■ Reviewer 1

Summary: This paper documents the tuning of relaxation parameters for an ocean DA based on EnKF and IAU. The methods used in this study are well established in other DA practice. Since ocean DA faces the challenge from dynamical imbalance with shorter cycling periods, the use of relaxation with IAU is a good approach and tuning results are meaningful for the ocean prediction community. I've found several issues in experimental design and result presentation, which I believe the authors should be able to address before the paper can be accepted.

We thank the reviewer for constructive comments, especially on the MULT parameter. We have investigated how much the inflation in the RTPP09+IAU experiment corresponds to the MULT parameter, as replied to the second major comment.

Major Issues:

1. Are cases with alpha<0.9 tested for RTPP/RTPS without IAU? As RTPP+IAU approaches NO INFL+IAU results in both balance and accuracy, I guess the RTPP cases will also approach NO INFL as alpha decrease. If you have tested several points (maybe RTPP05 and RTPP07) it would be interesting to add them in the plots. For example, could there be an alpha value for RTPP that beats RTPP09+IAU?

As described in subsection 3.3, we have performed the RTPP and RTPS experiments without the IAU only for $\alpha_{RTPP} = \alpha_{RTPS} = 0.9$ because of the limitation of computational resources. Since RTPP+IAU and RTPS+IAU experiments gradually approach the NO INFL+IAU experiment as the relaxation parameters decreases, it is likely that the RTPP and RTPS experiments without IAU also gradually approach the NO INFL experiment. Therefore, the RTPP and RTPS experiments would not surpass both dynamical balance and accuracy in the RTPP09+IAU experiment. We would appreciate your understanding of the matter of computational resources.

2. The choice of inflation factor in MULTI is more problematic. Since the multiplicative inflation is applied throughout the domain, it is more sensitive the the rho value. The relaxation methods have build-in spatial variations in inflation so I think it is not a fair comparison between MULTI and RTPP/RTPS methods. In regions where analysis increments are smaller (fewer observations) the inflation of spread can accumulate over time exponentially. Ideally using a spatial varying inflation (such as in adaptive MULTI algorithms) can help. So, if you choose to show MULTI results here the exact value of

rho is very important. Could you estimate an equivalent rho from the best RTPP/RTPS cases? You can averaged the (1-alpha) + alpha*prior_spread/posterior_spread over the domain and time (for RTPP) to estimate the equivalent rho, is it near 1.05 or much smaller?

We thank the reviewer for constructive comments on MULT. We have added the results of the estimated MULT parameter corresponding to the RTPP09+IAU experiment and the discussion of the adaptive MULT, respectively, to the first and second paragraph in Section 5 in the revised manuscript. In the first paragraph, we have indicated that the estimated MULT parameters for the SST, SSS, and SSH averaged over the whole analysis period and domain ($\rho_{est} = 1.08, 1.08, 1.11$, respectively) correspond well to the prescribed parameter ($\rho = 1.05^2 \approx 1.10$). In the second paragraph, we have described that the adaptive MULT would result in degradation, because Ohishi et al. (in review) demonstrated that adaptive observation error inflation (AOEI; Minamide and Zhang 2017), with opposite effects to the adaptive MULT, significantly improves dynamical balance and accuracy of temperature, salinity, and surface horizontal velocities.

As indicated by the reviewer, observations in the ocean interior are relatively sparse, and the ensemble spread might be exponentially inflated over time in MULT experiments. Consequently, the MULT might not be suitable for ocean data assimilation. However, we could not deny that appropriate MULT parameters do not exist. Therefore, we have described “Although it is difficult to find appropriate an MULT parameter as described in Section 5, it might be possible that MULT produces analyses with good balance and accuracy by tuning the inflation parameter.” in the second paragraph in Section 6 in the revised manuscript.

3. Tuning of relaxation can also be case-dependent, you also need to consider the density of observations and localization radius. In the method description maybe you should state more clearly how you tuned localization with this observing network (can you also show a map of observation density for reference?), and the results from tuning alpha in relaxation would likely change if one use another set of observations with different density and localization radius. A discussion in the conclusion would be nice.

Figure R1 shows the frequency of in-situ observations at 5° longitude×5° latitude bins in 2016. Except for coastal regions, 30 observations per month are broadly distributed over the whole domain, and there is no gap in the observation density over the offshore regions.

