Review of "Model-to-model data assimilation method for fine resolution ocean modelling" by Shapiro and Gonzalez-Ondina.

The paper describes a method to combine information from a high-resolution ocean model forecast with a lower resolution assimilating model forecast, building on a method called Stochastic-Deterministic Downscaling presented in a previous paper by Shapiro et al. (2021). This fits well with the journal topics, and the work presents application of a novel method. The method is demonstrated using some very idealised cases where the true field is known and a lower resolution version of the true field is combined with a high resolution version with random noise added. In these cases the SDDA method is shown reduce the error in the analysis. The SDDA method is also compared to a more conventional data assimilation approach.

There are a number of shortcomings in the paper including the simplified nature of the experimental set-up (no model is actually used so we can't tell how well the analysis would initialise a dynamical model and only one 2D variable is analysed); also the comparison to the "standard" data assimilation seems rather flawed since the set-up of the standard DA is very basic and is not described very clearly. These and other comments are listed below which should be addressed before publication so I recommend major revisions.

Comments on the motivation for the work and introduction

- The authors mention the availability of ocean models such as ROMS and NEMO, but do not mention the availability of data assimilation software such as DART, PDAF and others.
- I'm not sure from the description in the introduction if the issue being addressed by the new method is the need for expertise in DA, the lack of access to DA software, the lack of computational resources, or just that methods for improving the way high resolution models are initialised can be improved. This motivation could be made clearer.
- There seems to be a lack of a literature review on downscaling methods, for example the work of Katayouta and Thompson (2016) and von Storch et al. (2000). Generally the references to certain topics seem quite out of date.
- Line 67. The authors seem to be saying that variational methods and OI are the same which is not true.
- The overall structure of the paper seems appropriate but the section names, their titles and references to them in the text seem a bit confused sometimes.

Comments on the method

- I think it would be useful to clarify some aspects of the method in contrast to issues faced in standard DA. For instance, using the coarse model data as "observations" means that the spatial correlations in the coarse model errors need to be dealt with. One therefore needs to know the accuracy of the parent model which will depend on many factors including where observations were recently assimilated as well as the forecast error growth in the parent model. It is much more complicated to know the characteristics of the errors in the parent model than in the observations themselves.
- It would be good to address the question of whether the method aims to retain the spectral characteristics of the child model (e.g. as the spectral nudging method of Katayouta and Thompson)?

- Line 81: B is the background or forecast error covariance. The authors say “model” error covariance which is something different.
- Line 84-86: These statements aren’t specific enough. The size of matrix B is not affected by the number of observations. It is normally very large though, and methods have been developed to represent approximations of it efficiently. The size of matrix R is dependent on the number of observations, and most assimilation systems assume R to be diagonal (no observation error correlations) so that its inverse can be quickly/easily calculated. That approach is obviously not appropriate in the case where the “observations” coming from the parent model will have significant correlations.
- The statement on line 120-123 seems crucial to me. Clearly the high-resolution system does not have the same correlations as the coarse model (e.g. if one includes sub-mesoscale processes and one doesn’t then they would have very different correlations). If the downscaling operator S allows the “observations” S(y) to contain the same spatial correlations as the high-resolution model then this could be justified, but this needs some further comments and evidence to justify it. There is some discussion on this starting line 142 and the reader is referred to the separate paper on the SDD method.
- How would the method deal with islands in the child model which are not in the parent model, or different coastlines/bathymetries?
- I found it difficult to understand what the mean is referring to. Line 150-153 seems to be saying that ideally an ensemble would be used to generate an ensemble mean, but that instead a local spatial mean is used. Perhaps that could be made clearer.
- Line 157: the error variance of the mean could be obtained if there was an ensemble, e.g. using the ensemble spread.
- Line 160: setting the mean from the child model with the mean from the parent model seems problematic. Won't this remove the high-resolution spatial structure in the mean of the child model? In the case of a front, the high-resolution system would presumably have higher gradients in the mean compared with the parent model, and this benefit from having the child model will be removed.
- It would be useful to have a more stand-alone description of the SDD method which is the crux of the method used in this paper. At the moment the reader has to read a different paper to really understand what is being done.
- Line 188: A more recent reference is Janjic et al 2018 where they recommend using the term "representation error".

