Before the reply to the reviewer, we must inform that we found a bug in a code to calculate the RMSDs relative to the KEO buoy and then corrected it (Fig. 12 in the revised manuscript). Specifically, the number in the denominator was smaller in the RMSD calculation, and the corrected results show larger RMSDs for all experiments than the previous results. However, the corrected results are qualitatively the same as the previous results. Consequently, this correction has few impacts on the conclusion in this paper. We apologize for the above.

Referee #2

Thank you for the response and the revision of the manuscript. The response and small changes have filled-in some missing pieces and clarified some points. However, I have still some remarks that might need clarification within the manuscript, especially related to the proper tuning of the experiments:

We thank the reviewer for reviewing in detail. We have added the verification results between CTL and 1.5Terr runs (reply to major comments #1 and #2) and the discussion about under-dispersive ensemble spreads relative to the RMSDs (major comment #3), and have reduced the amount of the descriptions and figures (major comment #4).

#1) In your response to comment #2 from the previous review #2, you have stated that the scores of the 1.5Terr experiment are better than for the CTL experiment, why do you use then the CTL experiment as baseline experiment? This result also indicate that the static observational error is not properly tuned (the representation error part might be missing). Consequently, the experiments are in some sense an unfair comparison between using a static observational error and AOEI. Furthermore, the aggregated score for the 1.5Terr experiment is very similar to the scores of the AOEI experiment. In the end, one could wonder if the AOEI experiment is really so much better than using a static observational error, if properly tuned. What is then the advantage of using AOEI?

We have added the RMSDs of 1.5Terr and those relative to the Himawari-8 SSTs to Fig. 13 in the revised manuscript. The CTL run has significantly better accuracy for SSH and SST, whereas the accuracy of surface horizontal velocity and SSHA is degraded in the CTL run, but this is not significant. Since the Himawari-8 SSTs are not independent observations, we have also calculated the SST RMSDs relative to the independent KEO buoy. The results show the RMSDs of 0.52, 0.45, and 0.54 °C in the CTL, AOEI, and

1.5Terr runs, respectively, and therefore, the SST accuracy is better in the CTL run than in the 1.5Terr run. Namely, the 1.5Terr run does not have better accuracy for all variables compared with the CTL run. To highlight the impacts of the AOEI, we have chosen the experiment with temperature observation errors of 1.0°C rather than 1.5°C as the CTL run. We have added the related descriptions to the third paragraph in subsection 3.4.

Figure 13 in the revised manuscript shows that the AOEI run has the best accuracy for all variables except for SST. This paper demonstrates that the AOEI adaptively inflates the observation errors in the mid-latitude region, especially in the frontal region with the large representation errors, and suppresses the salinity degradation mechanisms in the CTL run. This results in better accuracy in the AOEI run than in the CTL and 1.5Terr runs with constant observation errors.

#2) Can you please insert the raw numbers for the 1.5Terr experiment in Fig. 17 and refer to this figure in line 367?

We have added the results from the 1.5Terr run to Fig. 13 in the revised manuscript (Fig. 17 in the previous manuscript). Reference of Fig. 13 has been described at the end of the first sentence in the third paragraph in subsection 3.4 in the revised manuscript.

#3) As can be seen in Fig. 17, the RMSE is more than three times larger than the ensemble spread. Comparing Fig. 9 and Fig. 18, the variance error is the main driver for the RMSE. As now correctly stated, AOEI assumes a correctly tuned ensemble spread, which is here clearly not the case. I would like to see a discussion of this case (a paragraph or so) in the AOEI method part or section 3.4, as it is highly relevant for this study. I think you gave a starting point for this discussion in your response to comment #2.