As discussed in subsection 3.2 in the original and revised manuscripts, we have prescribed the localization scale following Miyazawa et al. (2012) and Penny et al. (2013). The localization scale is not optimal and required for tuning, but this is an issue for future studies as described in the second paragraph in Section 5 (6) in the original (revised) manuscript.

As indicated by the reviewer, it is better to notice that readers are necessary to tune the RTPP parameters for their experimental setting. Consequently, we have added the discussion to the second paragraph in Section 6 in the revised manuscript.

4. I found time evolution of imbalance and errors to be important in this particular case. Since you used fixed values in inflation schemes, it is not guaranteed that the performance will be steady in time. Does the imbalance gradually increase or decrease over time for a chosen alpha value? A time series of spatially averaged delta NBE could be more convincing that the performance is steady. I would be also curious about how long the initial spin up period is for DA solutions to become steady.

The RTPP09+IAU experiment shows that it is not clear whether ΔNBE averaged over the domain have substantial trends or not (Fig. R2). Rather, it appears to undergo seasonal variations with larger (smaller) ΔNBE in summer (winter). This is the same for the other experiments except for the MULT+IAU and MULT experiments. We might be able to obtain insights into the spin-up period if we conduct the experiments for a longer period, but this is out of the scope of this study.

Minor Issues:

We thank the reviewer for carefully checking the manuscript. We have modified almost parts following your comments.

Line 153: abs denotes taking the absolute value, please use standard notation |x|.

We have modified the notation for the absolute function in Eq. (8).

Line 160: the same term IR is used for both RMSE and NBE?, maybe add a suffix to distinguish.

We have added suffix N and R to IR for ΔNBE and RMSD in Eqs. (9) and (10), respectively.

Line 220, Table 3: gross error check not "growth error check"?

We have replaced “growth error check” with “gross error check” in subsection 3.2 and Table 3.

Line 237: SSS nudging: could you provide more details of this approach, maybe a reference or technical report?

The SSS nudging during the data assimilation experiments is the same as the model spin-up. We have added “as in the model spin-up described in subsection 3.1” to the end of the first sentence of the second paragraph in subsection 3.3 in the revised manuscript.

Line 246: is every experiment tested against NO INFL for significance of improvement/degradation? If so, you should state this more clearly.

We have clarified the detail of the statistical analysis by modifying the descriptions in subsection 2.4.2 and the 2nd paragraph in subsection 3.3.

Line 264: this imbalance is substantially improved => reduced.

We have replaced “improved” with “reduced” in the last sentence of subsection 4.1.

Line 276: I guess the RTPP09 and RTPS09 cases are also tested against NO INFL for significance?

We have conducted statistical analyses for the RTPP09 and RTPS09 experiments relative to the NO INFL experiment. We have clarified the description of the statistical method, as is replied to the above comment regarding Line 246.

Figures 1, 3, 4, 7 and 8: you used hollow/solid circles to denote significant/non-significant improvements, but for degradation you used "x" which cannot show hollow/solid differences, maybe use another symobal (triangle?) so you can be consistent.

We have used closed circles for no significant improvement and degradation, open circles for significant improvement, and cross marks for significant degradation throughout the manuscript. If different symbols are used for no significant improvement and degradation, the figures might be hard to see. Therefore, we have maintained the symbols.

Line 551: confidence limit: do you mean confidence level (p value < 0.01)? and no significant difference has $p>0.01$?, if you used t-test just state the p value threshold to be clear.

We have replaced “confidence limit” with “confidence level” in the caption of Fig. 1. As described in subsection 2.4.2 in the original and revised manuscripts, we have used a bootstrap method rather than a t-test because of difficulties to accurately estimate the degree of freedom.

