Review from Referee #1

This paper is a resubmission of a paper I already made comments about as a reviewer (https://doi.org/10.5194/gmd-2020-327). I thank the authors for re submitting an updated version of their work, and congratulate them for the general improvement of the manuscript, and for taking into account most of my comments. Especially, I appreciate that they use longer time spans for the sensitivity tests. The statistical analysis of the results is also much improved.

The present paper is a case study investigating the effects of several parameterizations representing the impact of waves on the ocean surface layer (Langmuir mixing and Stokes-Coriolis force with entrainment) and atmosphere surface layer (change of roughness length, effect of surface currents on the turbulent fluxes). The CFS2.0 ocean-atmosphere climate model and the WAVEWATCHIII wave model (WW3) are used in coupled mode for global simulations at resolution 0.25° to 0.5° for two time periods of 53 days, in boreal summer and winter. Four different simulations enable to assess the different effects on the SST, ocean mixed layer depth (MLD), 10-m wind speed, significant wave height (SWH) and latent heat flux in an incremental way. The conclusion is that refining the CFS2.0 representation of the surface exchanges by including additional terms due to waves leads to an overall (although modest) improvement of the SST and MLD biases with respect to observations. The improvement is larger in the Southern Ocean in boreal winter. Some improvement is also obtained on the surface wind speed and SWH, compared to ERA5.

The results presented here are not especially new, as recent sensitivity studies using the same kind of modeling platforms showed similar effects (e.g. Shimura et al., 2017; Torres et al., 2018; Bao et al., 2020; Couvelard et al., 2020;). But the sensitivity of the system CFS2.0-WW3 to these wave effects has not been studied so far.

Nevertheless, I have several major comments about the description of the coupled system, the evaluation of the impact of the different parameterizations, and the interpretation of the results.

Response: We sincerely appreciate the reviewer for her/his constructive comments on the manuscript. Our responses are listed as follows in blue. Text are revised accordingly.

General comments

1- Part 2 describes the representation of several physical effects impact the wave-ocean or wave-atmosphere exchanges, which has been implemented in the coupled system. The effects of Stokes-Coriolis and Langmuir mixing come as additional terms in the Richardson number or turbulent velocity scale of the KPP mixing scheme, the wave effect on the atmospheric roughness length comes through a change of the Charnock parameter, and the effect of the surface currents corresponds to the use of the relative surface velocity in computing the turbulent fluxes. For the effect of the Langmuir mixing, the authors assume that wind and waves are aligned, arguing the effect of misalignment has been shown to be non significant by Li et al. 2016. However, other studies like Polonichko (1997), Van Roekel et al. (2012), and Li et al (2017) showed that the Langmuir cell intensity strongly depends on the alignment between the Stokes drift and wind direction. The latter study especially concluded that assuming alignment of wind and waves leads to excessive mixing, particularly in winter. As the strongest effect of the Stokes-Coriolis and Langmuir mixing parameterization is obtained on the Southern Ocean in winter, I suggest to mention the results of *these works in comparing the results of the VR12-AL-SC-EN experiment with respect to the CTRL one. Also Couvelard et al (2020) showed that there is a significant difference between annual averages of the module of the surface Stokes drift and of the part that is aligned with the wind (their Fig. 2). Please discuss.*

Also, the description of the exchanges of the different parameters between model compartments is unclear to me. I understand that all additional terms are computed in WW3, and that the Stokes drift and Langmuir mixing terms are transferred to MOM4, that the Stokes drift is transferred to GFS for computed the surface relative wind, and that the Charnock parameter is also transferred to GFS for computing the surface roughness (Fig.1). What is unclear is what is exchanged between GFS and MOM4? Especially, are the (regular) surface currents transferred from MOM4 to GFS and used for estimating a relative wind velocity in computing the turbulent fluxes by GFS? If so, is it consistent with the transfer and use of the Stokes drift from WW3? Please provide the corresponding information, with an update of Fig.1. What is the meaning of the blue arrow from the coupler to GFS in Fig.1?

About the effect of the surface current on the atmosphere: I guess from eq. 7 to 9 and section 2.4 that only the effect of the currents (and especially of the Stokes drift) on the turbulent fluxes is taken into account, and not the effect of the current on the surface wind through the tridiagonal matrix (see the work of Lemarié 2015). If so, the fact that the coupling is not complete should be clearly stated in section 2.4.

