

Author's responses to reviewer 2 comments on the MS 'Model-to-model data assimilation method for fine resolution ocean modelling' by Shapiro and Ondina

**Comment.** *This fits well with the journal topics, and the work presents application of a novel method.*

**Response** Thank you

**Comment.** *There are a number of shortcomings in the paper including the simplified nature of the experimental set-up*

**Response.** The purpose of this paper is to introduce a novel method and illustrate its quality in a synthetic case where the true solution is known. One of the problems in data assimilation is that the true solution is not known and therefore the error covariance matrices required to minimise the cost function are estimated based on various assumptions, see e.g. the review paper by Bannister (2008). Comparison with observations has its own drawbacks as observations are not perfect. Therefore, testing the new method in situations when the true solution is known should be considered as an advantage not shortcomings. This approach is not new, the testing of a DA method on a synthetic case was described e.g. in the (Carrassi et al, 2018) , and we followed this approach. Clarification is given in the text.

**Comment.** *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.*

**Response.** The procedure for the 'standard' data assimilation closely follows the E.Kalnay's data assimilation textbook to which a reference is given. The interested reader is directed to this book for details. In using a simplified synthetic case we followed the practice described in (Carrassi et al, 2018), see our response to the previous comment for more information.

**Comment.** *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.*

**Response.** Additional references are added as advised.

**Comment.** *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.*

**Response.** The existence of DA software does not mean that the DA is a simple process. The new method is aimed at small groups ( sometimes in developing countries) who would like to run high-resolution local ocean models, have limited computational resources but have access to freely available outputs from global/ basin scale models such as provided by EU Copernicus Marine Environment Monitoring Service (CMEMS) . The EU modelling community encourages the downscaling of CMEMS products by end users. As far as we know such groups just download or downscale (by nesting) the products from CMEMS but it is too difficult for them to use the complex

DA methods in the same way as in the world leading ocean forecasting centres. The SDDA method is simple and computationally efficient enough and would help these groups to use DA. For the period of 1.5 month up to 7 October, there were 353 views of the preprint of the paper registered by Ocean Science website which indicate that there is sufficient audience for the new method. The text is amended to clarify the motivation.

**Comment.** *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.*

**Response.** As in any research (not review) paper the number of references is limited. Following the reviewer's advice we have added a reference to Katavouta and Thompson (2016) as well as clarification of the differences between spectral nudging used in the above paper and the SDDA method. We prefer to give references to the original papers, and some classic results were obtained some time ago, for example the H-L method (1986) or Lorenc (1986) etc. This is not something new. The excellent paper by Bannister (2008) goes as far back in its reference list as the year of 1954.

**Comment.** Line 67. The authors seem to be saying that variational methods and OI are the same which is not true.

**Response.** The reviewer is not correct here. Lorenc (1986) showed that the OI solution is equivalent to a specific variational assimilation problem: Find the optimal analysis  $x_a$  field that minimizes a (scalar) cost function- see the original study by Lorenc which states: "Not surprisingly therefore, Wahba and Wendelberger (1980) have shown that OI can also be expressed in terms of the solution of a variational problem." More recent confirmation of this fact can be found in ([http://www.atmos.berkeley.edu/~inez/MSRI-NCAR\\_CarbonDA/lectures/Kalnay3\\_Ch53D\\_Var.pdf](http://www.atmos.berkeley.edu/~inez/MSRI-NCAR_CarbonDA/lectures/Kalnay3_Ch53D_Var.pdf); <http://twister.caps.ou.edu/OBAN2019/3DVAR.pdf> etc). The OI is also known as the Best Linear Unbiased Estimator (BLUE), see e.g. [https://docs.salome-platform.org/latest/gui/ADAO/en/ref\\_algorithm\\_Blue.html](https://docs.salome-platform.org/latest/gui/ADAO/en/ref_algorithm_Blue.html)

**Comment.** *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*

**Response.** From this comment it is not clear what exactly is confusing in the titles of section and what is the difference between the 'names' and the 'titles'.

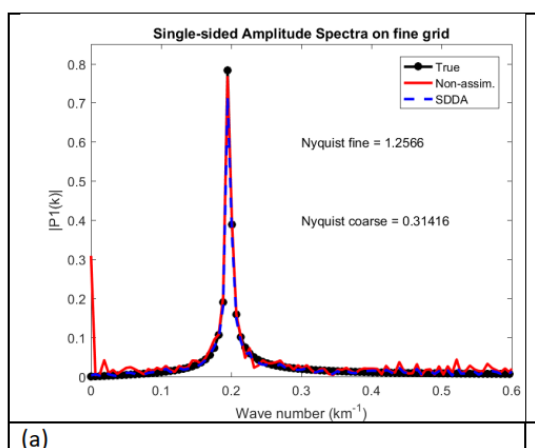
**Comment.** *I think it would be useful to clarify some aspects of the method in contrast to issues faced in standard DA. For instance, using the coarse model data as "observations" means that the spatial correlations in the coarse model errors need to be dealt with. One therefore needs to know the 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.*