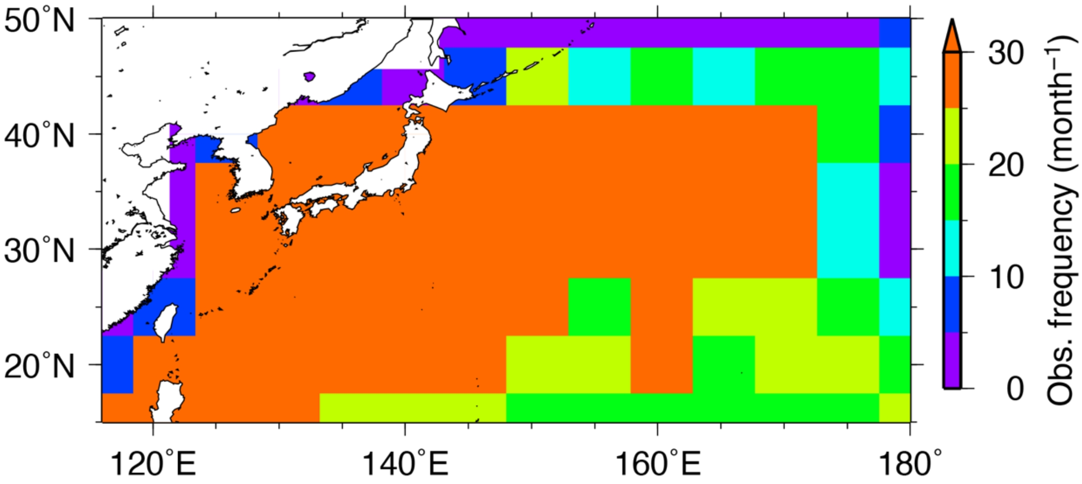

Figure R1: Frequency of in-situ observations at 5° longitude × 5° latitude bins in 2016.

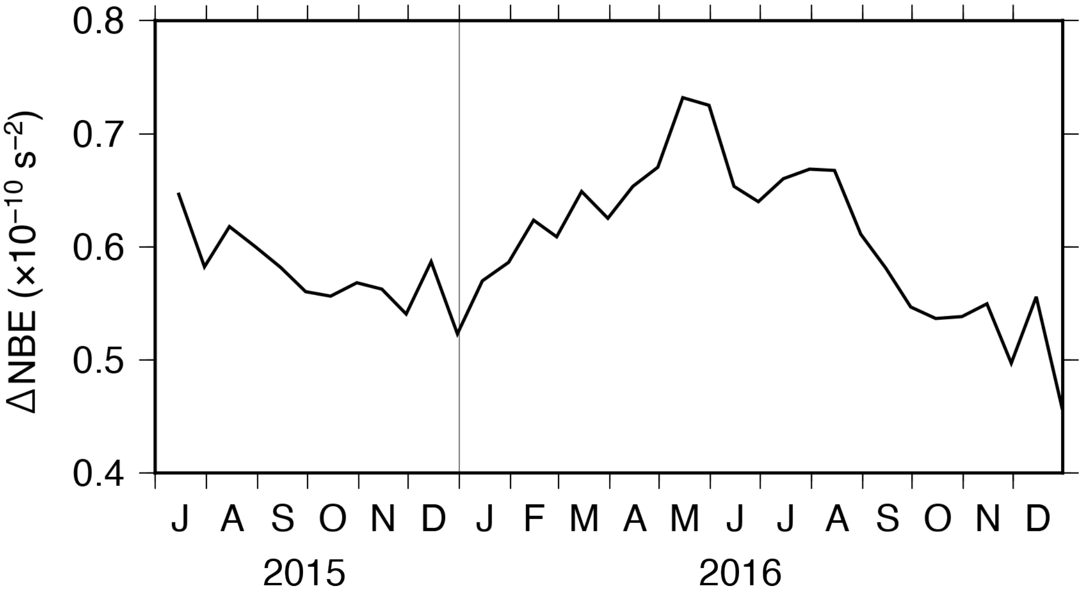

Figure R2: ΔNBE averaged over the whole domain in the RTPP09+IAU experiment.

■ Reviewer 2

The manuscript discusses a setup of an ocean circulation model (the Stony Brook Parallel Ocean Model, sbPOM) for the north-western Pacific region combined with an LETKF data assimilation step. Daily assimilation of satellite and in situ observations is applied and sensitivity experiments are performed with and without incremental analysis updates (IAU) in which the parameters of different covariance inflation methods (in particular RTPP and RTPS) are varied. In addition, a multiplicative inflation is tested with a single fixed inflation value. The study finds that IAU improves the balance of the model increments while the inflation schemes disturb the balance. In contrast IAU leads to higher estimation errors and less ensemble spread than the inflation methods. The multiplicative inflation is found to be failing by not reducing error enough. Parameter ranges are described in which the different methods yield the best assimilation results (low imbalance combined with low estimation errors) and the overall conclusion if that IAU in combination with RTPP with a parameter value of 0.8-0.9 provides the best configuration.

