REFEREE#1

Dear Dr. Small,

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.

Minor Comments

1) Regarding a response to another reviewer, in the WBCs the 1-year simulation is short due to the chaotic behavior in those regions and explains some of the noisiness in difference maps in those regions. It is worth mentioning this, if you have not already done so. It does not detract from the results of the paper.

We added a sentence about it at lines 258-260:
"It is worth mentioning that the one-year simulation might not be adequate to properly represent the mean state in WBCs regions due to the chaotic dynamics of these regions - this may explain some of the noise in the difference maps. However, this does not affect the robustness of the results."

2) Note the following differences are quite large, so it seems the CSWL does have a big impact on NCAR minus ECMWFG differences in the warm difference regions. The SST differences between ECMWF_NS and NCAR (Figure 6a) with respect to the SST differences between ECMWF_S and NCAR (Figure 4) present a reduction of the overall warm temperature differences.

We added a sentence about it at lines 241-242:
"*Nevertheless, the CSWL scheme has a large impact on the positive SST difference between ECMWF_S and NCAR.*"

3) Including gustiness in the ECMWF calculation produces the scalar wind differences in Figure 9a. Please clarify that differences in wind speed between ECMWF_NS and CdNC_CeEC_NS (Fig. 9a) can only be due to differences in gustiness as U does not depend on any other factors (i.e. does not depend on SST or drag coefficient).In fact it does not depend on using the ocean model so

you could remove reference to experiments ECMWF_NS and CdNC_CeEC_NS for Fig. 9a and just state that it is the difference due to gustiness. In other words it is an input to the model, not an output.

We rephrased at lines 275-277:
"Since U does not depend on SST and on C_D, including gustiness in the ECMWF calculation produces the scalar wind speed differences in Figure 9a."

4) Re Fig. 10a. Note that in the warm pool Equatorial regions there is some cancelling effect on wind stress of stronger U (Fig. 9a) but weaker CD (Figs. 9b,c).

We added reference to the Equatorial warm pool at lines 294, when we discussed Figure 9.

5) Indeed, a stronger acceleration (deceleration) of southeast trades north (south) of the equator in ECMWF_NS may lead to a stronger positive (negative) curl north (south) of the Equator (Chelton et al., 2001). Note also that Chelton et al explain changes in wind stress curl due to vertical mixing in the boundary layer above SST features. So, although WSC and Ekman pumping can affect SST, it can also work in the other direction, SST affecting the WSC.

Thank you. In the sentence you reported here, we are just referring to stronger wind stress that leads to stronger WSC, so that we removed the reference.

6) The reference for Levitus et al. 2013 is a bit strange. Perhaps use "Levitus, S. and coauthors

Thank you, we corrected it.

REFEREE #2

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.

Comments

1) The authors have addressed all the comments raised in the first round of review. There are a couple of issues that are still unclear concerning the response to the first round of comments (listed in the minor comments) and there is one main point (major comments) that I would ask the authors to address. In addition, I strongly recommend a deep revision of the use of English, because the level is still low for a scientific publication.

The manuscript has been revised for the use of English by a company specialized in English for academics (https://e4ac.com/). The revision is certified.

2) The presentation of the results is very hard to follow. I acknowledge that the subject is very technical and for this reason I think that a strong effort should be made in presenting the results in the clearest possible way. In particular, the figures show different metrics for different couples of experiments and the flow of the presentation goes back and forth talking about SST, tau, WSC and QT. In the set of experiments performed, the authors nicely change one aspect at a time to go from the ECMWF_S setup to the NCAR setup. In particular, between the ECMWF_S experiment and the NCAR experiments, all aspects considered in the work (skin/SST, Cd, Ce, gustiness) are different. The logical sequence to present the results to me is: ECMWF_S, ECMWF_NS (SST instead of skin), ECMWF_NS_NG (gustiness shut off), CdNC_CeEC_NS (which I strongly suggest renaming as ECMWF_NS_NG_CdNC, in which the Cd is computed with NCAR algorithm) and NCAR (which would be equivalent to