**Response.** In contrast to the standard DA (e.g H-L method), no assumption is made that the 'observations' are not spatially correlated. Actually, the data in the parent model are well correlated

as they are quite densely spaced. We agree that the accuracy of the parent model is an important factor. It is obvious that assimilating bad data would not improve the high-resolution model. In order to clarify this point, the MS considers a few examples (a to c) where the parent model is error-free (except errors caused by missing data due to insufficient resolution) and an example (d) where the parent model is noisy. As we know the true solution, we can estimate the errors i.e deviation from the known true solution of any data set: parent model, downscaled parent model, child model before and after SDDA. The knowledge of the true solution allows calculating the improvement of the child model output after the full SDDA cycle. This is done both in cases of error-free and noisy parent model. The SDDA method is aimed at a large community of users who utilise the outputs from global or basin scale models such freely available from CMEMS. Therefore, they trust that such products have the possible accuracy. The quality of CMEMS products is frequently monitored in Copernicus Marine Environment Monitoring Service (CMEMS) Ocean State Reports. A discussion is added as well as clarification on how much improvement comes from the first (SDD) step and how much from the second (DA) step.

**Comment.** *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)?*

**Response.** The answer to this question is given in Fig 14 of the original MS which shows the amplitude spectra of the child model before and after SDDA. The SDDA method does not retain the spectral characteristics of the non-assimilated child model in full. The SDDA removes the bias (which can be seen as a very long-wave component), this is somewhat similar to spectral nudging but more aggressive. Our model attempts to retain the “bias” of the coarse model, which in practice is a local spatial average with a length scale of the order of the correlation length scale. In contrast to spectral nudging by Katayouta and Thompson, the spectral components of noise are removed both in high-frequency and low frequency range, therefore the spectral component of the true signal is represented better after the SDDA. The left panel of Fig14 is reproduced here for convenience. The text is amended to clarify this issue.



**Comment.** *Line 81: B is the background or forecast error covariance. The authors say “model” error covariance which is something different.*

**Response .** Thank you, the text is corrected as advised.

**Comment.** *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.*

**Response .** We agree. However, no matter how good these methods of approximation are, it is still true that as B and R grow in size, the problem becomes larger and requires more time and computer resources to be solved. These time and resources are a significant barrier for many small groups. The text is corrected accordingly.

**Comment.** *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*

**Response.** This is correct. Therefore, the SDDA method does not assume that the R matrix is diagonal. Clarification is given in the text.

**Comment.** *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*

**Response.** This is true. The error covariance matrix B (for the child model) may be different from the error covariance matrix R (for the parent model) at very small distances not resolved by the parent model. It has to be noted that in any DA method, model error covariance matrices are not exact and are estimated based on using various approximations. In his excellent review paper, R. Bannister (2008) states ‘Owing partly to its very large rank, the **B**-matrix is impossible to use in an explicit fashion in an operational setting and so methods have been sought to model its important properties in a practical way... It is also common to assume that the **B**-matrix is static’. Whatever the approximation for the **B** (and **R**) matrix is used, its quality can be seen in how well the DA reduces the model errors. The SDDA method is shown to significantly reduce both bias and the RMSA, see Figs 2 to 14 of the original MS. It is also shown that the assumption of a similar spatial correlation structure between **B** and **R** matrices used in SDDA produces better results than the assumption of a non-correlated **R** matrix as used in the widely used H-L method, see Table 1 of the original MS. Clarification and a reference to Bannister (2008) is now added in the revised MS.

**Comment.** *How would the method deal with islands in the child model which are not in the parent model, or different coastlines/bathymetries?*

**Response.** A potential mismatch between fine and coarse grids due to finer features of the coastline and bathymetry can only relate to the first step of the SDDA method, namely the downscaling. The downscaling of the coarse model data onto the fine model grid is carried out using the SDD algorithm. As shown in (Shapiro et al, 2021) in the example of the coastal areas of the Red Sea, this mismatch is natively resolved during the SDD downscaling. Better resolution of the coastline/bathymetry could result in higher vorticity/enstrophy values in the downscaled field compared to the parent coarse model as shown in Fig 11-15 of the above paper. After downscaling the data from the parent and child model are available exactly on the same (fine) grid, therefore there is no grid mismatch at the DA step. Additional reference and clarification are given in the revised MS

**Comment.** *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.*

**Response .** A local spatial mean as well as standard deviation of fluctuations are calculated in the small trial area around each node. Clarification is given in the text.