Overall, I have large problems to find what is actually new in this study and what are relevant research results. Actually, while the authors write 'This study develops an ... (EnKF)-based regional ocean data assimilation system' (Abstract line 12), this system is certainly not new. Actually, Miyazawa et al. (2012) already described an LETKF in combination with the sbPOM model. This earlier publication did not use the same model configuration, but this implies that an actual LEKTF-sbPOM DA system already exists for 10 years and this leaves the impression that in the manuscript the authors (Y. Miyazawa is one of the co-authors) merely present some new model configuration. Even more, the applied methods IAU, RTPS and RTPP are established standard methods for ensemble data assimilation. Thus, it is unclear what new insight the experiments described in the manuscript actually provide. The given numbers like 'RTPP with the parameter of 0.8-0.9' (Abstract line 26) are not at all generalizable to other model configurations or other models. Further, the authors do not show any attempt to actually find explanations for their findings. As such it remains that they describe the behavior of a single data assimilation application when parameters of established standard methods are varied. For me, this is insufficient for a scientific publication. To this end, I can only recommend to reject the manuscript. Perhaps, the authors can then find a proper scientific question to assess with this ocean DA system and submit a new study that provides general insights.

We appreciate the reviewer for your comments. As indicated by the reviewer, Miyazawa et al. (2012) is the first paper to construct a regional data assimilation system, the sbPOM implemented with the LETKF, and performed experiments for a short period of about 1 month. It is because their system cannot provide realistic spatial patterns for temperature, salinity, and sea surface height if the experiment period extends over a few months, as added the description to the third paragraph in Section 1. This is similar to the results in the MULT and MULT+IAU experiments. Consequently, it is required to explore an appropriate setting for the sbPOM-LETKF system to represent accurate analysis fields.

We have substantially developed the sbPOM-LETKF system by implementing the perturbing atmospheric forcing method, IAU, RTPP, and RTPS, and have conducted the sensitivity experiments on the covariance inflation and IAU methods for a relatively long period of about 1.5 years. As a result, this study demonstrates that only the combination of the RTPP and IAU improves both accuracy and dynamical balance, and therefore it is the most suitable. To the best of our knowledge, only one among the IAU, MULT, and RTPS is adopted in existing ocean data assimilation systems (Table 1), and there are no studies to compare the impacts of covariance inflation and IAU methods on dynamical balance, accuracy, and ensemble spread. Therefore, we expect that this paper is helpful for readers to newly construct an EnKF-based ocean data assimilation system and improve the existing systems, although the suitable RTPP parameter would depend on tuning parameters and experimental setting. Since only the MULT is implemented with the LETKF source code on Github (https://github.com/takemasa-miyoshi/letkf), readers might choose MULT and face similar problems if readers do not find this paper.

Apart from the aspect of novelty and relevance, I have a few major comments:

1. The manuscript is submitted as a 'development and technical paper' and its title suggests that it might document particularities of the EnKF-sbPOM model system. However, the manuscript is missing detailed descriptions of the actual system.

The IAU, perturbed boundary conditions, RTPP, and RTPS are not incorporated into the system constructed by Miyazawa et al. (2012). This indicates that we have substantially developed the sbPOM-LETKF system. The detailed descriptions of the IAU, perturbed boundary conditions, RTPP, and RTPS have been included in Section 2, and the detailed setting of sbPOM and LETKF has been specified in Section 3 in the original and revised manuscripts.

2. The authors list EnKF-based ocean data assimilation systems in Table 1. Unfortunately, this list is very incomplete. E.g. there are EnKF/based system run operationally by the Copernicus Marine Service (CMEMS) for the global ocean and for the Baltic Sea (It is easy to find these systems via the CMEMS website marine.copernicus.eu). From the operational CMEMS systems, Table 1 only lists the TOPAZ4 system. There is also an operational EnKF-based system in Germany (the latest article about it is Bruening et al, 2021, but there are several publications about earlier versions dating back to the year 2012. This system uses 12-hourly analysis, thus even shorter than what is pointed out in the manuscript). Also there is an EnKF-based coupled system which focuses on the ocean (e.g. Tang et al. 2020). Overall the authors should perform a much more careful research on current systems. Publications dating back to 2011 or 2012 do most likely not describe the current status.

