Multivariate process-based evaluation of CMIP6 historical simulations over the tropical oceans

Miguel Nogueira
Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal
Corresponding author email address: mdnogueira@fc.ul.pt

Abstract
Significant biases have persisted throughout the latest phases of the Coupled Model Intercomparison Project (CMIP), with important implications for climate science and policy. Here the systematic errors in mean and variability of precipitation (P), total column water vapor (W) and sea surface temperature (SST) were analyzed over the tropical oceans for ten CMIP phase 6 (CMIP6) models. This is a particularly challenging region for climate simulations. A process-based error analysis was employed, grounded by state-of-the-art satellite observations and reanalysis datasets. The results revealed a generalized overestimation of P (up to +1.5 mm/day) and underestimation of W variability (up to -5%), and SST variability (up to -54%), largely due to issues in simulating the intensity and extension of deep convection. The P variability errors were further associated with uncertainties in precipitation sensitivity to SST and W, El-Niño Southern Oscillation (ENSO) variability, and atmospheric longwave and shortwave radiative interactions. The models showed both positive and negative W biases (up to 4.5 mm), resulting from differences in water vapor feedback and tropical subsidence. Most models overestimated W variability (up to +103%) due to misrepresentation of the water vapor feedback, tropical deep convection, and ENSO variability. Concerning SST, the biases were relatively low (below ±1.2 K), but the variability was generally overestimated (up to +110%). The SST errors showed multiple connections, including to water vapor and Bjerknes feedbacks, atmospheric radiative absorption and cooling, and ENSO. Overall, the present results support the importance of multivariate model evaluation frameworks, and the critical need for significant improvements in simulation of deep tropical convection and atmospheric-radiative interactions.

Keywords: CMIP6, Climate Models, Tropical oceans, Model Evaluation

1. Introduction
Many aspects of the tropical climate simulated by state-of-art Earth System Models (ESMs) are at odds with the observations. Stouffer et al. (2017) identified six important long-standing biases in the last three phases of the Coupled Model Intercomparison Project (CMIP). Four of them directly affect the tropical regions: the overestimation of tropical precipitation, deficiencies in the simulated atmospheric and oceanic structure of circulation systems in the tropics, the misrepresentation of low clouds over tropical oceans, and the overly deep tropical thermocline. These systematic errors have largely persisted throughout the different CMIP phases despite relevant inter-generation differences and improvements in model resolution and parameterization schemes (Flato et al., 2013; Knutti and Sedlácek, 2013; Stouffer et al., 2017; Tian & Dong, 2020; Fiedler et al., 2020; Bock et al., 2020). One major challenge in solving systematic errors in climate models is the difficulty in attributing a specific cause to each specific error (Eyring et al., 2019). This is largely due to the intricate couplings between multiple variables and processes. For example, temperature and atmospheric moisture content are tightly linked via the Clausius-Clapeyron mechanism at long (multi-year) timescales (Held and Soden 2006; Nogueira, 2019a). In turn, water vapor plays an important role in determining many aspects of global climate, including the greenhouse effect thus feedbacking with temperature (Pierrehumbert, 2000). The persistent errors over tropical oceans in CMIP models provide a good example of errors resulting from complex interactions involving multiple variables. Two well-known issues are the excess of hemispheric symmetry in the Inter-Tropical Convergence Zone (ITCZ) and equatorial Pacific cold tong bias (Mechoso et al. 1995; Lin, 2007; Li & Xie, 2014; Tian & Dong, 2020), which are not independent from each other. They have been associated with uncertainties in the simulation of tropical sea surface temperature (SST) variability and its interaction with near-surface wind speed and oceanic currents (Lin, 2007; de Szncoke & Xie, 2008; Zheng et al., 2011; Li & Xie, 2014; Lauer et al., 2018), but also with problems in the representation of clouds and downward solar radiation fluxes in the tropics and in the Southern Hemisphere midlatitudes (Hwang & Frierson, 2013; Li & Xie, 2014). In turn, these issues have tight links with the El Niño–Southern Oscillation (ENSO) and the Madden-Julian Oscillation (MJO) (van der Vaart et al., 2000; Latif et al., 2001; Ham & Kug, 2012; Flato et al., 2013; Fiedler et al., 2020). Additionally, deep convection is known to play a central role in the large inter-model spread in simulated tropical precipitation (Bony et al., 2013; Shepherd, 2014; Xie et al., 2015; Nogueira, 2017, 2020). Tropical precipitation and atmospheric moisture content are also tightly linked by an emergent non-linear relation exhibiting properties of critical behavior (Bretherton et al., 2004; Peters and Neelin, 2006). Furthermore, precipitation is energetically constrained by atmospheric radiative fluxes both at global and regional scales (Stephens & Ellis, 2008; Lambert & Webb 2008; Payerter & Ramaswamy, 2014; DeAngelo et al., 2015; Nogueira, 2019b, 2019c). Finally, tropical clouds have been identified as key uncertainty source for the simulated tropical and global climate, including significant impacts to radiative fluxes and circulation (Bony & Dufresne, 2005; Sherwood et al., 2014; Schneider et al., 2017).

The present works addresses the systematic errors over the tropical oceans simulated by models in the CMIP phase six ensemble (CMIP6, Eyring et al., 2016), contributing to the ongoing effort of rigorous and comprehensive evaluation of CMIP6. This is a critical task since this dataset will underly the upcoming Intergovernmental Panel on Climate Change Sixth Assessment Report (IPCC-AR6), playing a central role in climate science and policy over the upcoming years. Here the complex nature of the systematic errors in the mean and variability of precipitation (P), total column water vapor (W), and SST over the tropical oceans was tackled by employing a multivariate process-based approach. The analysis was grounded on last generation satellite-based and reanalysis products, which represent fundamental tools for model evaluation and development efforts (Balsamo et al., 2018; Levizzani & Cattani, 2019). The manuscript is
organized as follows: the considered datasets and analysis methodology are described in Section 2. The mean and variability error assessment of CMIP6 models is presented in Section 3, along with a process-based analysis of the relations amongst different models and variables. Finally, the main results are summarized and discussed in Section 4.