**Comment.** *Line 157: the error variance of the mean could be obtained if there was an ensemble, e.g. using the ensemble spread.*

**Response.** We do not calculate the variance of the mean, only the variance of fluctuations around the mean. Clarification is given in the text.

**Comment.** *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.*

**Response.** The replacement of the mean from the child model with the mean from the parent model is equivalent to an aggressive spectral nudging when only the largest wavelength (corresponding to the mean within the trial area) is nudged. It does not remove or modify any higher frequency components. The higher frequency components are improved by the DA of fluctuations as represented by Eq (10). The removal of bias and improvement of other frequencies is seen on the amplitude spectra shown in Fig. 14 of the original MS.

**Comment.** *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.*

**Response.** The SDD method is described in detail in the paper which is easily accessible for free to a reader from the journal website, so there is no need to repeat that paper in the MS.

**Comment.** *Line 188: A more recent reference is Janjic et al 2018 where they recommend using the term "representation error".*

**Response.** We prefer to reference the original papers, or the review papers / textbooks where necessary instead of secondary more recent papers.

**Comment.** 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.*

**Response.** The parent model is assumed to be of good quality and have small errors of whatever nature due to its own data assimilation procedure. These errors are spatially correlated as described in sub-section 'The algorithm' in particular see the part between Eq(4) and Eq(5) defining the correlation matrix  $C_R$ . The random noise is added to the parent model just as an illustration of the performance of the algorithm to the combination of correlated and non-correlated noise. If the parent model is of bad quality, it would not be prudent to use it for data assimilation, this is similar to rejection of bad observational data.

**Comment.** *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.*

**Response.** The noise in the child model represents its own errors (as every model has) not the real features not resolved by the parent model. The errors in the child model include both uncorrelated (random Gaussian noise) and spatially correlated (spatial shift) components in various combinations, see Figs 2,3 11 of the original MS. After SDDA the child model reveals the small scale features not resolved in the parent model, see for example Fig.10(a).

**Comment.** *Line 201: this is a typo as it says the child model is at the same resolution as the parent model*

**Response.** Thank you, the typo is corrected.

**Comment.** *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?*

**Response.** The correlation function is modelled by a Gaussian function in common with other DA methods. It is set to zero at distances greater than a double correlation radius as the value of Gaussian function beyond that distance is less than 2%. The spatial cut-off of the B matrix is common for DA methods (see e.g. Carrassi, 2018) . The details of computing the correlation matrix are given in (Shapiro et al, 2021).

**Comment.** *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?*

**Response .** The trial area is set to  $4L \times 4L$  where L is the correlation radius. The Gaussian function has a value of less than 2% beyond that and its contribution can be safely neglected. Clarification is added to the text.

**Comment.** *Line 214: Typo: is it supposed to be  $A=1$ ?*

**Response.** Thank you, the typo is now corrected.

**Comment.** Eq (14): how are x and y defined in this equation?

**Response.** The coordinates x and y are referenced to the centre of the domain in the zonal and meridional directions respectively. Clarification is added to the text.

**Comment.** *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?*

**Response.** The meridional and zonal sizes of features in example c) are intentionally different to reveal how SDDA performs in an anisotropic setting.

**Comment.** *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.*

**Response.** The SDDA algorithm is developed for use with a high-resolution deterministic (dynamic) child model run locally and combined with the outputs from a lower-resolution parent model which can be obtained either onsite or downloaded from respectful ocean forecasting services such as CMEMS. In common with other studies, see e.g. the review paper by Carrassi (2018), the method is illustrated using synthetic case when the true solution is known. The title uses words "model-to-model" to distinguish it from the usual "observations-to-model" DA. We think this is a fair and clear description of what this paper is about.

**Comment.** *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.*

**Response.** The noise in the child model represents its own errors (as every model has) not the real features not resolved by the parent model. It is not just white noise. The errors in the child model include both uncorrelated (random Gaussian noise) and spatially correlated (bias, spatial shift) components in various combinations, see Figs 2,3 11 of the original MS. The last two are very important and well known sources of error in real models. As our results show, the white noise is a "worst case scenario".

**Comment.** *Line 250: this is the first mention of adding bias I think. This should be mentioned in the experimental set-up section*

**Response.** The bias is mentioned in line 192 of the original MS in the experimental set-up section. The text is made more clear as advised.

**Comment.** Line 256-258: *This statement seems strange. The field is constant away from the front so it wouldn't be hard to represent it.*

**Response.** We agree that the parent model represents well the field away from the front. The original MS says the same: Even at a resolution of 10 km that does not resolve the structure of the front (half-width of 6 km) the parent model gives a reasonable representation of areas outside the front where the changes in the state variable are smooth- see Fig. 1(b)

**Comment.** Line 261: *this stand-alone sentence ought to be attached to the subsequent paragraph.*

**Response .** Thank you, amended as advised.

**Comment.** 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  $S(y)$ ?*

**Response.** The level of noise in the analysis is much reduced as compared to the child model output before the DA, which confirm that the SDDA method is reasonably formulated including the methods to compute matrices B and R. The algorithm to estimate the error variances in the parent model is explained in lines 150-153, 201-204 and 439-442 of the original MS and further clarified in the revised version.