We thank the reviewer for letting us know about the ocean and coupled data assimilation systems. As indicated by the reviewer, there are ocean and coupled data assimilation systems not listed in Table 1. However, this study does not aim to summarize all of them, and therefore we mainly refer to ocean data assimilation systems related to this study. Nevertheless, we have missed a regional system for the North Sea and Baltic Sea constructed by Bruening et al. (2021), in which only the satellite SSTs are assimilated at short interval of 12 hours. We have added the system to Table 1 and the description to the third paragraph in Section 1. Although we have carefully searched EnKF-based global ocean data assimilation systems on the CMEMS Web site, we could not find such systems.

We have found several EnKF-based coupled data assimilation systems (Brune et al. 2015; Chang et al. 2013; Counillon et al. 2016) in addition to Tang et al. (2020), but they do not assimilate all typical observations (SST, SSH, T, and S) at a frequent interval similar to the existing EnKF-based ocean data assimilation systems. We have added a brief description of coupled data assimilation systems at the end of the third paragraph in Section 1.

3. The authors express that their data assimilation setup is particular because of daily assimilation. However, when one has a sufficiently complete overview one sees that short assimilation cycles like daily are not that special. On the other hand there are good reasons for longer cycles. One particular reason is the repeat cycle of the altimetry satellite data. Further, while applying e.g. weekly analyses steps, systems like TOPAZ4 use asynchronous filtering, e.g. for SST. Thus, the system is able to also take some of the faster variability into account. The authors should take such characteristics of the DA

systems into account to provide a sound overview of EnKF-based ocean DA systems.

Using a regional atmospheric data assimilation system, Maejima and Miyoshi (2020) demonstrated that the accuracy for 3D-LETKF is better than 4D-LETKF, which is similar to asynchronous filtering, although the computation cost of 3D-LETKF is higher. The satellites now provide the huge amount of SST observations at a frequent interval, although the daily distribution of satellite SSH is sparse. To maximize the use of the satellite SSTs, data assimilation at a frequent interval by 3D-LETKF would be better than the 4D-LETKF with a 1-week window.

4. As mentioned above, IAU, RTPP and RTPS are standard methods in DA already for quite some years. As such it is surprising to still see a manuscript submission about these schemes. Unfortunately, the authors also miss to take into account the study by Yan et al. (2014), which discusses IAU in ocean data assimilation. However, also the CMEMS system for the global ocean uses IAU. Given that these methods are well established and well studied, I am quite skeptical that it is possible to find new general insights by just using standard methods and varying their parameters.

Using an EnKF-based regional ocean data assimilation system, Yan et al. (2014) investigated the impacts of the IAU on dynamical balance and accuracy in twin experiments with an idealized setting, whereas this study conducts the sensitivity experiment assimilating real satellite and in-situ observations. Although a global data assimilation system in CMEMS (https://resources.marine.copernicus.eu/product-detail/GLOBAL_MULTIYEAR_PHY_001_030/INFORMATION) with singular evolutive extended Kalman filter (SEEK) adopts IAU, the system does not reveal the effects of the IAU on the dynamical balance and accuracy. The results for dynamical balance and accuracy in this study are consistent with Yan et al. (2014), and we have added the descriptions to the first paragraphs in subsections 4.1 and 4.2.1.

Although the IAU, MULT, RTPP, and RTPS are now well used in data assimilation field, to the best our knowledge, there are no studies to evaluate their impacts on dynamical balance, accuracy, and ensemble spread, combining the IAU and covariance inflation methods in an EnKF-based ocean data assimilation system. As clear from Tables 1 and 4, the combination of the RTPP and IAU is the most suitable but has not adopted in the existing EnKF-based systems. Therefore, this study would be helpful for readers to newly construct an EnKF-based ocean data assimilation system or improve the existing systems. However, as indicated by Reviewer #1, the appropriate RTPP parameter might depend on other tuning parameters and experimental settings. We have added the description to the second paragraph in Section 6 in the revised manuscript.

5. The authors use a model spin-up of 4.5 years from an ocean in rest. This spin-up period looks far to short for properly spinning up the ocean unless one only looks at the upper layers.

If spatially uniform temperature and salinity are used for initial conditions, spinning up over several decades would be required. As described in subsection 3.1, however, observational monthly and seasonal temperature and salinity climatologies from the WOA18 are used for initial conditions with no motion, and the model is spun up with nudging temperature and salinity toward the monthly and seasonal climatologies with a 90-day timescale throughout the depth. Consequently, qualitatively similar results would be obtained regardless of the length of spin-up period. The long spin-up integration with 100 ensemble members is computationally expensive, and therefore we decide to choose a relatively short spin-up period.