2. Data and Methods

2.1. Datasets

Precipitation observations were obtained from the Global Precipitation Climatology Project (GPCP) version 2.3 (Adler et al., 2018), providing monthly precipitation estimates at 2.5° resolution from 1979-present. GPCP is produced by merging a variety of data sources, including data from rain gauges (including the Global Precipitation Climate Center, GPCC, dataset over land), passive microwave-based satellite retrievals from the special sensor microwave/imager (SSM/I) and the special sensor microwave sounder (SSMIS), and infrared rainfall estimates from geostationary and polar-orbiting satellites.

Observations of total column water vapor were obtained from Remote Sensing Systems (RSS) SSM/I version-7, extending from 1988 to present (Wentz, 1997; Wentz et al., 2007). This is a monthly dataset covering the ice-free ocean areas at 1° resolution, constructed by inter-calibrating the measurements from several different satellites for each channel at the radiance level, and then using a common algorithm to retrieve atmospheric moisture content.

Monthly mean fields of SST, zonal and meridional wind speed at 10-m (respectively u_{10} and v_{10}), vertical pressure velocity at 500hPa (\omega_500hPa), and atmospheric longwave and shortwave radiative fluxes were obtained from the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis product (ERA5). ERA5 combines a large number of observations across the globe with a recent version of the ECMWF Integrated Forecast System (IFS cycle 41r2) by employing a four-dimensional variation data assimilation (Hersbach et al., 2020). ERA5 is available from 1950 to present, covering the full globe with a horizontal resolution of around 31 km and with 137 vertical model levels.

CMIP6 is the latest generation set of multi-model global climate simulations (Eyring et al., 2016). Compared to its predecessors, CMIP6 models feature several improvements in different physical parameterizations (including in convection and clouds parameterization), the inclusion of additional processes and couplings, such as ice sheets and their dynamics, permafrost processes, and the nitrogen effects of terrestrial carbon uptake. Some models also feature increased grid resolution. Monthly mean fields of P, W, SST, u_{10}, v_{10}, \omega_500hPa, and longwave and shortwave radiative fluxes were obtained from ten different CMIP6 models (summarized in Table 1), for which all required variables were available for the historical experiment. Variant r1i1p1f1 was used whenever available, otherwise r1i1p1f2 was considered.

2.2. Analysis methodology

The present study assessed the systematic errors in different variables from CMIP6 historical simulations over the tropical oceans (20ºS to 20ºN). Given the distinct temporal extent of the considered products, the analysis focused on the period common to all datasets, covering 27 years from January/1988 to December/2014. Two error metrics were computed. The first was the seasonal cycle bias given by:

\[ \text{Bias}_m = \bar{f}_{m,o} - \bar{f}_{m,i} \]  

(1)

Where \( \bar{f}_{m,o} \) is an observed variable averaged over month m (\( \bar{f}_{m} = (1/27) \sum_{y=1988}^{2014} f_{m,y} \)), and \( \bar{f}_{m,i} \) is the same variable simulated by model i (i=1,..,10) averaged over month m. The second metric was the seasonal cycle interannual variability matching score, \( \sigma_m \), given by:

\[ \sigma_{m} = 100 \times \left( \frac{\sigma_{m,i}}{\sigma_{m,o}} - 1 \right) \]  

(2)

Where \( \sigma_{m,o} \) is the standard deviation of the observed variable for month m, and \( \sigma_{m,i} \) is the standard deviation of the same variable simulated by model i (with i=1,..,10) for month m. This metric provides a relative measure (in percentage) of the deviation between modeled and observed variability, with \( \sigma_n > 0 \) (< 0) when the model overestimates (underestimates) the observed variability, and \( \sigma_n = 0 \) when the model reproduces the observed variability exactly.

Both metrics were computed from a single time-series for each model and observational-based product, obtained by averaging over all tropical ocean grid-points. This step reduced the dimensionality of the problem and minimized the impact of spatial lags on the computed errors. While spatial displacements may be important (e.g., displacements in the average position of the ITCZ or of the Pacific cold pool), they are prone to dominate the systematic errors, masking relevant information on the models’ ability to reproduce other relevant properties of the main tropical climate features, such as the intensity, frequency and key dynamical processes.

