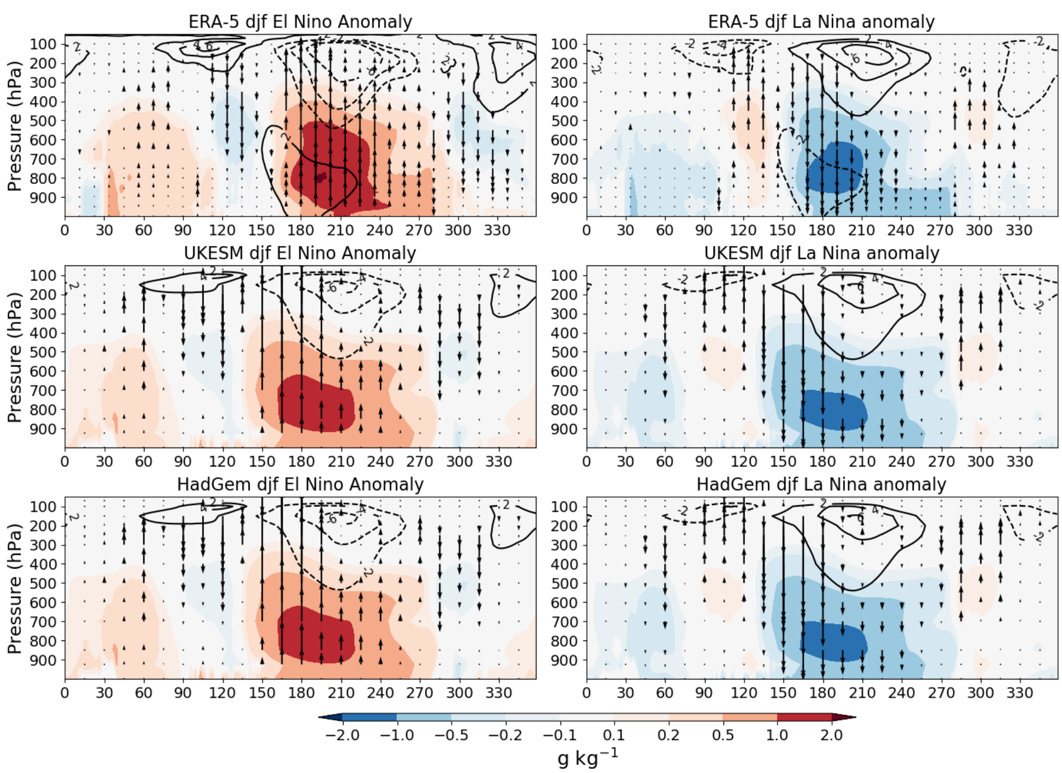

Figure 2: DJF Longitude-height Walker circulation anomalies of specific humidity (colour-contours), ω (vectors) and zonal wind (line-contours) during El Niño events (left) and La Niña events (right). Results are shown for ERA-5 (upper), UKESM-pi (middle) and HadGEM3 piControl (lower).

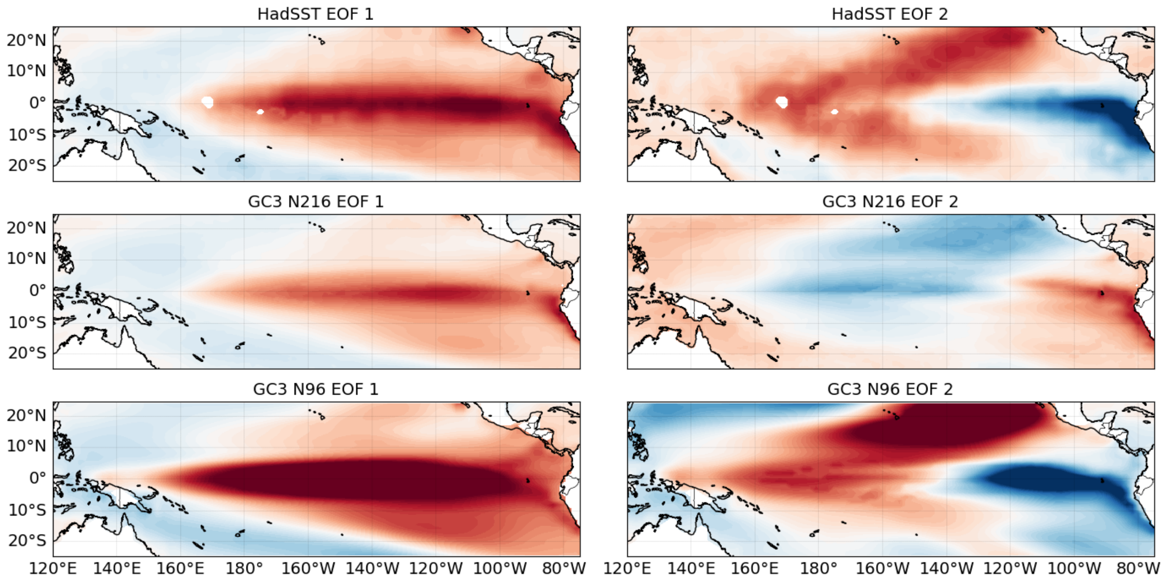

Figure 3: SST patterns [arbitrary units] of the two leading EOFs in HadSST, GC3 N216 and UKESM1.

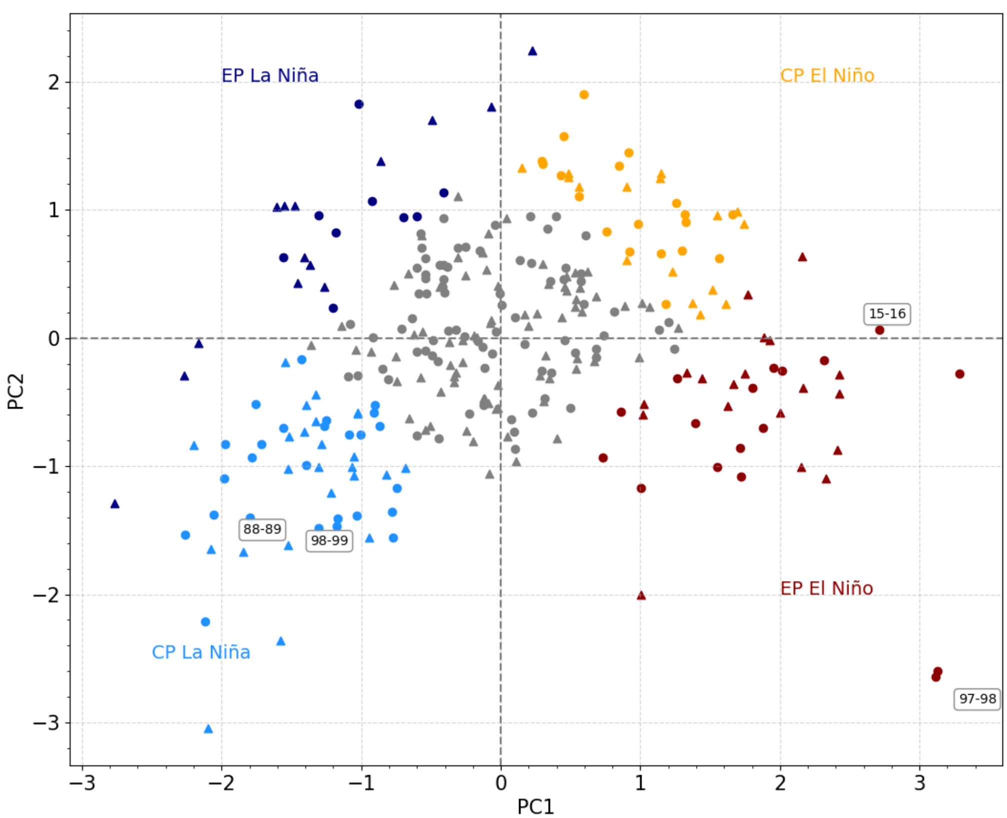

Figure 4: Principal component (PC) space of the first and second leading PCs of the deseasonalized Pacific SSTs diagram showing HadSST (circles) and GC3 N216 (triangles). The PCs are showing as DJF-means.

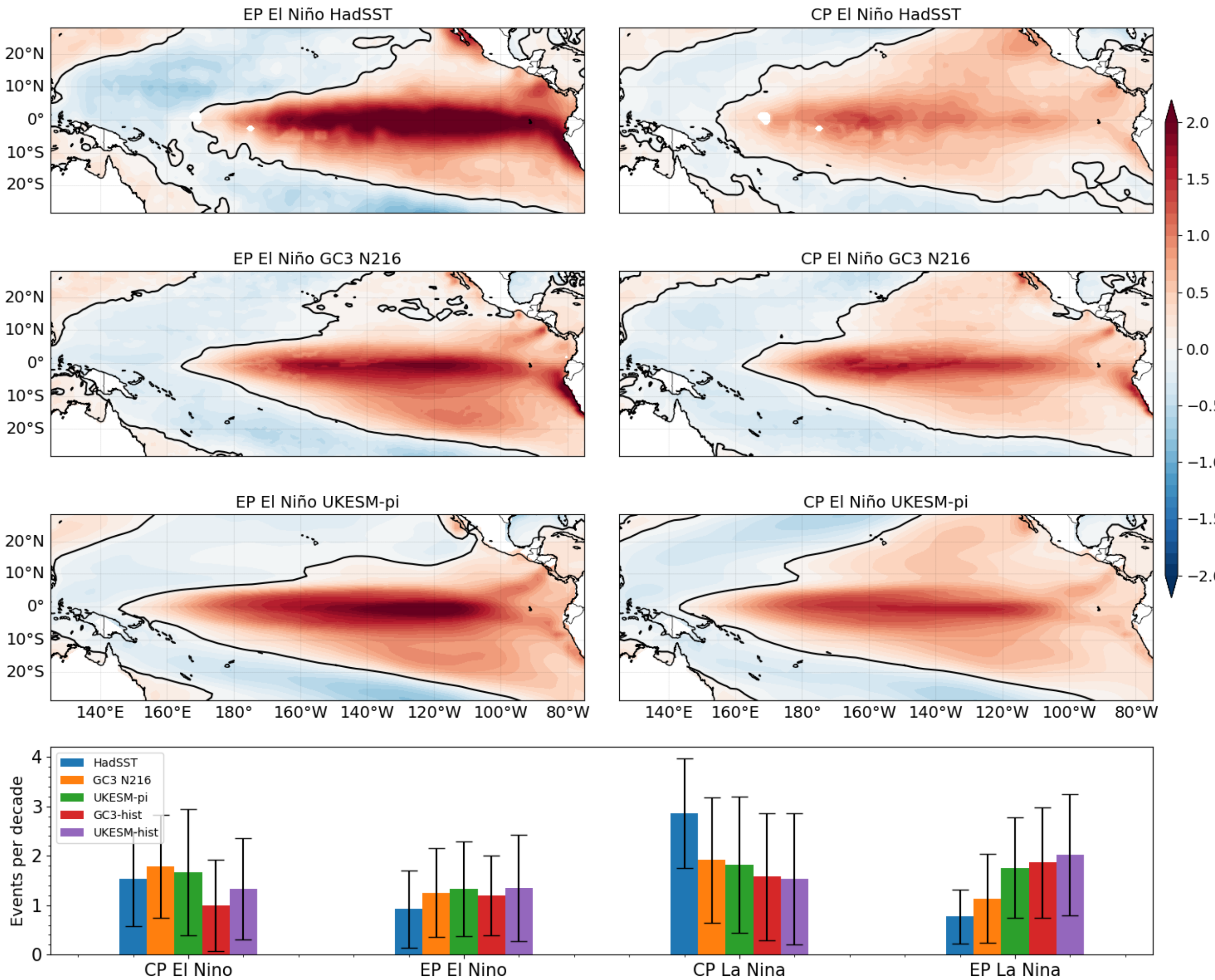

Figure 5: SST anomalies [K] for East Pacific (EP) and Central Pacific El Niño events in HadSST, GC3 N216 and UKESM piControl. EP (CP) events were defined where the E-index (C-index) was greater than 1. In the bottom panel, the frequency of events per decade (with standard deviation as error bar) is shown for HadSST and the simulations used in this study. The E-index is computed from $(PC1 - PC2)/\sqrt{2}$ and the C-index from $(PC1 + PC2)/\sqrt{(}2)$.

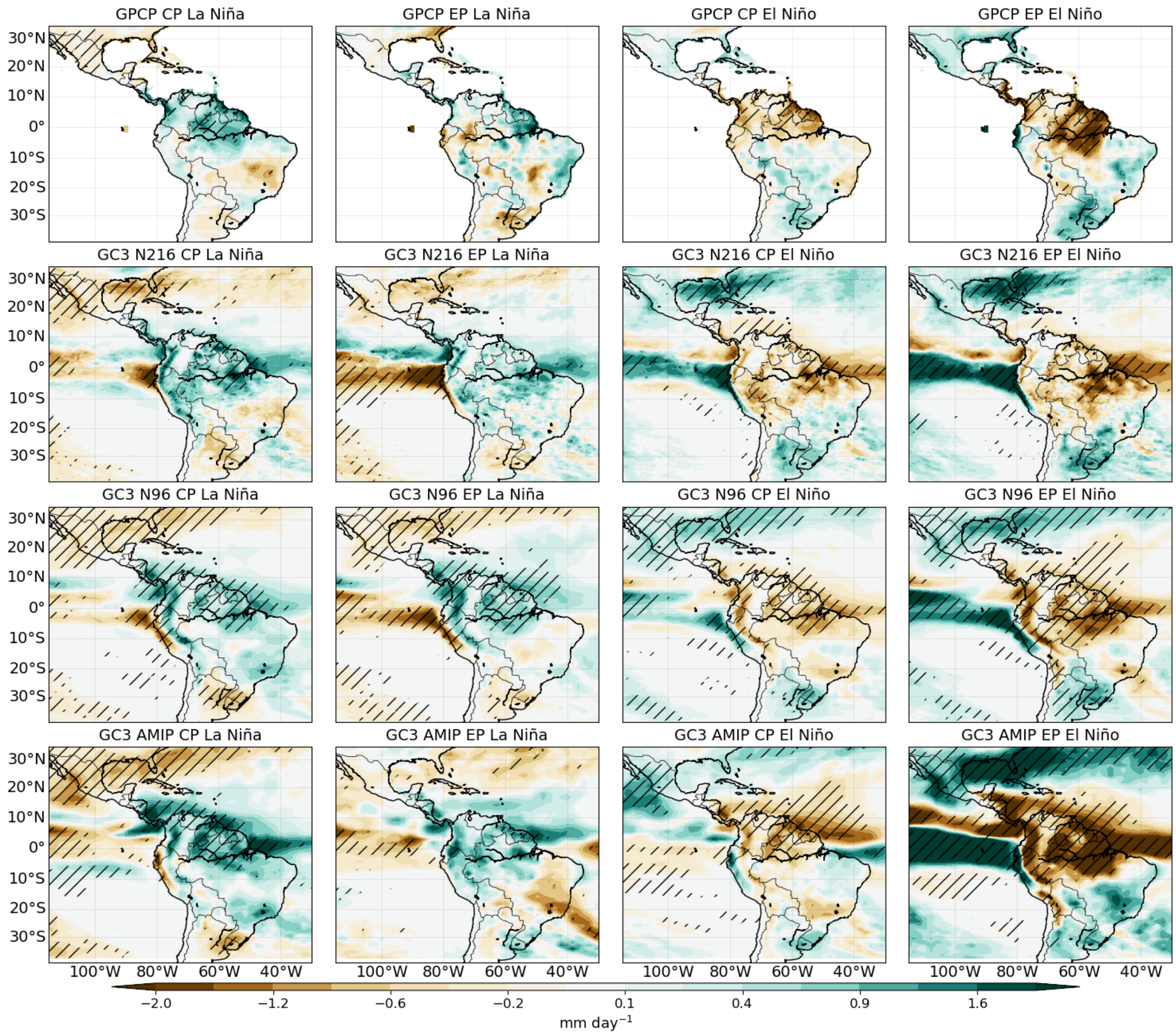

Figure 6: Precipitation anomalies in GPCC 1940-2013, GC3 N216, GC3 N96-pi and GC3 AMIP for the four different types of ENSO events, as defined by Cai et al. (2020). Statistically significant anomalies (95% confidence level) are hatched.

Response to Referee #2

García-Franco, JL., Gray LJ, Osprey S.

Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford. Parks Road Oxford, United Kingdom OX1 3PU email: jorge.garcia-franco@physics.ox.ac.uk

---

Dear Dr. David Adams,

Many thanks for your comments and suggestions on the manuscript. We have found your two main critiques to be very timely and useful to improve our study. We hope you'll find the changes made to manuscript satisfactory.

Regarding your two major comments we have made the following changes to the manuscript. First, we included the following in the introduction to motivate the study:

Climate research in recent decades has aimed to reduce uncertainty in climate projections by improving GCMs, but different approaches taken by modelling centres appear to be seemingly disconnected (Jakob, 2014). One approach is to reduce horizontal resolution down to km resolution to rely less on parametrizations and more on physical laws to represent clouds and convection (Palmer and Stevens, 2019). A second approach aims to model Earth System processes to better characterise complex land-atmosphere-ocean biogeochemical cycles that may provide a better constraint on climate sensitivity, a parameter that depends on the carbon cycle (Marotzke et al., 2017; Sellar et al., 2019; Andrews et al., 2019). Finally, recent arguments have also suggested to include stochastic parametrisations of sub-grid processes since this approach has improved seasonal forecasts and may therefore improve climate projections (Palmer, 2019).

To address your second comment regarding perhaps the parametrisations and their role for improved biases, as we mostly see an improvement of the biases in the South American Monsoon with increased resolution we argue that the oceanic component is very relevant, where perhaps the eddy heat flux parametrisations improve as resolution improves.

The results of this study showed that the medium resolution (GC3 N216) simulation improved upon some of the biases of the lower resolution simulations, such as most of the precipitation biases. This improvement in the medium resolution simulation may largely be due to the improved dynamics associated with relying less on model parametrisations and more on physical governing laws. The double-ITCZ problem and the Atlantic ITCZ biases have been shown to be directly affected by the convective scheme. Several parametrised processes associated with the convective scheme have been shown to the treatment of entraintment and moisture-cloud feedbacks (Oueslati and Bellon, 2013; Li and Xie, 2014). The resolution of the ocean sub-model is known to have an impact over the equatorial Atlantic SSTs and the ITCZ biases which are noticeably reduced in the medium resolution simulation (Kuhlbrodt et al., 2018), due to the improvement in the eddy heat flux and the 15associated heat uptake and transport of the ocean. The improvement in the ITCZ and the associated dynamics also improves the associated circulation biases in the South American Monsoon indicating that the oceanic resolution in these models improve the cross-equatorial transport, SST gradients and the land-sea circulation over the Amazon during austral summer.

1. Specific comments

Line 36 " A bimodal regime characterises the seasonal cycle of precipitation in southern Mexico, Central America and the Caribbean that is typically referred to as Midsummer Drought (MSD)" Perhaps for completeness you can include the more local reference terms for this phenomenon. In Central America it is often called "El Veranillo" and in southern and eastern Mexico "La Canícula". For example a bit more detail on the MSD can be found in these articles, Amador JA et al. 2016, Amador, J.A., et al., (2016), Durán Quesada et al (2017).

We have added the suggested references and more detail to this paragraph:

Line 41 . The complex interplay of moisture transport, evaporation and the dynamics..." When you say evaporation here you should probably clarify if you mean from the sea-surface or from land-surface or both, as terrestrial latent heat fluxes is a difficult quantity to measure and the effects on precipitation are unclear (e.g., moisture recycling).

The manuscript now states:

The complex interplay of SSTs, evaporation and moisture between the East Pacific Ocean and the Caribbean Sea are key for the spatial and temporal characteristics of the MSD (Amador et al., 2006; Herrera et al., 2015; Durán-Quesada et al., 2017; Straffon et al., 2019)

Line 51 "The date of monsoon onset is also region-dependent; in northern South-America convection is observed from early October, whereas convection in southeastern Brazil typically starts in mid- November or later (Marengo et al., 2001; Nieto-Ferreira and Rickenbach, 2011)." You probably want to clarify this. Do you mean deep convection and the associated rainy season? In the Central Amazon region, the rainy season begins late December and lasts until about April. Typically, in October, there may be intense deep convective events in the Central Amazon (see Adams et al 2013), but in terms of convective precipitation, January through April are very rainy (see Machado et al. 2004). How well models actually reproduce the geographic distribution of Amazon Basin rainfall is an important issue, you may want to discuss with a little more detail and citations.

Many thanks for these suggestions, the initial wording was indeed unclear. We hope you like the new wording in the manuscript. Now, when presenting the South American Monsoon:

In the central Amazon and northern South America, convective activity is observed from early October but the main rainy season extends from December to April (Machado et al., 2004; Adams et al., 2013), whereas convection in southeastern Brazil starts in November and peaks in January and February (Marengo et al., 2001; Nieto-Ferreira and Rickenbach, 2011).