6. The observation errors of 1.5degC for satellite SST and in situ temperature and of 0.2m for SSH are very large compared to what is commonly used today.

As seen in Ohishi et al. (in review), the low-salinity structure in the intermediate layer is degraded around the Kuroshio Extension if smaller temperature observation errors of 1.0 °C are applied. Consequently, temperature (SSH) observation errors are assumed to be 1.5 °C (0.2 m) in this study.

7. In lines 220-221 it is described that the localization settings are chosen following the studies by Miyazawa et al. (2021) and Penny et al., (2013). However, in these studies other model configurations with different resolutions are used and both use different localization radii. It is known that localization settings depend also on the model configuration. To this end, just selecting some settings from model configurations at other resolutions is not a reasonable approach. One can use values from other studies as a starting point for ones own tuning, but this tuning will be required as otherwise, there is a high risk that the DA system is suboptimal. Thus sub-optimality then also influences other DA parameters like those for the inflation.

As discussed in the 2nd paragraph in Sections 5 and 6 in the original and revised manuscript, respectively, we have noticed that the localization scale is a tuning parameter and might depend on other tuning parameters such as covariance inflation parameters. It is beyond the scope of this study and an issue in future studies to survey an appropriate localization scale.

8. In line 60 the authors describe the TOPAZ4 system with 'but with inflation of observation errors'. I'm unsure what the authors intend to express by 'but'. However, when the authors look carefully, the 'moderation of observation errors' used in TOPAZ4 is in fact a careful inflation that should have similar effect as a carefully tuning multiplicative inflation scheme.

As described in the first paragraph in subsection 3.2 in Sakov et al. (2012), to prevent filter divergence, the TOPAZ4 multiplies observation errors by a factor of 2 when the ensemble perturbations are updated. Since the MULT parameter is generally set larger than one (i.e., $\rho > 1$), this procedure deflates the forecast ensemble spread and has opposite effects to the MULT. Furthermore, there are no descriptions for tuning the factor, and the observation error matrices for analysis ensemble mean and perturbation update should be consistent in the formulation of EnSRF and ETKF. Although adaptive observation pre-screening method to prevent an excessive shock is described in the second paragraph in subsection 3.2 in Sakov et al. (2012), this appears not to follow any theories such as the statistic innovation (Desroziers et al. 2005). Therefore, we could not find reasonable descriptions for the observation error inflations, and have maintained the description "*the TOPAZ4 uses all types of observations but with inflation of observation errors.*" in the third paragraph in Section 1.

9. The multiplicative inflation schemes is described as 'not demonstrate sufficient skill'. This description is actually misleading and invalid. The authors only run a single experiment with a fixed inflation of 5%. Thus, any sensitivity assessment is missing. Actually, the data assimilation process in the system of the manuscript runs already stable with successful assimilation even without inflation as the figures show. This is a clear indication that 5% multiplicative inflation is too large.

We thank the reviewer for pointing out the description that might mislead the readers. To avoid it, we have specified the parameter of the MULT in Abstract.

Following the comments from Reviewer #1, we have estimated the MULT parameter corresponding to the RTPP09+IAU experiment, and have added the results as

Section 5 in the revised manuscript. The spatiotemporal averaged estimated MULT parameter for SST, SSS, SSH fields are 1.08, 1.08, and 1.11, respectively, and correspond well to the prescribed MULT parameter ($\rho = 1.05^2 \approx 1.10$). Nevertheless, the results are completely different between the MULT+IAU and RTPP+IAU experiments.

Adaptive MULT may be helpful to estimate appropriate MULT parameter. As described in Ohishi et al. (in review), however, the AOEI improves the dynamical balance and accuracy of the temperature, salinity, and surface horizontal velocities. Since the AOEI has opposite effects to the adaptive MULT, the adaptive MULT would degrade the dynamical balance and accuracy.

It is difficult to explore the suitable MULT parameter, but we cannot deny that the suitable MULT parameter does not exist. Therefore, as described in the third paragraph in Section 5 (6) in the original (revised) manuscript, the MULT might have suitable parameter to improve the dynamical balance and accuracy.

