

Response to Reviewer #1

We thank the reviewer for the thoughtful and thorough review of the manuscript. We have made a major revision to the manuscript to address all of the reviewer's comments, and we believe that we have addressed the concerns raised. We accept all of the reviewer's comments.

The paper analyses an implementation of the Incremental Strong Constraint Variational Data Assimilation scheme within the ROMS model system to generate a reanalysis for the Hawaiian Archipelago. In particular, the improvements in the system due to a newer version of the model source code and modifications in the data processing in comparison to a free running simulation are evaluated. However, the modifications to the model source that can be relevant to the results are not discussed and the modifications to the observations processing not clear. While the analysis techniques are appropriate, they are not novel. Although the work is interesting and provides information that can be useful for researchers developing similar systems for other regions, there are analysis flaws and misconceptions that must be addressed before it is ready for publication. The reanalysis evaluation is made against a free running simulation, and based on the same set of observations assimilated in the reanalysis. Therefore the results should be considered as an evaluation of the assimilation scheme, and not as improvements to the system. A comparison to previous versions of the reanalysis and/or global high resolution products (such as the GLORYS reanalysis) is necessary to evidence any ocean estate representation gain by the present simulation. This is useful to justify the effort of implementing and maintaining a regional operational system. Moreover, there is a lack of physical interpretation of the simulation misfits and the increments to the initial and boundary conditions. I recommend the authors are given the opportunity to improve their manuscript through a MAJOR REVISION before it is accepted for publication in the Geophysical Model Development Journal.

MAIN COMMENTS: While the paper is partly based on the upgrade to a newer version of ROMS, the latest version is not used. The study uses ROMS 3.6 while the version 3.7 was released over 4 years ago, and the current version is 3.8. The authors must justify why an older version of the model code is used.

Co-Author Powell is a ROMS code contributor, and the ROMS 3.6 codebase was used with modifications for the HF radar assimilation and others mentioned in Souza et al. (2015). The most current "stable" release of ROMS is the 3.6 branch as listed in the official release announcements at <https://www.myroms.org/forum/viewforum.php?f=13>.

The current development release is considered 3.7, and it includes the HF code developed by Co-Author Powell. The primary differences between ROMS 3.7 and ROMS 3.6 are the inclusion of the online nesting routines that are not

utilized in this paper.

Since an improvement is claimed for the current reanalysis, there should be a comparison against the previous version of the Hawai'i reanalysis in order to identify where and why such gain occurs.

The original text was rather vague in reference to “improvement”, rather the suggestion is that we are conducting a decadal reanalysis of the Hawaiian region using the configuration as detailed in citesouza2015. The test period used in Souza et al. (2015) was only an 18 month test case. This paper uses all available data over 10 years. The manuscript has been changed accordingly

Although the comparison against a free running model that constitutes a down-scaling of a global reanalysis is interesting, the manuscript would benefit from a comparison against the global reanalysis itself. This can provide an estimation of the improvement in relation to the solution used to force the open boundaries, and justify the effort of implementing a regional assimilation system.

The authors tried to gain access to the GLORYS data mentioned by the reviewer. The data, however, are not public and access comes with cost. Furthermore, the global systems use statistical, time-invariant assimilation schemes in an attempt to replicate observations. These schemes work to minimize the residual between the model and observations, but accomplish this at a cost of non-dynamical increments that typically worsen forecast skill. We have revised the manuscript to discuss that in 4D-Var, the purpose is to represent all available data in time and space not replicate. The only reason to withhold data is because the statistics do not fully account for the time variability. With state estimation techniques (4D-Var and ENKF), the statistics are time evolving by the dynamical system and all observations are relevant and used.

How are the super-observations generated? Is there a projection/average over the model grid? This is especially important for the HFR since it presents higher resolution than the model – and may therefore contain processes unresolved by the simulation.

For most datasets, we do not use “super-observations”. In a least-squares minimization, finding the minimum variance of co-located values is the same as

averaging them and assigning a larger variance. That said, for HFR, we did combine some due to the large number of observations; however, oftentimes, we have only one in a single grid cell. Regions where there are more than one HF radial in a single grid cell are those nearer to the HFR source. As such, the angles are very close, and we tested averaging the angle and radial value versus projecting them. Some of the details are found in Souza et al. (2015).