Line 69 You should probably clarify what hemisphere you mean here when you say "fall".

Done, we clarified to state "austral fall".

Line 73 "The next efforts to improve climate models include increased horizontal resolution, ..." This drive towards increased horizontal resolution is quite strong, down to the kilometer resolution for GCMs, you should refer to some of the literature I mention in the Major Comments section.

We have addressed this from your major comment.

Line 83 "The study documents the main biases in the simulated climate of UKESM1 and HadGEM3.0 and compares the effect of increased horizontal resolution and Earth System processes on the representation of the AMS climate. The analysis provides a framework for using these climate models in scenario studies, to highlight possible sources of model error that may be corrected and to further understand variability and teleconnections in this region."

Line 98 "GPCP, GPCC and CHIRPS are also used for their longer period, although arguably each of these datasets have shortcomings in either resolution or spatial coverage." You should probably include a few citation of studies that have used these data in similar context for the reader to consider, particularly studies where the shortcomings are discussed.

Reviewer 1 made a similar suggestion. Therefore, now we point to studies that have validated one or several or these datasets in a region of the AMS. However, a study intercomparing the different datasets and validating them against rain gauge data is not know to us. The manuscript now reads:

Some studies have validated one or several of these datasets in the AMS region (e.g. Franchito et al., 2009; Dinku et al., 2010; Trejo et al., 2016).

Line 113 "piControl", I assume you mean pre-industrial, but you should spell it out for the reader.

Done.

Line 190 "Afterwards, the ITCZ migrates northward reaching a peak latitude and mean rainfall at 10N by day 250, or May 30." I think you have made a mistake here, you probably mean early September.

Correct. We corrected this line.

Line 213 Write "Negative ω and low-level moisture biases in the central and East Pacific Ocean ..."

Done.

Line 221 "These are observed as negative zonal wind biases, indicative of significantly weaker upper-level westerlies resulting from the overturning circulation in the Pacific Ocean." This statement is a bit confusing, it sounds as if you are referring to the oceanic circulation within the Pacific Ocean.

This statement has now been removed as the Walker circulation figure has been moved to the Supplementary material and the discussion of the figure in the main paper has been made shorter.

Line 280 Rewrite using commas "The models also show a good representation of the transition from winter to summertime rainfall by representing, with relative skill, the smooth transition from 4 mm day-1 in September to 6 mm day -1 in November and close to 8 mm day -1 in late December."

Done.

Line 290 "these quantities characterise the strength and height of deep convection and the mid-level moisture." This idea is a little unclear, what you do mean "the mid-level moisture"? Specific humidity has a vertical distribution associated with instability and convection. And OLR for convective cloudiness would be associated with high levels in the atmosphere.

Yes, the wording in this paragraph was unclear. We use the free tropospheric specific humidity values (q at 500 hPa) and the vertical velocity at the same level, as well as OLR. And yes, these are technically different levels in the atmosphere we are looking at, and only a small picture of the vertical characteristics of convection, however, these metrics show interesting differences between model and reanalysis data.

Line 319 Check spelling "althougth"

Done.

The American Monsoon System in HadGEM3 and UKESM1

Jorge L. García-Franco[1], Lesley J. Gray[1,2], and Scott Osprey[1,2]

[1]Atmospheric, Oceanic and Planetary Physics, Department of Physics. University of Oxford.
[2]National Centre for Atmospheric Science, UK.

Correspondence: Jorge L García-Franco: jorge.garcia-franco@physics.ox.ac.uk

Abstract. The simulated climate of the American Monsoon System (AMS) in the U.K. models HadGEM3 GC3.1 (GC3) and the Earth System model UKESM1 is assessed and compared to observations and reanalysis. We evaluate the pre-industrial control, AMIP and historical experiments of UKESM1 and two configurations of GC3: a low (1.875°x1.25°) and a medium (0.83°x0.56°) resolution. The simulations show a good representation of the seasonal cycle of temperature in monsoon regions, although the historical experiments overestimate the observed summer temperature in the Amazon, Mexico and Central America by more than 1.5 K. The seasonal cycle of rainfall and general characteristics of the North American Monsoon are well represented by all the simulations, showing a noticeable improvement from previous versions of the HadGEM model. The models reasonably simulate the bimodal regime of precipitation in southern Mexico, Central America and the Caribbean known as the midsummer drought, although with a stronger than observed difference between the two peaks of precipitation and the dry period. Austral summer biases in the modelled Atlantic Intertropical Convergence Zone (ITCZ), cloud cover and regional temperature patterns are significant and influence the simulated regional rainfall in the South American Monsoon. These biases lead to an overestimation of precipitation in southeastern Brazil and an underestimation of precipitation in the Amazon. The precipitation biases over the Amazon and southeastern Brazil are removed in the AMIP simulations, highlighting that the Atlantic SSTs are key for representing precipitation in the South American Monsoon. El Niño Southern Oscillation (ENSO) teleconnections, of precipitation and temperature, to the AMS are well represented by the simulations. The precipitation responses to the positive and negative phase of ENSO in subtropical America are linear in both pre-industrial and historical experiments. Overall, the biases in UKESM1 and the low resolution configuration of GC3 are very similar for precipitation, ITCZ and Walker circulation, i.e., the inclusion of Earth System processes appears to make no significant difference for the representation of the AMS rainfall. In contrast, the medium resolution HadGEM3 N216 simulation outperforms the low-resolution simulations due to improved SSTs and circulation. biases in the dynamical core, shared with HadGEM3 GC3.1 dominate.

1 Introduction

The American Monsoon System (AMS) is the regional monsoon associated with summer rainfall in subtropical North and South America. The AMS is associated with the coupled rainfall and circulation response to the seasonal migration of the

Intertropical Convergence Zone (ITCZ) (Zhou et al., 2016), and is is typically subdivided into the North and South American Monsoon Systems (Vera et al., 2006). The North American Monsoon is the northernmost part of the AMS and the main source of rainfall in south-western North America, extending from central-west Mexico into the southwestern United States, with the core region located in northwestern Mexico (Adams and Comrie, 1997; Stensrud et al., 1997; Vera et al., 2006). The seasonal cycle is characterised by a wet July-August-September season and significantly drier conditions during the rest of the year (Adams and Comrie, 1997). Several features of the North American Monsoon are modulated by the East Pacific Ocean or the Gulf of Mexico, e.g., the frequency of Gulf Surges (Douglas et al., 1993; Adams and Comrie, 1997; Seastrand et al., 2015; Lahmers et al., 2016). Moisture in the North American Monsoon is mainly advected in the low-level flow from the Gulf of California and the East Pacific Ocean whereas moisture mixed in the mid-troposphere from the Caribbean Sea and Gulf of Mexico is a secondary, but relevant, source (e.g Stensrud et al., 1997; Pascale and Bordoni, 2016; Ordoñez et al., 2019).

The South American Monsoon is a primary source of precipitation for South America, especially in the Amazon region (Gan et al., 2004; Vera et al., 2006; Jones and Carvalho, 2013). During austral summer (DJF) monsoon rainfall accounts for over 60% of the total annual precipitation in the Amazon (Gan et al., 2004; Marengo et al., 2012), whereas austral winter rainfall accounts for less than 5% of the total annual rainfall (Vera et al., 2006). The spatial domain of the South American Monsoon generally includes central and southeastern Brazil, Bolivia, northern Argentina and Paraguay but this definition can vary amongst studies (e.g. Jones and Carvalho, 2002; Bombardi and Carvalho, 2011; Marengo et al., 2012; Yin et al., 2013). In the central Amazon, convective activity is observed from early October but the main rainy season extends from December to April (Machado et al., 2004; Adams et al., 2013), whereas convection in southeastern Brazil starts in November and peaks in January and February (Marengo et al., 2001; Nieto-Ferreira and Rickenbach, 2011). The mean-state and variability of the Atlantic, in particular the SSTs and the Intertropical Convergence Zone (ITCZ), greatly influences the South American Monsoon, as demonstrated in observations and climate models (see e.g. Giannini et al., 2004; Vera and Silvestri, 2009; Lee et al., 2011).

A bimodal regime characterises the seasonal cycle of precipitation in southern Mexico, Central America and the Caribbean, most commonly known as Midsummer Drought (MSD) (Magaña et al., 1999; Gamble et al., 2008), but also as "Veranillo" in Central America and "canícula" in southern Mexico (Dilley, 1996; Amador et al., 2016; Durán-Quesada et al., 2017). The seasonal cycle in these regions is characterised by two precipitation maxima, in June and September, that are separated by a drier period in July and August. The complex interplay of SSTs, evaporation and moisture transport between the East Pacific Ocean and the Caribbean Sea are key for the spatial and temporal characteristics of the MSD (Amador et al., 2006; Herrera et al., 2015; Durán-Quesada et al., 2017; Straffon et al., 2019). Although the regions with an MSD are not formally part of the North American Monsoon, several studies (e.g Vera et al., 2006; Wang et al., 2017; Pascale et al., 2019) have analysed aspects of the MSD in climate models and observations. This study uses the definitions for the North and South American Monsoons as previous studies (Vera et al., 2006; Marengo et al., 2012), with additional analysis on the MSD of southern Mexico and Central America (Magaña et al., 1999; Perdigón-Morales et al., 2018).

General Circulation Models (GCMs) are used to increase our understanding of monsoon dynamics and the current and future effect of greenhouse forcing on regional rainfall (see e.g. Arritt et al., 2000; Seager and Vecchi, 2010; Sheffield et al., 2013a; Ryu and Hayhoe, 2014; Colorado-Ruiz et al., 2018). In the AMS, studies have assesed how horizontal resolution modifies the simulated climate (Pascale et al., 2016) and how climatological model biases affect simulated teleconnections (Vera and Silvestri, 2009; Bayr et al., 2019). The CMIP5 simulations of the North American Monsoon misrepresented aspects of the seasonal cycle of precipitation and overestimated the peak monsoon rainfall (Geil et al., 2013; Sheffield et al., 2013a). Most CMIP5 models simulated an earlier onset date, but improved from CMIP3 since the onset date showed a clear separation of rainy and dry seasons in daily precipitation time-series. In contrast, the simulated retreat date was unclear in most models which highlighted problems for these models to simulate the regional changes during retreat stage (Geil et al., 2013; Sheffield et al., 2013a).

The majority of CMIP5 models were unable to represent the seasonal cycle of the MSD and the total annual rainfall in Central America and the Caribbean; most models did not show signs of two-peak bimodal distribution of precipitation (Ryu and Hayhoe, 2014; Colorado-Ruiz et al., 2018). However, some models such as HadGEM2 reasonably simulated the observed bimodal regime by showing a two-peak distribution of precipitation (Ryu and Hayhoe, 2014).

The accurate simulation of the geographic distribution and seasonality of rainfall in the Amazon rainforest is a relevant issue due to the impact of the rainforest on climate and society (e.g. Li et al., 2006; Malhi et al., 2009; Yin et al., 2013). In the South American Monsoon, CMIP5 models improved from CMIP3 in the simulated distribution of precipitation during monsoon maturity and exhibited an improved seasonal cycle (Jones and Carvalho, 2013; Yin et al., 2013). However, long-term biases in the South American Monsoon, e.g., the underestimation of rainfall in the central Amazon, persisted in CMIP5 (Yin et al., 2013). The geographic distribution of rainfall during austral fall and several characteristicis of the South Atlantic Convergence Zone were also poorly represented in CMIP5. Projections from CMIP5 consistently showed a longer wet season in the South American Monsoon with earlier onsets and later retreats (Jones and Carvalho, 2013).

Climate research in recent decades has aimed to reduce uncertainty in climate projections by improving GCMs, but different approaches taken by modelling centres are seemingly disconnected (Jakob, 2014). One approach is to reduce horizontal resolution down to km resolution to rely less on parametrizations and more on physical laws to represent clouds and convection (Palmer and Stevens, 2019). A second approach aims to model Earth System processes to better characterise complex land-atmosphere-ocean biogeochemical cycles that may provide a better constraint on climate sensitivity, a parameter that depends on the carbon cycle (Marotzke et al., 2017; Sellar et al., 2019; Andrews et al., 2019). Finally, recent arguments have also suggested to include stochastic parametrisations of sub-grid processes since this approach has improved seasonal forecasts and may therefore improve climate projections (Palmer, 2019). The new phase of the CMIP project will include a range of new submissions which will include models with higher resolution and more Earth System models (Eyring et al., 2016). A comparison and evaluation of simulations with increased horizontal resolution and Earth System models may suggest where modelling efforts are resulting in significant improvements in model representation of monsoons.

The main purpose of this study is to validate the U.K. models: UKESM1, an Earth System model, and HadGEM3, the latest generation of the Hadley Centre Global Environment model. In particular, we document the main biases in these models in the region of the AMS, comparing the effect of increased horizontal resolution and Earth System processes on the representation of the AMS. The analysis may provide a framework for using these climate models in scenario studies or to further understand variability and teleconnections in this region. The remainder of this paper is organised as follows: section 2 describes the

Table 1. Summary of the datasets used in this study. For each dataset, the acronym used hereafter, the period of coverage, the field used and the horizontal resolution are shown. Some datasets extend further back in time, but only the satellite-era period is used in most of the datasets. The variables used are: precipitation, surface-air temperature ($2mT$), sea-level pressure (SLP), SSTs, the x and y components of the wind (u, v), the lagrangian tendency of air pressure (ω), outgoing longwave radiation (OLR) and specific humidity (q).

Dataset/ Version	*Acronym*	*Variable*	*Period*	*Data type*	*Resolution*	*Reference*
Global Precipitation Climatology Project v2.3	GPCP	Precipitation	(1979-2018)	Surface station and satellite	2.5°x2.5°	(Adler et al., 2003)
Global Precipitation Climatology Centre	GPCC	Precipitation	(1940-2013)	Surface station	0.5°x0.5°	(Becker et al., 2011)
Climatic Research Unit TS v4.	CRU4	Surface temperature	(1979-2017)	Surface station	0.5°x0.5°	(Harris et al., 2014)
Climate Hazards Infrared Precipitation with Stations	CHIRPS	Precipitation	(1981-2018)	Surface station and satellite	0.05°x0.05°	(Funk et al., 2015)
Tropical Rainfall Measurement Mission 3B42 V7	TRMM	Precipitation	(1999-2018)	Surface station and satellite	0.25°x0.25°	(Huffman et al., 2010)
Hadley Centre SST3	HadSST	SST	(1940-2018)	Buoy and satellite	2.5°x2.5°	(Kennedy et al., 2011)
European Centre for Medium-Range Forecasting ERA-5	ERA-5	$2mT$, SLP, u, v, ω, OLR, q	(1979-2018)	Reanalysis	0.75x0.75°	(C3S, 2017)

observations, reanalyses and models used, section 3 compares modelled and observed climatological features such as the ITCZ. Section 4 analyses the spatial and temporal characteristics of rainfall and convection in the AMS while section 5 documents the simulated teleconnections of ENSO. Section 6 provides a summary and discussion.

2 Data and methods

2.1 Observations and reanalysis data

Table 1 summarises relevant information of the observations and reanalysis datasets used in this study. In short, surface and satellite observations were used where available, whereas other metrics were taken from reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF): ERA-5, downloaded from https://climate.copernicus.eu/climate-reanalysis. Four different precipitation datasets are used. The TRMM dataset has a high horizontal and temporal resolution and was used in several CMIP assessments (Geil et al., 2013; Jones and Carvalho, 2013) as a reliable source of precipitation (Carvalho et al., 2012). Therefore, we use TRMM as our best estimate for the spatial and temporal characteristics of the AMS rainfall. However, the period covered by TRMM (1998-2018) is too short to analyse statistically robust teleconnections or variability, so we use GPCP, GPCC and CHIRPS for their longer period. Although a thorough validation and comparison of these datasets across the AMS domain is missing, several studies have analysed one or more of these datasets in regions of the AMS (e.g. Franchito et al., 2009; Dinku et al., 2010; Trejo et al., 2016).

2.2 Model data

The MOHC has submitted the output of two models for CMIP6: HadGEM3 GC3.1 (hereafter GC3) is the latest version of the Global Coupled (GC) Met Office Unified Model (UM) and UKESM1, the new U.K. Earth System Model. The most substantial change from the version used in CMIP5 (HadGEM2-AO) is the inclusion of the new GC configuration 3.1 (Walters et al., 2019) with the updated components: Global Atmosphere 7.0 (GA7.0), Global Land 7.0 (GL7.0), Global Ocean 6.0 (GO6.0), and Global Sea Ice 8.0 (GSI8.0). The GC3.1 configuration runs with 85 atmospheric levels, 4 soil levels and 75 ocean levels; for details see Williams et al. (2018) and Kuhlbrodt et al. (2018). The GC3 model was run for CMIP6 deck experiments with two horizontal resolutions: a low resolution configuration, labelled as N96, with an atmospheric resolution of 1.875°x1.25° and a 1° resolution in the ocean model and a medium resolution configuration, labelled N216, with atmospheric resolutions of 0.83°x 0.56° and a 0.25° oceanic resolution (Menary et al., 2018).

The UKESM1 was recently developed aiming to improve the UM climate model adding processes of the Earth System (Sellar et al., 2019). These additional components include ocean biogeochemistry with coupled chemical cycles, tropospheric-stratospheric interactive chemistry which aim to better characterise aerosol-cloud and aerosol-radiation interactions (Mulcahy et al., 2018; Sellar et al., 2019). The physical atmosphere-land-ocean-sea-ice core of the HadGEM3 GC3.1 underpins the UKESM1, so that the UKESM1 and the HadGEM3 have the same dynamical core but the UKESM1 has the additional components mentioned above.

This study uses three CMIP6 deck experiments. First, the pre-industrial control (piControl) simulations, which are run with constant forcing using the best estimate for pre-industrial (1850) forcing of aerosols and greenhouse gas levels. The historical experiments are 164-yr integrations for 1850-2014 that include historical forcings of aerosol, greenhouse gas, volcanic and solar signals since 1850 (Eyring et al., 2016; Andrews et al., 2019). For further details, Andrews et al. (2020) extensively describes the historical simulations of HadGEM3-GC3.1.

In contrast to the pre-industrial control experiments, the historical experiments use time-varying aerosol and greenhouse gas emissions and land-use change (Eyring et al., 2016). In Latin-America, land-use change for agricultural purposes has dramatically decreased tree cover in Central America and south-eastern Brazil since the 1950s (Lawrence et al., 2012), thereby affecting the surface energy balance. The regional emissions of carbonaceous aerosols, nitrogen oxides and volatile organic compound in Latin America are also considered in the historical experiments. These emissions are noteworthy, e.g., due to the impact of black carbon emissions by increased biomass burning in the Amazon and northern Central America (Chuvieco et al., 2008).

The historical experiments of HadGEM3 and UKESM1 are composed of 4 and 9 ensemble members, respectively, but the results will be presented as the ensemble mean for the 1979-2014 period. These experiments will be referred to as GC3-hist and UKESM1-hist hereafter. Finally, we use the five ensemble members of the AMIP experiment from GC3 N96 covering 1979-2014. Table S1 summarises the main features of the experiments used in this study.

3 Climatological features

This section evaluates the simulated climatological temperature and low-level wind structure in the AMS region, as well as several characteristics of the ITCZ.

3.1 Temperature and low-level winds

The climatological representation of the near-surface air temperature and low-level winds in the models is compared to ERA5 in Figures 1 and 2. First, the climatology of DJF and JJA of ERA5 is shown in Figure 1a, b. The biases of the historical experiments, computed as the differences between the model and observed fields, are shown in Figures 1c, d) for GC3-hist and e, f) for UKESM1-hist. Only statistically significant differences are shown, according to a Welch t-test (Wilks, 2011), which accounts for the difference in sample size and variance between model and observations/reanalysis data. The significance for simulations with multiple ensemble members is estimated first for each ensemble member and then combined into a single probability or p-value using Fisher's method (Fisher, 1992). Pattern correlations and root-mean square error (RMSE) are shown in Figures 1c-f and in Table S2 for all seasons and more variables.

During DJF, the simulations show a colder-than-observed sub-tropical North America and a warm bias over the Amazon (≈ 3.5 K). The west coast of South America also shows a significant warm bias (> 4 K). The simulated circulation in austral summer in South America has a significant bias in the easterly flow coming from the equatorial and subtropical Atlantic. The biases in the low-level winds suggest a weaker easterly flow into southeastern Brazil but also a strong southward flow from northern to southern South America. The South America Low-Level Jet, the low-level northwesterly flow in Bolivia, observed in Figure 1a, is stronger in the simulations. This stronger than observed jet is suggestive of a stronger moisture transport to the La Plata Basin, with has been associated with a drying of the Amazon and positive precipitation anomalies at the exit region of the jet (Marengo et al., 2012; Jones and Carvalho, 2018). In turn in boreal summer (Figures 1d, f), positive biases are observed in southwestern North America (> 3.5 K), which are higher in UKESM1-hist than in GC3-hist. The easterly flow west of Central America has a negative bias in UKESM1 suggesting a biased flow that crosses from the Caribbean Sea into the East Pacific Ocean. Also in JJA, the simulated East Pacific surface temperatures are colder than observed for both historical experiments.