<u>Comments on the experimental set-up</u>

- Line 191: Only random noise is added to the parent model. In reality the parent model will contain errors which have spatial and temporal correlations and most likely biases.
- Only random noise is added to the child model which seems to be to represent features not resolved in the parent model. This seems rather unrealistic since the extra features in the child model will have structure, e.g. associated with sub-mesoscale processes.
- Line 201: this is a typo as it says the child model is at the same resolution as the parent model.
- Line 202: A correlation function is mentioned. How is the correlation function defined? How is it modelled? Why was it set to zero beyond some distance?
- Line 203: why was a region of 68 sq km chosen for the spatial averaging? Is there some justification for it, or analysis of the sensitivity of results to it?
- Line 214: Typo: is it supposed to be A=1?
- Eq (14): how are x and y defined in this equation?
- Line 232: The eddy size in the y direction of 105 km seems large compared to the range of sizes in the x direction. Is this justified?
- There doesn’t seem to be any model involved in the experiments. So how do the experiments relate to the title of the paper and the introduction which give the impression that this work is done in the context of assimilating data into a high-resolution model.

- What is the noise in the child model forecast meant to represent? In practice the child model will have errors of course, but it is unlikely they will be uncorrelated white noise.

Comments on the results

- Line 250: this is the first mention of adding bias I think. This should be mentioned in the experimental set-up section.
- Line 256-258: This statement seems strange. The field is constant away from the front so it wouldn't be hard to represent it.
- Line 261: this stand-alone sentence ought to be attached to the subsequent paragraph.
- Figs 2 & 3: There's a lot of noise in the analysis. I would have thought if the error variances were specified correctly that this noise would be largely removed. How do you estimate the error variance of the S(y)?
- Fig 4(b): the caption doesn't mention the red line.
- Fig 5: You don't show the true field on the parent grid in this example which would be nice to see.
- Fig 6: why does the RMSE start rising again for eddy sizes > about 25 km?
- Line 366: "…random noise in the non assimilated model". Not sure what you mean there.
- Fig 11: The green line seems to indicate that the method can't deal very well with random noise.
- Generally the method seems to be dealing well with reducing the bias which is calculated over the whole domain so that positive/negative differences will average out.
- Line 401: "the improvement is five to ten fold". Where do these numbers come from? They don't seem consistent with the ratio line in Fig. 12 (b).

Comments on the discussion and conclusions

- Line 437: you should include references for these methods for error covariance estimation and also reference the review by Bannister (2008).
- Line 438-439: I don't understand this sentence.
- Fig 14: The amplitude of the peak seems to be reduced in the SDDA method which should be mentioned.
- Line 467: It seems like this section comparing the SDDA method to the standard assimilation warrants a section of its own rather than being in the discussion section. The standard data assimilation chosen is only one of many available and the authors seem to have simplified its application significantly. Also, the SDDA method relies on the standard method producing a very good quality parent model solution.
- Line 478: In most applications I know of, the model and observation error variances are not assumed to be homogeneous as is stated here. They would be estimated as spatially varying fields. If this was done, how would your conclusions be affected?
- Line 497: Where were the observations located? Or did you assume observations to be available everywhere?
- Fig. 15(a). The quality of the figure is very poor and I found it hard to see the numbers or read any of the text. I couldn't tell where the observation error covariance was plotted as stated in the caption.
- Eq on line 525-526: K is usually used as the notation for the Kalman gain matrix.
- Fig 16: Could you show also the error (compared to the truth) in the standard method and the SDDA method?

- Line 542: It is stated that SDDA is more computationally efficient. Wouldn’t it be fairer to compare the standard method with the total cost of assimilating the data into the parent model and then doing the SDDA?

References:

- Katavouta and Thompson, 2016. Downscaling ocean conditions with application to the Gulf of Maine, Scotian Shelf and adjacent deep ocean. Ocean Model., 104 (2016), pp. 54-72, 10.1016/j.ocemod.2016.05.007
- von Storch, H., Langenberg, H., & Feser, F. (2000). A Spectral Nudging Technique for Dynamical Downscaling Purposes, Monthly Weather Review, 128(10), 3664-3673.
- Janjić, T, Bormann, N, Bocquet, M, Carton, JA, Cohn, SE, Dance, SL, Losa, SN, Nichols, NK, Potthast, R, Waller, JA, Weston, P. On the representation error in data assimilation, Q J R Meteorol Soc. 2018; 144: 1257– 1278. https://doi.org/10.1002/qj.3130