Response: We agree that the misalignment between the Stokes drift and wind direction is important for the intensity of Langmuir cell (Polonichko, 1997; Van Roekel et al., 2012; Li et al., 2017), particularly in the Southern Ocean (Couvelard et al., 2020). In the comparison of the VR12-AL-SC-EN experiment with respect to the CTRL, we add "Since we didn't consider the misalignment between the Stokes drift and wind direction for the Langmuir mixing, which is important in the Southern Ocean (Couvelard et al., 2020), the effects of the Stokes drift might be overestimated". Without the misalignment, however, the simulation still underestimated the mixing, and so we didn't consider misalignment in the study.

To clarify the exchanges between model components, we updated the original Fig. 1 as shown in Fig. R1. Between GFS and MOM4, originally GFS receives SST from MOM4, and sends fluxes of heat, momentum, and freshwater to MOM4 via mlc_coupler. In the study, the (regular) surface currents from MOM4 are transferred to GFS via C_Coupler2 to estimate the relative wind velocity for the turbulent fluxes (Eqn. 7-10) in GFS. Similarly, the Stokes drift from WW3 is also transferred to GFS. The blue arrows in Fig. R1 indicate the surface currents transferred from MOM4 to GFS and WW3. The text and Fig. 1 are revised accordingly.

To complete the coupling, we add the surface current and Stokes drift to the tridiagonal matrix (Lemarié 2015) in CFS for the FLUX experiment. The experiments are re-run and the associated figures and text are revised.

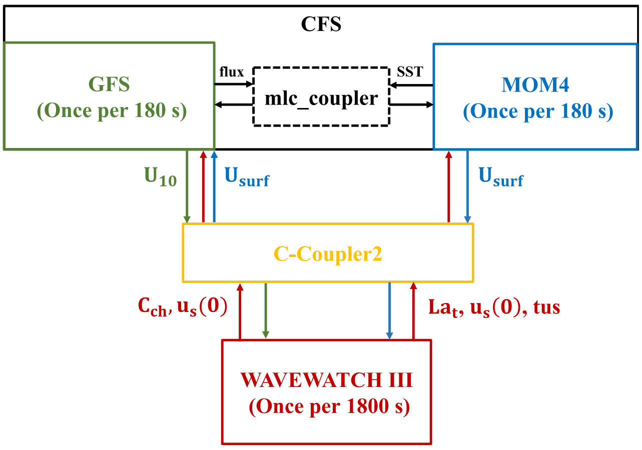

Figure R1. A schematic diagram of the atmosphere-ocean-wave coupled modeling system. The arrows indicate the coupled variables that are passed between the model components. In the diagram, C_{ch}, La_t, $u_s(0)$, tus, U_{10}, and U_{surf} are Charnock parameter (red arrows), turbulent Langmuir number (red arrows), surface Stokes drift velocity (red arrows), Stokes drift transport (red arrows), 10-m wind (green arrows) and surface current (blue arrows), respectively.

2- The statistical analysis of the different sensitivity tests is much clearer and easier to understand than in the previous version of the paper. I still feel rather uncomfortable with the different diagnostics used by the authors. For instance, the correlation between the bias reduction and the absolute bias shown in Fig. 3,4, 8 to 11 is almost never commented, and I am not sure about its meaning: from the text, I guess that its corresponds to the correlation between the relative change between CTRL and ALL (the so-called PRD) and the absolute bias, but only when the time evolution of the bias corresponds to an increase. Is it so? What is the additional information with respect to the PRD as shown elsewhere? Please elaborate.

For most of the parameters compared in this study, the maps represent the relative improvement (PRD). For the 10-m wind speed and MLD however, differences with the CTRL are given and I find these maps easier to read. Please justify why you use different diagnostics or homogenize. The relative improvement (PRD) depends strongly on the initial value of the bias. Why not showing maps of the biases for the different simulations? It would help to appreciate where the biases have been corrected or not. Please give the values of the final biases (and RMSE) for every parameter/experiment, in addition to the PRD.