Figures 2a-d compare the GC3 piControl simulations with ERA5. In DJF, the piControl simulations show a smaller positive bias in the Amazon than the historical experiments, as well as a similar bias in the circulation in South America, with the smallest biases in GC3 N216. The inclusion of Earth System processes appears to make no improvement on the low-level circulation biases. Figures 2e, f show the difference between the historical and piControl experiment of GC3, illustrating the response to historical forcing in GC3. The temperature response in austral summer in South America is observed as 1.5 K whereas in JJA in North America temperatures were 4 K higher in the historical experiment than in the piControl. A very similar temperature pattern response to historical forcing was observed for UKESM1 (not shown) although of slightly different magnitude. The only difference in low-level winds seem to be the easterlies in the East Pacific Ocean during JJA.

The seasonal cycle of temperature in key regions of the AMS is shown in Figure 3 which provides a better comparison of the temperature field in these experiments. These regions are illustrated in Figure 1a. The temperature in the North American Monsoon region ranges from the boreal winter 12°C to a maximum in June close to 27°C. Although colder than observed in the piControl and warmer in the historical experiments throughout the year, the models accurately reproduce the seasonal cycle, which may be relevant for the simulated monsoon onset timing and strength (Turrent and Cavazos, 2009). The piControls show a colder-than-observed winter in southern Mexico and northern Central America whereas the historical experiments show a warming signal of about 1.5 K in winter and 2 K in the summer when compared to the piControls. In spite of these biases, both types of experiments follow closely the seasonal cycle in North and Central America.

However, the seasonal cycle in South America is poorly represented in these models (Figures 3 c, d). The simulations show a stronger than observed seasonal cycle, especially the historical experiments. For example, the modelled temperature difference between late austral winter and spring was ≈4 K whereas the observed varies by less than 1 K in the same period. The models show a warm bias in the Amazon region (Fig. 3 d) which peaks in austral spring (SON), during the development of the monsoon (Marengo et al., 2012). In southeastern Brazil, the seasonal cycle is reasonably well reproduced but with a significant cold bias throughout the year which maximizes during austral winter (JJA), as models (e.g. UKESM1) simulate a temperature 4 K lower than observed. In all panels of Figure 3, the historical experiments show a significant warming signal as a response to historical forcing, which is generally stronger in UKESM1 than in GC3.

3.2 The ITCZ

The AMS is intertwined with the seasonal migration of the East Pacific and Atlantic ITCZ and influenced by the Walker circulation through teleconnections (Zhou et al., 2016; Cai et al., 2019). Figure 4 shows the observed and modelled climatological rainfall and the ITCZ climatological position. Three simulations are shown: the ensemble-mean UKESM1-historical, the ensemble mean GC3 AMIP and GC3 N216-pi. Other simulations are not shown as all the coupled low resolution simulations, historical and piControl, showed very similar precipitation and ITCZ characteristics.

The observed ITCZ (Figure 4a) is found, on average, at 8°N in the East Pacific and at 6°N in the Atlantic. All the simulations reasonably represent the climatological position of the East Pacific ITCZ; however, the modelled Atlantic ITCZ near the coast of Brazil is found south of the equator at 3°S in the coupled model simulations. The GC3 N216-pi ITCZ and spatial distribution of rainfall is more consistent with the climatological position of the ITCZ of the TRMM dataset than the UKESM-hist. Rainfall near the Amazon river mouth is significantly larger in the low resolution simulations than in the TRMM dataset. However, the GC3 AMIP shows the best agreement with TRMM in ITCZ position and rainfall distribution.

The seasonal cycle of the ITCZ, precipitation rates and low-level winds in both basins are shown in Figure 5, for TRMM, UKESM1-hist, the GC3 AMIP GC3 N96-pi and GC3 N216-pi. The East Pacific (EP) ITCZ in observations (Fig. 5a) migrates southwards during the first days of the year. The EP ITCZ reaches minimum precipitation and its southernmost position at 5°N around day 100 (mid-April). During boreal spring, the ITCZ migrates northward reaching a peak latitude and maximum rainfall at 10°N by day 250, or early September. The low-level winds are predominantly easterlies, which are stronger away from the ITCZ and weaker and convergent near the ITCZ position. The position and seasonal migration of the East Pacific ITCZ is reasonably well represented in the four simulations (Figs. 5c, e, g, i), but a noticeable bias is observed in the boreal winter precipitation south of the equator in the coupled model simulations. The modelled low-level wind in the coupled model structure shows significant biases near the ITCZ. These wind biases are observed as stronger wind vectors converging toward the ITCZ during boreal summer and spring and stronger wind vectors diverging away from the equator during boreal winter.

The observed Atlantic ITCZ (Figure 5b) has a similar seasonal cycle to the EP ITCZ. The Atlantic ITCZ is close to 4°N at day 1 and migrates southwards at the start of the year reaching its southernmost position at 0° at the end of March. During boreal spring, the Atlantic ITCZ migrates north, reaching 8°N at the start of boreal summer. The boreal winter position of the modelled ITCZ is displaced with respect to the observations. The simulated ITCZ cross south of the equator during boreal winter covering 10-0°S with rainfall rates above 12 mm day $^{-1}$. After boreal spring, the modelled ITCZ crosses back north of the equator and matches the observed ITCZ reasonably well for boreal summer and fall. Low-level wind biases near the Atlantic ITCZ (Figures 5f and h) show that north of the equator the models show a stronger than observed northward wind, and a stronger northerly flow south of 10°S. The biases in the Atlantic ITCZ can also be observed in the overturning circulation (Figure S1) and the associated Walker circulation as significant negative ω and q biases just north and south of equatorial South America indicative of weaker convective activity. The Atlantic Ocean shows a biased strong ascent south of the equator and a biased weak ascent north of the equator in the low resolution simulations. These biases described above were found to be of similar magnitude in the coupled model simulations run at N96 resolution, both historical and piControl experiments, however, these biases improved in the medium resolution GC3 N216-pi and in the AMIP simulations (Figures 5f, j).

The South Atlantic Convergence Zone (SACZ) is a nortwest-southeast oriented band of convection and is a prominent influence on the South American Monsoon mean and extreme rainfall (Carvalho et al., 2004; Marengo et al., 2012). UKESM1 and GC3 appear to reasonably simulate the spatial pattern of active SACZ days and the seasonal cycle of SACZ activity (Figure S2).

4 The American Monsoon System

4.1 Mean seasonal precipitation

The austral summer (DJF) rainfall distribution and biases in South America are shown in Figure 6 for GC3 N216, UKESM-hist and GC3 AMIP. The maximum austral summer rainfall in TRMM (Figure 6a) is found as a nortwest-southeast oriented band of precipitation from the core Amazon region into southeastern Brazil. The coupled simulations (e.g. Figure 6b, c) overestimate rainfall in southeastern Brazil and underestimate rainfall in the core Amazon region.

The biases are illustrated (Figures 6e-h) as the precipitation difference between the simulations and TRMM. The coupled simulations show three main biases. Rainfall in the Atlantic ITCZ in these simulations is displaced southwards, observed as positive (+5 mm day^{-1}) biases south of the equator and negative biases (-5 mm day^{-1}) north of the equator in the Atlantic. Second, the models underestimate rainfall in the core Amazon basin by -3 mm day^{-1} on average, whereas rainfall in southeastern Brazil is overestimated by more than +5 mm day^{-1}, approximately +100% of the observed rainfall in this region. The precipitation biases are associated with a stronger northerly flow in South America, transporting moisture from the Amazon into southeastern Brazil and the La Plata Basin. The magnitude of these biases is smaller in GC3 N216 (Figure 6f) than in the low resolution simulations, such as UKESM1-hist. The ensemble mean GC3 AMIP (Figure 6d) shows a better representation of the austral summer rainfall patterns, removing the main biases (Figure 6g) of the coupled simulations. The response to historical forcing, illustrated by the difference between UKESM1-hist and UKESM1-pi (Figure 6h), is much weaker than the magnitude of the biases.

The modelled and observed JJA mean rainfall and biases for Mexico and Central America are shown in Figure 7. The main feature (Figure 7a) is the East Pacific ITCZ which extends north to 15°N near the western coast of Mexico as a broad band of rainfall (>11 mm day^{-1}). The North American Monsoon can be observed as a band of precipitation mainly across western Mexico. In the core monsoon region, near the Sierra Madre Occidental (Adams and Comrie, 1997; Zhou et al., 2016), the JJA-mean rainfall is higher than 6 mm day^{-1}.

The modelled East Pacific ITCZ (Figures 7e, f, g) rainfall is overestimated by more than 5 mm day^{-1}, even more so in GC3 AMIP. This wet bias is associated with an easterly bias in the low-level circulation, suggesting a weaker flow from the Caribbean into the East Pacific. The low-resolution simulations (Figure 7e) underestimate rainfall (-5 mm day^{-1}) over land in southern Mexico, Guatemala and Belize. Rainfall in the Caribbean islands and Florida is underestimated (-1 mm day^{-1}) in all simulations. The distribution of rainfall in the North American Monsoon region is relatively well represented in all the simulations, showing only a small wet bias (+2 mm day^{-1}) in western Mexico. The northernmost part of the North American Monsoon (southwestern US) is best simulated by GC3 N216-pi. In most cases for JJA in this region, the precipitation and wind biases were reduced in the high-resolution simulation (Figure 7f). The precipitation response to historical forcing is much lower than the biases (Figure 7h) with no significant precipitation differences over land.

4.2 The annual cycle of rainfall

Figure 8 shows the pentad-mean cycle of rainfall over the North American Monsoon, the Midsummer drought (MSD), the Amazon and Eastern Brazil regions. The correlation between TRMM and model and reanalysis data is also shown in each panel. The seasonal cycle of precipitation in the MSD region in the simulations is well represented as all the simulations show the characteristic bimodal distribution. However, the characteristics of the simulated MSD are different to observations. For example, the magnitude of the first peak in the simulations is higher than TRMM by 4 mm day^{-1}. Similarly, the differences between the first peak and the MSD and between the MSD and the second peak are more pronounced in the coupled simulations. The timing of the MSD period is different in the models, as the simulations show the driest period taking place 10 days after TRMM and ERA5. All the simulations show different magnitudes of the first and second peak and the MSD precipitation, including the AMIP simulation that overestimates the second maximum of rainfall by 2-3 mm day^{-1}.

Rainfall in the North American Monsoon (Figure 8b) show a sharp increase of rainfall around mid-June in models and observations, suggesting that onset timing and strength is well represented in these models. Moreover, the modelled and the observed mean rainrates during monsoon maturity is 4 mm day^{-1}, from mid-July until early September. The historical simulations show a shorter wet season characeterised by an earlier retreat of the rainfall and, as all the simulations, a positive boreal fall rainfall bias (+1 mm day^{-1}), a feature that has been shown in these models in CMIP5 (Geil et al., 2013).

The seasonal cycle of precipitation in eastern Brazil is characterised by a very wet summer ($\sim$8 mm day^{-1}) compared to a very dry ($\sim$0.2 mm day^{-1}) winter (Figure 8c). Austral summer rainfall in the observations consistently shows that maximum rainfall is found in early January ($\sim$8 mm day^{-1}). Rainfall in this region decreases to $\sim$6 mm day^{-1} by late March as the monsoon migrates northward and sharply decreases in austral fall. The models (Figure 8c) show a positive bias during monsoon maturity. This bias was found to be of +4 mm day^{-1} and +2.5 mm day^{-1} for the low and medium resolution simulations, respectively. The bias in the seasonal cycle is consistent with the biases shown in Figure 6, which showed that rainfall in southeastern Brazil is overestimated, especially in the low resolution coupled simulations. In contrast to the coupled simulations, GC3 AMIP shows a very good agreement with the observed maximum summer rainfall and the seasonal cycle (r=0.978) throughout the year.

Finally, the simulated rainfall in the Amazon in the coupled simulations show a dry bias in the austral summer and a good agreement with the observations during austral winter (Figure 8d). The models also represent, with reasonable skill, the transition from early austral spring (4 mm day^{-1} in September) to summertime rainfall (6 mm day^{-1} in November). However, peak summertime rainfall is underestimated by the coupled model simulations, particularly the historical experiments. Rainfall in the Amazon from January to March, in both TRMM and ERA-5, is close to 10 mm day^{-1}, yet the low resolution simulations show rainfall rates of 8 mm day^{-1} in mid-February. GC3 N216-pi shows a better agreement with observations but still underestimates summertime rainfall by 1 mm day^{-1}. The dry Amazon bias has been a know feature of GCMs, including the MOHC models since CMIP3 (Li et al., 2006; Yin et al., 2013). In these simulations the dry Amazon bias is only alleviated in GC3 AMIP whose seasonal cycle and maximum summer rainfall agree well with observations.

4.3 Characteristics of convective activity

The seasonal cycles of out-going longwave radiation (OLR), vertical velocity (ω) and specific humidity (q) are key features of a monsoon since these quantities characterise the strength and height of deep convection. Figure 9 shows the pentad-mean annual cycle of OLR, q and ω at the 500-hPa level in four regions of the AMS. For the North American Monsoon the seasonal cycle of OLR, q and ω is relatively well represented in the simulations. During late boreal winter and early spring, OLR increases steadily as a result of surface warming. However, in early June, near the onset date (Douglas et al., 1993; Geil et al., 2013), OLR sharply decreases reaching a minimum value of 246 W m^{-2} by mid-July. The vertical velocity decreases steadily from January to a minimum in August, indicating ascent from May 1st until September 15th. The models show similar seasonal cycles but overestimate the summertime OLR by $\approx$ 6 W m^{-2} and underestimate mid-level moisture by 0.3 g/kg and ω by 0.01 Pa s^{-1}. The simulated shallower convection and drier mid-troposphere is seemingly compensated by stronger ascent.

In the MSD region, OLR and q show signs of convective activity from mid-April, as OLR sharply decreases and moisture increases. The characteristic MSD bimodal distribution of precipitation can also be observed as two peaks of low OLR, high q and low ω. These periods are separated by a period of relatively higher OLR, lower q and weaker ascent from June 15 until late August. Arguably with a small dry bias with shallower convection after mid-July, the simulations follow closely the observed seasonal cycle. The simulated first peak of rainfall has similar OLR and mid-level moisture but stronger ascending motions, which may explain the positive rainfall bias in this period showed in Figure 8a. In the period between the first peak and the MSD, the simulated OLR increase more sharply than observations from 220 W m^{-2} (June 15) to 250 W m^{-2} (early August), with similar behaviour in ω and q. The period during the second peak of rainfall in September shows signs of shallower convection and a drier mid-level when compared to ERA5.

In southeastern Brazil, the simulations reasonably follow the annual cycle of OLR, q and ω of the reanalysis, particularly during austral winter. The observed q in the dry seasons of austral fall, winter and spring in ERA5 is very similar to the simulated q. However, during austral summer, the coupled model simulations show significant biases characterised by stronger ascent and increased specific humidity in the mid-levels, although the height of convection (OLR 225 W m^{-2}) is only modestly higher in the simulations.

The simulated OLR, q and ω exhibit the highest biases in the Amazon. During austral summer, particularly January and February, the simulated convective activity is shallower (OLR bias of +25 W m^{-2}) and weaker (positive ω bias +0.02 Pa s^{-1}) and the mid-level troposphere is drier (-0.5 g/kg) than in ERA5. In spite of biases in the magnitude of OLR, q and ω during peak convective activity, the seasonal variation is very well simulated so that convective activity, as evidenced by these metrics, starts and ends in the simulations within one or two pentads of the reanalysis. The smallest biases in coupled simulations are those of GC3 N216-pi, for this and the other regions. The simulated OLR, q and ω by GC3 AMIP in southeastern Brazil and the Amazon show a much better agreement with the reanalysis during austral summer than the rest of the observations.

5 ENSO Teleconnections

El Niño-Southern Oscillation (ENSO) teleconnections are the prominent source of interannual variability in the AMS (Vera et al., 2006). This section shows the temperature, sea-level pressure (SLP) and precipitation responses to observed and simulated ENSO events in the AMS. ENSO events were defined when the DJF-mean Niño 3.4 index was above or below 0.65 (Trenberth, 1997). Other indices, including the use of a 5-month running mean (Trenberth et al., 1998), were tested without significantly changing the results. Previous studies (e.g. Menary et al., 2018; Kuhlbrodt et al., 2018) showed that the MOHC models reasonably simulate several characteristics of ENSO such as the period and SST patterns.

The temperature and SLP response to ENSO events is shown in Figure 10 for model and ERA5 data. The modelled warm anomaly during El Niño events in the East Pacific Ocean does not extend to the east as much as the observed warm anomaly and the cold anomalies during La Niña events in the Central Pacific are colder in the simulations. However, the simulated and observed teleconnection patterns to South America, e.g., the cold anomalies during La Niña events in northern South America are seemingly well simulated. The teleconnection to southern North America, i.e., colder (warmer) conditions in southern (northern) North America during El Niño events are relatively well simulated even though the low resolution simulations showed a broader and stronger than observed response in southeastern US.

The SLP response in the northern Pacific and North America, known as the Pacific North-American pattern, is linked with a displacement of the subtropical jet affecting the eastward propagating wave activity that reaches the North Atlantic (e.g. Bayr et al., 2019; Jiménez-Esteve and Domeisen, 2020). During ENSO events, the Aleutian Low is strengthened in ERA5, with a strong SLP anomaly (-4 hPa) off the coast of California. The models show a similar but smaller SLP response in the same region. Positive ENSO events are typically associated with a negative phase of the North Atlantic Oscillation (NAO), with an opposite response for La Niña events. While the models seem to be able to capture this response of the NAO, the simulated response is weaker than observed. A sensible representation of the ENSO-NAO tropospheric teleconnection may be relevant to then simulate the effect of the NAO on Central American and northern South American rainfall (Giannini et al., 2000, 2004).

The rainfall anomalies associated with ENSO events are shown in Figure 11. Three regions in the AMS have a significant precipitation response to ENSO events in the observations and simulations. In southern North America, rainfall increases (decreases) during El Niño (La Niña) events due to the effect of Rossby waves on the subtropical jet and wintertime midlatitude disturbances (Vera et al., 2006; Bayr et al., 2019). The GPCP dataset (Figure 11a, b) shows significant boreal winter rainfall increases in southeastern US and the Gulf of Mexico during El Niño events, and an opposite response to La Niña phases. All the simulations reproduce this teleconnection rainfall pattern.

The anomalies in the Amazon show the strongest response to ENSO events in the observations. This teleconnection works through the coupling of ENSO with the Walker circulation (Vera et al., 2006; Cai et al., 2019), illustrated in Figure S3. Significant positive (negative) rainfall anomalies during the negative (positive) phase of ENSO in northern South America are observed in GPCP. All the simulations show a very similar and statistically significant response. The models also simulate the observed response in southeastern South America (SESA) of positive anomalies during El Niño and negative anomalies during La Niña events.

Figure 12 shows the observed and simulated precipitation responses in four regions of the AMS to different magnitudes of ENSO events, essentially showing the degree of linearity of ENSO teleconnections to the AMS. While the observed response shows some degree of linearity for El Niño events in South America (panels c, d), the majority of the observed responses, particularly to La Niña phases, are not linear. However, the simulations show several signs of linearity; for instance the historical experiments exhibited a linear response in precipitation to ENSO events in North America and SESA. However, some simulated responses, e.g. to La Niña phases in South America in the piControl simulations, show signs of non-linearity.

The different observed SST patterns for each ENSO event are a source of non-linearity of ENSO impacts over South America (Sulca et al., 2018; Cai et al., 2020). Principal component analysis has shown that ENSO events may be separated into two categories: Central Pacific (CP) and East Pacific (EP) events (Cai et al., 2020). Figure S4 shows that both UKESM1 and GC3 reasonably simulate the observed SST patterns associated with EP and CP El Niño events. Furthermore, Figure 13 compares the precipitation anomalies for each type of ENSO event in observations with three simulations: GC3 N96-pi, GC3 N216 and GC3 AMIP.

The observed precipitation response in the GPCC dataset to EP La Niña over equatorial South America is not significant and is smaller than the observed strong positive precipitation response to CP La Niña events in the same region. However, the simulated response in GC3 N96-pi and GC3 N216 during La Niña events appears to be independent of the type of event. In contrast, GC3 AMIP does show a positive, and significant, anomaly for CP La Niña events and weaker and not significant anomalies during EP events. The observed response to El Niño events in GPCC is also dependent on the type of event. EP

EL Niño events show significant negative anomalies over the Amazon and positive anomalies over SESA whereas CP events only show significant anomalies (-1 mm day^{-1}) over northeastern South America. While the coupled models (GC3 N96-pi and GC3 N216) do show a stronger response to EP EL Niño events than to CP events, the patterns of the response are very similar. In contrast, GC3 AMIP shows a very strong negative response to EP El Niño events in the Amazon but the response to CP events is much weaker and is only significant in northeastern South America. In other words, GC3 AMIP agrees well with the observed non-linear teleconnection patterns whereas the teleconnections in the coupled models do not depend on the type of ENSO event.