We thank the reviewer for indicating the large differences between RMSDs and ensemble spreads. As described in the third paragraph in subsection 2.2, Ohishi et al. (in review) performed the sensitivity experiments of covariance inflation and IAU methods and demonstrated that the combination of the IAU and RTPP with relaxing the analysis perturbations toward the forecast ensemble perturbations by 90% is the best for dynamical balance and accuracy. The large relaxation parameter plays a role in maintaining the ensemble spreads inflated by the perturbed atmospheric and lateral boundary conditions. Kurihara et al. (2016), for example, show that the RMSDs of the Himawari-8 SSTs relative to the buoys are about 0.5°C and larger in the higher latitude with a larger zenith angle, and therefore, observation error variances might have substantial contributions to the RMSDs. Nevertheless, the ensemble spreads are much smaller than the RMSDs as indicated by the reviewer, likely being under-dispersive in this system.

We are now constructing analysis products using an eddy-resolving system with higher horizontal resolution of 0.1°, and the verification results show that the temporally averaged RMSDs of the surface horizontal velocity roughly correspond to the ensemble spread in the mid-latitude region, especially around the Kuroshio Extension region, whereas they do not in the subtropical region (figure not shown). Therefore, methods to inflate the ensemble spread more, especially in the subtropical region, are necessary but this will be a future topic. We have added the above discussion to the final paragraph in subsection 3.4.

#4) The added paragraphs at the beginning of the result sections have made them more readable. Nevertheless, I would highly recommend streamlining the discussion of the results. By its vast amount of information, the results are still difficult to follow. For example, by concentrating on the important parts of the Figures, Fig. 3 could be merged with Fig. 10. There is even a Figure (Fig. 18) that is not used at all.

We have merged Figs. 3, 4, and 10 in the previous manuscript, which shows the zonal and meridional sections of temperature and salinity, into Fig. 3 in the revised manuscript, moved the results of the salinity budget analysis to Appendix, and removed relatively not crucial figure (Fig. 18 in the previous manuscript), equations [Eqs. (11), (13), and (14) in the previous manuscript], and related descriptions.

Minor points:

#1) Line 135: Please add a line break between the description of the assimilated observations and the description of the used parameters for the LETKF as the paragraph can be otherwise confusing.

We have separated the descriptions of assimilated observation and parameters in LETKF, respectively, into the second and third paragraphs in subsection 2.2.

#2) Line 139: The source/citation for the observational errors is still missing, as indicated in the previous review.

We have added the reference of Miyazawa et al. (2012). We note that the larger salinity observation errors are prescribed because the satellite observations appear to have large measurement errors.

#3) Line 384: Two tenses are used within that paragraph. I see the point in switching to past here for the summary within the summary. Nevertheless, I would stick to the present tense as the study still demonstrates the positive impact of AOEI.

We have corrected from the past to present tense in the last sentence in the second paragraph of the revised manuscript.

#4) The chosen colormaps for continuous data can be still misleading and is difficult to see for colour-blind persons.

We have modified the shading color of Figs. 1–3, 5–8, and 11 using scientific color maps based on Fabio Crameri (https://docs.generic-mapping-tools.org/6.2/cookbook/cpts.html).

#5) Fig. 13 (c-e): different blue tone in legend than in figures.

We have modified the blue tone to be consistent with the lines in Fig. 9 in the revised manuscript (Fig. 13 in the previous manuscript).

#6) Fig. 14: Where is the heat flux within the figures? If it is covered by other lines (e.g., constant 0), please indicate this in the caption.

We have added the description of "We note that the shortwave penetration gradient term is almost zero and is behind the diffusion gradient term in (a)–(c)." at the end of the caption of Fig. 10 in the revised manuscript (Fig. 14 in the previous manuscript).

#7) Fig. 17: The figures can be misleading. By plotting the ensemble spread together with the errors and using another axis, the reader could believe that the ensemble spread is similar to the errors, which is not the case.

If the same ranges are used for the RMSDs and ensemble spreads, the differences between experiments are hard to be seen. We have added the description of "However, the ensemble spreads are much smaller than the RMSDs for all variables (Fig. 13).” and the notes, respectively, in the last paragraph in subsection 3.4 and at the end of the caption of Fig. 13 in the revised manuscript (Fig. 17 in the previous manuscript).

#8) Fig. 18: Please use the same limits in the colormaps for the RMSE and spread; the figures are otherwise not comparable.

Following major comment #4 from the reviewer, we have removed Fig. 18 in the previous manuscript.