The case for salinity is curious in Figure 4. There are “negative improvements” even during the period covered by Aquarius mission. Figures 7 and 8 reinforce the idea that there is no benefit in assimilating SSS. Moreover, I wonder why other SSS observations were not explored after the end of the Aquarius mission (SMOS). Since the assimilation of SSS is not a common feature in ocean reanalysis and operational systems, a more detailed analysis of the reasons behind this relatively poor performance would improve the interest on the manuscript.

We thank the reviewer for calling this out. Due to the loss of two SeaGliders in 2014, the SeaGlider data (temperature and salinity) stopped in mid-2014. After this period, only sporadic Argo as well as the SSS fields were available. The SSS fields are highly noisy (0.2 ppt) with coarse spatial coverage (less than 100 observations used for our region). As such, there were very few salt measurements available after mid-2014. These have little impact on the cost-function to minimize; hence, there are—in a practical sense—no constraints for salt. We have updated the manuscript accordingly.

In sections 3.2 and 3.3 you show the a priori SST errors are overestimated for both the observations and the background model solution. Please elaborate on how this impacts your reanalysis results. Did you make any experiments changing the errors estimation method? The same applies to the HFR (underestimated errors).

It is true that the errors were overestimated, which means that there is an “underfitting” to the data (that is, it is less trusted); however, this is somewhat intentional as neither SST or HFR observations are linearly independent. The purpose is to assimilate the observations with the errors that are ascribed to them. In the case of the SST, if the data source attaches a 0.4K error to the value, but the posterior analysis suggests that is higher than it should be, it is because of the lack of linear independence. As shown in Matthews et al. (2011), SST observations within approximately 100km in this region are correlated. It is true that changing the errors would provide a different analysis; however, one has to be careful that the point is not to tune or optimize the posterior error estimates, but to optimize the representation of the observations as a whole in space

and time. More can be found in a number of papers by Desroziers et al. (2001, 2005, 2009). We have update the manuscript to point out the work by Desroziers.

Although the comparison against the assimilated observations is interesting, it basically diagnoses the assimilation is working. However, it says nothing about any improvement in the representation of the regional dynamics. For this purpose, the model must be compared against independent observations (not assimilated). In special, observations in under-sampled areas/variables would be interesting. Perturbations generated during the estate estimate process can degrade the solution in regions where no constraint is provided due to the lack of observations. Are there any other observations that can be used for such (WOD, drifters, SMOS, etc)? If no independent observations can be found, the authors should clearly state the limitations of the provided comparisons.

In general, we agree with the reviewer. However, for a robust validation a significant number of independent observations would be required. A handful of e.g. CTD profiles would not be reliable. All data available to the authors have been used for the assimilation. The authors refrained from reserving data sets for validation as this is an arbitrary procedure. This limitation is now clearly mentioned in the revised manuscript in Section 4 and in the Summary.

It is interesting to note there is a good agreement between the variability of both RMSA and CCA from the forecast and the analysis for all observations in Figures 7 and 8 (especially the low frequency). This suggests a physical process may be absent from the background solution, represented by the forecast. Although the assimilation reduces the mismatch, it does not introduce the missing physics. Please comment on this.

Please recall that the assimilation is 4 days and the figures span 10 years. The “variability” is the variations in the number of observations, locations, etc. that are different between each window. The lower frequency in SST, particularly, the twice yearly increase in RMSA and drop in CCA (relatively low amplitudes), are due to the warming/cooling of the seasons, and disconnect between the HYCOM boundary conditions and the SST used. There are some physical processes that are not resolved that commonly are addressed that can be seen in Figure 9. In the lee of the Big Island, the lack of winds (physically resolved) and cloud cover (not physically resolved) drives a strong diurnal cycle in the SST. This is captured in the observations, but the modeled diurnal amplitude is smaller.

In the evaluation of the analysis of increments in the initial conditions, I don't understand why you averaged over the top 100m. By one side most of the data is in the surface (SST), by the other side there is great interest in what happens below the mixed layer. Especially after you show the larger mismatches are at the thermocline. For both sections 6.1 and 6.2, a deeper interpretation of the physical meaning of the EOFs is lacking. This kind of analysis can give you an insight into processes that are not well represented by the background solution, and fuel improvements in the system. The inclusion of the temporal components of the EOFs should help such interpretation.

Mixed Layer Depth estimates in the region (both from the model and observations, see Matthews et al. (2012) vary between 120 and 80m throughout the seasons. As such, we settled for a 100 m average. We have revised the discussion of the EOFs. However, it must be kept in mind that by computing the EOFs of the increments and adjustments, we are dealing with EOFs of residuals. These residuals have non-physical variance. In the revised manuscript we discuss the spatial patterns of EOFs but clearly state that no physical modes came out of our EOF analysis.

