Comments from Reviewer 1

At the outset, the manuscript is very well written and is suitable for publication. The author has raised pertinent issues which are scientifically relevant. The methodology adopted is well suited to solve this problem. The key issues raised by the authors were:

What is the impact of subsurface data assimilation into a forecast model? The authors have considered assimilation with all parameters included, and have also considered subsets of it to study the sensitivity of different parameters using the observing systems experiments approach. There are a few minor concerns that the authors should address, which will make the work reproducible. The following are the minor concerns:

Fig 1 looks cluttered with too many features. The authors can either split them into subplots or make different plots to make it more clear.

Dear Srinivasa Ramanujam Kannan,

The authors thank you for providing relevant comments that improved the quality of the manuscript.

We agree that Fig. 1 contained many features. We have removed vectors and numbers representing geotrophic currents and Argo locations, respectively. Now Fig. 1 has been simplied.

Page 3, Section 2.1: Authors should provide additional information about the ROMS model used in the present study, such as various parameterization schemes available and suitable for the considered study. Even though the sensitivity of various parameterization schemes is beyond the scope of the present work, the authors must highlight significant results from the literature. Suitable modifications are also required in Line numbers 41-45 on Page 2.

In response to your concern about Section 2.1, we have included references to parameterisation of bottom drag and vertical mixing of tracers and momentum, which adds to the parameterisation of atmospheric fluxes already in the text (lines 90-93). We refrained from adding more information about other schemes to avoid making the manuscript extensive. However, we do acknowledge other parameterisations used in the literature. We also removed some references that made the text between lines 41-45 on Page 2 quite lengthy. Now this part of the manuscript reads better.

Page 6, Eq. 1: How is the cost function minimized? The author should briefly explain the algorithm used to minimize the cost function.

Regarding Eq. 1 on Page 6, we have added a citation (Moore et al., 2011a) to the algorithm that minimises the cost function used in 4D-PSAS. The second sentence of paragraph four now reads: "The method chosen in the present work is the physical-space statistical analysis system (4D-PSAS) and the algorithm that minimises the cost function is shown in Fig. 2 on Moore et al. (2011a)". Lines 50-54 in the new manuscript version.

Comments from Reviewer 2

The manuscript explores the impacts of assimilating different types (altimetry SSH, satellite SST, mooring-observed T,S,U,&V) of measurements on reconstructing the ocean state near EAuC, using ROMS-4DVAR system and OSEs. The authors conducted six (or seven) experiments to examine the relative importance of different observations on estimating the ocean state. This is the first step towards an operational ocean forecasting system. However, there are still many deficiencies that need to be addressed before publication, especially the comparisons and explanation of the final results. In addition, I would suggest the authors to edit the manuscript by a person with a better command of English.

In the manuscript, the authors didn't develop (at least didn't document) the ROMS-4Dvar system. Considering the scope of GMD and the content of this manuscript, I would suggest the authors transfer/resubmit the manuscript to more

appropriate Journals such as Ocean Science, Ocean Modelling, QJRMS, MWR, JAMES, if possible.

Dear reviewer,

Thank you for your feedback on our manuscript. We appreciate your insightful comments and have carefully considered your suggestions. We have revised the paper and corrected the manuscript in response to the thoughtful suggestions received. Below, we detail our responses to your suggestions, addressing the concerns raised.

We acknowledge we did not developed ROMS 4D-Var which has been established over the years, and has been compiled and described in Moore et al., 2011a,b,c. Instead, we developed a model of the East Auckland Current (EAuC) which can assimilate surface and subsurface data considering different data availability scenarios. Similar works have been published on previous editions of GMD, such as: Kerry et al., 2016, Gwyther et al., 2022,2023, and de Souza et al., 2023 (references' list below). The manuscript has also been transferred to "Technical and development paper" as suggested by GMD's Executive Editor. We believe that the manuscript is a suitable fit for GMD, as it adds to the body of knowledge in the field of ocean modeling and data assimilation and presents important implications for future observational strategies in the region.

Kerry, C., Powell, B., Roughan, M., and Oke, P.: Development and evaluation of a high-resolution reanalysis of the East Australian Current region using the Regional Ocean Modelling System (ROMS 3.4) and Incremental Strong-Constraint 4-Dimensional Variational (IS4D-Var) data assimilation, Geoscientific Model Development, 9, 3779–3801, 2016.

Gwyther, D. E., Kerry, C., Roughan, M., & Keating, S. R. (2022). Observing system simulation experiments reveal that subsurface temperature observations improve estimates of circulation and heat content in a dynamic western boundary current. Geoscientific Model Development, 15(17), 6541-6565.

Gwyther, D. E., Keating, S. R., Kerry, C., & Roughan, M. (2023). How does 4DVar data assimilation affect the vertical representation of mesoscale eddies? A case study with observing system simulation experiments (OSSEs) using ROMS v3. 9. Geoscientific Model Development, 16(1), 157-178.

de Souza, J. M., Suanda, S. H., Couto, P. P., Smith, R. O., Kerry, C., & Roughan, M. (2023). Moana Ocean Hindcast a >25-year simulation for New Zealand waters using the Regional Ocean Modeling System (ROMS) v3. 9 model. Geoscientific Model Development, 16(1), 211-231.

I have commented the manuscript in the attached pdf files. And below are some general comments on on individual Sections:

Section 1 Introduction:

I suggest the authors to rewrite these paragraphs from 45-60, including the comparisons of different assimilation schemes, motivating the current study. The introduction to OSEs is too much, can be move to before the last paragraph of Introduction. The authors should motivate and highlight the current study better here.