6 Summary and discussion

This study analysed results from the MOHC models, HadGEM3 and UKESM1 in their pre-industrial control, historical and AMIP experiment contributions to CMIP6. In particular, we focused on evaluating the modelled climate of the AMS comparing the effect of including Earth System processes or increasing resolution for representing regional rainfall. A schematic in Figure 14 shows the primary components of the AMS climate and summarizes the main biases in these simulations.

Rainfall in the North American Monsoon was particularly well simulated by the models. The seasonal cycle, peak monsoon rainfall rates and timings of monsoon onset and retreat in the simulations agreed well with TRMM. The historical experiments overestimate the mean temperature in most of the Americas by 1.5 K, but particularly in boreal summer in southwestern North America (+4 K). In spite of this warm bias, the temperature seasonal cycle is well represented by these models. These results suggest model improvement on the simulation of the North American Monsoon from previous versions of the MOHC models (Arritt et al., 2000), and most of the model cohorts of CMIP3 and CMIP5 (Geil et al., 2013). For example, most of CMIP5 models showed a very wet bias during monsoon maturity whereas rainfall during monsoon maturity in all the experiments of this study are within 1 mm day^{-1} of observations. However, these models continue to show biases during monsoon retreat as rainfall does not decrease as sharply as in observations after mid-September.

The Midsummer Drought (MSD) of southern Mexico and Central America is a regional feature of precipitation that most of CMIP5 models had difficulty capturing, except for instance for the MOHC models (Ryu and Hayhoe, 2014). The MSD in UKESM1 and GC3 continues to be relatively well represented, although with some differences in the timing and strength of the bimodal cycle. The models simulate a wetter-than-observed first peak of precipitation and a drier MSD period, therefore simulating a higher difference between the first peak and the dry period. The so-called second peak of precipitation found in late August is simulated in close agreement with TRMM, except in the AMIP experiment. Rainfall during the first peak has been too wet in these models since CMIP3, suggesting a persistent wet bias in this region associated with the East Pacific ITCZ (Ryu and Hayhoe, 2014; Mulcahy et al., 2018).

The East Pacific ITCZ migration and position was shown to be relatively well represented by the models (Figs. 4 and 5). However, the models showed an overestimation of boreal summer rainfall near the coast of Central America (Figure 8). These biases are associated with an easterly bias in the low-level wind, suggesting a bias in the flow from the Caribbean Sea into the Eastern Pacific which is relevant for moisture transport and controlling the SSTs (Herrera et al., 2015; Durán-Quesada et al., 2017). The simulations also showed a biased Atlantic ITCZ that was displaced south of the observed ITCZ position during boreal winter (Figure 5), particularly in the low resolution coupled simulations.

In the Amazon, the simulations showed a warm bias (+2 K) during austral spring and summer, a typical feature of previous models (Jones and Carvalho, 2013), and a colder than observed southeastern Brazil. These biases were linked with decreased cloud cover and less rainfall over the Amazon and more high clouds and rainfall in southeastern Brazil (Figures 7 and 9). The low cloud cover, warm and dry Amazon biases are intertwined with the low-level circulation from the Atlantic into the South American continent. The biases in the circulation during austral summer were observed as a northerly flow anomaly over the central and southern Amazon, a feature that has been associated with a stronger moisture transport away from the Amazon (Marengo et al., 2012; Jones and Carvalho, 2018). During the period of maximum mean rainfall rates in February, the simulations can overestimate rainfall by 3 mm day^{-1} in southeastern Brazil and underestimate rainfall in the Amazon by a similar rate. The historical experiments showed a small drying response to historical forcing in the Amazon therefore slightly increasing the magnitude of this dry bias. The AMIP simulation with the SST biases removed improved the Atlantic ITCZ representation and the precipitation, cloud cover and temperature biases over the South American Monsoon. The improvement in the circulation and precipitation biases in the AMIP simulation suggest that the origin of the dry Amazon bias are the biases in the Atlantic SSTs.

The canonical teleconnection responses of temperature, SLP and precipitation in the AMS to ENSO events are well represented in these models. The positive (negative) anomalies observed in northern Mexico and South Eastern South America during El Niño (La Niña) events are well represented in these experiments. Similarly, the teleconnection to the Amazon is well represented for both phases of ENSO, in spite of relevant biases in the region. ENSO teleconnections in these simulations were found to be approximately linear, i.e., the precipitation response is linearly related to the magnitude of the SST perturbation in the EN 3.4 region. In this model framework, positive and negative phases produce the opposite and equivalent precipitation response in the AMS. In contrast to observations and the GC3 AMIP simulation, the precipitation response in the coupled models appears to be independent of separating ENSO events into Central and East Pacific events. The fact that these models show a reasonable representation of ENSO diversity in SST patterns but the models do not replicate the observed non-linear dependance to ENSO events warrants further analysis.

The main biases are smaller in the medium resolution GC3 N216 compared to the low resolution experiments. In contrast, including Earth System processes in the UM model only affects the surface temperature response to historical forcing and not the dynamical biases that drive the precipitation and ITCZ biases. In short, the main dynamical biases in UKESM1 are very similar to those in GC3 N96 as these two models share the same dynamical core and only when resolution is increased are these biases reduced significantly. In spite of not improving the dynamic representation of the AMS, UKESM1 does show a stronger temperature response to forcing, as this model has a higher climate sensitivity than GC3 (Andrews et al., 2019; Sellar et al., 2019). A relevant difference between UKESM1 and GC3 is that warming over the historical period in Mexico and the Amazon is higher in UKESM1 than in GC3. This warming may be a consequence of the land-use change in these regions playing a role in the UKESM1 representation of soil-atmosphere feedbacks.

The improvement in the medium resolution simulation may largely be due to the improved dynamics of the ocean or the atmosphere. For example, the Atlantic ITCZ biases have been shown to be directly affected by processes in the convective scheme (Bellucci et al., 2010), such as the treatment of entraintment and moisture-cloud feedbacks (Oueslati and Bellon, 2013; Li and Xie, 2014). The resolution of the ocean model has been shown to impact the eddy heat flux parametrisation and the associated heat uptake and transport of the ocean (Kuhlbrodt et al., 2018). The improvement in the Atlantic SSTs and ITCZ and the associated dynamics also improves the associated circulation biases and moisture transport in the South American Monsoon. In other words, the oceanic resolution may play an important role in the cross-equatorial heat and moisture transport, SST gradients and the land-sea circulation over the Amazon during austral summer that is key for representing the geographic distribution of rainfall in South America.

These CMIP6 simulations of HadGEM3 and UKESM1 show some signs of model improvement, particularly in the North American Monsoon and may be used to better understand the response to current and future response to anthropogenic forcing. Furthermore, several aspects of the climate of the AMS that are well simulated by these models, such as a good representation of the MSD and a resonable representation of ENSO diversity, may suggest further use of these simulations to address outstanding questions of climate variability in this region across different temporal scales.

Author contributions. JLGF conducted the analyses, LJG and SO directed the research. All authors were fully involved in the revisions and the preparation of the paper.

Competing interests. The authors declare that there are no competing interests.

Data availability. ERA5 data was made available by Copernicus at https://cds.climate.copernicus.eu whereas the model data is available in the CMIP6 Earth System Grid Federation (ESGF) at https://esgf-index1.ceda.ac.uk/projects/cmip6-ceda/.

Acknowledgements. JLGF was funded by an Oxford-Richards Scholarship.The authors thank David Adams and one anonymous reviewer for their very helpful comments that have improved this manuscript. The authors would also like to thank Robin Chadwick, Tim Woollings, Antje Weisheimer and Gabriel Martins Palma Perez for their very helpful suggestions and useful discussions.

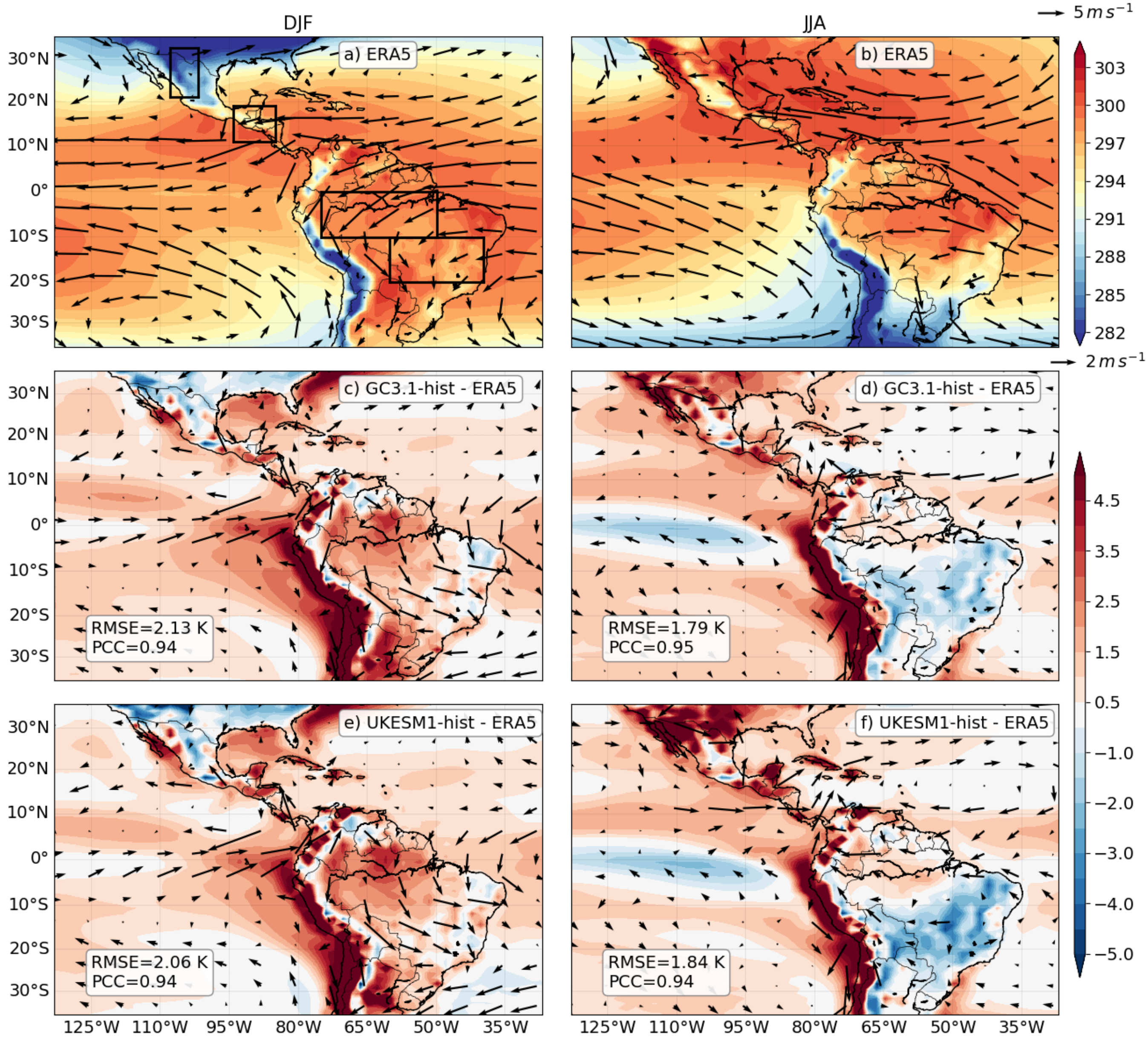

Figure 1. (a, b) Temperature (color-contours in K) and wind speed (vectors) at 850 hPa DJF and JJA climatogies in ERA5. The biases are shown as the differences between the ensemble mean from the historical experiment of (c, d) GC3 and (e, f) UKESM1 and ERA5. The climatogies and biases are shown for (a, c, e) boreal winter (DJF) and (b, d, f) boreal summer (JJA). Only differences statistically significant to the 95% level are shown, according to a Welch t-test for each field. The key for the size of the wind vectors is shown in the top right corner of panels b) and d). The root-mean square error (RMSE) and pattern correlation coefficient (PCC) are shown on the bottom left of c-f.

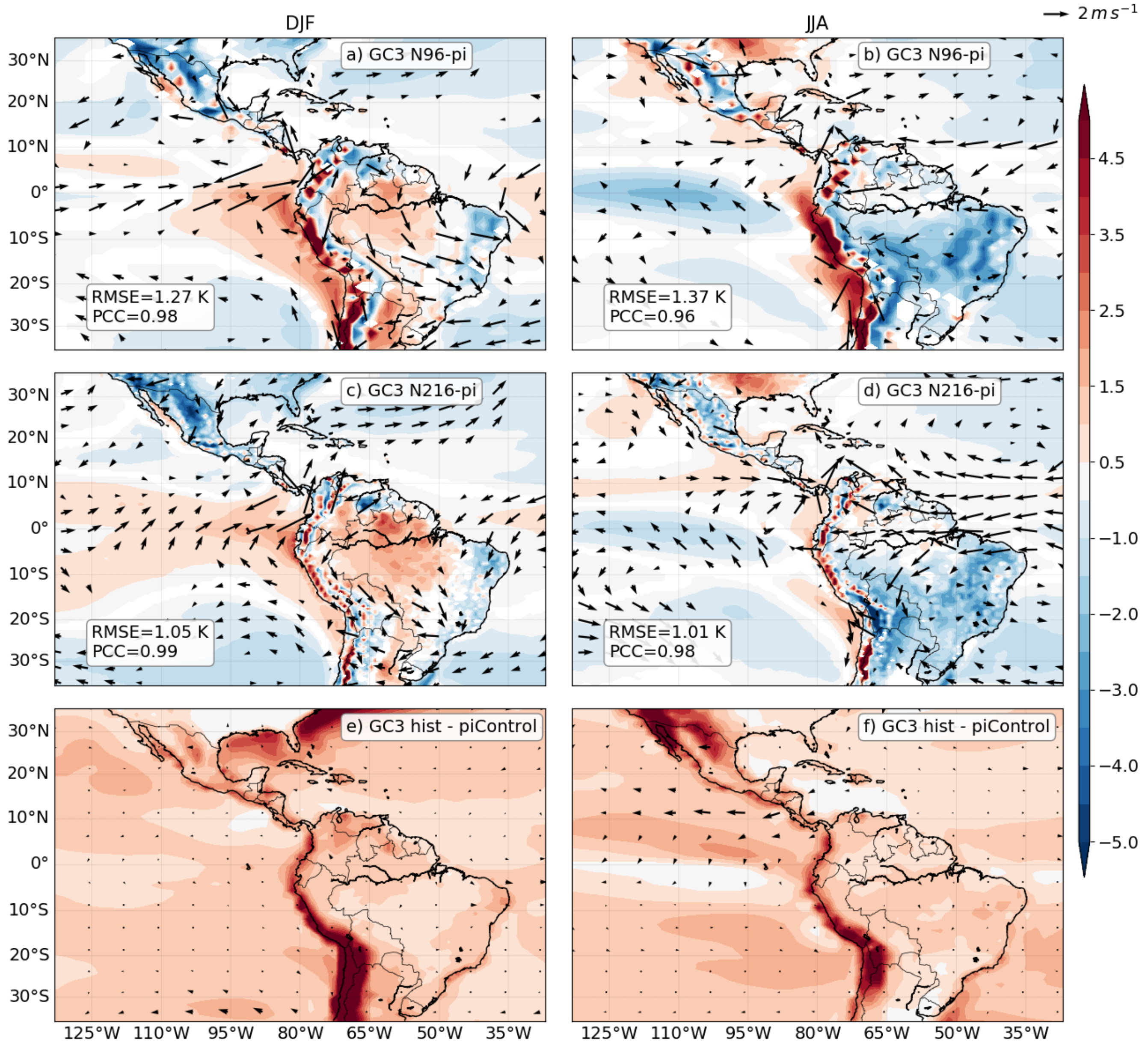

Figure 2. As in Figure 1, but showing the differences between the piControl simulations of (a, b) GC3 N96-pi and (c, d) GC3 N216-pi, and ERA5. (e, f) show the statistically significant differences between the historical (1979-2014) and piControl experiments of GC3. The RMSE and PCC are shown on the bottom left of a-d.

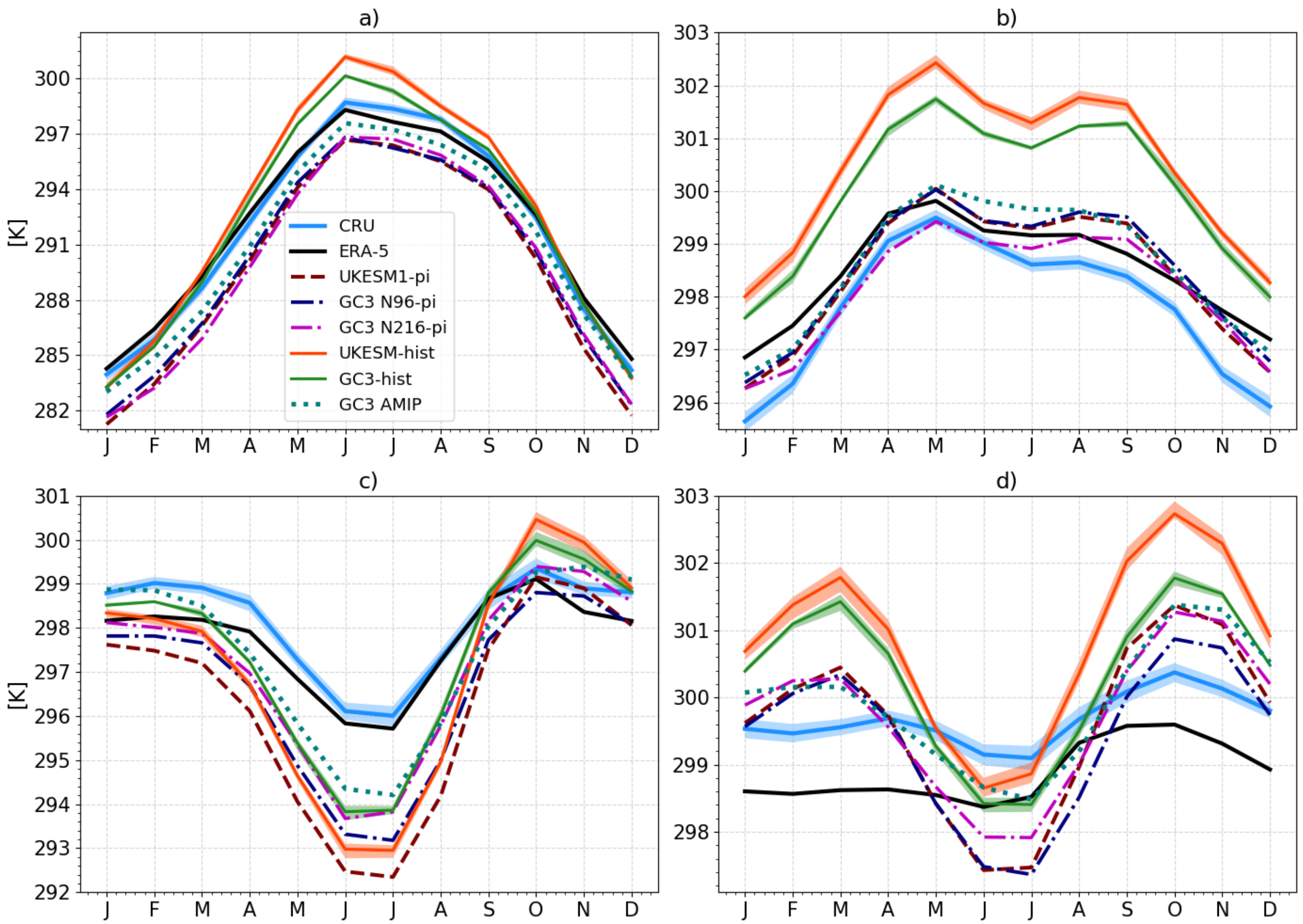

Figure 3. Monthly-mean temperature in the (a) North American Monsoon [19-35°N,110-103°W], (b) the Midsummer drought [11-19°N,95-85°W] (c) Eastern Brazil [20-10°S,60-40°W] and (d) the Amazon basin [-10-0°S,75-50°W] regions. The shadings for the CRU dataset represents the observational uncertainties and for the historical simulations the shading is the ensemble spread. The regions for this plot are shown in Figure 1a.