DETAILED COMMENTS: Lines 106 – 113: Please comment on the compatibility between the 2 sources for atmospheric forcing used. Are there discontinuities in the fluxes from the different sources? How does the ocean model respond to this change?

We have revised the manuscript to better explain the generation of the pre-WRF atmospheric dataset. They are both made to be consistent with the NCEP/NCAR reanalysis; hence, though the local variability may be different the large-scale structure and fluxes are the same.

Line 122: Please explain in more details how the observations errors (uncertainty) are calculated and the assumptions made. There are several possible methods for doing so, and the outcomes can greatly impact the assimilation results. (you already started this in lines 194-196)

As noted above, we do not use “super-observations” for most of the observational types (because the resolution of the model is similar or greater than many of the observational platforms. For the in situ observations, they are combined in the vertical every 10 m. The prescribed errors are described in the manuscript,

as the error of representativeness is calculated using long-term analyses of the vertical profiles.

Line 129: Were sensibility tests conducted to estimate the optimum number of inner/outer loops so J is properly minimized? Can the authors present a plot (probably as a supplementary figure) to justify their choice?

These were conducted in earlier experiments, see Matthews et al. (2012) and Souza et al. (2015)

Lines 140-141: What is considered as “close to the boundary” and “shallow water”? The values should be given.

Due to the sponge layer at the boundary of the domain, observations within one Rossby radius (~ 80 km) of the domain’s boundary are neglected. This is now specified in the revised manuscript. The word “shallow” has been removed.

Lines 143-160: Please justify the change in the SST data source to a coarser product in April 2008.

Though the product is coarser, it is more accurate and preferred. The NAVO SST product is higher quality than the OSTIA interpolated product. The OSTIA product is overly smooth and removes ocean structure. Unfortunately, the NAVO SST was not available prior to 2008.

Line 205: What do you mean by that the HFR data is less reliable at the range limit. Does it mean there was a different treatment for such data? Or are you referring to a smaller number of observations?

As the returns are further away from the source, the quality of the return decreases. As such, as described in the manuscript, the errors associated with those radials is much higher (as computed spectrally). There are also a smaller number of observations as the radial spreading of the return lines away from the

source.

Lines 231-233: You should briefly explain how the breakdown of the cost-function is made, and offer references for further details.

The manuscript has been updated. ROMS reports the contributions to the cost-function as a function of the state variables considered by the I4D-Var scheme. The sum of those terms provides the total cost function. We present them individually to show how each state variable is contributing to the cost (in terms of observational residual and increments). The total cost function is detailed in the references provided in the manuscript.

Line 232: Probably you meant model variable by “observation type”.

Corrected

Line 233: change “is” for “it”

Corrected.

Lines 259-261: It is not clear what the “optimality value” is. If it is the result of equation 1, would this mean that a optimality of 1 refers to the model perfectly representing the observations? Please be more straightforward about it.

We improved the explanation of the minimum of the cost function and the associated optimality value. An optimality value of one does refer to a “perfect” representation of the observations for the given number of observations that determine the expected value of the minimum.

Paragraph bellow line 307 (has no line numbers): Are the initial conditions for the analysis and the forecast the same? You specifically say that “the boundary and atmospheric forcing are altered as part of the state estimate solution”.

However, the modification of the initial conditions is usually the most important factor impacting the model solution. Please clarify what are the initial conditions used by the forecast (updated or otherwise).

The line numbers were re-established. The initial conditions come from the end of the previous analysis. They are used for the forecast. Then, increments are made to the initial conditions by the I4D-Var method and the altered initial conditions are subsequently used for the analysis run.

Line 393: The explanation of the skill metrics is confusing. What is the persistence assumption? Please explain it better. Moreover, if the assimilation window is 4 days how do you calculate the analysis skill for 8 days?

The manuscript has been updated for clarity.

Lines 410-412: The fact there is a reduction in skill once per day for the SST is highly dependent on how you define the persistence. What would happen if it was defined from the mean of the first day? Would it be a more balanced option? Please comment.

The daily reduction is due to the prior mentioned diurnal cycle (which can be 1K in the lee of the island); however, this is the tropics and SST varies little from day to day.

Lines 420-422: Could you please provide some evidence the perturbations of the boundary conditions don't influence the model solution? Or at least references of similar systems that provide such evidence.