We re-wrote part of the text fragment and reduced the number of citations. The major findings are described in the next paragraph together with their OSEs' descripition. We kept the highest motivation point right before the main study goal to highlight its importance (lines 45-69 in the new manuscript version). Please see commented supplementary material attached.

Section 2

Please consider listing all experiments here in a table, showing their names, assimilated observations, number of observation.

Besides, more information about the background terms (covariance patterns) should be provided. If the authors follow previous studies, the authors should also provide some figures about different decorrelation scales, covariance patterns here or in the discussion, since it is likely the background terms play dominating roles in determining the sensitivity rather than the adjoint model part. I feel like the heat flux increments are weird. More details can be found in the attached pdf files.

Figure 1 should be simplified. Please point out regions with high eddy activities (SSH STD) and the model-data differences. Also schematic circulation pattern should be shown here. This could help in the results part.

We have included tables describing the observations assimilated and the different assimilation experiments run (Tables 1 and 2). We have also included more information about the background covariance terms and added citations to Moore et al., 2011a,b which fully describes the method and to de Paula et al., 2021 which shows an example of convolution of a unit impulse function with the horizontal (vertical) decorrelation length scale set to 100 km (50 m) (lines 164-174).

We agree that the increments in the heat flux are large and look odd but the actual net heat fluxes (now included in the manuscript – Fig. 13) that forced the models have heterogeneous spatial variability . We argue that heat fluxes increments are large due to the variances used to compute the forcing background error covariance matrix which includes solar daily cycle and cloud coverage. This increased uncertainty and, consequently, the size of the increments.

We removed the Argo locations and the arrows of geostrophic velocity to make Fig. 1a less cluttered. Fig. 1c now shows the 20-year average geostrophic currents which highlight the main circulation patterns in the region. A citation to Chiswell et al. (2015) was included to guide readers to a nice circulation schematic of the region (lines 26 - 28).

Section 3:

I would prefer better presentations of the assimilation results.

Here I give a few suggestions, details can be found in the pdf files.

I feel like results of NoUVTS (or NoUV) is not necessary be shown here. The authors mostly concentrate on the impacts of with/without TS profiles. Maybe one of them is enough.

What is the radius of meso-eddies in the model domain? I cannot see any mesoscale eddies in Figures 4, 6,7. If the authors want to discuss impacts of assimilation on mesoscale process, I suggest the authors plot the horizontal structure of the eddies in specific time, for instance when there are large errors which are corrected by assimilation. Otherwise, it is not clear at all.

I also encourage explore more on salinity which was degraded. Since the author want to develop this system into an operation system, the authors need to understand why the degradations occur.

We have removed NoTS results from Fig. 6 and Fig. 7 and included results from HYCOM-NCODA and ASFUVTS-2days. We have also generated a new figure (Fig. 2) comparing SSH fields from AVISO observations, ASFUVTS and NoDA runs and identified four mesoscale eddies previously studied in Santana et al., 2021. We also included more figures (Fig. 8 and 9) and tables (Table 3) validating model salinity from most of the experiments.

One main concern from the reviewer regards the different results from ASFUVTS and ASFUVTS-2days. The latter experiments generated better results because of the combination of more frequent increments (2-day assimilation window) and doubled decorrelation length scales which impacts a wider region in comparison to using 100 km (m) as horizontal (vertical) decorrelation length scale. This configuration, however, leads to larger temperature and salinity rmsd if subsurface tracer data is not assimilated (e.g. experiment NoUVTS-2x in new Fig. 10).

Section 4 discussions

In this part, the authors should explain why 2-d experiment looks better than 7-d experiment? Is it because of background terms? Comparison between background terms, adjoint model terms, even just for a specific assimilation cycle should be shown here to explain the differences.

The 2-day assimilation wind run (ASFUVTS-2days) generated better results due to the impact on a wider region (double decorrelation length scale of tracers) and more frequent increments added to the initial conditions. In contrast, the 7-day window runs had less frequent increments and were more dependent on corrections added to the atmospheric forcing. For instance, ASFUVTS had larger average increment in the wind stress curl in *comparison to the 2-day window run (ASFUVTS-2days) which probably caused degradation of the velocity results in the 7-day window run. Nevertheless, more frequent increments in ASFUVTS-2days increased salinity variance around 200 m at M4 and M5. Higher salinity variance was also found in HYCOM-NCODA which assimilates data every day. Horizontal and vertical decorrelation length scales of temperature at the surface and 200 m depth were computed (Fig. 15 in the manuscript). It highlights the heterogeinity in the actual length scales which are different from the fixed values defined in the 4D-Var scheme. At the surface, actual temperature decorrelation length scales are wide spread (hundreds of kilometres) and limited to the first 20 m depth. At 200 m depth, however, the correlation spans ~50 km across and ~350 m at depth.*

Section 5 Conclusions

Since the authors want to develop the model to an operational system, the authors should list details of the computational cost here.

We have included computational cost in paragraph two of the "Conclusion" section. It states: the computational cost to produce one day of reanalysis is about 52 min, and 7 min to generate a 7-day forecast using 80 cores on the NeSI supercomputer (https://www.nesi.org.nz/), a Cray XC50. Lines 540-542.

Once again, we appreciate your valuable comments and suggestions, that have ensured a high quality manuscript with reproducible results.