Responses to Reviewer 2/editor comments

Dear Authors
Thank-you again for your revisions. I now have the referee 1 second set of comments, copied at the end of this in case you have not seen them. I am asking for "minor revision" in which you should please address my "Editor comments" and those of referee 1. In view of other correspondence between myself and the referee as well as with yourselves, the "Editor comments" raise a few points which should be brought into the open review process (on eventual publication all comments in the system are made public).
Yours sincerely, John Huthnance (Editor)

Editor comments
Comment.

I believe that we all (Referee 1, yourselves and myself) could agree on "optimal interpolation" (both words together) as a descriptor. Or maybe "objective analysis"? "Stochastic" and "deterministic" next to each other seems self-contradictory. "Stochastic" implies random (albeit obeying a distribution) whereas "deterministic" implies no randomness.

Response. We were thinking of using 'objective analysis' instead of 'optimal interpolation', however , there are some pitfalls there. We agree that the term 'optimal interpolation' may be confusing without a proper explanation as it is of a very different nature than the usual deterministic interpolation methods (linear, polynomial, spline, inverse distance etc) where the weighting coefficients are determined by the location of points, not by the data themselves. . In contrast the OI method calculates the weights based on statistical properties of the data. As to the term 'objective analysis', it has already been occupied in the original publication by Cressman (1959) for his deterministic interpolation method. Therefore we decided to follow a convention from literature and use the term 'optimal interpolation' even though it is not strictly interpolation. It is known that OI method is a kind of minimum variance estimator that is algorithmically similar to Kalman filtering.

We agree that "Stochastic" and "deterministic" are antonyms, and we replaced the term Stochastic Deterministic by Stochastic-Deterministic in the MS. This format (using an 'en-dash') would be similar to the convention used in general physics e.g. 'wave-particle duality of light'. **Further clarification is given in the revised MS lines 57-63.**

Comment. *"Double penalty as an issue" mainly applies to forecasts rather than analyses. Lines 337-344 nicely describe the "double penalty". However, this raises the question of the aim of your work; (i) interpolation or (ii) enhanced presence of small-scale features? For (i) you want to avoid "double penalty" if faithfulness to the parent model is important. For (ii) if you want small-scale but probably stochastic features with realistic frequency (spatial and temporal) and intensity, however obtained, then RMSE with liability to double jeopardy is an inappropriate measure.*

Response. The aim of the paper is to present an alternative downscaling method, i.e. better reveal small scale features of the fields by using stochastic properties of the data rather than use a 'brute force' method, i.e. run the same deterministic model but with more densely placed grid nodes. The reduction of the double penalty is just a nice side effect.

Comment. *This all relates to (a) clarifying the aims of your work and (b) using corresponding terminology. The present line 93 is relevant but aims should not be "buried" in a section headed "The algorithm". [It seems to me that your method may do best in going from eddy-permitting resolution where the desired features are "already" there embryonically and guided by assimilation (e.g. as in CMEMS) to somewhat finer resolution so that the embryonic features can be properly represented. I don't think your process has a basis for generating (ii) if the realistic frequency (spatial and temporal) and intensity of probably stochastic smaller-scale features are greater than what is already embryonic in the coarser-resolution product.]*

Response. Thank you for this comment which gives a very good and concise formulation on the applicability of the SDD method. **Assuming you do not mind, we copy-paste your words at the beginning of the Intro section,** to avoid them being buried in the Algorithm section.

Referee 1

Second Review of "High resolution stochastic downscaling method for ocean forecasting models and its application to the Red Sea dynamics" by Georgy I. Shapiro, Jose M. Gonzalez-Ondina, Vladimir N. Belokopytov

Overall comments

The authors have given good responses to most of my previous comments. This review repeats the previous comments which I do not think have been adequately addressed and explains why I am concerned about them.

I have made three additional comments to those from my earlier review. I apologise that I did not make these comments in the first round but these points have only crystallised in my mind following more study of the paper and authors' responses.

I still think this paper is interesting and should be suitable for publication after significant revision provided the authors are willing to re-consider some of their claims.

Responses to additional comments by reviewer 1

Previous comments.

Comment. *As I read it, the main idea in this paper is that optimal interpolation (OI) is the best method for interpolating a field f of values specified on a set of regular grid points* x*, to another set of points. This may seem obvious from the name "optimal interpolation", but OI*

is not usually thought of in this context. It's usually thought of as a method for combining a number of irregularly spaced noisy observations with a climatology or model background to produce an optimal analysis. Figure 2 of the paper shows that, for some regularly gridded fields, OI provides much better interpolation results than standard methods like bi-linear or bi-cubic interpolation. If this is correct, which I think it is, it could be important for several reasons. First, in the data assimilation context, interpolating the model to the observations accurately is widely acknowledged to be a key step. So that doing that more accurately should improve the results. Second, as shown in figure 14, maps of fields with a lot of fine scale structure, such as the vorticity, may be rendered with greater fidelity using OI for interpolation rather than other interpolation methods.
The authors did not respond to this (first) paragraph of my review which contained quite a number of relevant assertions and suggestions. My main point was that the proposed technique is simply an improved method for interpolation of fields based on the theory of optimal analysis. As the authors illustrate, this method can be successful even when the statistics of the correlation functions used are not very accurate (the statistics were very anisotropic and assumed to be isotropic). Optimal interpolation used in such a context is usually termed objective analysis. The method proposed is clearly not a dynamical downscaling methodology. Although, as the authors explain, optimal analysis can be considered to be founded on concepts derived from homogeneous turbulence and related to stochastic methods, describing the method as a stochastic downscaling will give most readers the impression that the method is based on multiple realisations with higher resolution, which is not really the case.