Robust and long (multi-decadal) observational products covering the tropical oceans are not available for many relevant climate variables. Here this limitation was overcome by considering data from ERA5 as the reference for model evaluation. It is important to notice that the agreement between ERA5 fields and observations depends on the specific variable. For example, SST is an observational-based forcing in ERA5, and thus may be considered as an observation. Tropospheric humidity and near-surface zonal and meridional wind fields are tightly constrained by data-assimilation of satellite, in situ and radiosonde observations (see Hersbach et al., 2020 for details). As a result, these fields have been shown to compare well with observations in recent studies (Belmonte-Rivas and Stoffelen, 2019; Olason, 2018; Kalverla et al., 2019; Hersbach et al., 2020; Soares et al., 2020). In contrast, vertical velocity and surface and TOA radiative fluxes are less constrained by data assimilation, although they benefit from the assimilation of relevant quantities combined with a state-of-the-art numerical model with high horizontal and vertical grid-resolution compared to CMIP6 models. Indeed, a recent
study has demonstrated that improved representation of tropical rainfall in ERA5 compared previous generation reanalysis results from improved representation of the tropical circulation (Nogueira, 2020). Nonetheless, these differences amongst different ERA5 variables must be considered when interpreting the comparisons against CMIP6 models. Besides quantifying bias and $\sigma$, the present study explored the relations in these error metrics computed for different variables. These relations were quantified by the Pearson correlation coefficients. All correlations were grounded by physical arguments (including dynamical, thermodynamical and radiative processes). This physical-based analysis relied on $P$, $W$, SST and seven additional variables, which were computed for the CMIP6 and ERA5 datasets. The first two were the average upward and downward velocity intensity, respectively denoted $\omega_{500hPa}$ and $\omega_{850hPa}$ (and respectively computed by averaging the vertical pressure velocity over tropical ocean grid-points with negative and positive $\omega_{500hPa}$ values only). The third was the fraction of tropical area occupied by the upward motion, denoted FTo$^+500hPa$ (computed as the area where $\omega_{500hPa}>0$ divided by the total area in the tropical ocean domain). Additionally, three different measures of radiative exchanges were considered: i) the atmospheric shortwave absorption (SWA, given by the difference between the net downward radiative flux at the top of the atmosphere (TOA) and the net downward radiative flux at the surface); ii) the atmospheric longwave cooling (LWC, given by the sum of the outgoing longwave radiation at TOA and the net downward longwave radiation at the surface); and iii) the downwelling longwave radiative flux at the surface (DLR). Finally, the Ni34 index was computed as the variability of the SST averaged over the Niño-3.4 region (5ºS to 5ºN and 170ºW to 120ºW). This type of process-based analysis allows to mitigate the risk of misinterpreting relations arising from spurious correlations rather than from an intrinsic underlying process which the model is expected to reproduce (Eyring et al., 2019; Maloney et al., 2019).

3. Results

3.1. Systematic errors over tropical oceans in CMIP6 simulations

All ten CMIP6 models considered here systematically overestimated the average precipitation over tropical oceans throughout the full seasonal cycle when compared to GPCP (Fig. 1a). The monthly averaged $P$ biases ranged between 0.5 and 1.5 mm/day. IPSL-CM6A displayed the overall better performance, while MPI-ESM1-2-HR showed the largest average bias. ERA5 also overestimated precipitation over tropical oceans, although its average bias was 0.6 mm/day while the average bias in CMIP6 models ranged between 0.7 and 1.1 mm/day.

The bias in $W$ averaged over the entire tropical oceans was within ±2 mm throughout the entire annual cycle (Fig. 1b) for most models. But there were two exceptions: CNMR-CM6-1 showed a systematic underestimation of around -3 mm throughout the entire seasonal cycle, while CAMS-CSM-1 showed a systematic overestimation of around +4 mm throughout the entire seasonal cycle. ERA5 displayed a $W$ bias of around -1 mm throughout the entire seasonal cycle. The systematic differences between CMIP6 and ERA5 for SST averaged over the tropical oceans for the 1988-2015 period ranged between -0.3ºC and +1.2ºC depending on simulation and month (Fig. 2a). The closest agreement with ERA5 was found for UKESM-0-LL and IPSL-CM6A-LR.

Besides the diversity in yearly averaged bias for $P$, $W$ and SST, Fig. 1a, 1b and 2a also evidence significant differences in the seasonal cycle amongst different models. While some models displayed pronounced bias seasonality, other models displayed nearly constant bias for all months. Moreover, the bias seasonality is not coherent amongst variables for each model. This illustrates the complexity of the systematic error sources in different models which will be further discussed in Section 3.2.

The systematic differences in $u_{10}$ between eight of the CMIP6 models and ERA5 (Fig. 2b) are within ±0.5 m/s for all months. The other two models (ACCESS-ESM1-5 and CAMS-CSM1-0) showed a larger magnitude overestimation of zonal wind speed, up to +1.1 m/s. The systematic differences in $v_{10}$ over the tropical oceans between CMIP6 and ERA5 ranged between -1.1 m/s and +0.6 m/s (Fig. 2c). The $v_{10}$ biases showed pronounced seasonality, being predominantly negative during boreal winter, spring, and autumn, but with a noticeable tendency towards positive values during boreal summer.

Seven CMIP6 models showed a systematic negative differences compared to ERA5 in $\omega_{500hPa}$ averaged over the tropical oceans (Fig. 2d). In contrast, ACCESS-ESM1-5, IPSL-CM6A-LR and MIROC6 displayed nearly zero $\omega_{500hPa}$ average systematic differences to ERA5, but with relatively large seasonality in the latter case. Seven CMIP6 models showed overall negative $\omega_{500hPa}$ differences compared to ERA5 (Fig. 2e), corresponding to a stronger average intensity of the upward motion over tropical oceans. Notice that the latter set of seven models did not entirely coincide with the set of seven models displaying negative systematic differences in $\omega_{500hPa}$. Concerning the remaining three models, one showed only small differences in $\omega_{500hPa}$ when compared to ERA5, while the other two displayed systematic underestimation of the average upward motion intensity over the tropical oceans.

Six CMIP6 models showed slightly positive differences in $\omega_{850hPa}$ compared to ERA5 (Fig. 2g), corresponding to a small overestimation of the downward motion intensity over tropical oceans. Two other models slightly underestimate this average downward motion intensity. Larger differences were found for INM-CM4-8 and CNRM-CM6-1, with the former largely overestimating the downward motion average intensity (and underestimating the upward motion intensity, cf. Fig. 2e) motion compared to ERA5, while CNRM-CM6-1 overestimated the average downward motion intensity (and also overestimated the upward motion intensity, Fig. 2e). In addition to these differences in vertical motion intensity, the results also showed relevant differences in the areal coverage regions showing average upward (and thus also negative) motion between different CMIP6 models and ERA5, which can be as large as 10% of the total tropical oceans area (Fig. 2f).