ECMWF_NS_NG_CdNC_CeNC). First, I would check whether the annual bias of each couple of experiments (considered in the order suggested above) sum up to give the annual bias between ECMWF_S and NCAR. This could be done for the various quantities of interest (SST, tau, WSC, QT) and would explicitly show whether all these differences in the algorithms sum up linearly (at least in the annual bias). If this is the case, then, one could compute the relative importance of each correction, showing which one contributes the most in the various regions (either with a fractional variation or the correlation coefficient with respect to the full difference ECMWF_S - NCAR). For example, the annual SST bias of both couples ECMWF_S - ECMWF_NS (figure 5a) and CdNC_CeEC_NS - NCAR (figure 6c) seems to be comparable to the full bias ECMWF_S - NCAR over WBCs. I would consider modifying the presentation of the results with two figures (one for SST and QT, and another for tau and WSC) where the maps of the following differences are shown in this order:

1. ECMWF_S - ECMWF_NS;

2. ECMWF_NS - ECMWF_NS_NG;

3. ECMWF_NS_NG - ECMWF_NS_NG_CdNC (currently named CdNC_CeEC_NS);

4. ECMWF_NS_NG_CdNC - NCAR.

In this way, the role of each modification might appear more clearly and ease the interpretation of the results.

An alternative approach to assess the importance of each difference between the algorithms would be to change one thing at a time (in some new experiments that would be ECMWF_NG, ECMWF_CdNC and ECMWF_CeNC, in addition to the ECMWF_NS experiment), compare them with ECMWF and see which one resemble the most to the NCAR-ECMWF difference. But this would require new experiments.

Thanks for the suggestions, we computed all the TASFs (i.e. τ and QT) and WSC annual mean differences for each pair of experiments. The annual differences not discussed in the manuscript are reported in the supplementary material. In particular, we added Plot 2 (Annual mean differences of wind stress, and wind stress curl between ECMWF_S and ECMWF_NS) as Figure S1 for ECMWF_S - ECMWF_NS (section 3.2), and Plot 3 (Annual mean differences of wind stress, and wind stress curl between ECMWF_S and ECMWF_NS) as Figure S3 for CdNC_CeEC_NS - NCAR (section 3.3). We also checked that the annual bias of each pair of experiments sums up to give the annual bias between ECMWF_S and NCAR. This is verified for all the variables. We added a sentence on this at lines 221-223 in Section 3.1: "*It is worth mentioning that the annual mean differences (i.e. SST, τ, WSC and QT) of each pair of experiments discussed in the following sections sum up linearly to give the annual mean difference between ECMWF_S and NCAR (not shown).*"