Response. The aim of the paper is to present an alternative approach to the commonly used 'brute force' method which uses the same ocean model but with more densely placed computational nodes in order to resolve small-scale features. The 'brute force' method is very expensive computationally at high (fine) resolution, while the SDD is computationally efficient. Therefore the main idea of the paper is to use a data assimilation method (we use the OI as an example, however could have used the Kalman filter or anything else) to create a finer -resolution field from a coarser -resolution parent model. As we explained earlier, despite its name, Optimal Interpolation is not a deterministic interpolation method like linear, inverse distance etc but minimum variance estimator that is algorithmically similar to Kalman filtering used in data assimilation. The SDD is a form of data assimilation with many similar features to the methods used in operational models. We have re-worded the introduction to make this clearer.

We have not found any evidence that statistics were anisotropic so we cannot comment on this statement. The statistics are assumed to be locally isotropic in line with many data assimilation methods.

The lack or reduction of double penalty error is an additional benefit of the SDD but it was not the primary purpose of the SDD development. The SDD method employs statistical properties of the data and hence we agree it is not purely deterministic, but it does not mean it is not dynamical. Quantum physics is a good example where the dynamical properties (e.g. movement of electrons or even nuclear explosion) are described statistically using the wave- (aka psi-) function which is a complex-valued probability amplitude.

We understand that the use of the original terminology (Optimal Interpolation) could cause confusion to the reader, and sometimes it is replaced with a 'objective analysis. However an issue has to be taken into account that the term 'objective analysis' was originally occupied (eg Cressman 1959, Vasquez, 2003) a method which is a pure deterministic interpolation with weightings based on geometrical locations not on the statistical properties. It would have been more logical to call OI an 'objective analysis' and to call Cressman's 'objective analysis' a version of inverse distance interpolation. However we cannot change the history and use the terminology as it was introduced by its authors. We think that we could enhance confusion rather than remove it by changing the established terminology in our paper. **Clarification of this issue is given in Line 57-63 of the revised MS.**

Comment. *Title: I wonder whether the words Optimal interpolation (or objective analysis?) should be in the title. For example: "Extraction of near grid-scale dynamical information from model fields using optimal interpolation." I'm a bit concerned that the main point of the paper is not evident from the title.*
The authors did not respond to this comment either. My earlier comment implies that "objective analysis methods" would be a better wording than "optimal interpolation"

Response. We use the term 'Optimal interpolation' rather than more recent terms 'objective analysis' or 'Kalman filtering' in the body of the MS for historical reasons, namely we use the terms from the original literature of 1950s. We explained the method in detail in the paper trying to avoid some confusion in terminology (caused by the term interpolation) which was resolved 60 years ago. In modern literature, 'interpolation' generally refers to methods that allow to estimate magnitudes at new data points, based on the geometrical relationships between nearby points. In this case the weighting coefficients are computed based on points locations not on the data themselves. When these coefficients only depend on the point locations and, perhaps, a few parameters, the name 'interpolation' is appropriate.

The term Optimal interpolation is widely used in data assimilation schemes for ocean models (see e.g. Pinardi, N., Allen, I., Demirov, E., De Mey, P., Korres, G., Lascaratos,A., Le Traon, P.Y., Maillard, C., Manzella, G., Tziavos, C., 2003. The Mediterranean ocean forecasting system: first phase of implementation(1998–2001). Annales Geophysicae 21, 3–20) whilst the term 'Interpolation' is not used in the title.

The use of the term 'interpolation' for a method in which the coefficients depend also on the data can be misleading. This is why we (and other researchers) use this term only in a fixed combination 'optimal interpolation'. The SDD calculates the weights based on the data which is similar to modern data assimilation schemes both from a fundamental point of view and from the mathematical and computational methods employed.

We believe that the title correctly represents the content of the paper. The method used is clearly described in the body and briefly in the abstract, and hence there is no need to include everything in the title.