Seven CMIP6 models overestimated the surface DLR compared to ERA5 (Fig. 2h), with four of them displaying small magnitude differences (+2 Wm$^{-2}$), while the other three showed larger differences (between +5 and +12 Wm$^{-2}$). Contrasting, two models showed a slight underestimation of DLR (~1 Wm$^{-2}$), and CNRM-CM6-1 showed a systematic underestimation of DLR of around -4.5 Wm$^{-2}$ compared to ERA5. All CMIP6 models underestimated SWA over tropical oceans compared to ERA5, with overall values ranging...
3.2. Exploring relations between systematic model differences

Section 3.1 evidenced relevant systematic differences across numerous variables between CMIP6 and satellite-based observations or ERA5 reanalysis over the tropical oceans. However, no clear relations between the systematic differences across variables could be immediately identified from Figures 1 and 2. For example, a better (worse) performance in reproducing the satellite-observed W climatology over the tropical oceans (Fig. 1b) did not imply a better (worse) performance in reproducing the satellite-observed P climatology over the same domain (Fig. 1a), and vice-versa. Another example was provided by the systematic negative difference in $\omega_{500hPa}$ between CMIP6 models compared to ERA5 (Fig. 1e). This result was coherent with the overall larger overestimation of rainfall in CMIP6 models compared to ERA5. However, a comparison between Figure 1a and Figure 2d does not reveal a one-to-one relation between differences in both variables for each model. Indeed, the lowest differences for $\omega_{500hPa}$ were obtained for ACCESS-ESM1-5, MIROC6, and IPSL-CM6A-LR, but these three models are characterized by very distinct P bias.

A more comprehensive and quantitative analysis of the relations between the systematic differences in precipitation and other potentially relevant variables between CMIP6 and ERA5 is presented in Figure 3. The results showed no simple relation between the systematic differences in P and SST (Fig. 3b), $u_{10}$ (Fig. 3c), $v_{10}$ (Fig. 3d), DLR (Fig. 3i), SWA (Fig. 3j), or LWC (Fig. 3k). This was quantified by the respective Pearson correlation coefficients, which were $\leq 0.2$ for all these cases. A weak linear correlation ($r=0.3$) emerged between the systematic differences in P and W (Fig. 3a). However, the frailness of this relation was clearly illustrated by the results for CNRM6-CM6-1 in Fig. 3a, where W was underestimated but P overestimated.

A strong negative correlation ($r=-0.7$) emerged between the systematic differences in P and $\omega_{500hPa}$ (Fig. 3e). Relevant correlations were found for the systematic differences between P and $\omega^*_{500hPa}$ ($r=-0.4$, Fig. 3h), and between P and $P$ and $\text{Fro}^*_{500hPa}$ ($r=0.3$, Fig. 3g), but not between P and $\omega^*_{500hPa}$ ($r=-0.1$, Fig. 3f). In other words, the results revealed that model biases in simulated average precipitation over the tropical oceans resulted from systematic uncertainties in average vertical motions, both in the intensity and areal extent of the tropical ascending regions.

The correlation coefficient between systematic differences in W and SST was $r=0.5$ (Fig. 4a), supporting the relevance of the Clausius-Clapeyron relationship at the regional scale over the tropical oceans. But SST was not the single dominant factor for W bias in CMIP6 models. This fact was clearly illustrated by the fact that INM-CM4-8 and MIROC6 displayed negative systematic differences for SST but positive ones for W. Moreover, distinct systematic differences in W were found for similar systematic differences in SST amongst different models (e.g. BCC-CSM2-MR vs CAMS-CSM1-0).

A correlation coefficient of $-0.5$ was found between systematic differences in W and $\omega^*_{500hPa}$ (Fig. 4f). In contrast, the correlation coefficient between the systematic differences in W and $\omega^*_{500hPa}$ was $r=0.0$ (Fig. 4h). Additionally, the correlation coefficient between systematic differences in W and $\omega_{500hPa}$ was $r=-0.4$ (Fig. 4e), and between W and $\text{Fro}^*_{500hPa}$ was $r=0.3$ (Fig. 4g). Notice that the fractional areal coverage of the downward motion is complementary to $\text{Fro}^*_{500hPa}$. Hence, systematic differences in the former will be characterized by the same correlation coefficients as the latter. In this sense, the above results point out the importance of uncertainties in subsidence over the tropical oceans to atmospheric water vapor content systematic errors.

The correlation coefficient between systematic differences in W and DLR was $r=0.4$ (Fig. 4i), evidencing the well-known greenhouse effect of water vapor. Interestingly, the results showed relevant correlation between systematic differences in W and SWA ($r=0.6$, Fig. 4j), but not between W and LWC ($r=0.0$, Fig. 4k). This suggests that inter-model differences in W biases were related to atmospheric shortwave absorption, but not to longwave cooling.

Besides the above discussed link between systematic differences in SST and W (Fig. 4a), we found relevant correlation coefficients between systematic differences in SST and DLR ($r=0.3$, Fig. 5g), SST and SWA ($r=0.7$, Fig. 5h) and SST and LWC ($r=0.4$, Fig. 5i). These correlations reflected bulk measures of the complex relationships between surface temperature and its radiative exchanges with the atmosphere. The correlation between systematic differences in SST and DLR was notoriously low. However, one must notice that MIROC6 was largely for the obtained low correlation coefficient, due to its relatively large difference to ERA5 in DLR, but a relatively low difference for SST. In fact, when MIROC6 was removed from the analysis, the correlation coefficient between SST and DLR increased to $r=0.7$.

