Responses to comments of Anonymous Referee #1

We want to thank reviewer #1 for providing helpful comments on our manuscript. Below we present the reviewer's comments in black, while our responses are in blue and directly follow each comment.

Comments

Could you elaborate on why the observational error is constructed to be dependent on the model background error in this way? I didn't find how the background error covariance B is modelled, but I imagine it has a very different spatial structure than the satellite SST error.

A description of how the **B** matrix is constructed has been added to Sect. 2.1.

We agree that the spatial structure of the satellite SST errors is probably different from those of the standard deviations of the model. In addition to the errors that are related to the instrument themselves, an important contribution to the observational error is the error of representation which may be a larger contributor to the total observational error than the instrumental error. For the case of SST assimilation, the error of representation is not only made up by the mismatch in spatial resolution but should also reflect any mismatch in measurement depth and the location of the uppermost model layer. The SSTs are sub-skin temperatures, valid at a depth of approximately 1 mm, while the vertical position of the uppermost model layer in our model setup is on average 60 cm below the surface. Specifying an appropriate observational error is thus not straightforward.

As the main focus of our study is to demonstrate the potential benefits of using an observation operator that takes the spatial resolution mismatch into account, the observational errors of all assimilated products were constructed in the same way to be able to fairly assess the different experiments. As pointed out by Levin et al. 2019, who use a similar specification of the observation errors, this also has the benefit of ensuring a similar impact of the assimilation in e.g. eddy-rich areas (high values in **B**) as in areas characterized by more steady flow (lower values in **B**).

Sec. 2.2: I think it would be good to include a discussion on how IR sensors intrinsically measure the skin temperature and how their measurements are converted to sub-skin temperature for the different products. For most products this is done by the data creator, but for Sentinel-3A the conversion is done by the authors by adding a constant offset of 0.17 °C. Since Sentinel-3A is subsequently used as the reference satellite, this simple offset could possibly degrade measurements from satellites that use more sophisticated solutions.

Including such a discussion is a good idea and we have modified the paragraph in Sect. 2.2 where we introduce SLSTR Sentinel-3A SSTs accordingly.

Tab. 3: Is it possible to include the metrics for the free model run in this table? The fact that PWM2 is similar to the free model in terms of spectrum (L330) makes me curious to see how much the impact of assimilating with the supermod operator really is.

Thanks for the suggestion. The metrics for the free model run are now added to Tab. 3. Error statistics calculated for the free model run show that PMW2 validates better than the free model run – both when satellite and in situ SSTs are used as reference.

L263: Are the reduced increments for PMW2 compared to PMW1 not simply a result of the choice of the parameter alpha?

A lower (higher) alpha, and thus a lower (higher) observation error, would indeed create stronger (weaker) increments in PMW2. However, as demonstrated in Fig. 6, the supermod operator decreases the amplitude of the increments for increasing footprint sizes when the observation error is kept constant.

Note that in the updated manuscript we use the same alpha for all satellite products in all experiments because we want to ensure a fair comparison between the experiments.

L270: Is the fact that PMW SST fills in the gaps of IR SST sufficiently represented in the performance of the combined experiments? If the comparison with satellites is using IR SST, it does not take into account situations where only PMW SST is available and you may be underestimating the performance.

Thank you for pointing this out. We have changed this part of the manuscript such that we now calculate error statistics of the background for each cycle by using reference observations where we had clouds during the previous cycle (which is the cycle that created the initial conditions which the background was initiated from). Similarly, clear-sky conditions are evaluated by evaluating areas where we had good IR SST coverage during the previous cycle.

References

Levin, J., Arango, H. G., Laughlin, B., Wilkin, J., and Moore, A. M.: The impact of remote sensing observations on cross-shelf transport estimates from 4D-Var analyses of the Mid-Atlantic Bight, Adv. Space Res., 68, 553–570, https://doi.org/10.1016/j.asr.2019.09.012, 2021.