The comparison of the 10-m wind speed and SWH with the NDBC follows some of my previous recommendation, and I thank the authors for that. I think, however, that the way this comparison is presented could be greatly improved. I suggested that maybe, the wind speed can influence the bias and the difference between CTRL and ALL, and this comment is still valid. There is some effect of the value of the bias with CTRL, even though the current presentation of the results makes it difficult to apprehend. Rather than a table giving the relative difference for different quantiles of biases, I would suggest using a graph comparing directly the 10-m wind speed of the simulation outputs (y-axis) with the 10-m wind speed of the NDBC buoys (x-axis) in wintertime (same in summertime, and for the SWH), every dot on the graph representing a buoy (4 graphs in total). The results of the different simulations can be plotted in the same graph, with different colors. This would enable a direct comparison, including the effect of the wind speed (x-axis) and of the bias (distance to the y=x line). The changes between the different simulations can be given by the mean biases and *standard deviations with respect to observations, rather than the relative mean changes.*

Response: The correlation between the absolute CTRL biases (CTRL minus OISST/ERA5) and PRD in ALL is replaced by the correlation between the absolute CTRL biases and the differences of absolute biases between ALL and CTRL ($|\widehat{y_a} - y| - |\widehat{y_c} - y|$, where y is OISST or ERA5, $\widehat{y_c}$ ($\widehat{y_a}$) is simulated result in CTRL (ALL)). In the updated figures (e.g. Fig.R2b-R4b for boreal winter), all values with statistically significant at 95% confidence level are shown. Note that the absolute bias of CTRL increases with time in most areas (Fig. S3 in the supplementary), where the negative (positive) correlation values indicate that the simulated bias in CTRL decreases (increases) with time in ALL experiment. While in areas where the absolute bias of CTRL decreases with time, the situation is reversed.

To clarify, we have replaced all maps of PRD with the differences between sensitivity experiments and CTRL. The global averaged biases and RMSEs are shown in the upper right of each map (e.g. Fig.R2-R4 for boreal winter). The figures for boreal summer are also updated. The text is revised accordingly.

As suggested, the comparisons of the 10-m wind speed and SWH with the NDBC buoys are shown in Fig. R5. The x-axis is WSP10/SWH of buoys, and the y-axis is the simulated WSP10/SWH in Jan-Feb, 2017 (Fig. R5a&b) and Aug-Sep, 2018 (Fig. R5c&d) for all experiments. The mean biases with standard deviations and RMSEs for every experiment are shown in Table R1. The differences between 4 sensitivity experiments and CTRL are all statistically significant at 95% confidence level. In CTRL, the WSP10/SWH is generally overestimated in both winter and summer with positive mean bias (Table R1). Except the FLUX experiment in Jan-Feb, 2017, all other experiments lead to a decrease of mean bias by reducing WSP10 and SWH. With the increase of WSP10, the reduction of overestimation is more obvious. In ALL experiment, the RMSEs decrease compared with CTRL.

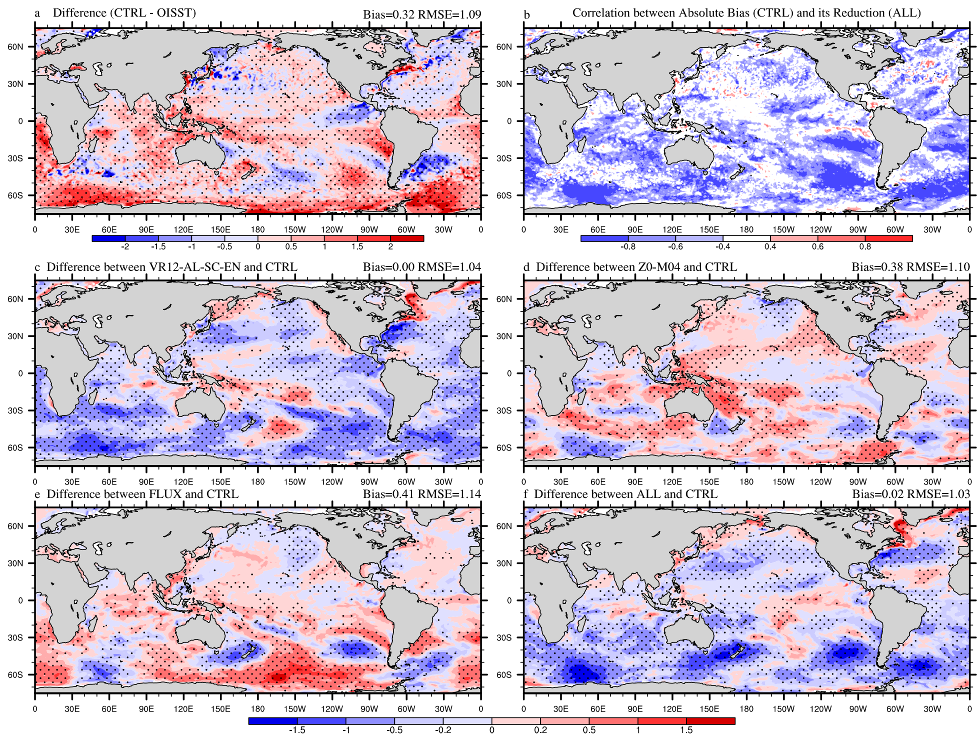

Figure R2. The 53-day average SST (°C) bias in CTRL (a), the time correlation between the absolute CTRL biases and the differences of absolute biases between ALL and CTRL (b), and the differences between VR12-AL-SC-EN (c)/Z0-M04 (d)/ FLUX (e)/ ALL (f) and CTRL in Jan-Feb,