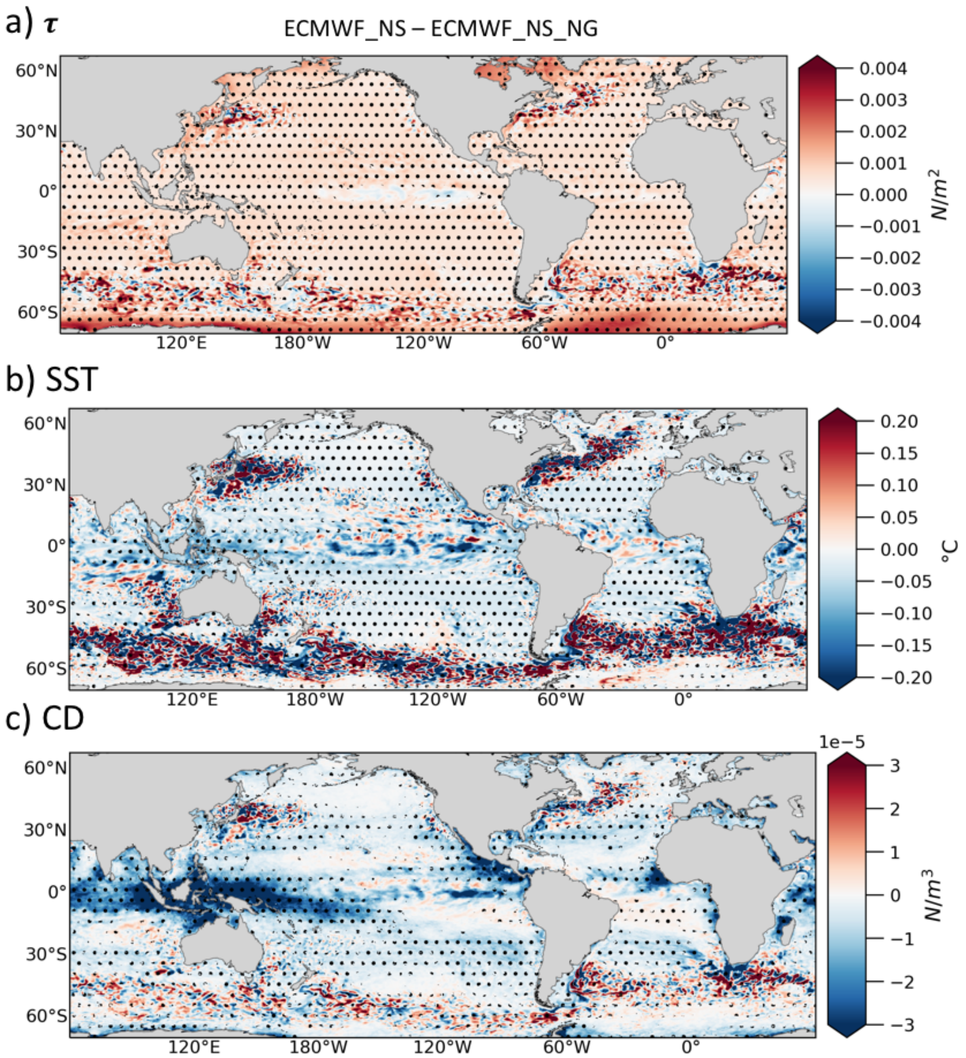
Nevertheless, we decided to not include the ECMWF_NS - ECMWF_NS_NG as a separate couple of experiments. The wind stress, C_D and SST differences between ECMWF_NS - ECMWF_NS_NG (Plot 1) are not relevant enough (less than 10% of

the difference ECMWF_S - NCAR, Figure 3) to have a dedicated paragraph in the manuscript.
The couples of experiments in the manuscript are still:

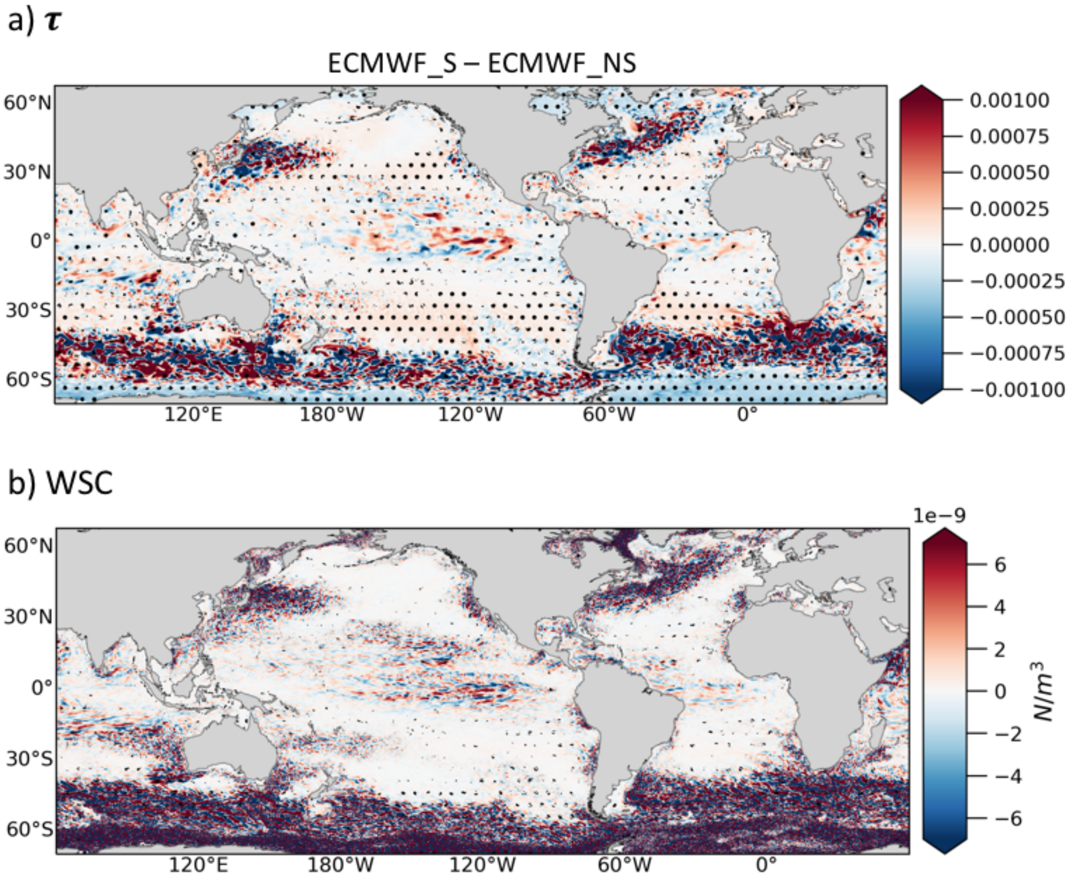
1. ECMWF_S - ECMWF_NS (section 3.2);
2. CdNC_CeEC_NS - NCAR (section 3.3);
3. ECMWF_NS - CdNC_CeEC_NS (section 3.4).

However, the difference in wind stress between ECMWF_NS and ECMWF_NS_NG is still discussed in section 3.4 and shown in Figure S2.
Moreover, the name ECMWF_NS_NG_CdNC seems to refer just to ECMWF parameterization, when, actually, the used bulk algorithm is a mixture of ECMWF and NCAR parameterizations. We therefore prefer to retain the name CdNC_CeEC_NS.