Finally, the correlation coefficients between systematic differences W and $u_{10}$ and W and $v_{10}$ were respectively $r=0.4$ (Fig. 5c) and $r=-0.2$ (Fig. 4d). The correlation coefficients between systematic differences SST and $u_{10}$ (Fig. 5a) and SST and $v_{10}$ (Fig. 5b) were respectively $r=0.2$ and $r=-0.4$. The interpretation of the correlations with zonal and meridional wind speed is far from trivial because multiple mechanisms are involved, including impact of near surface wind speed to surface turbulent heat and moisture fluxes, to upwelling ocean circulations, or the advection of air masses with different temperature and humidity. A detailed analysis of all these interactions is beyond the scope of the present research.

3.3. Evaluation of interannual variability over tropical oceans in CMIP6 simulations

Nine CMIP6 models underestimated the observed precipitation variability over tropical oceans (Fig 6a), with yearly averaged $\sigma_n$ values ranging between -54% and -4%. The only exception was MIROC6 which displayed a yearly averaged $\sigma_n$ value of +24%. MIROC6 also showed the largest seasonality of $\sigma_n$, reaching an overestimation of nearly +75% during July to September. ERA5 slightly overestimated the variability of precipitation in GPCP on the yearly average ($\sigma_n=+7\%$), but with relevant seasonality of $\sigma_n$ (ranging between ±20%).
MPI-ESM1-2-HR showed the overall best performance in simulating the observed precipitation variability over the tropical oceans, with a yearly averaged $\sigma_n$ value of -4%.

Concerning W, ERA5 showed an overall $\sigma_n$=8% (Fig. 6b), only outperformed by CNRM-CM6-1 ($\sigma_n$=2%). Nine out models overestimated W variability compared to satellite observations, with $\sigma_n$ ranging between +2% and +103% (again the largest overestimation was for MIROC6). The only exception to this generalized overestimation of W variability was INM-CM4-8, which underestimated the variability for both W and P. A comparison of Figures 6a and 6b shows that the best performing CMIP6 model in terms of W variability was not the best performing CMIP6 model in terms of P variability, and vice-versa. However, MIROC6 was the worst performing model, largely overestimating the variability for both variables.

The differences in variability were also quantified for SST, $u_{10}$, $v_{10}$, $\omega^5_{500hPa}$, $\omega^0_{500hPa}$, $\omega^0_{500hPa}$, Fr0$_{500hPa}$, DLR, SWA, LWC and Ni34 using ERA5 as reference. Nine CMIP6 models overestimated SST variability compared to ERA5 (Fig. 7a), with INM-CM4-8 being the only exception displaying a slight (-4%) underestimation. The overestimation magnitude varied significantly amongst different models, ranging between +7% for CAMS-CSM1-0 and +110% for MIROC6. In fact, MIROC6 stood out from all other models for its large overestimation of the variability for P (Fig. 6a), W (Fig. 6b), SST (Fig. 7a), $\omega_{500hPa}$ (Fig. 7d), Fr0$_{500hPa}$ (Fig. 7f), DLR (Fig. 7h), SWA (Fig. 7i) and Ni34 (Fig. 7k). This overestimation was particularly notorious for $\omega_{500hPa}$ with $\sigma_n$=+31% for MIROC6 clearly contrasting with the negative values (ranging between -59% and -5%) for all other models. Similarly, MIROC6 overestimated the variability of Fr0$_{500hPa}$ compared to ERA5 ($\sigma_n$=+34%), while all other models showed negative $\sigma_n$ values (ranging between -37% and -2%). Identical exceptional behavior was identified above for P variability (Fig. 6a), suggesting a link between these model differences, which is supported by the dynamical control of tropical precipitation variability.

MIROC6 showed the largest overestimation of variability for SST (Fig 7a), DLR (Fig. 7h), and Ni34. Eight other models also displayed positive $\sigma_n$ values for these three variables, while INM-CM4-8 showed negative values. This situation was identical to the behavior reported for $\sigma_n$ computed for P presented above (Fig. 6b), suggesting a link between these model differences in these four variables. This result was supported by the ENSO modulation of tropical SST, the coupling between SST and near-surface air temperature, the impact of W to the greenhouse effect, and the coupling between SST and W associated with the Clausius-Clapeyron relationship. These connections are further explored in Section 3.4.

Most CMIP6 models overestimated $u_{10}$ variability over tropical oceans during boreal spring but underestimated during boreal autumn and summer (Fig. 7b). INM-CM4-8 was the exception to this behavior - although the seasonal cycle of $\sigma_n$ had a similar shape to the other models, $\sigma_n$ remained negative in all months except in April. Concerning $v_{10}$, the results showed large inter-model spread for both magnitude and sign of $\sigma_n$ (Fig. 7c). Finally, all CMIP6 models overestimated the LWC variability compared to ERA5 (Fig. 7j), with average $\sigma_n$ values ranging between +9% for INM-CM4-8 and +74% for UKESM1-0-LL.