**Comment.** Fig 4(b): *the caption doesn't mention the red line*

**Response.** Thank you. The caption is now corrected

**Comment.** Fig 5: You don't show the true field on the parent grid in this example which would be nice to see.

**Response.** The true field on the parent grid is now added to Fig5(d).

**Comment.** Fig 6: *why does the RMSE start rising again for eddy sizes > about 25 km?*

**Response.** The increase in the Assimilated RMSE to Non-Assimilated RMSE ratio for eddy sizes >25 km is very small. It is likely due to the increasing role of the correlated component of child model errors (spatial shift). In any case the improvement in RMSE after SDDA is greater than 10- fold across the whole range of sizes which shows good skill for SDDA.

**Comment.** Line 366: *"...random noise in the non assimilated model". Not sure what you mean there.*

**Response.** Thank you. This is a typo. We meant 'assimilated' rather than 'non-assimilated'. The text is corrected accordingly.



**Comment.** Fig 11: *The green line seems to indicate that the method can't deal very well with random noise.*

**Response.** When the errors in the child model are completely uncorrelated (random noise only) the improvement produced by SDDA is approximate 2-fold. This is the most difficult case for the SDDA. However, the actual child model is unlikely to have purely uncorrelated errors. The SDDA method is more efficient in a more realistic example of a combination of correlated ( spatial shift, bias) in uncorrelated (random noise) errors as shown in Figs. 2(d), 3(d), 4(b) , 6, 11.

**Comment.** *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.*

**Response.** The bias is actually calculated in a per-node basis. The reduction of bias is due to aggressive nudging of the bias in the child model to the bias of the parent model, which ( parent model) is assumed to be of good quality, see Eq (12).

**Comment.** 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).

**Response.** Fig 6 shows up to 5-fold ( even slightly better) improvement in RMSE after SDDA . The phrase is corrected to state ‘ the improvement is up to five-fold’

**Comment.** Line 437: you should include references for these methods for error covariance estimation and also reference the review by Bannister (2008).

**Response .** The text is amended as advised.

**Comment.** *Line 438-439: I don't understand this sentence*

**Response .** The confusing sentence is removed as the following sentence conveys the message more clearly.

**Comment.** *Fig 14: The amplitude of the peak seems to be reduced in the SDDA method which should be mentioned*

**Response .** The text amended as requested.

**Comment.** *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.*

**Response.** The comparison of SDDA with the ‘standard’ method is a small amount of text and therefore fits well, in our opinion, into the discussion section along with the comparison with the

spectral nudging method added in the revised MS as advised by the reviewer. The ‘standard’ method is selected as most widely used (Stewart et al, 2014; Carrassi et al 2018) and follows by a letter the textbook by Kalney (2003).

**Comment.** *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?*

**Response .** In his review paper, R.Bannister states that modern DA methods ‘make assumptions about the way that background errors are spatially correlated (e.g. homogeneity and isotropy in the horizontal)’. In the SDDA method, the global homogeneity and isotropy are not assumed, but only the local one, in a small vicinity of each node, similar to the theory of turbulence, see e.g. a classical text by Tennekes and Lumley (Tennekes H., Lumley J.L, 1972. A First Course in Turbulence, MIT press, 300pp). To clarify, a new figure is added to the revised MS showing the spatial variations of the diagonal elements of B and R matrix.

**Comment.** *Line 497: Where were the observations located? Or did you assume observations to be available everywhere?*

**Response .** The ‘observations’ are located everywhere on the parent mesh. Clarification is added to the text.

**Comment.** *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.*

**Response.** The figure is updated and it is more clear in the revised MS. The authors will provide the production quality figures when and if the paper is accepted according to the journal publication policy.

**Comment.** *Eq on line 525-526: K is usually used as the notation for the Kalman gain matrix.*

**Response.** The letter K is often (but not always) used to denote the Kalman gain matrix. The MS uses the same notation (W) as in the referenced textbook by Kalney to avoid confusion.

**Comment.** *Fig 16: Could you show also the error (compared to the truth) in the standard method and the SDDA method?*

**Response.** The errors compared to truth are presented in Table 1 which seems to be adequate for the purpose.

**Comment.** *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?*

**Response.** The SDDA method is about assimilation of available data from the parent model into a high-resolution child model. The difficulty and cost of running and assimilating data into the parent model is beyond the scope of this paper. The SDDA method is intended for small groups that will not run the parent model. For them, the parent model output is the input data.