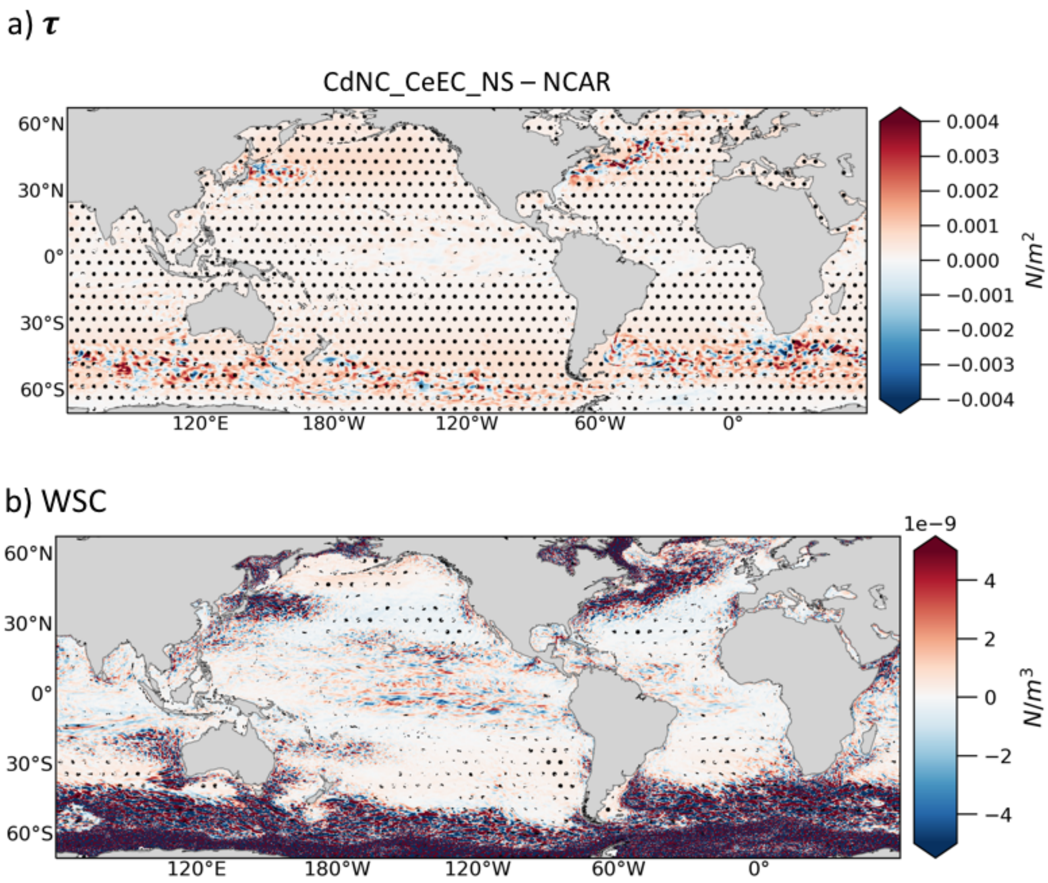
We clarified our logical flow at lines 177-187:
"*Here we discuss the parameterization-related discrepancies in the control experiments in terms of TASFs (*i.e. τ and QT*), WSC, SST and meridional heat transport (section 3.1). Then, we try to analyze the contribution of various aspects of the parameterizations in driving these SST and meridional transport discrepancies. In particular, the comparison between ECMWF_S and ECMWF_NS is used to determine the skin temperature contribution (section 3.2), while the comparisons between* CdNC_CeEC_NS *and NCAR (section 3.3) and between* CdNC_CeEC_NS *and ECMWF_NS (section 3.4) teach us about the Bulk Transfer Coefficients contribution. In section 3.4, we also compare ECMWF_NS_NG and ECMWF_NS experiments to show the effect of the inclusion of convective gustiness in the wind speed calculation on wind stress computation (shown in the supplementary material). For each couple of experiments, we only show the differences in TASFs and their components (e.g. U, C_D, C_E) which are relevant to understand the SST or meridional heat transport discrepancies. The complementary TASFs differences are reported in the supplementary material. We analyze annual mean differences between experiments and assess their statistical significance using t-test.*"



Plot 1: Annual mean differences of a) wind stress (τ), b) Sea Surface Temperature (SST) and c) wind stress transfer coefficient (CD) between ECMWF_NS and ECMWF_NS_NG. Hatching indicates significant values (95% confidence level)



Plot 2: Annual mean differences of a) wind stress (**τ**), and b) wind stress curl (WSC) between ECMWF_S and ECMWF_NS. Hatching indicates significant values (95% confidence level)



Plot 3: Annual mean differences of a) wind stress (**τ**), and b) wind stress curl (WSC) between CdNC_CeEC_NS and NCAR. Hatching indicates significant values (95% confidence level)

3) Add somewhere in the methods how you computed the statistical significance of the differences.