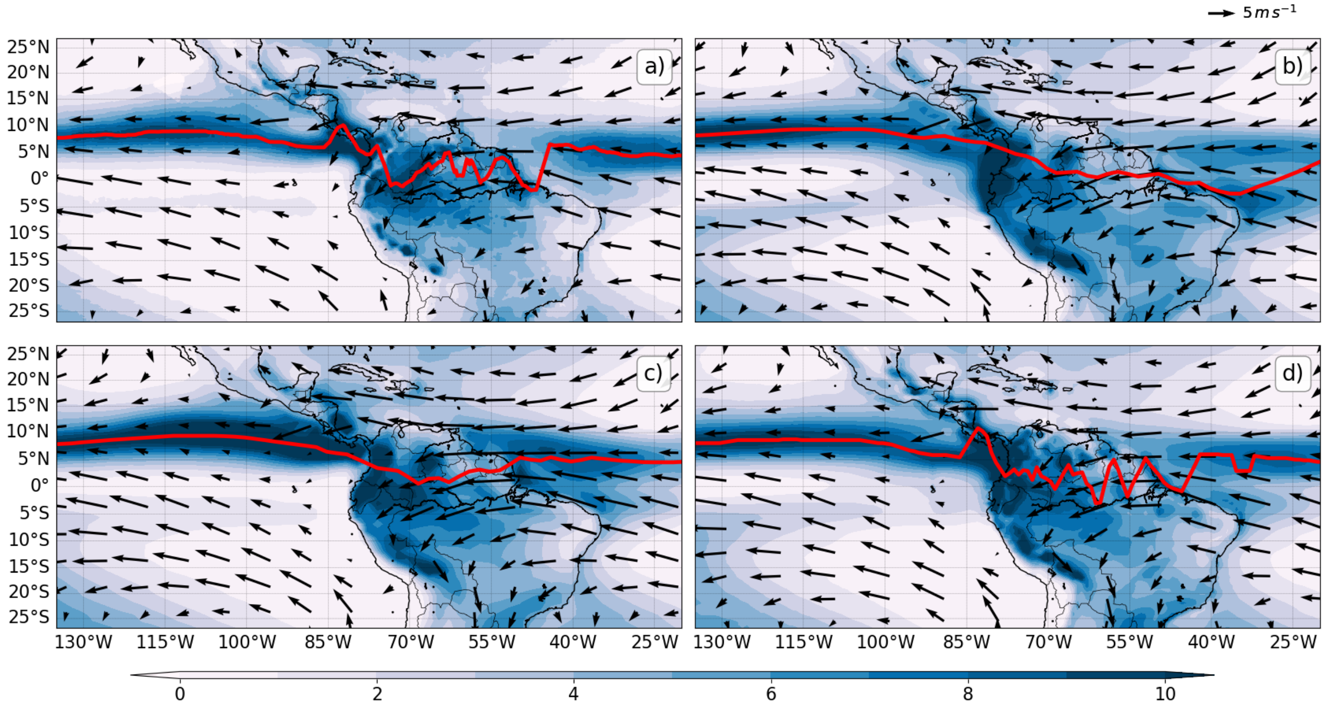

Figure 4. Climatological rainfall [mm day^{-1}] and low-level wind speed (850-hPa) in (a) TRMM and ERA-5, (b) the ensemble-mean UKESM-historical, (c) GC3 AMIP and (d) GC3 N216-pi. The red line highlights the maximum rainfall for each longitude as a proxy for the position of the ITCZ.

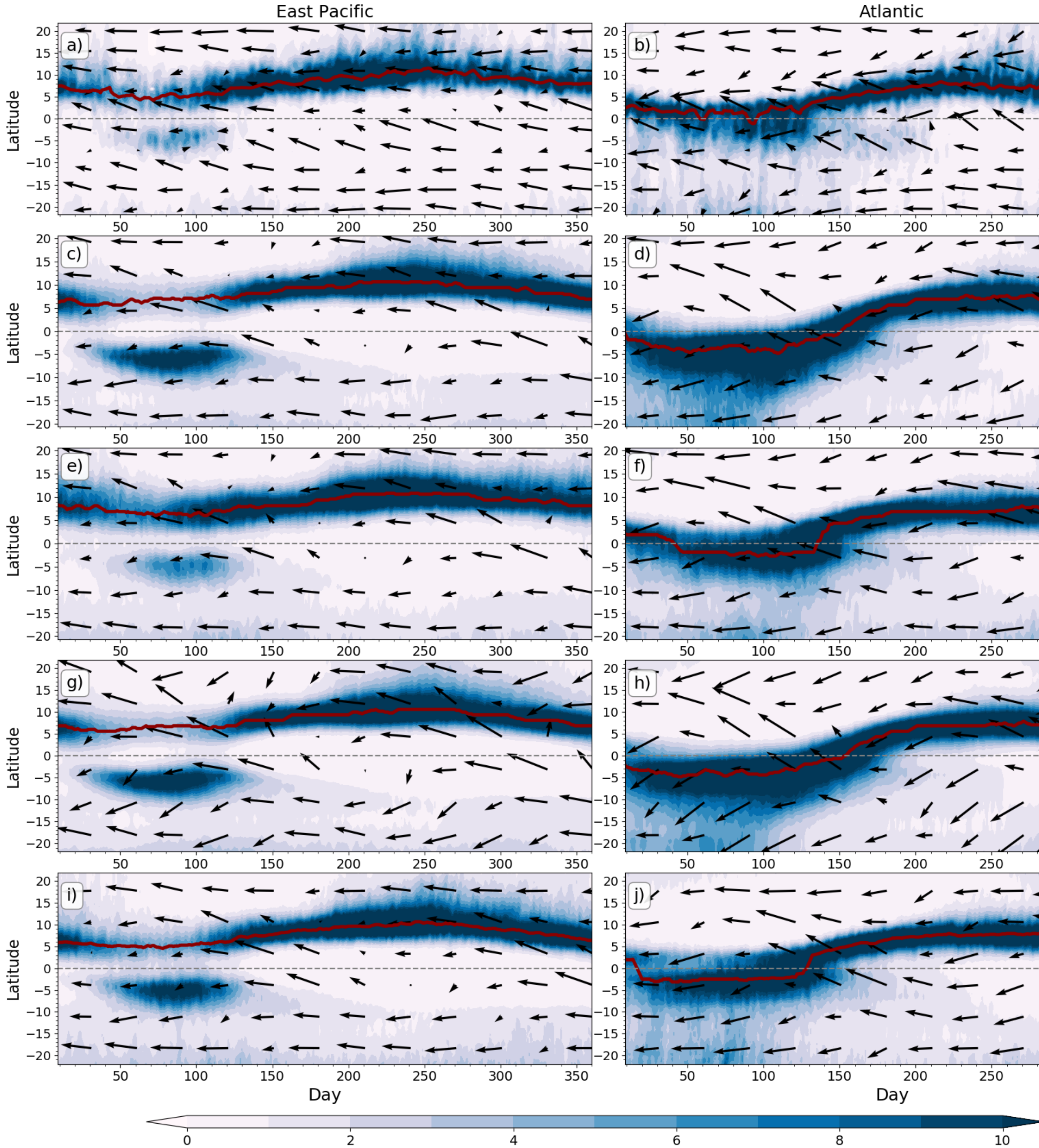

Figure 5. Time-Latitude plot of daily mean rainfall (colour contours) and low-level wind speed (850 hPa) longitudinally averaged over the (a, c, e, g) East Pacific [150°W-100°W] and (b, d, f, h) Atlantic [40°W-20°W] Oceans. (a, b) show rainfall from TRMM and winds from ERA-5, (c, d) the ensemble-mean UKESM-historical, (e, f) GC3 AMIP, (g, h) N96-pi and (i, j) GC3 N216-pi. The red solid line shows the ITCZ as the latitude of maximum precipitation.

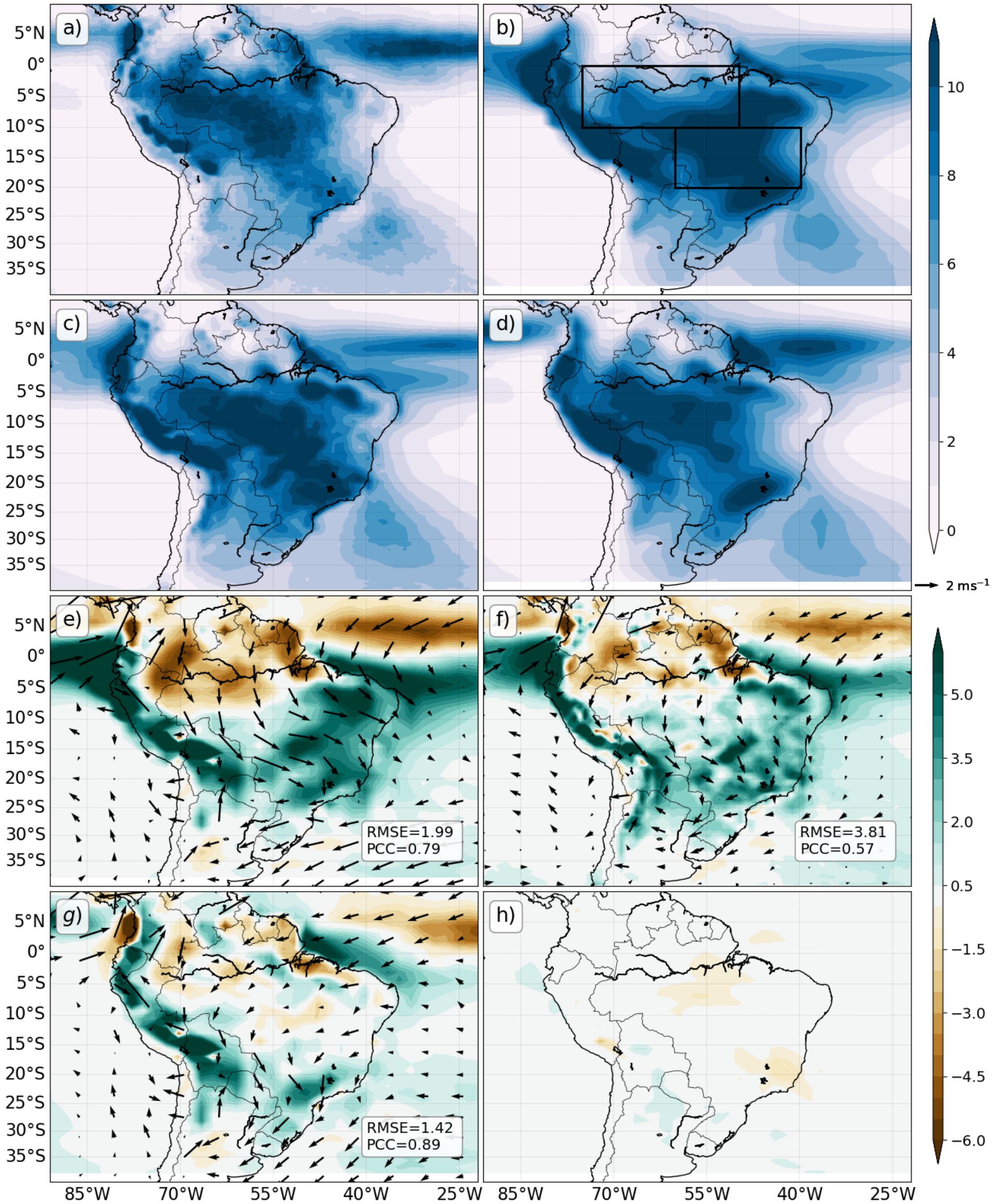

Figure 6. DJF mean rainfall [mm day^{-1}] from (a) TRMM, (b) UKESM1-historical, (c) GC3 n216 and (d) GC3 AMIP. (e, f, g) show the statistically significant differences between panels (b, c ,d) and (a) TRMM, respectively. (h) Precipitation difference between UKESM-historical and UKESM1-pi, only statistically significant differences (95%) confidence level is shown.

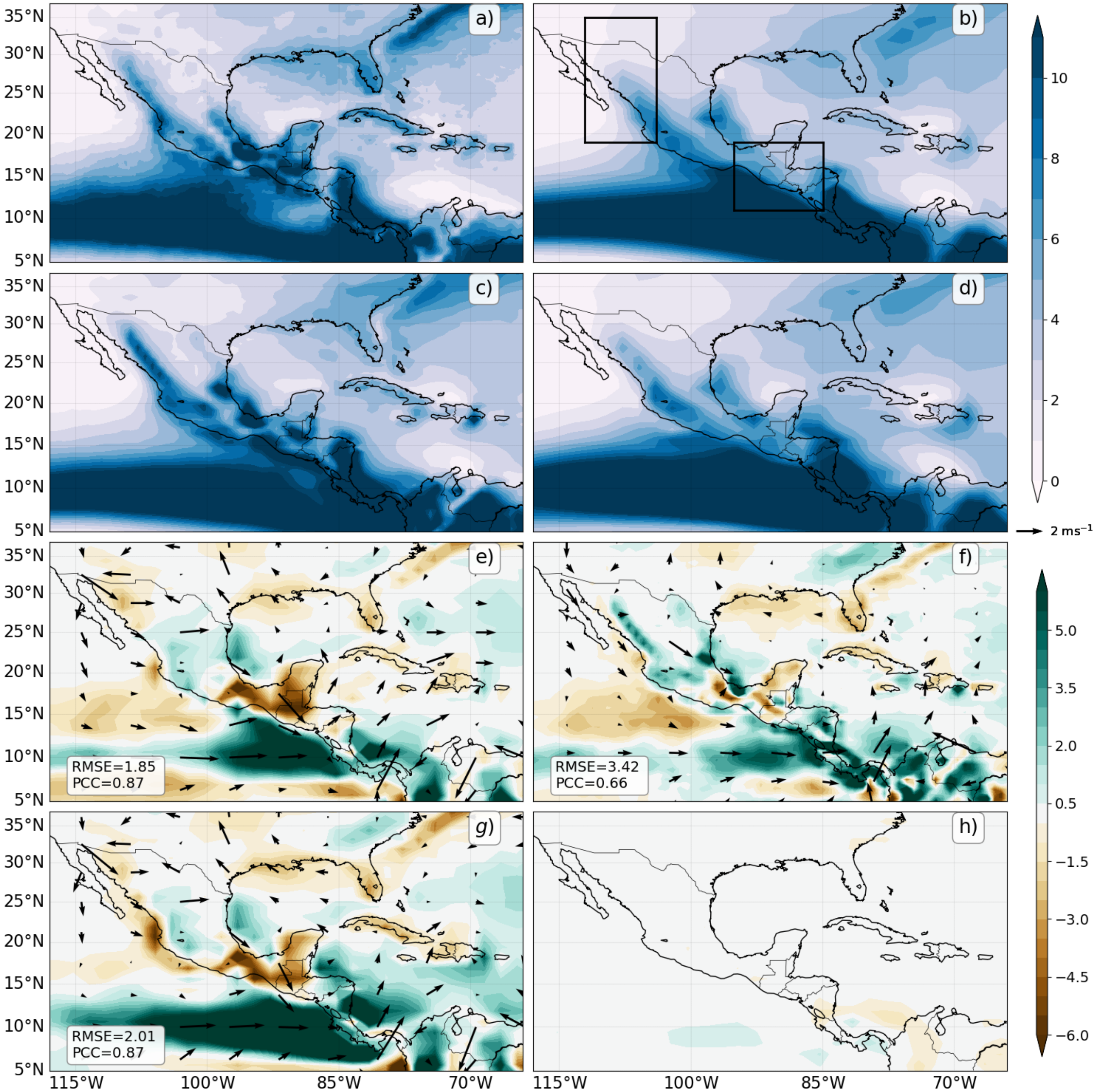

Figure 7. As in Figure 6 but for JJA in the northern part of subtropical America.

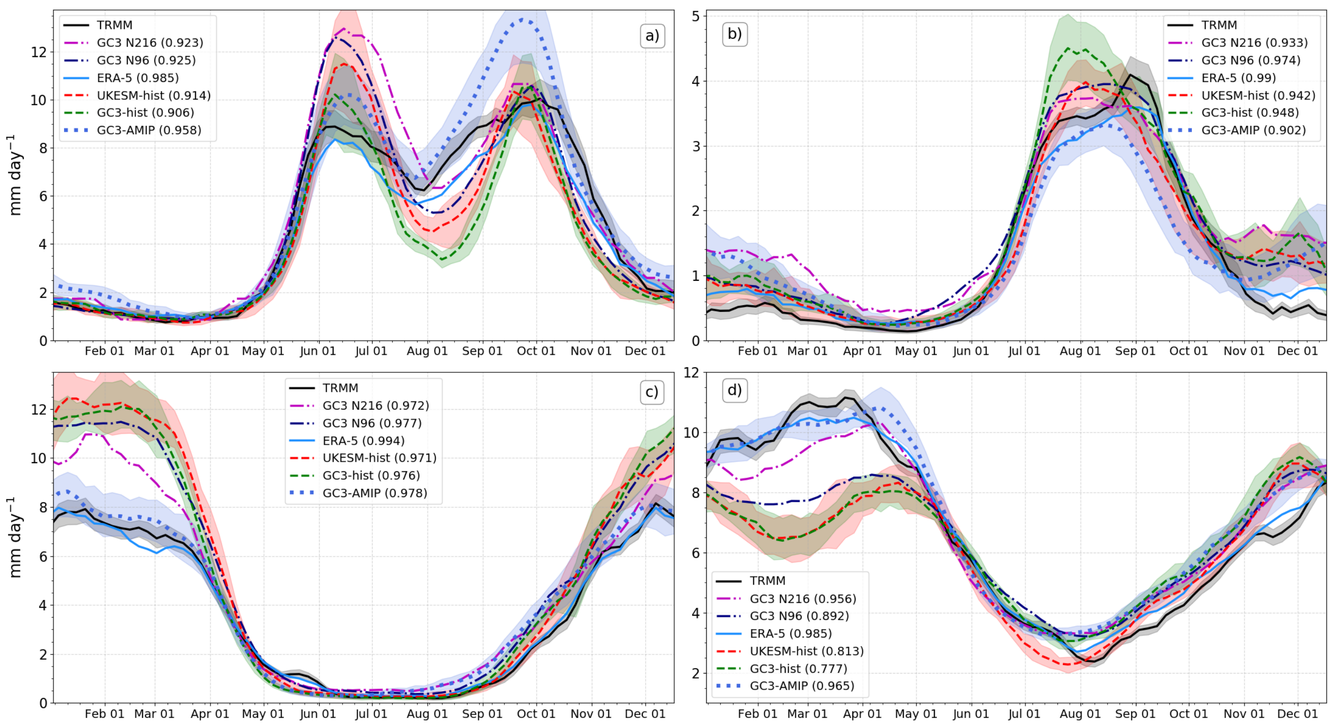

Figure 8. Annual cycle of pentad-mean rainfall in the regions (a) the Midsummer drought, (b) the North American Monsoon, (c) Eastern Brazil and (d) the Amazon Basin. The regions are defined as in Figure 3 and are illustrated in Figure 7b and Figure 8b. The shaded regions represent observational uncertainty for TRMM and ensemble spread for the historical experiments. The correlation coefficient for each of the model-driven seasonal cycles with TRMM is given in brackets in each panel.

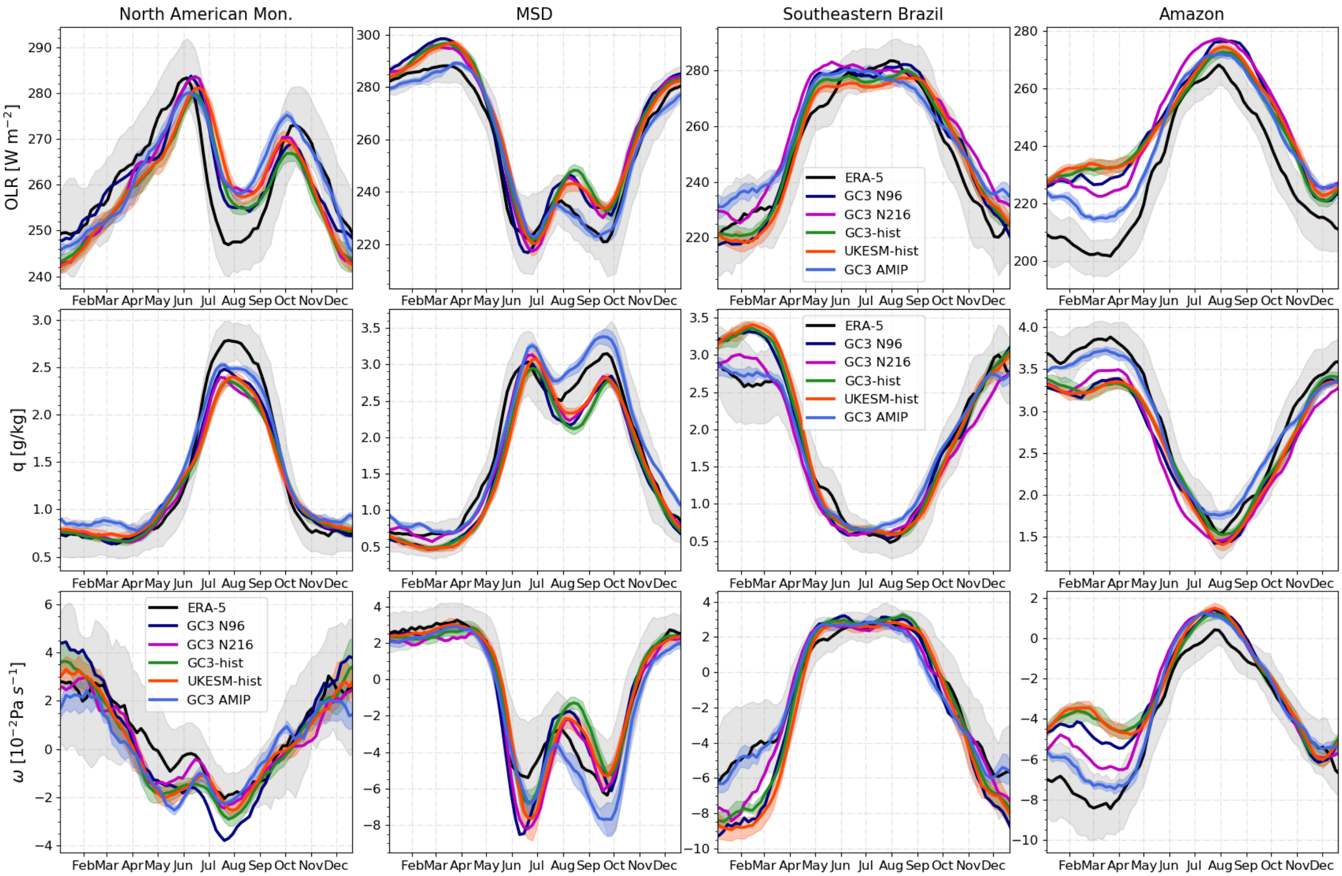

Figure 9. Pentad-mean (upper) out-going longwave radiation (OLR), (middle) specific humidity at 500-hPa and (lower) ω 500-hPa. These are shown from left to right for the North American Monsoon, the Midsummer drought, southastern Brazil and the core Amazon. The uncertainty in ERA-5 data, shown as faint gray shading was estimating by bootstrapping with replacement the ERA-5 record 10,000 times.