2017. The first 3-day simulation is discarded. The dotted areas are statistically significant at 95% confidence level.

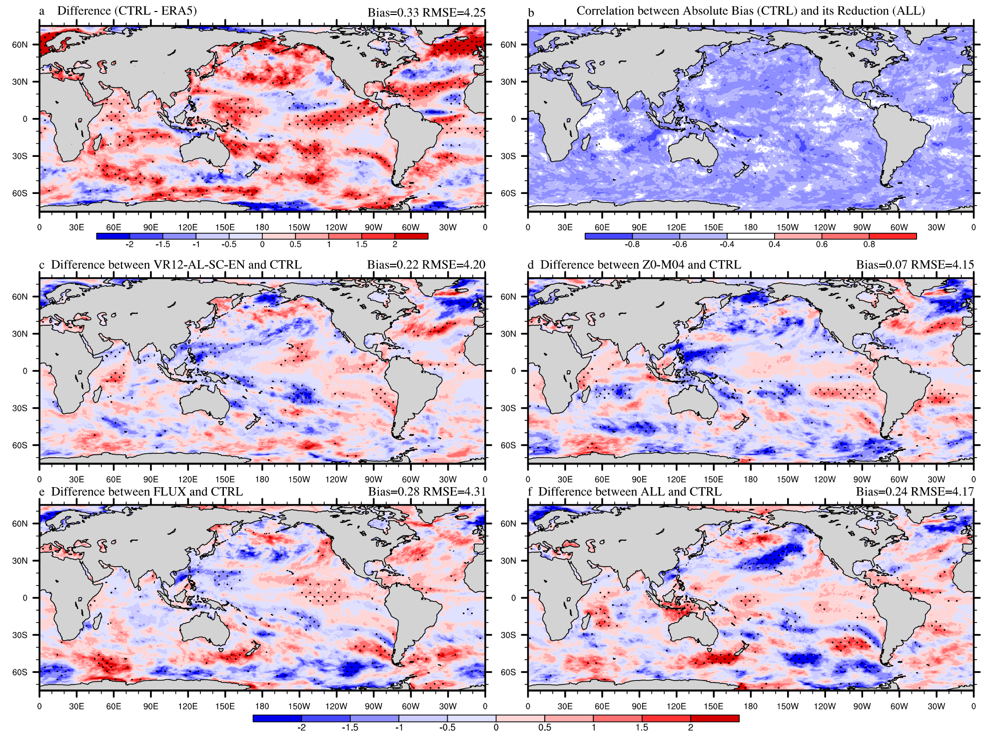

Figure R3. The same as Figure R2 but for WSP10 (m/s).