10. The residual of the nonlinear balance equation {¥Delta}NBE (Eq. 8) is not normalized. As such it is unclear whether any of values shown in Fig. 1 and described in the text (like 2.11x10^{-10} for MULT+IAU in line 249) is actually significant.

In the first paragraph in subsection 4.1, spatiotemporal averaged ΔNBE is compared among the sensitivity experiments, and ΔNBE=2.11×10^{-10} and 5.22×10^{-10} s^{-2} in MULT and MULT+IAU experiments, respectively, is much larger than the other experiments. This study has shown the raw values of ΔNBE to directly compare among the sensitivity experiments.

References:

Bruening, T., Li, X, Schwichtenberg, F., Lorkowski, I. (2021) An operational, assimilative model system for hydrodynamic and biogeochemical applicatios for German coastal waters. Hydrographische Nachrichten, 118, 6-15, doi:10.23784/HN118-01

Tang, Q., Mu, L., Sidorenko, D., Goessling, H., Semmler, T., Nerger, L. (2020) Improving the ocean and atmosphere in a coupled oceanâ□□atmosphere model by assimilating satellite sea surface temperature and subsurface profile data. Q. J. Royal Metorol. Soc., 146, 4014-4029 doi:10.1002/qj.3885

Yan, Y., Barth, A., Beckers, JM. (2014) Comparison of different assimilation schemes in

a sequential Kalman filter assimilation system, Oce. Mod. 73, 123-137, doi:10.1016/j.ocemod.2013.11.002

References:

Brune, S., Nerger, L. and Baehr, J.: Assimilation of oceanic observations in a global coupled Earth system model with the SEIK filter, Ocean Model., 96, 254–264, doi:10.1016/j.ocemod.2015.09.011, 2015.

Chang, Y. S., Zhang, S., Rosati, A., Delworth, T. L. and Stern, W. F.: An assessment of oceanic variability for 1960-2010 from the GFDL ensemble coupled data assimilation, Clim. Dyn., 40(3–4), 775–803, doi:10.1007/s00382-012-1412-2, 2013.

Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L. and Bentsen, M.: Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model, Tellus, Ser. A Dyn. Meteorol. Oceanogr., 68(1), doi:10.3402/tellusa.v68.32437, 2016.

Maejima Y, Miyoshi T (2020) Impact of the window length of four-dimensional local ensemble transform Kalman filter: A Case of convective rain event. SOLA 16:37–42. https://doi.org/10.2151/sola.2020-007

Ohishi, S., Miyoshi, T., and Kachi, M.: An EnKF-based ocean data assimilation system improved by adaptive observation error inflation (AOEI), Geosci. Model Dev. Discuss. [preprint], https://doi.org/10.5194/gmd-2022-91, in review, 2022.

■ Reviewer 3

Recommendation: Major revision

Summary

This manuscript describes the local ensemble transform Kalman filter (LETKF) implemented in the Stony Brook Parallel Ocean Model (sbPOM), with daily assimilation of satellite and in-situ observations. Sensitivity experiments with IAU and various multiplicative inflation methods are performed. Results show that the application of IAU improves the analysis balance, but degrades the analysis accuracy and also reduces ensemble spread. The constant multiplicative inflation with or without IAU had much larger imbalances and errors than the other configurations. RTPP and RTPS with IAU had improved balances and smaller errors when the inflation parameter is tuned. The presentation of the manuscript is fine, and the lessons of inflation and IAU with influences on imbalance and accuracy are useful for the ocean DA community. But the results need further clarifications and explanations. Please see my comments below.

We thank the reviewer for constructive comments, especially on the IAU method.

1. It is confusing about the impact of IAU on the assimilation results. Compared to NOINFL, IAU in NOINFL+IAU degrades the accuracy. Why IAU degrade the accuracy for ocean assimilation that has longer time scale than atmosphere?

The main difference between without and with the application of the IAU is directly updated the SSH or not. Temperature, salinity, horizontal velocities, and SSH analyses are used for the initial conditions for model integration within the assimilation window if the IAU method is not applied, whereas the analysis increments of temperature, salinity, horizontal velocities except for the SSH are distributed if the IAU method is applied. Therefore, the direct update of the SSH would result in higher accuracy of the SSH, SSHA, and surface horizontal velocities in the experiments without the IAU.