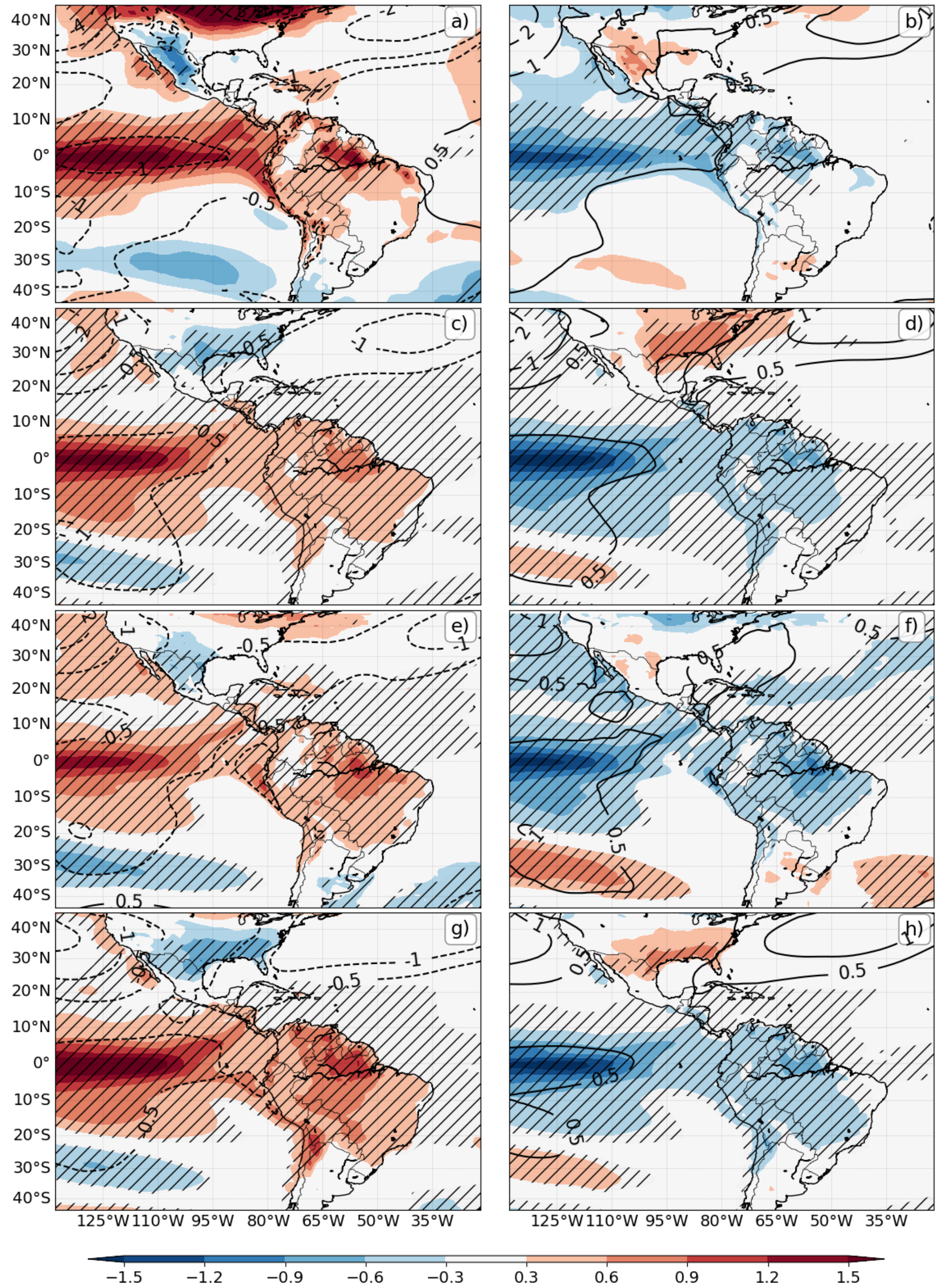

Figure 10. DJF Temperature anomalies (colour contours in K) and SLP (line contours in hPa) during (a, c, e, g) El Niño and (b, d, f, h) La Niña events. Results are shown for (a, b) ERA-5, (c, d) UKESM1-historical, (e, f) GC3 N96-pi and (g, h) GC3 N216-pi. The hatched regions denote 99% significance from a Welch t-test for the temperature field.

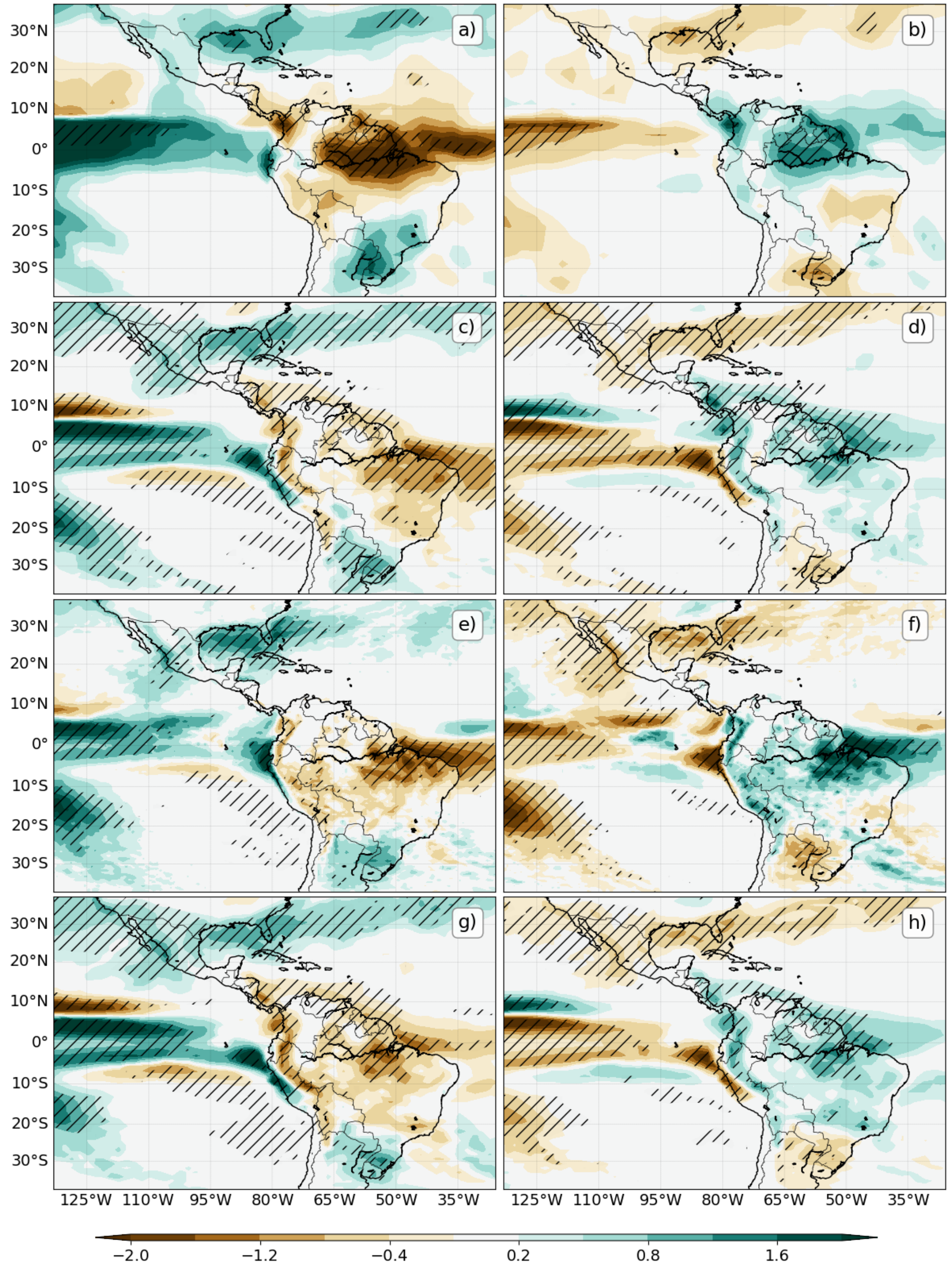

Figure 11. As in Figure 10 but for the rainfall response [mm day^{-1}] using GPCP as the observational dataset.

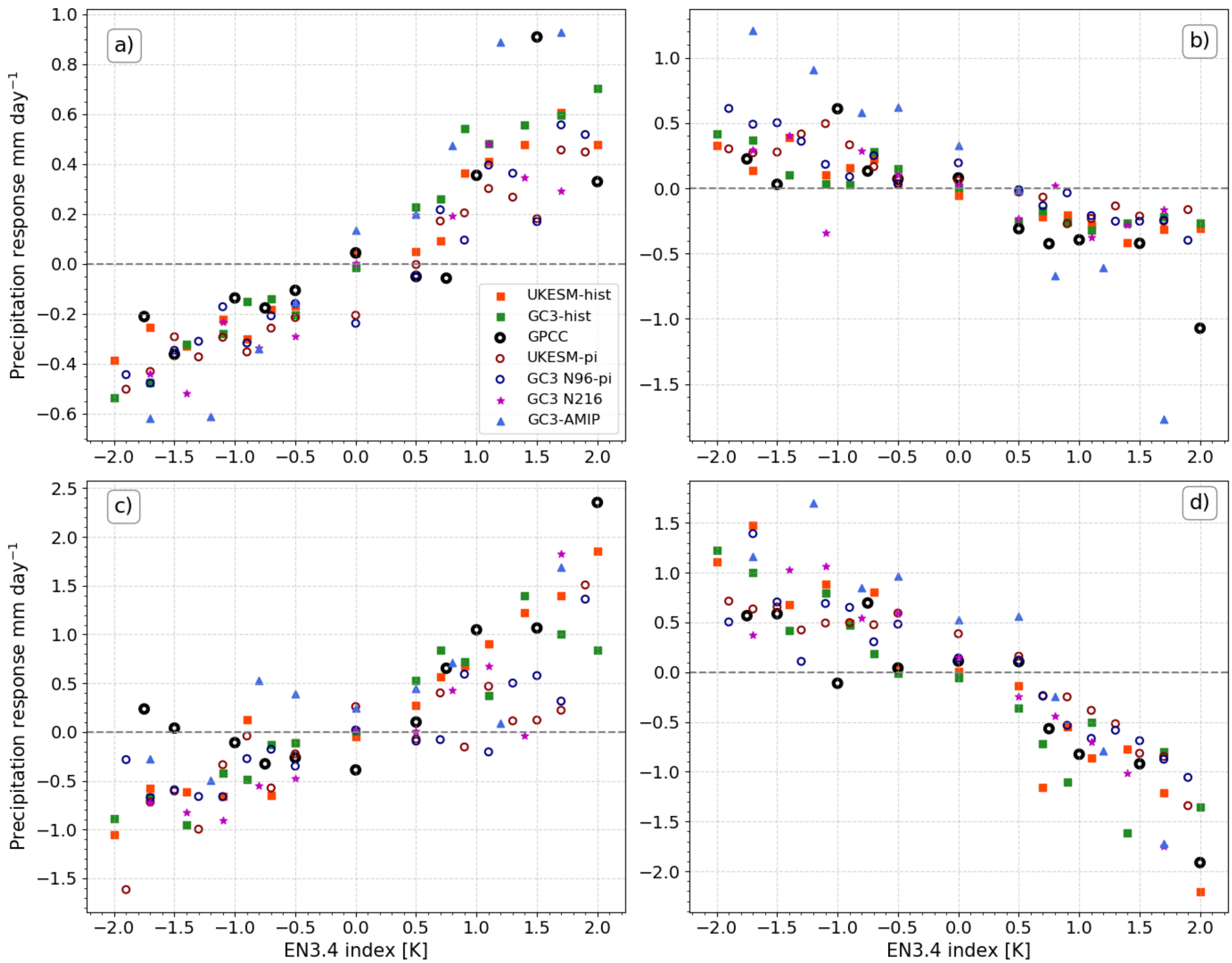

Figure 12. Precipitation response [mm day^{-1}] as a function of the El Niño 3.4 index (see text) for (a) southwestern North America [20-37°N, 112-98°W], (b) Central America and southern Mexico [5-19°N, 95-83°W],, (c) South Eastern South America [35-25°S, 60-50°W], and (d) the Amazon [10-0°S, 70-45°W]. The observation scatter points are from GPCC in the period of 1940-2013.

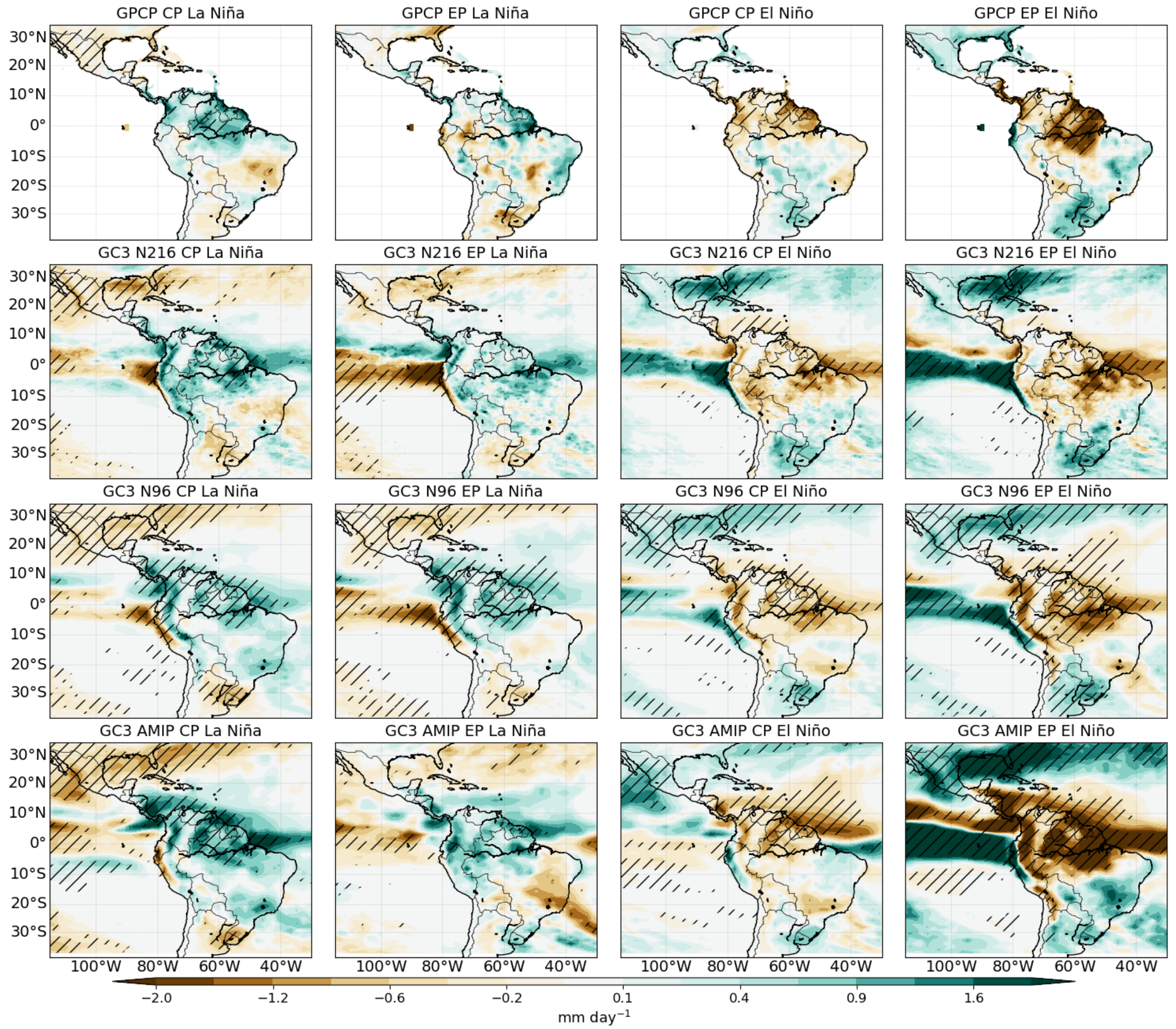

Figure 13. Precipitation anomalies in GPCC 1940-2013, GC3 N216, GC3 N96-pi and GC3 AMIP for the four different types of ENSO events, as defined by Cai et al. (2020). Statistically significant anomalies (95% confidence level) are hatched.

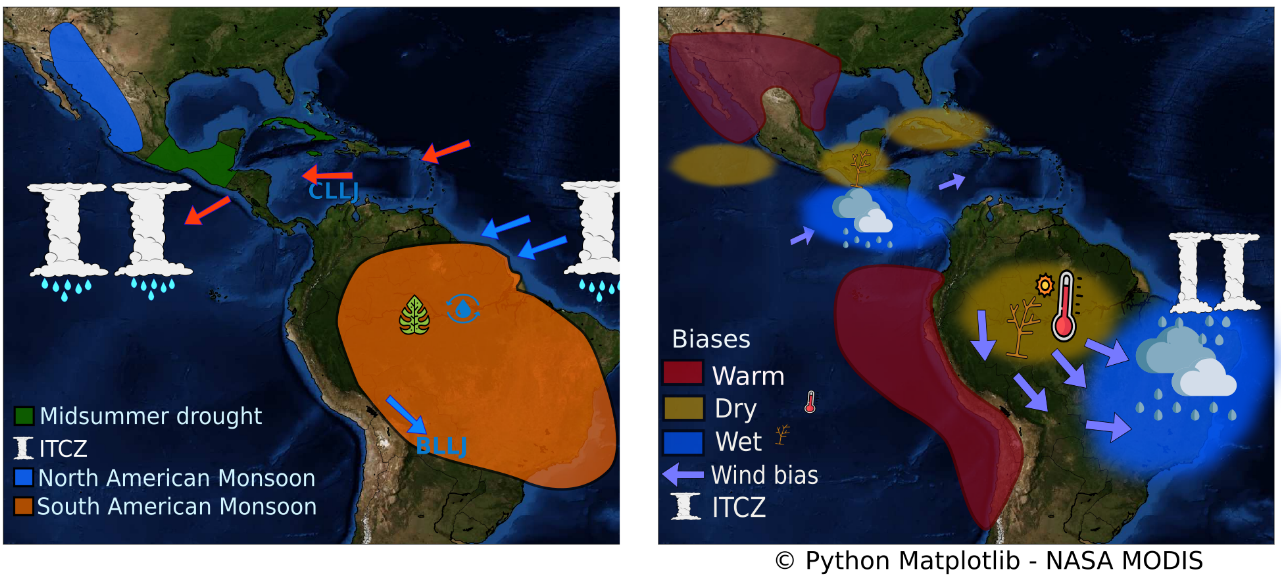

Figure 14. Schematics of (left) the main features in the AMS and (right) the main biases in UKESM1 and HadGEM3. In (a) the boreal summer easterlies (red) and austral summer circulation (blue) are shown with the Caribbean and Bolivian Low-level Jets (CLLJ and BLLJ, respectively). In (b) the biases are shown for the respective northern and southern Hemisphere summers. The ITCZ bias in (b) refers to the southward displacement bias of the Atlantic ITCZ in the simulations.

References

Adams, D. K. and Comrie, A. C.: The north American monsoon, Bulletin of the American Meteorological Society, 78, 2197–2214, 1997.