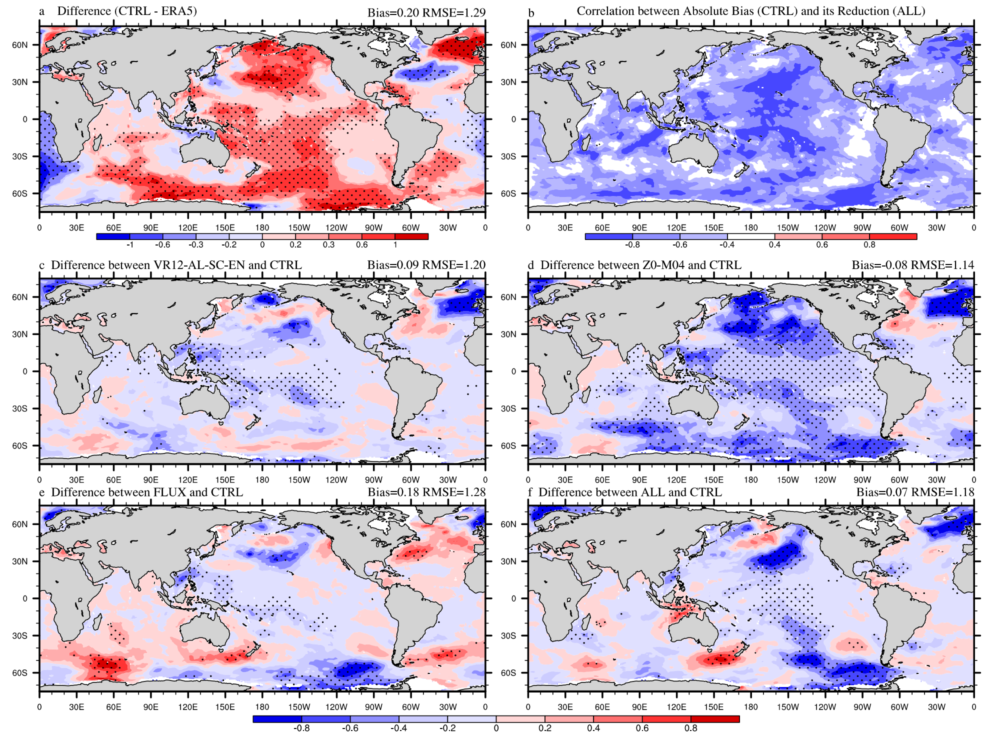

Figure R4. The same as Figure R2 but for SWH (m).

