REFEREE #3

Dear Referee,

we would like to thank you for the careful reading of the manuscript and the constructive comments that substantially helped to improve and clarify the paper. Answers to all your comments are detailed hereafter. Corrections to the English grammar were adopted in the revised version of the manuscript according to the reviewer's recommendations, but are not reported or discussed here. All authors agree with the modifications made to the manuscript. The comments by the referee are reported in bold followed by our response (in blue). The text added to the revised manuscript is reported in italic font. The revised manuscript that includes track changes and line numbers is provided in pdf format.

In the following answers, we use 'Figure' to identify the figures in the updated manuscript and we use 'Plot' to identify the figures in this document.

The name of the experiments have been slightly modified, as reported in Table 1. They are used in the following answers and in the updated manuscript.

Experiment name	sea surface temperature used (T_s)	computation of C_D	computation of C_E and C_H	convective gustiness
COARE_S	SSTskin	COARE3.6	COARE3.6	Yes
ECMWF_S	SSTskin	ECMWF	ECMWF	Yes
NCAR	SST	NCAR	NCAR	No
ECMWF_NS	SST	ECMWF	ECMWF	Yes
CdNC_CeEC_NS	SST	NCAR	ECMWF	No
ECMWF_NS_NG	SST	ECMWF	ECMWF	No

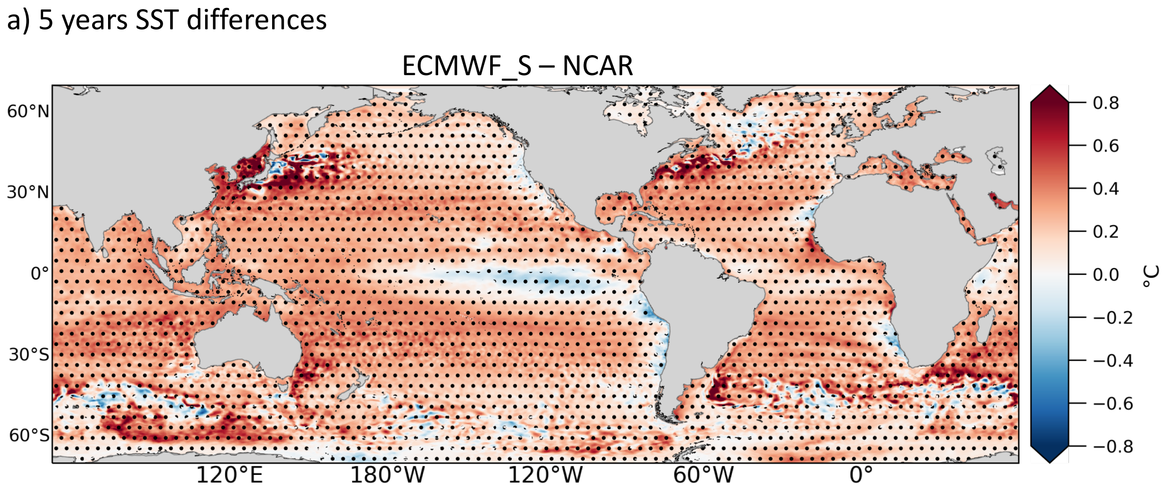
Table 1. Summary of the numerical experiments.

Major comments

1) A first important caveat of this study is the duration of the simulations, and consequently the significance of the results presented here. Even if the SST adjusts quickly and locally to surface turbulent fluxes modifications (from a few hours to a few days), the large-scale patterns and differences presented here might need more than one year to spin up and reach an new equilibrium state. The simulated interannual variability can also be modified between the sensitivity experiments, which can be misleading when interpretating the simulation differences. Another less important consequence is that spatial figures are quite noisy, which make them less readable.

Hence I would suggest to extend the different simulations to at least a 5-year period to make the results presented here more robust. As a comparison, Brodeau et al 2017 simulations which are referred in this manuscript discussions cover a 30-year period. If it not possible to extend the simulations for practical/technical reasons, I recommend to extend at least one simulation and compare the simulated turbulent fluxes and SST between the 1-year and the 5-year simulations to make this study more convincing.