Adams, D. K., Gutman, S. I., Holub, K. L., and Pereira, D. S.: GNSS observations of deep convective time scales in the Amazon, Geophysical Research Letters, 40, 2818–2823, https://doi.org/10.1002/grl.50573, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50573, 2013.

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present), Journal of hydrometeorology, 4, 1147–1167, 2003.

Amador, J. A., Alfaro, E. J., Lizano, O. G., and Magaña, V. O.: Atmospheric forcing of the eastern tropical Pacific: A review, Progress in Oceanography, 69, 101–142, 2006.

Amador, J. A., Durán-Quesada, A., Rivera, E., Mora, G., Sáenz, F., Calderón, B., and Mora, N.: The easternmost tropical Pacific. Part II: Seasonal and intraseasonal modes of atmospheric variability, Revista de Biología Tropical, pp. 23–57, 2016.

Andrews, M., Ridley, J., Wood, R., Andrews, T., Blockley, E., Booth, B., Burke, E., Dittus, A., Florek, P., Gray, L., Haddad, S., Hardiman, S., Hermanson, L., Hodson, D., Hogan, E., Jones, G., Knight, J., Kuhlbrodt, T., Misios, S., and Sutton, R.: Historical simulations with HadGEM3-GC3.1 for CMIP6, Journal of Advances in Modeling Earth Systems, https://doi.org/10.1029/2019MS001995, 2020.

Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., Menary, M. B., Ridley, J., Ringer, M. A., Sellar, A. A., Senior, C. A., and Tang, Y.: Forcings, feedbacks, and climate sensitivity in HadGEM3-GC3. 1 and UKESM1, Journal of Advances in Modeling Earth Systems, 11, 4377–4394, 2019.

Arritt, R. W., Goering, D. C., and Anderson, C. J.: The North American monsoon system in the Hadley Centre coupled ocean-atmosphere GCM, Geophysical Research Letters, 27, 565–568, 2000.

Bayr, T., Domeisen, D. I., and Wengel, C.: The effect of the equatorial Pacific cold SST bias on simulated ENSO teleconnections to the North Pacific and California, Climate Dynamics, pp. 1–19, 2019.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC full data reanalysis version 6.0 at 1.0: monthly land-surface precipitation from rain-gauges built on GTS-based and historic data, Global Precipitation Climatology Centre (GPCC): Berlin, Germany, 2011.

Bellucci, A., Gualdi, S., and Navarra, A.: The double-ITCZ syndrome in coupled general circulation models: The role of large-scale vertical circulation regimes, Journal of Climate, 23, 1127–1145, 2010.

Bombardi, R. J. and Carvalho, L. M.: The South Atlantic dipole and variations in the characteristics of the South American Monsoon in the WCRP-CMIP3 multi-model simulations, Climate Dynamics, 36, 2091–2102, 2011.

C3S: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate., Copernicus Climate Change Service Climate Data Store (CDS), date of access: 18 March 2019, https://climate.copernicus.eu/climate-reanalysis, 2017.

Cai, W., Wu, L., Lengaigne, M., Li, T., McGregor, S., Kug, J.-S., Yu, J.-Y., Stuecker, M. F., Santoso, A., Li, X., Ham, Y.-G., Chikamoto, Y., Ng, B., McPhaden, M. J., Du, Y., Dommenget, D., Jia, F., Kajtar, J. B., Keenlyside, N., Lin, X., Luo, J.-J., Martín-Rey, M., Ruprich-Robert, Y., Wang, G., Xie, S.-P., Yang, Y., Kang, S. M., Choi, J.-Y., Gan, B., Kim, G.-I., Kim, C.-E., Kim, S., Kim, J.-H., and Chang, P.: Pantropical climate interactions, Science, 363, eaav4236, 2019.

Cai, W., McPhaden, M., Grimm, A., Rodrigues, R., Taschetto, A., Garreaud, R., Dewitte, B., Poveda, G., Ham, Y.-G., Santoso, A., Ng, B., Anderson, W., Wang, G., Geng, T., Jo, H., Marengo, J., Alves, L., Osman, M., Li, S., and Vera, C.: Climate impacts of the El Niño–Southern Oscillation on South America, Nature Reviews Earth & Environment, 1, 215–231, 2020.

Carvalho, L. M., Jones, C., and Liebmann, B.: The South Atlantic convergence zone: Intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall, Journal of Climate, 17, 88–108, 2004.

Carvalho, L. M. V., Jones, C., Posadas, A. N. D., Quiroz, R., Bookhagen, B., and Liebmann, B.: Precipitation characteristics of the South American Monsoon System derived from multiple datasets, Journal of Climate, 25, 4600–4620, https://doi.org/10.1175/JCLI-D-11-00335.1, https://doi.org/10.1175/JCLI-D-11-00335.1, 2012.

Chuvieco, E., Opazo, S., Sione, W., Valle, H. d., Anaya, J., Bella, C. D., Cruz, I., Manzo, L., López, G., Mari, N., González-Alonso, F., Morelli, F., Setzer, A., Csiszar, I., Kanpandegi, J. A., Bastarrika, A., and Libonati, R.: Global burned-land estimation in Latin-America using MODIS composite data, Ecological Applications, 18, 64–79, https://doi.org/10.1890/06-2148.1, https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/06-2148.1, 2008.

Colorado-Ruiz, G., Cavazos, T., Salinas, J. A., De Grau, P., and Ayala, R.: Climate change projections from Coupled Model Intercomparison Project phase 5 multi-model weighted ensembles for Mexico, the North American monsoon, and the mid-summer drought region, International Journal of Climatology, 38, 5699–5716, 2018.

Dilley, M.: Synoptic controls on precipitation in the Valley of Oaxaca, Mexico, International Journal of Climatology: A Journal of the Royal Meteorological Society, 16, 1019–1031, 1996.

Dinku, T., Ruiz, F., Connor, S. J., and Ceccato, P.: Validation and intercomparison of satellite rainfall estimates over Colombia, Journal of Applied Meteorology and Climatology, 49, 1004–1014, 2010.

Douglas, M. W., Maddox, R. A., Howard, K., and Reyes, S.: The mexican monsoon, Journal of Climate, 6, 1665–1677, 1993.

Durán-Quesada, A. M., Gimeno, L., and Amador, J.: Role of moisture transport for Central American precipitation, Earth System Dynamics, 8, 147–161, https://doi.org/10.5194/esd-8-147-2017, https://www.earth-syst-dynam.net/8/147/2017/, 2017.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development (Online), 9, 2016.

Fisher, R. A.: Statistical methods for research workers, in: Breakthroughs in statistics, pp. 66–70, Springer, 1992.

Franchito, S. H., Rao, V. B., Vasques, A. C., Santo, C. M., and Conforte, J. C.: Validation of TRMM precipitation radar monthly rainfall estimates over Brazil, Journal of Geophysical Research: Atmospheres, 114, 2009.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Scientific data, 2, 150 066, 2015.

Gamble, D. W., Parnell, D. B., and Curtis, S.: Spatial variability of the Caribbean mid-summer drought and relation to north Atlantic high circulation, International Journal of Climatology, 28, 343–350, 2008.

Gan, M., Kousky, V., and Ropelewski, C.: The South America monsoon circulation and its relationship to rainfall over west-central Brazil, Journal of climate, 17, 47–66, 2004.

Geil, K. L., Serra, Y. L., and Zeng, X.: Assessment of CMIP5 model simulations of the North American monsoon system, Journal of Climate, 26, 8787–8801, 2013.

Giannini, A., Kushnir, Y., and Cane, M. A.: Interannual variability of Caribbean rainfall, ENSO, and the Atlantic Ocean, Journal of Climate, 13, 297–311, 2000.

Giannini, A., Saravanan, R., and Chang, P.: The preconditioning role of tropical Atlantic variability in the development of the ENSO teleconnection: Implications for the prediction of Nordeste rainfall, Climate Dynamics, 22, 839–855, 2004.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset, International journal of climatology, 34, 623–642, 2014.

Herrera, E., Magaña, V., and Caetano, E.: Air–sea interactions and dynamical processes associated with the midsummer drought, International Journal of Climatology, 35, 1569–1578, 2015.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Nelkin, E. J.: The TRMM multi-satellite precipitation analysis (TMPA), in: Satellite rainfall applications for surface hydrology, pp. 3–22, Springer, 2010.

Jakob, C.: Going back to basics, Nature Climate Change, 4, 1042, 2014.

Jiménez-Esteve, B. and Domeisen, D. I. V.: Nonlinearity in the tropospheric pathway of ENSO to the North Atlantic, Weather and Climate Dynamics, 1, 225–245, https://doi.org/10.5194/wcd-1-225-2020, http://www.weather-clim-dynam.net/1/225/2020/, 2020.

Jones, C. and Carvalho, L. M.: Active and break phases in the South American monsoon system, Journal of Climate, 15, 905–914, 2002.

Jones, C. and Carvalho, L. M.: Climate change in the South American monsoon system: present climate and CMIP5 projections, Journal of Climate, 26, 6660–6678, 2013.

Jones, C. and Carvalho, L. M.: The influence of the Atlantic multidecadal oscillation on the eastern Andes low-level jet and precipitation in South America, npj Climate and Atmospheric Science, 1, 1–7, 2018.

Kennedy, J., Rayner, N., Smith, R., Parker, D., and Saunby, M.: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization, Journal of Geophysical Research: Atmospheres, 116, 2011.

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., Hill, R., Graham, T., Ridley, J., Blaker, A., Calvert, D., Copsey, D., Ellis, R., Hewitt, H., Hyder, P., Ineson, S., Mulcahy, J., Siahaan, A., and Walton, J.: The low-resolution version of HadGEM3 GC3. 1: Development and evaluation for global climate, Journal of advances in modeling earth systems, 10, 2865–2888, 2018.

Lahmers, T. M., Castro, C. L., Adams, D. K., Serra, Y. L., Brost, J. J., and Luong, T.: Long-term changes in the climatology of transient inverted troughs over the North American monsoon region and their effects on precipitation, Journal of Climate, 29, 6037–6064, 2016.

Lawrence, P. J., Feddema, J. J., Bonan, G. B., Meehl, G. A., O'Neill, B. C., Oleson, K. W., Levis, S., Lawrence, D. M., Kluzek, E., Lindsay, K., and Thornton, P. E.: Simulating the Biogeochemical and Biogeophysical Impacts of Transient Land Cover Change and Wood Harvest in the Community Climate System Model (CCSM4) from 1850 to 2100, Journal of Climate, 25, 3071–3095, https://doi.org/10.1175/JCLI-D-11-00256.1, 2012.

Lee, J.-E., Lintner, B. R., Boyce, C. K., and Lawrence, P. J.: Land use change exacerbates tropical South American drought by sea surface temperature variability, Geophysical Research Letters, 38, 2011.

Li, G. and Xie, S.-P.: Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems, Journal of Climate, 27, 1765–1780, 2014.

Li, W., Fu, R., and Dickinson, R. E.: Rainfall and its seasonality over the Amazon in the 21st century as assessed by the coupled models for the IPCC AR4, Journal of Geophysical Research: Atmospheres, 111, 2006.

Machado, L., Laurent, H., Dessay, N., and Miranda, I.: Seasonal and diurnal variability of convection over the Amazonia: A comparison of different vegetation types and large scale forcing, Theoretical and Applied Climatology, 78, 61–77, 2004.

Magaña, V., Amador, J. A., and Medina, S.: The midsummer drought over Mexico and Central America, Journal of Climate, 12, 1577–1588, 1999.

Malhi, Y., Aragão, L. E. O. C., Galbraith, D., Huntingford, C., Fisher, R., Zelazowski, P., Sitch, S., McSweeney, C., and Meir, P.: Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest, Proceedings of the National Academy of Sciences, 106, 20 610–20 615, https://doi.org/10.1073/pnas.0804619106, https://www.pnas.org/content/106/49/20610, 2009.

Marengo, J. A., Liebmann, B., Kousky, V. E., Filizola, N. P., and Wainer, I. C.: Onset and end of the rainy season in the Brazilian Amazon Basin, Journal of Climate, 14, 833–852, 2001.

Marengo, J. A., Liebmann, B., Grimm, A. M., Misra, V., Silva Dias, P. L., Cavalcanti, I. F. A., Carvalho, L. M. V., Berbery, E. H., Ambrizzi, T., Vera, C. S., Saulo, A. C., Nogues-Paegle, J., Zipser, E., Seth, A., and Alves, L. M.: Recent developments on the South American monsoon system, International Journal of Climatology, 32, 1–21, 2012.

Marotzke, J., Jakob, C., Bony, S., Dirmeyer, P., O'Gorman, P., Hawkins, E., Perkins-Kirkpatrick, S., Quéré, C., Nowicki, S., Paulavets, K., Seneviratne, S., Stevens, B., and Tuma, M.: Climate research must sharpen its view, Nature climate change, 7, 89–91, 2017.

Menary, M. B., Kuhlbrodt, T., Ridley, J., Andrews, M. B., Dimdore-Miles, O. B., Deshayes, J., Eade, R., Gray, L., Ineson, S., Mignot, J., Roberts, C. D., Robson, J., Wood, R. A., and Xavier, P.: Preindustrial Control Simulations With HadGEM3-GC3. 1 for CMIP6, Journal of Advances in Modeling Earth Systems, 10, 3049–3075, 2018.

Mulcahy, J. P., Jones, C., Sellar, A., Johnson, B., Boutle, I. A., Jones, A., Andrews, T., Rumbold, S. T., Mollard, J., Bellouin, N., Johnson, C. E., Williams, K. D., Grosvenor, D. P., and McCoy, D. T.: Improved aerosol processes and effective radiative forcing in HadGEM3 and UKESM1, Journal of Advances in Modeling Earth Systems, 10, 2786–2805, 2018.

Nieto-Ferreira, R. and Rickenbach, T. M.: Regionality of monsoon onset in South America: a three-stage conceptual model, International Journal of Climatology, 31, 1309–1321, 2011.

Ordoñez, P., Nieto, R., Gimeno, L., Ribera, P., Gallego, D., Ochoa-Moya, C. A., and Quintanar, A. I.: Climatological moisture sources for the Western North American Monsoon through a Lagrangian approach: their influence on precipitation intensity, Earth System Dynamics, 10, 59–72, 2019.

Oueslati, B. and Bellon, G.: Convective entrainment and large-scale organization of tropical precipitation: Sensitivity of the CNRM-CM5 hierarchy of models, Journal of climate, 26, 2931–2946, 2013.

Palmer, T.: Stochastic weather and climate models, Nature Reviews Physics, 1, 463–471, 2019.

Palmer, T. and Stevens, B.: The scientific challenge of understanding and estimating climate change, Proceedings of the National Academy of Sciences, 116, 24 390–24 395, 2019.

Pascale, S. and Bordoni, S.: Tropical and extratropical controls of Gulf of California surges and summertime precipitation over the southwestern United States, Monthly Weather Review, 144, 2695–2718, 2016.

Pascale, S., Bordoni, S., Kapnick, S. B., Vecchi, G. A., Jia, L., Delworth, T. L., Underwood, S., and Anderson, W.: The impact of horizontal resolution on North American monsoon Gulf of California moisture surges in a suite of coupled global climate models, Journal of Climate, 29, 7911–7936, 2016.

Pascale, S., Carvalho, L. M., Adams, D. K., Castro, C. L., and Cavalcanti, I. F.: Current and Future Variations of the Monsoons of the Americas in a Warming Climate, Current Climate Change Reports, pp. 1–20, 2019.

Perdigón-Morales, J., Romero-Centeno, R., Ordóñez, P., and Barrett, B. S.: The midsummer drought in Mexico: perspectives on duration and intensity from the CHIRPS precipitation database, International Journal of Climatology, 38, 2174–2186, 2018.

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP piControl, https://doi.org/10.22033/ESGF/CMIP6.6294, https://doi.org/10.22033/ESGF/CMIP6.6294, 2018.

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP amip, https://doi.org/10.22033/ESGF/CMIP6.5853, https://doi.org/10.22033/ESGF/CMIP6.5853, 2019a.

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6109, https://doi.org/10.22033/ESGF/CMIP6.6109, 2019b.

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-MM model output prepared for CMIP6 CMIP piControl, https://doi.org/10.22033/ESGF/CMIP6.6297, https://doi.org/10.22033/ESGF/CMIP6.6297, 2019c.

Ryu, J.-H. and Hayhoe, K.: Understanding the sources of Caribbean precipitation biases in CMIP3 and CMIP5 simulations, Climate dynamics, 42, 3233–3252, 2014.

Seager, R. and Vecchi, G. A.: Greenhouse warming and the 21st century hydroclimate of southwestern North America, Proceedings of the National Academy of Sciences, 107, 21 277–21 282, 2010.

Seastrand, S., Serra, Y., Castro, C., and Ritchie, E.: The dominant synoptic-scale modes of North American monsoon precipitation, International Journal of Climatology, 35, 2019–2032, 2015.

Sellar, A. A., Jones, C. G., Mulcahy, J., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and evaluation of the UK Earth System Model, Journal of Advances in Modeling Earth Systems, 0, https://doi.org/10.1029/2019MS001739, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001739, 2019.

Sheffield, J., Barrett, A. P., Colle, B., Nelun Fernando, D., Fu, R., Geil, K. L., Hu, Q., Kinter, J., Kumar, S., Langenbrunner, B., Lombardo, K., Long, L. N., Maloney, E., Mariotti, A., Meyerson, J. E., Mo, K. C., David Neelin, J., Nigam, S., Pan, Z., Ren, T., Ruiz-Barradas, A., Serra, Y. L., Seth, A., Thibeault, J. M., Stroeve, J. C., Yang, Z., and Yin, L.: North American climate in CMIP5 experiments. Part I: Evaluation of historical simulations of continental and regional climatology, Journal of Climate, 26, 9209–9245, 2013a.

Stensrud, D. J., Gall, R. L., and Nordquist, M. K.: Surges over the Gulf of California during the Mexican monsoon, Monthly Weather Review, 125, 417–437, 1997.

Straffon, A., Zavala-Hidalgo, J., and Estrada, F.: Preconditioning of the precipitation interannual variability in southern Mexico and Central America by oceanic and atmospheric anomalies, International Journal of Climatology, 2019.

Sulca, J., Takahashi, K., Espinoza, J.-C., Vuille, M., and Lavado-Casimiro, W.: Impacts of different ENSO flavors and tropical Pacific convection variability (ITCZ, SPCZ) on austral summer rainfall in South America, with a focus on Peru, International Journal of Climatology, 38, 420–435, 2018.

Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., Walton, J., and Jones, C.: MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6113, https://doi.org/10.22033/ESGF/CMIP6.6113, 2019a.

Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., Walton, J., and Jones, C.: MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP piControl, https://doi.org/10.22033/ESGF/CMIP6.6298, https://doi.org/10.22033/ESGF/CMIP6.6298, 2019b.

Trejo, F. J. P., Barbosa, H. A., Peñaloza-Murillo, M. A., Moreno, M. A., and Farías, A.: Intercomparison of improved satellite rainfall estimation with CHIRPS gridded product and rain gauge data over Venezuela, Atmósfera, 29, 323–342, 2016.

Trenberth, K. E.: The definition of El Nino, Bulletin of the American Meteorological Society, 78, 2771–2778, 1997.

Trenberth, K. E., Branstator, G. W., Karoly, D., Kumar, A., Lau, N.-C., and Ropelewski, C.: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures, Journal of Geophysical Research: Oceans, 103, 14 291–14 324, 1998.

Turrent, C. and Cavazos, T.: Role of the land-sea thermal contrast in the interannual modulation of the North American Monsoon, Geophysical Research Letters, 36, 2009.

Vera, C. and Silvestri, G.: Precipitation interannual variability in South America from the WCRP-CMIP3 multi-model dataset, Climate dynamics, 32, 1003–1014, 2009.

Vera, C., Higgins, W., Amador, J., Ambrizzi, T., Garreaud, R., Gochis, D., Gutzler, D., Lettenmaier, D., Marengo, J., Mechoso, C. R., Nogues-Paegle, J., Dias, P. L. S., and Zhang, C.: Toward a unified view of the American monsoon systems, Journal of Climate, 19, 4977–5000, 2006.

Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T., Bushell, A., Copsey, D., Earnshaw, P., Edwards, J., Gross, M., Hardiman, S., Harris, C., Heming, J., Klingaman, N., Levine, R., Manners, J., Martin, G., Milton, S., Mittermaier, M., Morcrette, C., Riddick, T., Roberts, M., Sanchez, C., Selwood, P., Stirling, A., Smith, C., Suri, D., Tennant, W., Vidale, P. L., Wilkinson, J., Willett, M., Woolnough, S., and Xavier, P.: The Met Office Unified Model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations, Geoscientific Model Development, 12, 1909–1963, 2019.

Wang, P. X., Wang, B., Cheng, H., Fasullo, J., Guo, Z., Kiefer, T., and Liu, Z.: The global monsoon across time scales: Mechanisms and outstanding issues, Earth-Science Reviews, 174, 84–121, 2017.