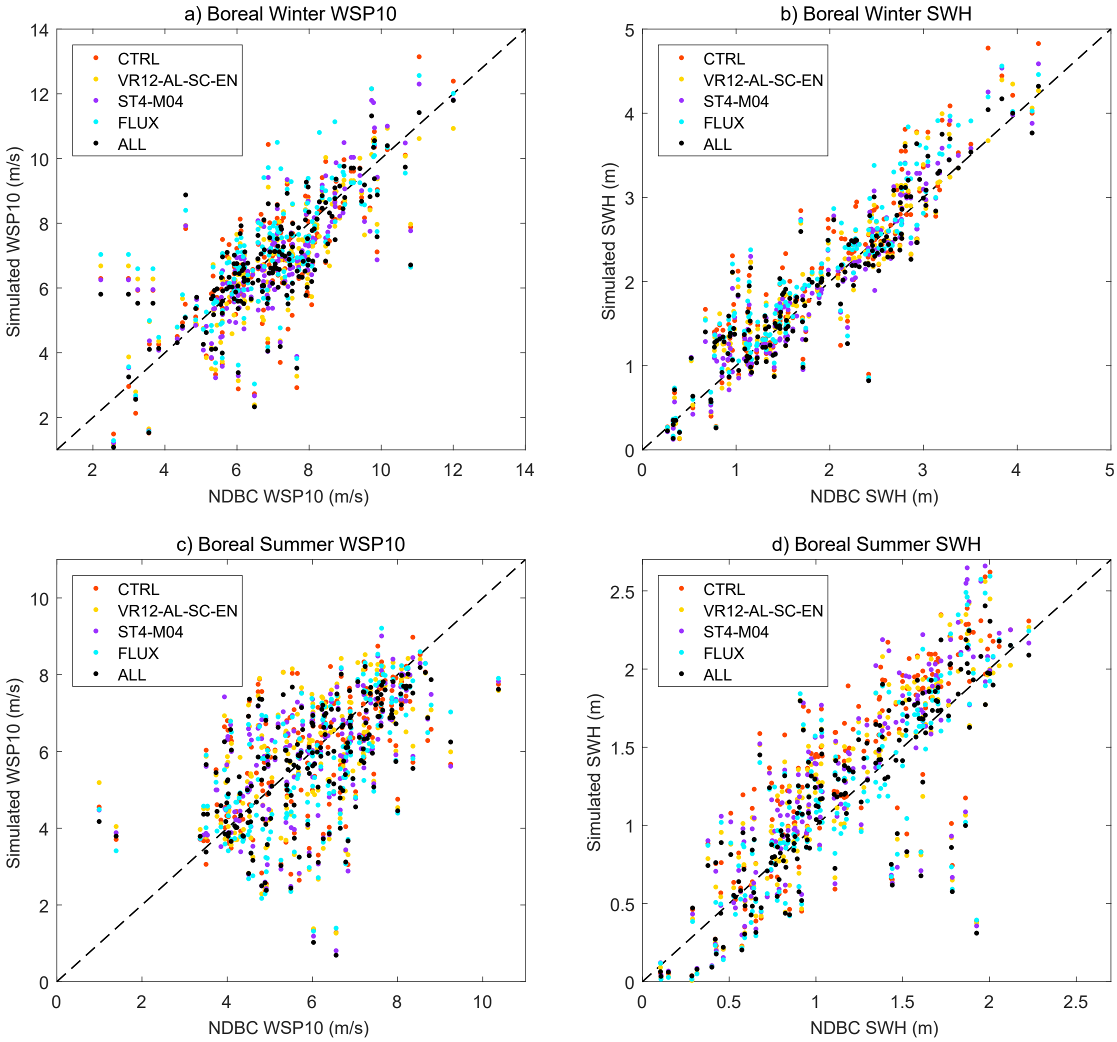

Figure R5. Scatter plots of simulated WSP10/SWH (y-axis) vs buoy WSP10/SWH (x-axis): (a) the WSP10 in Jan-Feb, 2017, (b) the SWH in Jan-Feb, 2017, (c) the WSP10 in Aug-Sep, 2018, and (d) the SWH in Aug-Sep, 2018. The dotted line is y=x.

Table R1. The 53-day mean bias with standard deviation (STD) and RMSE for WSP10 and SWH compared with NDBC buoy observation: the bias is calculated as simulation minus NDBC.

Boreal Winter WSP10	**Bias with STD**	**RMSE**
CTRL	0.15±1.27	1.28
VR12-AL-SC-EN	0.02±1.18	1.18
Z0-M04	-0.01±1.13	1.13
FLUX	0.39±1.25	1.30
ALL	0.05±1.11	1.11
Boreal Winter SWH	**Bias with STD**	**RMSE**
CTRL	0.21±0.38	0.44
VR12-AL-SC-EN	0.14±0.35	0.37
Z0-M04	0.10±0.30	0.32
FLUX	0.24±0.34	0.42
ALL	0.12±0.34	0.36
Boreal Summer WSP10	**Bias with STD**	**RMSE**

CTRL	0.17±1.26	1.27
VR12-AL-SC-EN	0.06±1.33	1.33
Z0-M04	0.02±1.24	1.24
FLUX	-0.12±1.22	1.22
ALL	-0.10±1.18	1.18
Boreal Summer SWH	**Bias with STD**	**RMSE**
CTRL	0.28±0.25	0.38
VR12-AL-SC-EN	0.20±0.24	0.31
Z0-M04	0.22±0.26	0.34
FLUX	0.14±0.25	0.29
ALL	0.12±0.21	0.24

3- *Interpreting the results could, again, be made in a more accurate and concise way. For instance, the discussion in section 4.3 is rather long and not very easy to follow. Probably the effects in boreal summer with respect to boreal winter could be presented more briefly.*

Overall, I am not sure commenting in details improvements of a few percent is meaningful, but adding information about the correlation between the changes of different parameters can help to interpret the results. For instance, what is the correlation between the 2-m temperature and the SST changes? It confirm that the SST change is actually at the origin of the 2-m temperature change. Also, the correlation between the bias changes in 10-m wind speed and SWH is probably high. Please give values and discuss.

At some parts, the interpretation is not complete. For instance, the part about the latent heat flux does not lead to clear, concise results.

From Fig. S3, I understand that the time evolution of the absolute value of the biases of the different parameters considered is overall positive, both in winter and in summer. At most places, these trends are significant, and even large, like more than 0.02°C/day (corresponding to more than 1°C difference for the 53-day simulated period) or 0.02 m/day for SWH (more than 1 m difference). What is the implication for the mean biases, and their changes from CTRL to ALL? Does it mean that the simulations are drifting in time from their initial state, because there is no assimilation of data? Or, conversely, that their stationarity is not reached yet because the simulated period are too close from the initial state, despite the hot start? Please comment on that.

I specifically asked about the possible effects of including the parameterization of processes related with waves on the turbulent heat fluxes. The authors added a section about that effect, and I thank them for investigating it. However, it appears not significant, at least not for the time scales considered in this study.

Response: We will modify the Section 4, especially the Section 4.3, to make the discussion concise. As suggested, the correlation coefficients between the SST and the T02 changes (ALL minus CTRL) in boreal winter and summer are calculated. The values are 0.60 and 0.51 respectively, significant at 99% confidence level, indicating the SST change is actually the origin of the 2-m temperature change. The correlation coefficients between WSP10 difference and SWH difference in ALL compared with CTRL are 0.77 and 0.73 in boreal winter and summer, significant at 99% confidence level, indicating the SWH change is originated from wind speed. The text is revised accordingly.

For Figure S3 in CTRL, we think the simulations are drifting in time from the initial state due to no data assimilation, and the effects in ALL partially limit this. In Figure 5 of the main text, the RMSE of SST increases in the first twenty days and then decreases with time. So the stationarity has been reached in the simulation.