### 3.4. Exploring relations between model differences in interannual variability

The relations between interannual variability in different variables are explored in this section. Inter-model differences in $\sigma_n$ for W were tightly correlated with differences in $\sigma_n$ for SST (r=0.9, Fig. 8a), as expected from the Clausius-Clapeyron relationship. The differences in $\sigma_n$ for SST were also strongly correlated to differences in $\sigma_n$ for DLR (r=0.9, Fig. 9g), differences in $\sigma_n$ for SWA (r=0.7, Fig. 9h), and in Ni34 variability (r=0.7, Fig. 9j). In fact, SST variability showed correlation coefficients greater or equal to 0.3 with the variability of nearly all considered variables: r=0.5 for LWC (Fig. 9i), r=0.5 for Fr0$_{500hPa}$ (Fig. 9e), r=0.4 for P (Fig. 10b), r=0.4 for $\omega_{500hPa}$ (Fig. 9c), r=0.4 for $u_{10}$ (Fig. 9a), r=0.3 for $v_{10}$ (Fig. 9b), and r=0.3 for $\omega^0_{500hPa}$ (Fig. 9f). A slightly lower correlation coefficient of r=0.2 was obtained between the differences in $\sigma_n$ for SST and for Fr0$_{500hPa}$ (Fig. 9d). These results evidence the intricate ocean-atmosphere interactions, involving multiple thermodynamic, dynamic, and radiative processes.

The strongest correlation for the inter-model differences in $\sigma_n$ for P was with differences in $\sigma_n$ for $\omega_{500hPa}$ (r=0.9, Fig. 10e). The correlation coefficients between the variability of P and the variability of average ascending intensity and areal coverage were respectively r=0.6 (Fig. 10h) and r=0.7 (Fig. 10g). These correlation coefficients translate the tight link between the variability in the ascending branch of tropical circulation and precipitation. Indeed, the present results showed the dominant role of these dynamical links over the thermodynamical link of precipitation variability with SST variability (r=0.4, Fig. 10b), the link with ENSO variability (r=0.5, Fig. 10i), and the link with W variability (r=0.6, Fig. 10a). The results showed relatively weak connections between $\sigma_n$ for precipitation and $\sigma_n$ for average tropical subsidence (r=0.2, Fig. 10f), $\sigma_n$ for $u_{10}$ (r=0.1, Fig. 10c), and $\sigma_n$ for $v_{10}$ (r=0.2, Fig. 10j). The dynamical controls of precipitation variability were also stronger than the radiative ones. Namely, +0.4 correlation coefficients were obtained between both $\sigma_n$ for P and DLR (Fig. 10i) and $\sigma_n$ for P and SWA (Fig. 10j). The correlation between $\sigma_n$ for P and for LWC was 0.2 (Fig. 10k). This means that the well-known global energetic constraints on precipitation are present, but weaker, at the regional tropical ocean scale.

As pointed out above, inter-model differences in W variability were tightly linked with differences in SST variability. Relevant correlation coefficients were also found between $\sigma_n$ for W and for the upward vertical circulation variability - r=0.6 for Fr0$_{500hPa}$ (Fig. 8d), r=0.4 for $\omega^0_{500hPa}$ (Fig. 8g), and r=0.6 for Fr0$_{500hPa}$ (Fig. 98). Additionally, ENSO variability also played a strong modulating effect for W variability inter-model differences over tropical oceans (r=0.7, Fig. 8k). Lower correlation coefficients were obtained between $\sigma_n$ for W and $\omega^0_{500hPa}$ (r=0.2, Fig. 8e), $u_{10}$ (r=0.2, Fig. 8b), and $v_{10}$ (r=0.3, Fig. 8c). Strong links were found between the atmospheric water vapor and radiative fluxes variability. Specifically, the correlation coefficient between the $\sigma_n$ for W and for DLR was r=0.9 (Fig. 8h), which resulted from the strong greenhouse effect of atmospheric water vapor. The correlation between the $\sigma_n$ for W and for SWA was
The fourth was the 
longwave energy constraints of tropical vertical circulation for precipitation and water vapor and, in turn, to SST found here supports the need for convective permitting model-grid resolution in climate models, which also accounted for model differences in atmospheric circulation and radiative exchanges.

The observed precipitation averaged over tropical oceans was systematically overestimated by CMIP6 and ERA5 (ranging between +0.7 and +1.1 mm/day for CMIP6, and 0.6 mm/day for ERA5). Nine CMIP6 models underestimated precipitation variability (ranging between -54% to -4%), while the other model and ERA5 overestimated precipitation variability (respectively by +24% and 7%). The inter-model spread in precipitation bias and variability were tightly related to differences in tropical deep convection intensity and areal coverage. This result provided a different perspective on the previously reported dominant role of atmospheric circulation to tropical precipitation uncertainty in climate models (Bony et al., 2013; Shepherd, 2014; Xie et al., 2015; Nogueira, 2017, 2020).

Five additional error sources for precipitation variability in CMIP6 models emerged from the analysis. The first was the difference in SST variability, reflecting the thermodynamic constraints of precipitation response associated with the Clausius-Clapeyron relationship (Allen and Ingram, 2002; Held & Soden, 2006; Schneider et al., 2010; Nogueira, 2019a,c). The second was the difference in Ni34 index, corresponding to the well-known ENSO modulation of tropical precipitation (Trenberth, 1997; van Oldenborgh et al. 2005; Nogueira, 2019c). The third was the difference in atmospheric water vapor content variability, reflecting an emergent relation between P and W over the tropical oceans previously found in models and observations (Bretherton et al., 2004; Peters and Neelin, 2006; Rushley et al., 2018; Kuo et al., 2020). The fourth was the difference in DLR variability, in agreement with the longwave energy constraints of precipitation at global scale (Allen & Ingram, 2002; Stephens & Ellis, 2008; Nogueira, 2019b) and at regional scale over the tropical oceans (Nogueira, 2019c). The fifth was the difference in atmospheric shortwave absorption, in agreement with DeAngelis et al. (2015) hypothesis that the underestimation of shortwave absorption by climate models leads to an overestimation of precipitation. The latter may be explained by a reduction of the latent heat required to maintain a balanced atmospheric energy budget (Takahashi, 2009; O’Gorman et al., 2012). Indeed, the present analysis found a generalized underestimation of SWA in CMIP6, in line with previous findings for CMIP5 (DeAngelis et al., 2015). Moreover, the present work also confirmed that the SWA underestimation is tightly related to variations in atmospheric water vapor (Paynter & Ramaswamy, 2014; DeAngelis et al., 2015). Cloud-radiation interactions have also been previously identified as key model error sources in climate models (Bony & Dufresne, 2005; Randall et al., 2007; Chen et al., 2013). Interestingly, these five error sources showed little impact on P bias. In other words, these five mechanisms represented relevant uncertainty sources to the simulated precipitation variability but not to the (circulation dominated) precipitation mean errors.