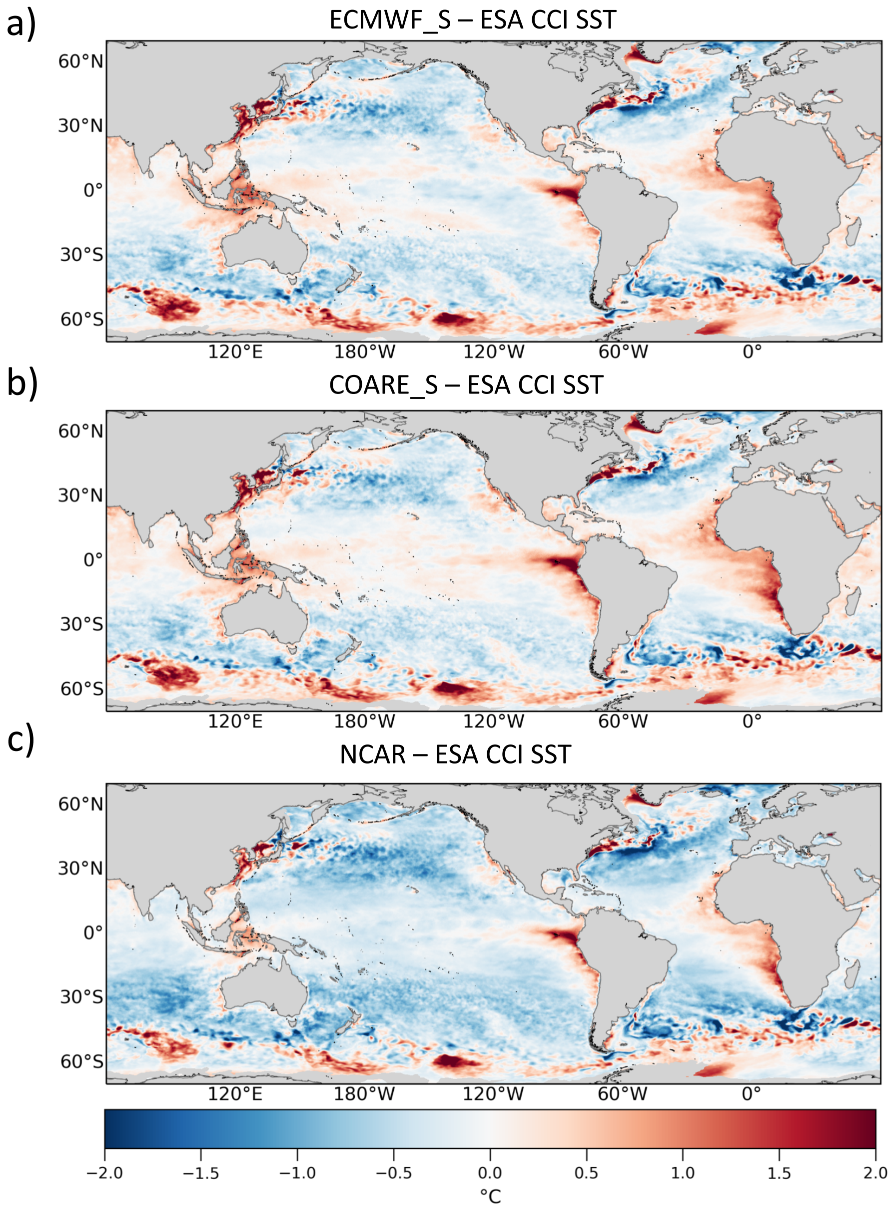
We agree with the referee, and we extended the ECMWF_S and NCAR control experiments (i.e. experiments which do not include modification in the bulk parameterization) to 5-year simulations to assess if the SST differences noticed with 1-year simulations are robust. Model results after 5 years confirm the presence of colder SST at the equator and over the EBUS in CdNC_CeNC_NS (Plot 1). We added the following text in the manuscript (lines 211-212): *"This spatial pattern of SST differences persists when extending the simulations up to 5 years (not shown).*"



Plot 1: Differences of the 5 year mean SST between ECMWF_S and NCAR experiments. Hatching indicates significant values (95% confidence level).

2) Another important issue is that the manuscript does not contain any validation of the simulated oceanic state, and especially the SST. I understand that a detailed validation is out of the scope of this study, but SST is the only assessed oceanic variable here, and because it is a key variable in STHF computation, we should know about the potential model biases compared to observations, and how these SST biases can modify STHF estimates (through air- sea temperature and humidity differences), and more importantly the turbulent fluxes sensitivity. SST is a well observed variable, especially at global scale and over large time period, so it would not require too much work to include an observed SST map over the same period as a reference. The idea here is not to classify the "best" bulk parameterizations, but to have a global idea of model SST biases.