The turbulent heat fluxes change is primarily resulted from the change of 10-m wind speed, and it is non-significant due to the relative short simulation period. So the related discussion is removed in the text.

Detailed comments

l. 26-29: are you sure that the SST change is at the origin of the 10-m wind speed change? I understand from 4.3 that the change of z0 also plays a role. Please check.

Response: We apologize for the confusion. The corresponding lines are changed as "The largest regional mean SST improvement occurs in the Southern Ocean. For WSP10 and SWH, the wave-related processes generally lead to reduction of biases in regions where wind speed and SWH are overestimated. The decreased SST stabilizes marine atmospheric boundary layer, weakens wind speed and then SWH. The increased roughness length due to waves leads to reduction in the originally overestimated wind speed and SWH. Meanwhile, the effects of Stokes drift and current on air-sea fluxes also play a role in change of WSP10 and SWH."

l. 65-66: the studies cited here are at climate scale, not for numerical prediction. The only model including wave effects and used for numerical prediction is ECMWF (IFS-WAM).

Response: We agree. The text is revised to "The overall effects of wave-related processes have been shown to be important in global coupled systems for both climate research (e.g., Law-Chune and Aouf, 2018; Bao et al. 2019; Couvelard et al. 2020) and numerical prediction (Breivik et al. 2015)".

l. 112-114: this set of experiment follows one of my previous question (is there any impact of the coupling frequency?). Table S1 brings some statistics but they are not commented in the text. Please justify why 1800s was chosen as the coupling frequency.

Response: The coupling frequency indeed influences the simulation results slightly. To quantify, the RMSEs of SST, SWH and WSP10 with different coupling steps for the ALL experiment are shown in Table R2. From Table R2, the 10_STEP_WW3 experiment has the closest RMSEs to the 1_STEP_ALL for SST, SWH and WSP10, and has the relatively small runtime. Therefore, the time steps of 10_STEP_WW3 are selected to compromise computer time consumption and the model RMSE. Note that the original one time step in CFS is 180 s, and 10 time steps are 1800 s. Text is revised to clarify.

Table R2. The 28-day global RMSEs and daily runtime for SST, SWH and WSP10 in the ALL experiment with different coupling steps in Jan, 2017. In 1_STEP_ALL experiment, three model components are coupled every time step, and in 5_STEP_ALL (10_STEP_ALL) they are coupled every 5 (10) steps. Particularly, in 10_STEP_WW3, only the WW3 is coupled every 10 time steps, whereas the GFS and the MOM4 remain the one time step coupling frequency as the original settings in CFS.

Experiments	SST (℃)	SWH (m)	WSP10 (m/s)	Daily runtime (s)
1_STEP_ALL	0.88	1.09	3.94	25677
5_STEP_ALL	0.88	1.09	3.88	19546

10_STEP_ALL	0.88	1.23	4.25	19012
10_STEP_WW3	0.88	1.10	3.95	19171

l. 122-124: "the daily initial fields at 00:00 UTC.." I guess it is rather the initial field of the first day of each experiment? Or is the model re initialized every 24h from the operational analysis? Please specify.

Response: Yes, the initial field here refers to the first day of each experiment, not re-initialized every 24h from the operational analysis. The text is revised to "the initial fields at 00:00 of the first day in each experiment".

l. 285 and following: Couvelard et al (2020) also obtained improvement of SST/MLD biases in the Southern Ocean. Please discuss your results against theirs.

Response: Based on the coupling of NEMO and WW3, Couvelard et al. (2020) considered the wave-supported stress, the modification of Charnock parameter, the Stokes drift-related forces, the Langmuir cell and the modified turbulence kinetic energy, and the SST and MLD biases were largely improved. The main differences between our study and Couvelard et al. (2020) are three aspects: (1) Couvelard et al. (2020) considered the misalignment of wind and waves for Langmuir cells, which leads to weaker mixing than assuming alignment; (2) the modification of wave-supported stress and turbulence kinetic energy; (3) the initial SSTs are generally underestimated in Couvelard et al. (2020; Fig. 13a&14a), especially in the Southern Ocean. While in our system, we don't have the second parametrizations, and even the application of VR12-AL (assuming alignment of wind and waves) cannot produce strong enough mixing in the Southern Ocean, and subsequently cannot large reduce the warm bias of SST and the shallow bias of MLD. Therefore, we further include the effects of Stokes–Coriolis force and entrainment (VR12-AL-SC-EN) to enhance mixing. Our results also show that the VR12-AL-SC-EN experiment leads to reasonable improvement of SST and MLD, especially in the Southern Ocean in boreal winter. Text is revised to clarify the differences between our results and theirs.

l. 315-316: "generally consistent", please quantify.