CMIP6 models showed both positive and negative biases in W (up to 4.5 mm). These biases were tightly correlated to biases in SST and in average subsidence at 500 hPa. The first relation was related to the water vapor feedback – in simple terms, on the one hand the Clausius-Clapeyron relationship implies that higher temperatures result in increased atmospheric water vapor, and on the other hand the increased atmospheric water vapor increases the greenhouse effect, resulting in increased surface temperature. The link found between systematic errors in W and DLR provided additional robustness to this interpretation. The second relation reflected the large impact of changes in the intensity of tropical subsidence to atmospheric water vapor content in the tropical atmosphere (Pierrehumbert, 2000).

The water vapor feedback also played an important role in explain the overestimation of W variability in nearly all CMIP6 models (up to 103%). Inter-model differences in the simulated variability of tropical deep convection and of the SST over the Niño-3.4 region were additional error sources to simulated W variability.

Finally, CMIP6 models displayed relatively small positive and negative SST bias values (up to ±1.2 K) compared to an observational-based product, and a generalized overestimation of SST variability (up to 110%). The inter-model differences in SST bias and σn showed relevant correlations with W, DLR, SWA, LWC, and v10. Furthermore, the errors in simulated SST bias was also correlated with differences 500 hPa subsidence variability, while SST variability errors were also linked to differences in σu for u10, ω500 Pa, Fr0500 Pa, Ni34, and P. The emergence of these multiple correlations highlighted the intricate ocean-atmospheric coupling at interannual (and longer) timescales (Lovejoy et al., 2017; Nogueira, 2020b), involving the water vapor feedback, the Bjerknes feedback and the tropical ENSO teleconnections, along with other intricate interactions with atmospheric and oceanic circulation and energy fluxes. In summary, significant systematic uncertainties in the mean and variability of multiple relevant variables persist over the tropical oceans in CMIP6 models. There are tight links amongst the errors in different variables, which has two important implications. First, model evaluation and development must be carried in a multivariate framework, and informed by physical principles, to ensure that increased accuracy that may be attained for a particular variable is coherent across the multiple aspects of the simulated climate. This implies taking full advantage of the rapid development of Earth observation capabilities over the past decades. Second, the central importance of tropical vertical circulation for precipitation and water vapor and, in turn, to SST found here supports the need for convective-permitting model-grid resolution in climate models, which is likely the best path to achieve improved simulation of deep tropical convection (Stevens et al., 2019; Weber & Mass, 2019; Palmer & Stevens, 2019; Schär et al. 2020). Significant benefits to the tropical climate may also be expected from improved representation of atmosphere-radiation interaction in climate models.

Funding
This study was funded by the Portuguese Science Foundation (F.C.T.) under project UID/GEO/50019/2019-Instituto Dom Luiz.
Author Contributions
MN conceptualized the study, collected the data, performed the analysis, and wrote the manuscript.

Availability of data and material
All datasets used in the present study are publicly available as detailed below.
The author acknowledges the World Climate Research Programme (WCRP) which coordinated and promoted CMIP6, the climate modeling groups for producing and making available their model output through the Earth System Grid Federation (ESGF), and the multiple funding agencies who support CMIP6 and ESGF.
The author also acknowledges ECMWF for producing ERA5, which was obtained from the Copernicus Climate Change Service Information website.

Availability of data and material
GPCP was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at http://www.esrl.noaa.gov/psd.
RSS monthly mean precipitable water product was provided by Remote Sensing Systems Version-7 Microwave Radiometer Data, Santa Rosa, CA, USA, from their website www.remss.com.

Code availability
Not applicable

Ethics declarations
Ethical approval and consent to participate
Not applicable

Conflicts of Interest
The author declares that he has no conflict of interest.