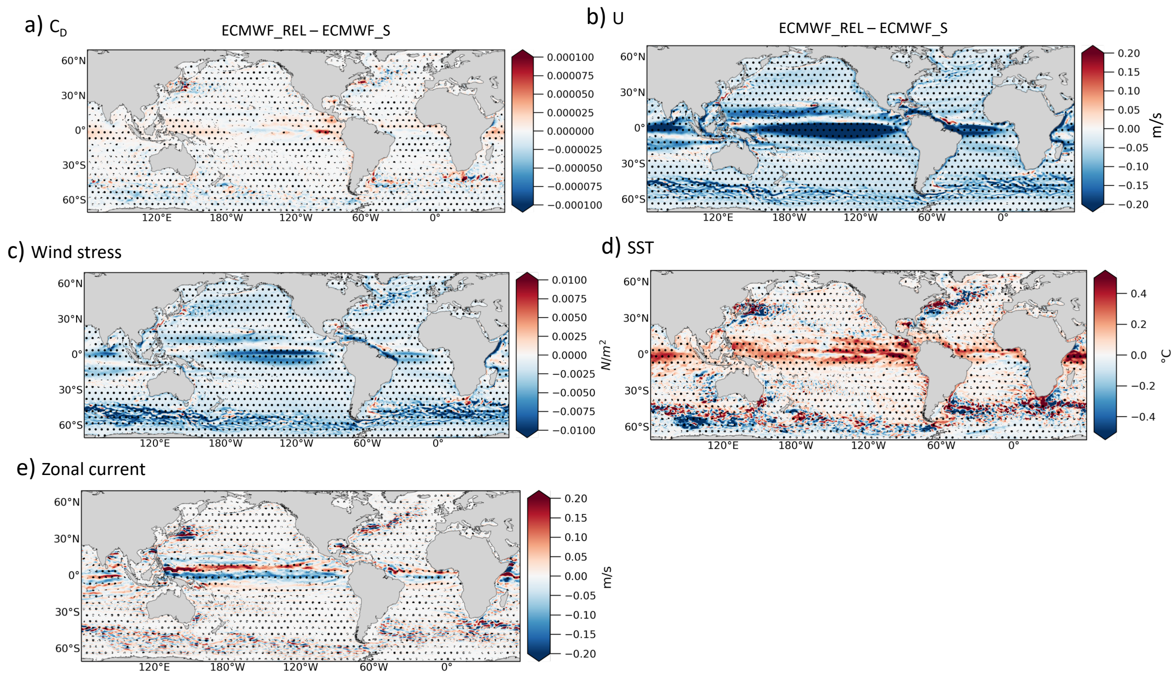
We agree with the referee. The manuscript can benefit from the evaluation against observation, so we included some in the revised manuscript. We compared SST from the "control experiments" against the *European Space Agency (ESA) Climate Change Initiative (CCI) SST dataset v2.0* (ESA CCI SST, Merchant et al. 2019). Results are presented in Plot 2 which was added as Figure 2 to the revised manuscript. Text has been added from lines 193 to 201: "*We compare the SST simulated by the ECMWF_S, COARE_S and NCAR control experiments with the European Space Agency (ESA) Climate Change Initiative (CCI) SST dataset v2.0 (hereinafter ESA CCI SST dataset) which consists of daily-averaged global maps of SST on a 0.05° x 0.05° regular grid, covering the period from September 1981 to December 2016 (Merchant et al., 2019). All the control experiments present a warm bias in the Eastern Pacific, in the Eastern Boundary Upwelling systems (EBUS), in the Western Boundary Currents (WBCs) and in the Antarctic Circumpolar Current (ACC) region. The SST reproduced by COARE_S and ECMWF_S shows a cold bias of about -1°C in the North Atlantic open ocean at mid- latitudes, and a warm bias of about 0.5°C in the Indian Ocean and the Western Pacific (Figure 2a,b); NCAR SST is also colder than observations, with a larger bias of about -2°C in the North Atlantic (Figure 2c). The bias is generally higher compared with other two experiments and covers wider areas.* "



Plot 2: Annual mean SST differences between a) ECMWF_S, b) COARE_S, and c)NCAR against ESA CCI SST.