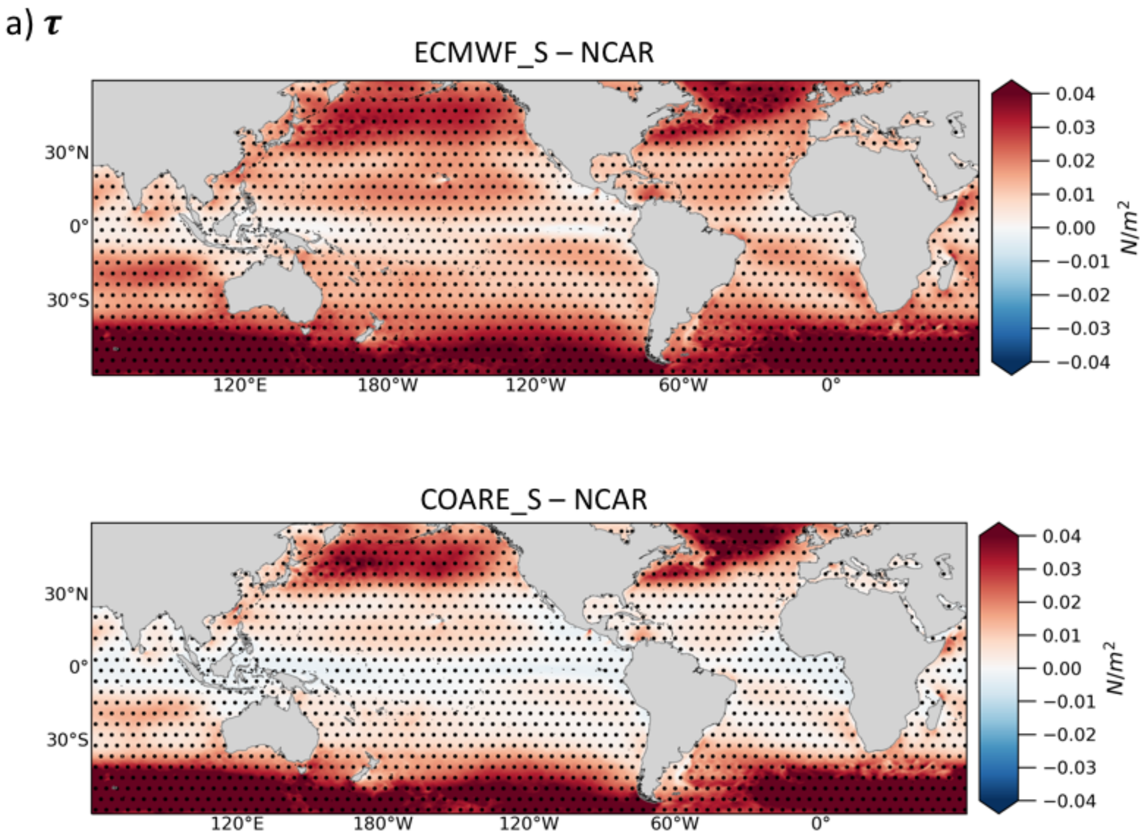
Thank you, we added a sentence on statistical significance at line 186-187: *"We analyze annual mean differences between experiments and assess their statistical significance using t-test."*

4) In equation (1a), according also to Brodeau et al. (2017), u should be replaced with U (i.e. it should be capitalized, to be consistent with the other formulae.

Thank you, we corrected it.

5) Figure 3: Especially in panels a) and b) I would suggest using a colormap that is symmetric about zero. As it is now, in fact, it enhances the negative differences, which is misleading in the interpretation of the figure.

Thank you for the suggestion. With the symmetric colorbar (Plot 4), we cannot appreciate the **τ** difference between the two couples of experiments in the equatorial band, an important feature used to discuss COARE_S - NCAR in section 3.4. Therefore we decided to retain the Figures with the non symmetric colorbar.



