Comments to gmd-2021-436.R1
"The bulk parameterizations of turbulent air-sea fluxes in NEMO4: the origin of Sea Surface Temperature differences in a global model study" by Bonino et al.

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 list of language corrections provided in the minor comments is not exhaustive. Line numbers in this file refer to the tracked-change manuscript.
Overall, the quality of the work is good and, thus, I suggest minor revision.

Major comments

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.

Minor comments

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

There are still many English mistakes in the text, such as (to mention some):

- L6: drives -> drive
- L44: does only partially close -> only partially closes
- L55: estimations -> estimates
- L146: on three of bulk -> on three bulk
- L164: and to lower extend to -> and, to a lower extent, to
- L170: weak -> weak-wind
- L192: two -> three
- L261: as component -> as a component
- L279: are -> is
- L283: SST differences -> SST difference
- L297: suggests that the QT drive -> suggests that QT drives

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.

LL111-112: “... which may be referred to the ocean currents.” -> “... which may be absolute or relative to the ocean currents.”

L128: “The form of the exchange coefficients” -> “The dependence of the exchange coefficients on the wind speed”

L247: Remove “cool-skin/warm layer”

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.

Figure 4: Add in the caption the letters c) and d)

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?

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.

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.