3) My last major point concerns the surface current effect into the surface stress computation. Considering absolute or relative winds in stress formula in forced ocean simulation is still debated, but an additional sensitivity experiment using relative wind could give additional insight (as it is done for wind gustiness for example) to this manuscript compared to Brodeau et al. 2017. From my understanding, as the prognostic SST does not influence surface stress (or very weakly through stability functions), your sensitivity experiments using different Cd is totally similar to Brodeau et al. 2017, and hence leads to the same already-known results. This additional experiment would allow to assess the current-stress negative feedback (as it is done for SST-STHF feedback in 3.5), and how it changes the stress sensitivity to the bulk choice. This would substantially enrich the 3.4 section of the manuscript, which is currently of limited interest.

We thank the referee for this interesting comment. We performed an extra experiment, 1 year long, in which we applied the relative wind, instead of absolute wind, in the ECMWF_S bulk parameterization. We refer to the new experiment as ECMWF_REL . Plot 3 presents the results. As expected, the wind stress is reduced by the inclusion of the surface ocean velocity in the bulk formula, with respect to the absolute wind simulation: the wind speed in ECMWF_REL is weaker (up to -0.2 m/s) than ECMWF_S in the equatorial band (Plot 3b). As expected from the dependencies between C_D and the wind speed (Figure 1b of the manuscript), we find higher values of C_D in ECMWF_REL in the area of calm wind conditions and weaker values elsewhere. Differences of C_D and U between experiments are reflected onto the resulting wind stress field after bulk calculation (Plot 3c): the ECMWF_REL wind stress is weaker than ECMWF_S, especially where the U differences are higher (e.g. equatorial band). This difference in wind stress also leads to the SST differences (Plot 3d), hence ECMWF_REL results are warmer than ECMWF almost everywhere. Changes in wind stress also affect the current (Plot 3e): due to the weaker wind stress along the equator, the ECMWF_REL zonal currents are weaker than ECMWF ones. Even though the results provide insight into the effects tha bulk modifications can have in the upper ocean , we think that the current-stress negative feedback needs more and longer experiments (i.e. one for each bulk parameterization) to be properly assessed. We do not include a proper analysis in the manuscript, but we consider the effect of relative vs. absolute wind in the manuscript. Text was added in section 2. (lines 122-126): " *The effect of the ocean current interaction/feedback in the bulk formulation has been widely explored in the literature (e.g. Renault et al., 2019a, b; Sun et al., 2019). Although many previous studies highlighted the substantial difference in the surface input to the ocean between calculations that use absolute vs. relative wind, we have preferred to leave this aspect to further work since the implementation of this correction does substantially depend on the characteristics of the forcing fields (Renault et al., 2020).*"



Plot 3: Annual mean differences of a) drag coefficient (C_D), b) wind speed (U), c) Wind stress, d) SST and e) zonal current between ECMWF_S and ECMWF_REL.

Minor comments

1) A lot of English typos can be found in the text. A careful check is needed. Some punctuations are also missing.

Spatial figures must be improved to reach publication quality requirements. Here is some recommendations to improve them:

Spatial figures color extremes are often too much saturated and iso-contours are too thick. They are also very noisy due to the short experiments length. All these aspects make them hardly readable. Latitudes should also be extended from 70°S to 70°N as in Brodeau et al. to facilitate the comparison between those 2 studies.

A longitudinal average would also greatly improve and simplify figures interpretation as most of the results are mainly latitude dependant.

Some figures have resolution issue and appears blurry when zoomed in.

We agreed with the referee and we carefully checked the text and the punctuation. We also reproduced all the figures at higher resolution, and extended the latitudinal range from 70°S to 70°N. Palettes, saturation and contours were modified following the referee comments. New figures are clearer and it is easier to interpret. We preferred to maintain the lat-lon maps which show the spatial variability.

REFERENCES:

Merchant, C. J., Embury, O., Bulgin, C. E., Block, T., Corlett, G. K., Fiedler, E., Good, S. A., Mittaz, J., Rayner, N. A., Berry, D., et al.: Satellite-based time-series of sea-surface temperature since 1981 for climate applications, Scientific data, 6, 1–18, 2019.