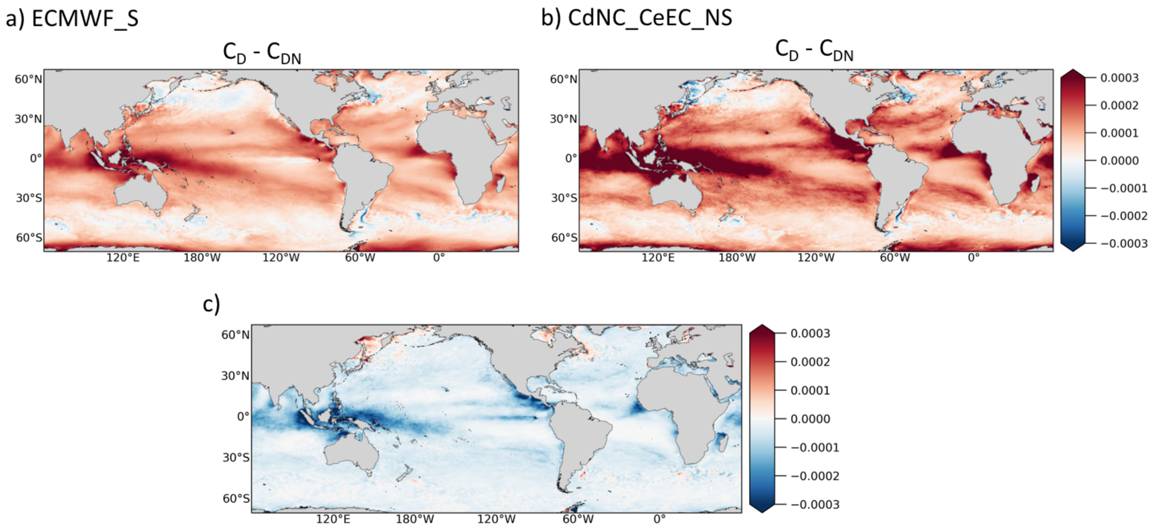
Plot 4: Annual mean differences between experiments of a) wind stress (τ) between ECMWF_S and NCAR experiments (top panel) and COARE_S and NCAR experiments (bottom panel). Hatching indicates significant values (95% confidence level)

6) Figure S1: why only such a small area is selected? The pdfs over the entire WBC using high frequency data would be more informative. What is the time frequency of the data in the scatter plot? 5 days?

The analysis was performed on a set of different areas in the WBC regions. Results are all consistent. We specified that the plot shows the relationship between total turbulent fluxes and the air-sea virtual temperature difference for selected grid points in one of the boxes as a matter of clearness. Yes, 5 days is the frequency of the data.

7) LL317-319: Have you tried to compute the difference between CdN and Cd? Because it looks that its pattern might also resemble the pattern of U. Thus, I would not be sure that the differences between parameterizations are only due to the different ways to compute Cd, as stated in the text.

In lines 317-319, since the pattern differences in Figure 9b and Figure 9c are very similar, we are suggesting that the differences in C_D between parameterizations are related to the computation of neutral coefficient (C_{DN}) calculation rather than to its stability correction (term to add to C_{DN} to get C_D coefficients). Plot 5 shows the difference between C_D and C_{DN} for (a) ECMWF, (b) NCAR and c) shows a-b. The patterns of C_D - C_{DN} differences are really similar between experiments. The stability correction of NCAR is stronger than the stability correction of ECMWF everywhere. Plot 5c basically shows the contribution of the atmospheric stability function to the C_D differences pattern (Figure 9b). Even though stability has for sure its role in determining the C_D values, the pattern of C_D differences between experiments (positive and negative areas) are related to the different C_{DN} computation.



Plot 5: Annual mean differences of CD - CDN for a) ECMWF_S and b) CdNC_CeEC_NS; c) differences between a) and b).

8) L312-313 - in the comparison between CdNC_CeEC_NS and ECMWF_NS there are two changes: the use of a different algorithm to compute Cd and the inclusion of gustiness in the stress computation. One should first show the effects of including the gustiness (comparison between ECMWF_NS and ECMWF_NS_NG) and then compare CdNC_CeEC_NS to ECMWF_NS_NG to show the impact of a different algorithm to compute Cd.

See answer to comment 2 above.