Consent for publication
Not applicable

References


Table 1. Summary of the CMIP6 simulations considered here. The resolution of the native grid of each model for the atmosphere is provided.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Variant label</th>
<th>Native Nominal Resolution</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS-ESM1-5</td>
<td>rliplf1</td>
<td>250 km</td>
<td>Ziehn et al. (2020)</td>
</tr>
<tr>
<td>BCC-CSM2-MR</td>
<td>rliplf1</td>
<td>100 km</td>
<td>Wu et al. (2018)</td>
</tr>
<tr>
<td>CAMS-CSM1-0</td>
<td>rliplf1</td>
<td>100 km</td>
<td>Rong et al. (2018)</td>
</tr>
<tr>
<td>CNRM-CM6-1</td>
<td>rliplf2</td>
<td>250 km</td>
<td>Voldoire (2018)</td>
</tr>
<tr>
<td>EC-Earth3-Veg</td>
<td>rliplf1</td>
<td>100 km</td>
<td>EC-Earth (2019)</td>
</tr>
<tr>
<td>INM-CM4-8</td>
<td>rliplf1</td>
<td>100 km</td>
<td>Volodin et al. (2018)</td>
</tr>
<tr>
<td>IPSL-CM6A-LR</td>
<td>rliplf1</td>
<td>250 km</td>
<td>Boucher et al. (2018)</td>
</tr>
<tr>
<td>MIROC6</td>
<td>rliplf1</td>
<td>250 km</td>
<td>Tatebe et al. (2019)</td>
</tr>
<tr>
<td>MPI-ESM1-2-HR</td>
<td>rliplf1</td>
<td>100 km</td>
<td>Muller et al. (2018)</td>
</tr>
<tr>
<td>UKESM1-0-LL</td>
<td>rliplf2</td>
<td>250 km</td>
<td>Tang et al. (2019)</td>
</tr>
</tbody>
</table>
Figure 1. Seasonal cycle of bias averaged over the 1988-2015 period for a) P and b) W using satellite observation products as reference. The average ± standard deviation of the bias over all months for each model are represented on the r.h.s. of each panel.
Figure 2. Seasonal cycle of the bias differences between CMIP6 models and ERA5 averaged over the 1988-2015 period for a) SST, b) $u_{10}$, c) $v_{10}$, d) $\omega_{500\text{hPa}}$, e) $\omega^\downarrow_{500\text{hPa}}$, f) $\omega^\uparrow_{500\text{hPa}}$, g) $\omega^2_{500\text{hPa}}$, h) DLR, i) SWA, and j) LWC. The average ± standard deviation of the bias over all months for each model are represented on the r.h.s. of each panel.
Figure 3. Differences in P monthly bias between each CMIP6 model and ERA5 against the respective bias differences (model – ERA5) in a) W, b) SST, c) u$_{10}$, d) v$_{10}$, e) $\omega_{500\text{hPa}}$, f) $\omega_{\downarrow}$500hPa, g) Fr$_{\uparrow}$500hPa, h) $\omega_{\uparrow}$500hPa, i) DLR, j) SWA, and k) LWC. Different models are represented by different colors. The zero bias lines are represented by black dashed lines. The Pearson correlation coefficient computed from all months in all models is provided.
Figure 4. Differences in W monthly bias between each CMIP6 model and ERA5 against the respective bias differences (model – ERA5) in a) SST, b) \( u_{10} \), c) \( v_{10} \), d) \( \omega_{500hPa} \), e) \( \omega_{500hPa} \), f) \( \omega_{500hPa} \), g) \( F_r \), h) LWC, i) DLR, and j) SWA. Different models are represented by different colors. The zero bias lines are represented by black dashed lines. The Pearson correlation coefficient computed from all months in all models is provided.
Figure 5. Differences in SST monthly bias between each CMIP6 model and ERA5 against the respective bias differences (model – ERA5) in a) $u_{10}$, b) $v_{10}$, c) $\omega_{500hPa}$, d) $\omega_{500hPa}$, e) $\omega_{500hPa}$, f) $\omega_{500hPa}$, g) DLR, h) SWA, and i) LWC. Different models are represented by different colors. The zero bias lines are represented by black dashed lines. The Pearson correlation coefficient computed from all months in all models is provided.
Figure 6. Seasonal cycle of $\sigma_n$ averaged over the 1988-2015 period for a) P and b) W using satellite observation products as reference. The average ± standard deviation of the bias over all months for each model are represented on the r.h.s. of each panel.
Figure 7. Seasonal cycle of the $\Delta$ differences between CMIP6 models and ERA5 averaged over the 1988-2015 period for a) SST, b) $u_{10}$, c) $v_{10}$, d) $\omega_{500}$, e) $\omega_{500}$, f) $F_{\omega}$, g) $\omega_{500}$, h) DLR, i) SWA, and j) LWC. The average ± standard deviation of the bias over all months for each model are represented on the r.h.s. of each panel.
Figure 8. Differences in monthly averaged $\sigma_n$ for W between CMIP6 models against the respective $\sigma_n$ differences in a) SST, b) $u_{10}$, c) $v_{10}$, d) $\omega_{500hPa}$, e) $\omega_{\uparrow 500hPa}$, f) Fr, g) Fr $\downarrow 500hPa$, h) DLR, i) SWA, j) LWC, and k) Ni34. Different models are represented by different colors. The Pearson correlation coefficient computed from all months in all models is provided. ERA5 was considered as reference.
Figure 9. Differences in monthly averaged $\sigma_n$ for SST between CMIP6 models against the respective $\sigma_n$ differences in a) $u_{10}$, b) $v_{10}$, c) $\omega_{500\text{hPa}}$, d) $\omega^\downarrow_{500\text{hPa}}$, e) $\text{Fra}^\uparrow_{500\text{hPa}}$, f) $\omega^\uparrow_{500\text{hPa}}$, g) DLR, h) SWA, i) LWC, and j) Ni34. Different models are represented by different colors. The Pearson correlation coefficient computed from all months in all models is provided. ERA5 was considered as reference.
Figure 10. Differences in monthly averaged $\sigma_n$ for P between CMIP6 models against the respective $\sigma_n$ differences in a) W, b) SST, c) $u_{10}$, d) $v_{10}$, e) $\omega_{500\text{hPa}}$, f) $\omega^{\uparrow}_{500\text{hPa}}$, g) Fr$\omega^{\uparrow}_{500\text{hPa}}$, h) $\omega^{\downarrow}_{500\text{hPa}}$, i) DLR, j) SWA, k) LWC, and l) Ni34. Different models are represented by different colors. The Pearson correlation coefficient computed from all months in all models is provided. ERA5 was considered as reference.