Wilks, D. S.: Statistical methods in the atmospheric sciences, vol. 100, Academic press, 2011.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.: The Met Office global coupled model 3.0 and 3.1 (GC3. 0 and GC3. 1) configurations, Journal of Advances in Modeling Earth Systems, 10, 357–380, 2018.

Yin, L., Fu, R., Shevliakova, E., and Dickinson, R. E.: How well can CMIP5 simulate precipitation and its controlling processes over tropical South America?, Climate Dynamics, 41, 3127–3143, 2013.

Zhou, T., Turner, A. G., Kinter, J. L., Wang, B., Qian, Y., Chen, X., Wu, B., Liu, B., Zou, L., and Bian, H.: GMMIP (v1. 0) contribution to CMIP6: global monsoons model inter-comparison project, Geoscientific Model Development, 9, 3589–3604, 2016.

Table S1. Summary of the CMIP6 simulations in this study. For each simulation the acronym used hereafter, the experiment and the horizontal resolution are shown. The first 100 years of the piControl simulations are used and for historical experiments the period 1979-2014 is used.

Model	Experiment	Resolution	*Acronym*	*Reference*
Hadley Centre Global Environment Model version 3 (HadGEM3)	Pre-industrial control	N96 1.875°x1.25°	GC3.1 N96-pi	(Menary et al., 2018; Ridley et al., 2018)
HadGEM3	Pre-industrial control	N216 0.83°x0.56°	GC3 N216-pi	(Menary et al., 2018; Ridley et al., 2019c)
HadGEM3	Historical	N96 1.875°x1.25°	GC3-hist	(Andrews et al., 2020; Ridley et al., 2019b)
HadGEM3	Atmospheric Model Inter-comparison (AMIP)	N96 1.875°x1.25°	GC3-AMIP	(Ridley et al., 2019a)
United Kingdom Earth System Model version 1 (UKESM1)	Pre-industrial control	1.875°x1.25°	UKESM-pi	(Tang et al., 2019b)
UKESM1	Historical	1.875°x1.25°	UKESM-hist	(Tang et al., 2019a)

Table S2. Root-mean square error (RMSE) and pattern correlation coefficients (PCC) for each season and each model experiment. Near surface air temperature ($t2m$), wind components (u and v) and mean-sea level pressure ($mslp$) are assessed against ERA-5 and precipitation (pr) against TRMM.

Variable	*Model experiment*	DJF RMSE	DJF PCC	MAM RMSE	MAM PCC	JJA RMSE	JJA PCC	SON RMSE	SON PCC
t2m	GC3 N96	1.28	0.98	1.3	0.96	1.38	0.96	1.31	0.96
t2m	GC3 N216	1.05	0.99	1.07	0.98	1.02	0.98	0.98	0.98
t2m	GC3 Hist	2.06	0.94	1.75	0.93	1.73	0.94	2.05	0.92
t2m	UKESM-hist	2.03	0.94	1.77	0.93	1.8	0.94	2.0	0.93
t2m	GC3 AMIP	1.17	0.98	1.12	0.97	1.2	0.97	1.2	0.97
u	GC3 N96	0.78	0.99	0.59	0.99	0.9	0.98	0.87	0.98
u	GC3 N216	0.78	0.99	0.59	0.99	0.9	0.98	0.87	0.98
u	GC3 Hist	1.02	0.98	1.04	0.97	0.92	0.98	0.84	0.98
u	UKESM-hist	1.04	0.98	1.01	0.97	0.91	0.98	0.82	0.98
u	GC3 AMIP	0.96	0.98	0.77	0.99	1.18	0.97	1.09	0.96
v	GC3 N96	0.75	0.93	0.66	0.93	0.65	0.95	0.59	0.94
v	GC3 N216	0.6	0.96	0.5	0.95	0.57	0.96	0.54	0.94
v	GC3 Hist	0.76	0.94	0.72	0.92	0.66	0.95	0.59	0.94
v	UKESM-hist	0.75	0.93	0.69	0.92	0.65	0.95	0.6	0.93
v	GC3 AMIP	0.67	0.95	0.52	0.95	0.68	0.94	0.61	0.93
mslp	GC3 N96	1.33	0.96	1.03	0.97	1.15	0.96	0.95	0.97
mslp	GC3 N216	1.11	0.97	0.9	0.97	1.1	0.96	0.89	0.97
mslp	GC3 Hist	1.31	0.97	1.12	0.96	1.08	0.96	0.94	0.97
mslp	UKESM-hist	1.4	0.97	1.15	0.96	1.14	0.95	0.99	0.97
mslp	GC3 AMIP	1.15	0.97	0.87	0.97	1.09	0.96	0.93	0.97
pr	GC3 N96	2.02	0.79	2.24	0.71	1.62	0.9	1.69	0.86
pr	GC3 N216	1.58	0.88	1.72	0.85	1.4	0.93	1.57	0.89
pr	GC3 Hist	2.05	0.78	2.49	0.64	1.69	0.88	1.69	0.86
pr	UKESM-hist	1.96	0.8	2.39	0.66	1.71	0.88	1.62	0.87
pr	GC3 AMIP	1.42	0.9	1.61	0.88	1.95	0.88	1.8	0.88

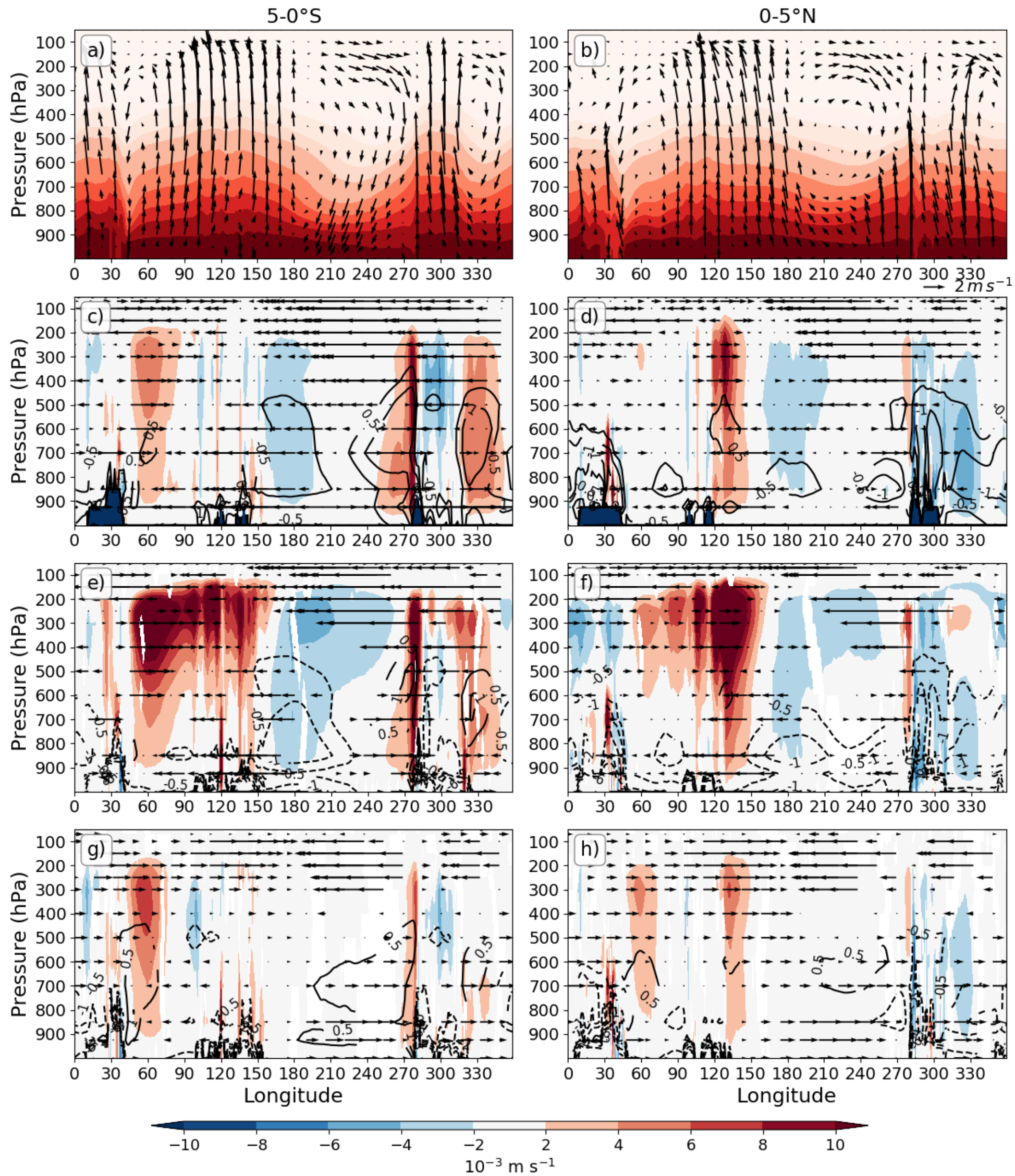

Figure S1. Longitude-pressure level plots of the mean DJF (a, b) specific humidity (color contours) and zonal and vertical velocities (vectors) in ERA5. (a) is latitudinally averaged in 5-0°S and (b) in 0-5°N. (c, d, e, f, g, h) show the bias in vertical velocity (color-contours), zonal wind (vectors) and specific humidity (line contours). Biases are shown for (c, d) UKESM1-historical, (e, f) GC3.1 N96-pi and (g, h) GC3.1 N216-pi. Only biases statistically significant to the 95% confidence level are shown, according to a Welch t-test between model and ERA5 data for all fields.

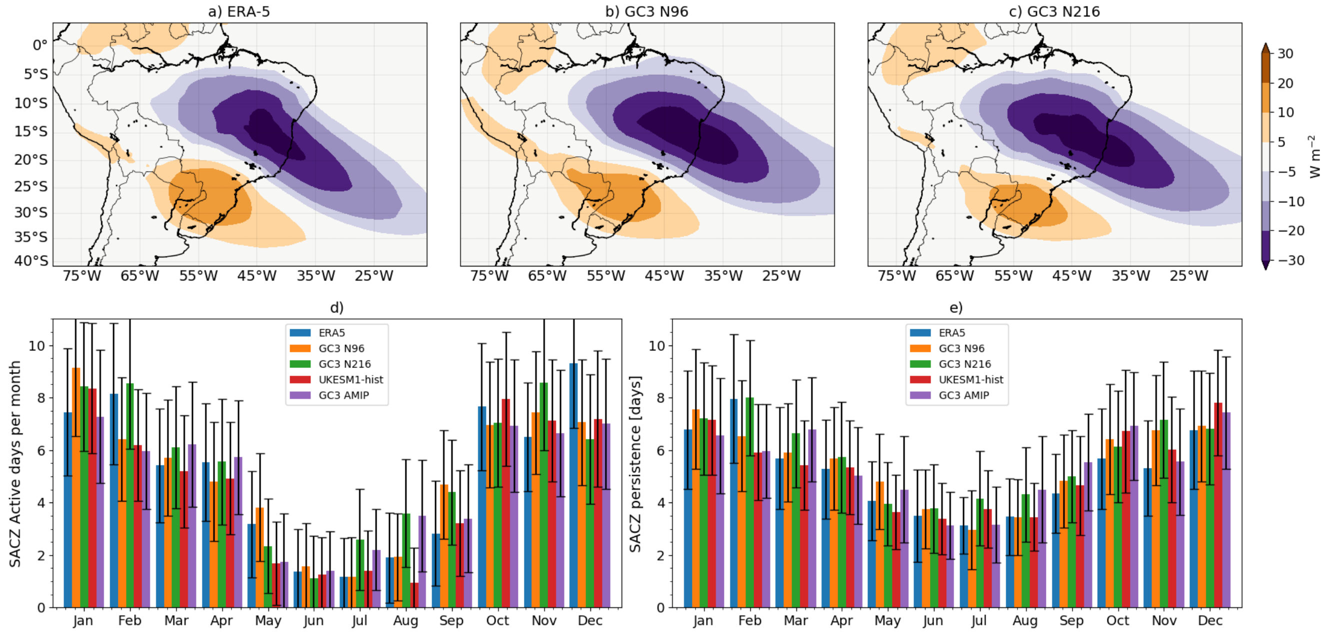

Figure S2. (a, b, c) OLR anomalies during active South Atlantic Convergence Zone (SACZ) events. (d, e) Frequency of active SACZ days and length of active SACZ events in reanalysis and model data, the standard deviation is shown as the error bar. The SACZ active days are constructed by first computing the first EOF of the monthly-mean deseasonalized OLR and then the daily OLR, previously filtered to remove periods higher than 99 days, is projected on the EOF pattern to produce a time-series of pseudo-principal components. Active SACZ days are found when this time-series of pseudo-PCs is greater than 1, and the persistence is measured as the number of continous days where the time-series is greater than 1.

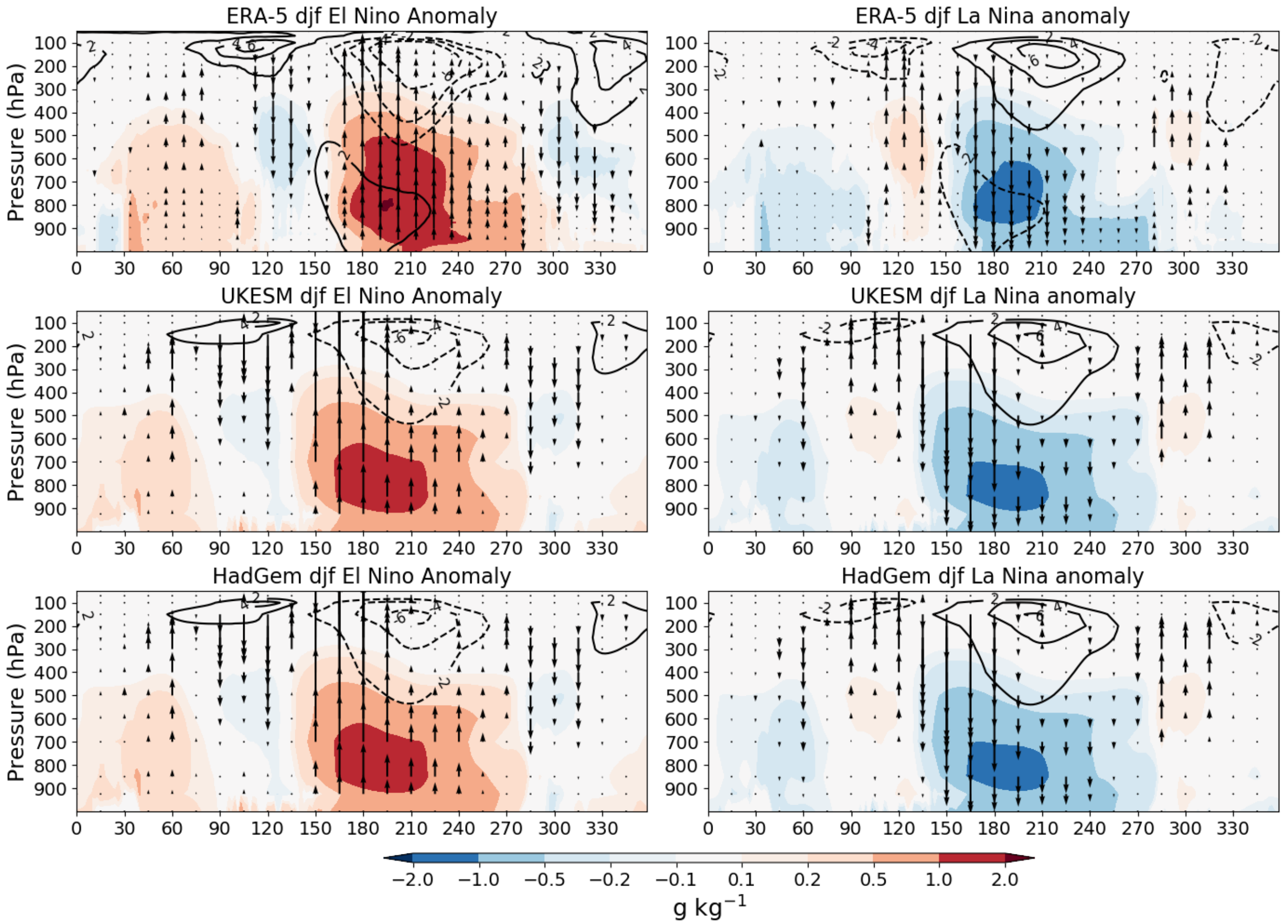

Figure S3. DJF Longitude-height Walker circulation anomalies of specific humidity (colour-contours), ω (vectors) and zonal wind (line-contours) during El Niño events (left) and La Niña events (right). Results are shown for ERA-5 (upper), UKESM-pi (middle) and HadGEM3 piControl (lower).

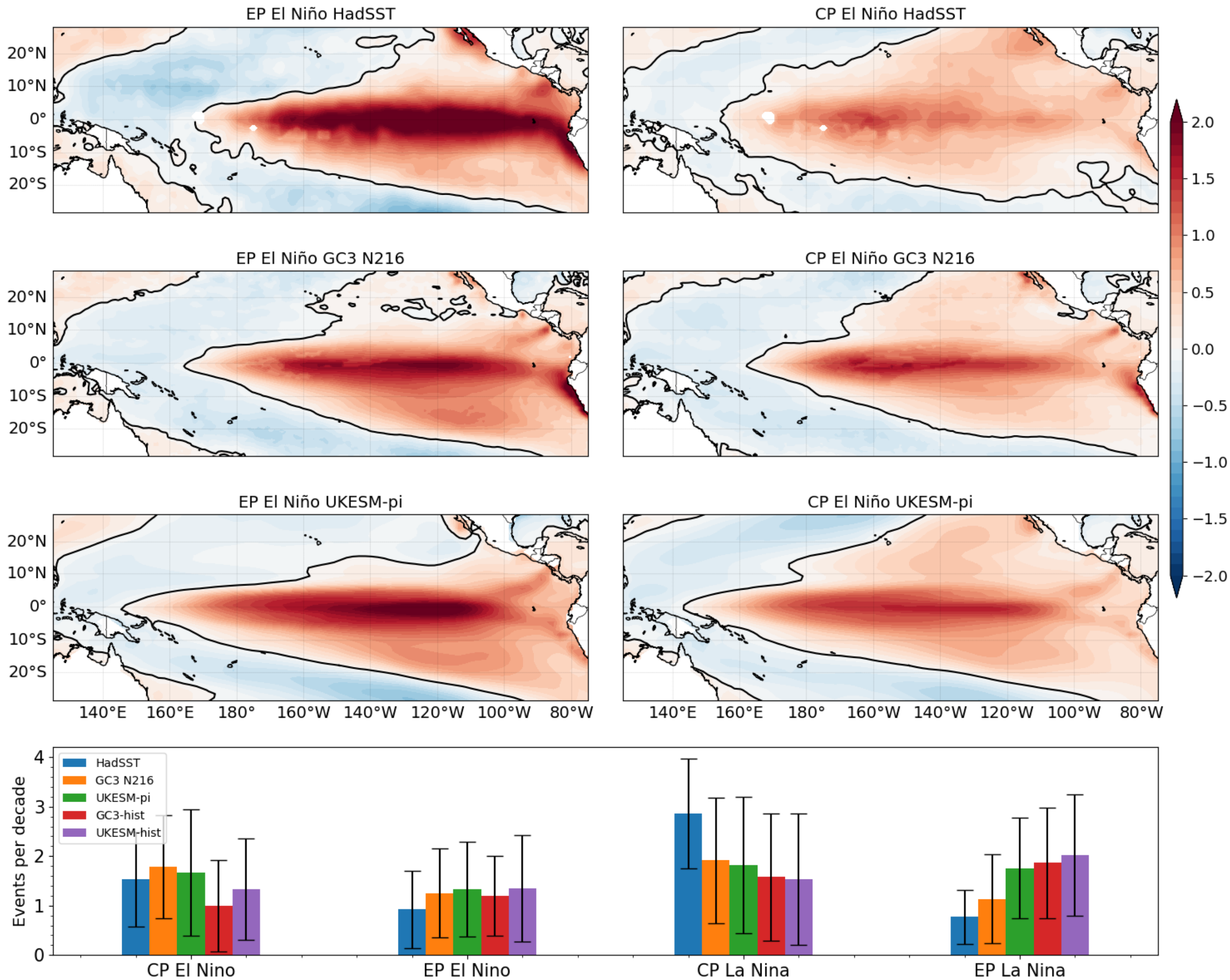

Figure S4. SST anomalies [K] for East Pacific (EP) and Central Pacific El Niño events in HadSST, GC3 N216 and UKESM piControl. EP (CP) events were defined where the E-index (C-index) was greater than 1. In the bottom panel, the frequency of events per decade (with standard deviation as error bar) is shown for HadSST and the simulations used in this study. The E-index is computed from $(PC1-PC2)/\sqrt{2}$ and the C-index from $(PC1+PC2)/\sqrt{(2)}$.