Response: To quantify, the spatial correlation coefficients between the SST and the T02 change in ALL relative to CTRL are calculated, which are 0.60 and 0.51 in boreal winter and summer respectively, significant at 99% confidence level.

l. 332 and following: please give the correlation coefficient between ALL MLD and the observations, so they can be compared with the 0.55 value given for CTRL.

Response: The correlation coefficient of MLDs south of 45°S (north of 45°N) with Argo observations in CTRL is 0.55 (0.68), while the correlation coefficient of MLDs in ALL enhances to 0.69 (0.79). Text is revised accordingly.

l. 348: I do not know how to interpret the "negative trends of bias". What is the meaning of that?

Response: We apologize for the confusion. We updated Figure 8b&9b. The "negative trends of bias" is deleted.

l. 349-351: I would rather say the opposite: the biases in SWH are directly related to the biases of *the 10-m wind speed.*

Response: Text is revised as suggested.

l. 390 and following: see general comments. A graph showing the biases (model vs obs) in the different simulations would be easier to understand. Also, listing the buoy numbers in the SI is probably not useful, especially without additional information (position, number of observations). Please indicate, for every comparison, the number of buoys used.

Response: As suggested, we replace Table 2 with Fig. R5 and Table R1. The number of buoys used for every figure and corresponding buoy identifiers with longitude and latitude are listed in the revised supplementary.

Section 4.4: investigating the heat exchanges is nice at climate scale, but probably not relevant at the time scale of 2 months. I asked to authors to check for that, but it seems that the latent heat flux is directly influenced by the 10-m wind. Plus, the discussion in this part does not lead to clear results (to me). What would be your conclusion, beyond "the latent heat flux depends on the wind speed only"?

Response: Yes, we only conclude that the latent heat flux change depends on the wind speed. Since the wave-related effect on the turbulent heat fluxes is non-significant for the 2-month simulation, we remove the Section 4.4.

References

Bao, Y., Song, Z., & Qiao, F. (2020). FIOâ□□ESM version 2.0: Model description and evaluation. Journal of Geophysical Research: Oceans, 125, e2019JC016036. https://doi.org/10.1029/2019JC016036

Couvelard, X., Lemarié, F., Samson, G., Redelsperger, J. L., Ardhuin, F., Benshila, R., & Madec, G. (2020). Development of a two-way-coupled ocean–wave model: assessment on a global NEMO (v3. 6)–WW3 (v6. 02) coupled configuration.GMD, 13(7), 3067-3090.

LemarieÌ□, F., 2015: Numerical modification of atmospheric models to include the feedback of oceanic currents on air–sea fluxes in ocean–atmosphere coupled models. INRIA Grenoble-Rhône-Alpes Tech. Rep. RT-464, 10 pp. [Available online at https://hal.inria.fr/hal-01184711/document.]

Li, Q., Webb, A., Fox-Kemper, B., Craig, A., Danabasoglu, G., Large, W. G., & Vertenstein, M. (2016). Langmuir mixing effects on global climate: WAVEWATCH III in CESM. Ocean Modelling, 103, 145-160.

Li, Q., Fox-Kemper, B., Breivik, Ø., & Webb, A. (2017). Statistical models of global Langmuir mixing.Ocean Modelling, 113, 95-114.

Polonichko, V.: Generation of Langmuir circulation for nonaligned wind stress and the Stokes drift, J. Geophys. Res., 102, 15773– 15780, https://doi.org/10.1029/97JC00460, 1997.

Shimura, T., N. Mori, T. Takemi, and R. Mizuta (2017), Long-term impacts of ocean wave-dependent roughness on global climate systems, J. Geophys. Res. Oceans, 122, 1995–2011, doi:10.1002/2016JC012621.

Torres, O., Braconnot, P., Marti, O., & Gential, L. (2019). Impact of air-sea drag coefficient for latent heat flux on large scale climate in coupled and atmosphere stand-alone simulations. Climate Dynamics, (3), 2125-2144.

Van Roekel, L. P., Fox-Kemper, B., Sullivan, P. P., Hamlington, P. E., and Haney, S. R.: The form and orientation of Langmuir cells for misaligned winds and waves, J. Geophys. Res., 117, C05001, https://doi.org/10.1029/2011JC007516, 2012.