We have not performed a full dynamical analysis; however, the boundary increments are typically less than one-tenth of the increments made to the initial field. The advective speeds of the NEC and NECC (the two major inflows into the domain) are generously 20 cm/s, such that over the 4 day cycle, the increments would propagate around 70km into the domain. Furthermore, due to the SPONGE layer applied, the increments are diffusive. Similarly, in (Powell, 2017), he performed an observational impact analysis and found that the boundary condition increments have no significance to the solution of the metrics used

in the paper.

Lines 451-452: How was the update interval for the perturbations in the forcing chosen? What was the reasoning behind the 6 hours value?

Six hours were chosen because the atmospheric forcing from NCEP is on 6 hour increments.

Lines 470-472: Was a flux correction (QCORR) applied to ROMS?

Yes, the correction was applied because the WRF model was run with an independent SST estimate, which created a heat flux imbalance between the atmosphere and ocean. Because we are nested within Global HYCOM, the QCORRECTION was calculated based on the WRF 10 and 2m fields with the HYCOM SST.

Line 475: Please substitute “vertically” and “horizontally” by “meridionally” and “zonally”.

Corrected.

Section 5: It is surprising how the analysis represents a small improvement in performance when compared to the forecast. Also, it seems the forecast skill is very stable along the 8 days window. Please elaborate on the reasons behind this.

The RMSAs are increasing over the course of the forecast period. However, we define the “skill” as a ratio of RMSAs, that is with respect to the RMSA of the persistence.

Sections 6.1 and 6.2: Please include the temporal (different cycles) component of the EOFs. These are indispensable for the interpretation of the results.

We added the principal components to all EOFs shown in the manuscript. According to our understanding, however, our analysis represents EOFs of residuals with non-physical variance. As such the PCs do not yield a physical interpretation of the increments made to the forcing or the initial conditions respectively.

Figure 3: Do you know the reason for the “sinusoidal” variation on the number of Aquarius observations?

This is from the Aquarius orbit precession and missing data that is variable in time each day.

Figure 11: Interesting how the assimilation can't reduce the RMSA or improve the CCA below 500m for the Argo and HOT data. But the same doesn't happen for the gliders data. Can you articulate about why this happens?

The background error covariance is low below 500m, so it requires strong observational evidence to overcome the penalty in increments there. The sparse Argo provides an occasional profile; whereas, the SeaGliders (when available) provided many profiles per day in the same region.

Figure 12: The standard deviation around the mean SS should be presented. Also, a grid would make the plot easier to read.

The figure has been updated. The standard deviations and a grid have been added.

Figures 13 and 14: Please include the time component of the EOFs.

The figures have been updated including the principal components. Please see above response on the non-physical variance of the PCs.

References

- G. Desroziers, O. Brachemi, and B. Hamadache. Estimation of the representativeness error caused by the incremental formulation of variational data assimilation. *Q.J.R. Meteorol. Soc.*, 127(575), 2001. doi: 10.1002/qj.49712757516.
- G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Q.J.R. Meteorol. Soc.*, 131(613), 2005. doi: 10.1256/qj.05.108.
- G. Desroziers, L. Berre, V. Chabot, and B. Chapnik. A posteriori diagnostics in an ensemble of perturbed analyses. *Monthly Weather Review*, 137(10), 2009. doi: 10.1175/2009MWR2778.1.
- D. Matthews, B. S. Powell, and R. Milliff. Dominant spatial variability scales from observations around the Hawaiian islands. *Deep-Sea Research*, 58(10): 979–987, 2011. doi: 10.1016/j.dsr.2011.07.004.
- D. Matthews, B. S. Powell, and I. Janeković. Analysis of four-dimensional variational state estimation of the Hawaiian waters. *Journal of Geophysical Research: Oceans*, 117:C03013, 2012. doi: 10.1029/2011JC007575.
- BS Powell. Quantifying how observations inform a numerical reanalysis of hawaii. *Journal of Geophysical Research: Oceans*, 122(11):8427–8444, 2017.
- J. M. A. C. Souza, B. S. Powell, A. C. Castillo-Trujillo, and P. Flament. The vorticity balance of the ocean surface in Hawaii from a regional reanalysis. *Journal of Physical Oceanography*, 45(2):424–440, 2015. doi: 10.1175/JPO-D-14-0074.1. URL <http://dx.doi.org/10.1175/JPO-D-14-0074.1>.