2. The authors state that IAU reduces the spread and accuracy of DA. But MULT, RTPP and RTPS have totally different impacts on the spread and accuracy when IAU is applied. Why MULT that also inflate the ensemble spread has the opposite impacts on spread and accuracy than RTPP and RTPS? Since the results with different inflation methods are inconsistent, it would be helpful to understand the roles of

different inflation methods, especially the interactions with IAU.

As indicated by Ohishi et al. (in review), the exceedingly large temperature and salinity increments result in the degradation of the temperature, salinity, and surface horizontal velocities, because they induce exceedingly strong vertical diffusion through weakening density stratification around the Kuroshio Extension region. Therefore, such large increments are not favorable for maintaining the stratification.

The RTPP and RTPS relax the analysis ensemble perturbations toward the forecast ensemble perturbations. This implies that the analysis increments in the RTPP and RTPS would be smaller than the MULT, and the above degradation process might be suppressed.

3. Previous studies of IAU (e.g., Lei and Whitaker 2016, He et al. 2020) showed that IAU has more advantages for variables that are more influenced by imbalances that variables that are less influenced by imbalances. However, results here are inconsistent with the previous findings. IAU improves the accuracy of wind field more than the accuracy of height field (Figures 3 and 4). Please provide explanations or insights for these counter-intuitive results.

Figures 3 and 4 indicate the degradation of the accuracy of the SSH and surface horizontal velocities by the IAU rather than the improvement. The degradation of the accuracy by the IAU is consistent with He et al. (2020) who demonstrated that the accuracy of most variables is worser in the 3D-IAU experiment than experiment without the IAU when the assimilation windows are short of 1 and 3 hours [See table 3 of He et al. (2020)]; Lei and Whitaker (2016) who indicated that the accuracy of temperature and wind speed is worser in the 3D-IAU experiment than the experiment without the IAU using NCEP GFS experiments with assimilation of real observations [See fig. 8 of Lei and Whitaker (2016)]; and Yan et al. (2014) who showed that the IAU degrades the accuracy in twin experiments using an EnKF-based ocean data assimilation system [See table 3 of Yan et al. (2014)].

4. Details of how the verification are done are needed. Which time is the imbalance deltaNBE computed at? Is it the prior or posterior at middle of DA window? The RMSD is computed for the prior or posterior? How the RMSD is computed for experiments with IAU?

Since ΔNBE can be calculated only at the assimilation timestep, it is calculated at the beginning (middle) of the data assimilation window in the experiments without (with) the IAU. As described in the last paragraph in subsection 3.3 in the original and revised manuscripts, “We estimate ΔNBE from ensemble analysis increments on days 1 and 16 of each month, the RMSDs from the daily averaged ensemble mean, and the ensemble spread from the daily-mean ensemble.”

5. Since assimilation is conducted at a daily frequency, both the daily prior and free forecast at longer forecast lead times worth to check.

We thank the reviewer for constructive comments. To perform a free forecast after every assimilation cycle, all experiments must to be integrated again, and the huge amounts of the computational resources are required. Consequently, this is an issue in future studies.

Minor comments:

L90, for the IAU configuration here, is the analysis computed at the middle of an DA window or not? The 1.5 times computational cost is compared to the standard method with or without IAU? It is not clear why analysis is performed at the beginning of an DA window.

As described in subsection 2.1 in the original and revised manuscripts, the assimilation is conducted at the middle of window, and the computational costs with the IAU are 1.5 times those of the standard method (i.e., without the IAU).

The time lag between the forecast and observation becomes large if the assimilation is conducted at the beginning of an assimilation window, and therefore we have chosen the middle of the window for an assimilation timestep as proposed by Bloom et al. (1996).

References

He, L. Lei, J. S. Whitaker, and Z.-M. Tan, 2020: Impacts of Assimilation Frequency on Ensemble Kalman Filter Data Assimilation and Imbalances. J. Adv. Model. Earth Syst., 12, e2020MS002187.

Lei, L., and J. S. Whitaker, 2016: A four-dimensional incremental analysis update for the ensemble Kalman filter. Mon. Wea. Rev., 144, 2605-2621. doi:

http://dx.doi.org/10.1175/MWR-D-15-0246.1.

Yan, Y., Barth, A. and Beckers, J. M.: Comparison of different assimilation schemes in a sequential Kalman filter assimilation system, Ocean Model., 73, 123–137, doi:10.1016/j.ocemod.2013.11.002, 2014.