The method we use is related to both dynamical and static downscaling, as it can be applied both to varying (dynamic) and static (stationary) ocean fields. The terms 'stochastic' or 'probabilistic' are not the antonyms to 'dynamic'. 'Stochastic' (based on data) can be considered as opposite to ’deterministic’ (based on equations) and both approaches are used in physics and mathematics. For example, in quantum mechanics the dynamic processes such as movement of electrons or even nuclear explosions are expressed in probabilistic terms. The main function describing the properties of elementary particles, so called wave function (aka psi-function) is a complex-valued probability amplitude. The probabilistic approach using ensemble modelling is used in data assimilation in weather /ocean forecasting (e.g. Zanna et al, 2018, https://doi.org/10.1002/qj.3397). One of the major textbook on the subject has both ‘random’ (aka ‘stochastic’) and ‘dynamic’ in the title-L.Arnold, 1998, Random dynamical system, Springer, DOI: 10.1007/978-3-662-12878-7. **Clarification is given in lines 89-96**

Comment. The literature on methods for post-processing of model outputs using Kalman filters should probably be discussed in the introduction. I think the main idea being pursued in this paper is somewhat different from the main ideas in that literature but the techniques are clearly related.
My point here is that there is a literature on post-processing that is different from that on data assimilation. The authors could explore that by simply googling “Kalman filter post processing”.

Response. The SDD method, in common to other data assimilation techniques, can be used both in the attached and detached modes. In the attached mode the downscaling is carried out on the same computer which solves the equations of ocean dynamics at the same time as the forecast advances. We use such mode for the operational Persian Gulf model which was developed jointly with the Met Office (not presented in this paper due to sensitive nature of the data it produces). Programmatically, in the attached mode the SDD is contained within the same executable module as all other elements of the model and is applied regularly as the model advances in time.   On the other hand, in the detached mode, the SDD is applied after the forecast has been completed by the parent model. This mode was used in SMORS. In this case the SDD (or any data assimilation) can be considered as post-processing. Some scientists consider data assimilation being not part of the model but part of the modelling system. The same applies to SDD. **We amended the text and added two references on the use of data assimilation (Kalman filter) as post-processing in the revised MS as advised.**

Comment. *One might ask whether the method proposed is a post-processing of model output or a statistical model in its own right. It is described both as a Statistical Model (in SMORS) and a Stochastic Deterministic Downscaling (SDD) method. Personally I would view it as a postprocessing method but do not feel strongly about this semantic issue.*
Authors’ response: We prefer to term the SDD method as part of the model based on how it is implemented in the code. This is shown in the flowchart in Fig.4.
From my previous comments it should be clear that I feel more strongly now that the method should be viewed as a post-processing step.

Response. The SDD can be used both as part of the model run or at a post-processing stage. It seems that the boundaries between the core ocean model and pre-or post -processing are somewhat vague. Let us consider the NEMO model as an example. The vertical grid is generated by the model during the runtime using the subroutine domzgr.F which is supplied as an integral part of the model distribution code. However, philosophically, creation of the grid could be considered as pre-processing. The NEMO subsystem called XIOS (stands for XML-input-output server) can be run in two modes: attached and detached. In the attached mode the XIOS library is compiled into the same single executable file as the rest of NEMO, and each NEMO process acts also as a XIOS server. Hence it is considered as part of the model. In the detached mode every NEMO process runs as a XIOS client. Output is collected and collated by external, stand-alone XIOS server processors. In this case XIOS can be regarded as post-processing. Another example: is data assimilation an intrinsic part of a model or is it post-processing? The answer depends on what someone means by the 'ocean model'. If it is just the set of primitive equations, then the answer is 'no'. If a model is wider collection of algorithms to simulate ocean processes, then probably 'yes'. Some modellers say that data assimilation is not part of the model but it is part of the overall modelling system. As SDD is a form of data assimilation, the same terminology can be applied to SDD. **We have modified the text clarifying that SDD can be used in both attached and detached (post-processing) mode in lines 89-96.**

Comment. *Lines 296-298: It seems strange to use nearest neighbour values in the ARGO intercomparison. With the OI method one can do much better interpolations! Some readers may be concerned that the nearest neighbour method could somehow account for the lack of a double penalty (see previous paragraph).*
I'm quite concerned about this point. For many observations the nearest neighbour is further away on the coarse resolution grid. So the coarse grid value can be expected to be less accurate than the fine grid value. I think this gives the fine resolution grid model a significant advantage over the coarse grid one and could well be disguising a resolution penalty.

Response. We agree with the statement '*So the coarse grid value can be expected to be less accurate than the fine grid value. I think this gives the fine resolution grid model a significant advantage over the coarse grid one and could well be disguising a resolution penalty.'* This is why we do the downscaling- to utilise the advantages of finer resolution grid.

We use the nearest neighbour method for compatibility reasons, as it is used for validation of MyOcean / Copernicus Marine Environment Monitoring Service products, see e.g. Delrosso, D., Clementi,E., Grandi,A., Tonani,M., Oddo, P., Feruzza, G., Pinardi,N. 2016. Towards the Mediterranean forecasting system MyOcean v5: numerical experiments results and validation, 2016. INGV technical report, No 345, ISSN 2039-7941. The methodology presented in the report was developed by world leaders in the subject, and we follow it as good practice. The full text of this report is available from https://www.researchgate.net/publication/303973468_Towards_the_Mediterranean_Forecasting_System_MyOcean_V5_numerical_experiments_results_and_validation. **Clarification and an additional reference are given in lines 378-386 and 403-404.**

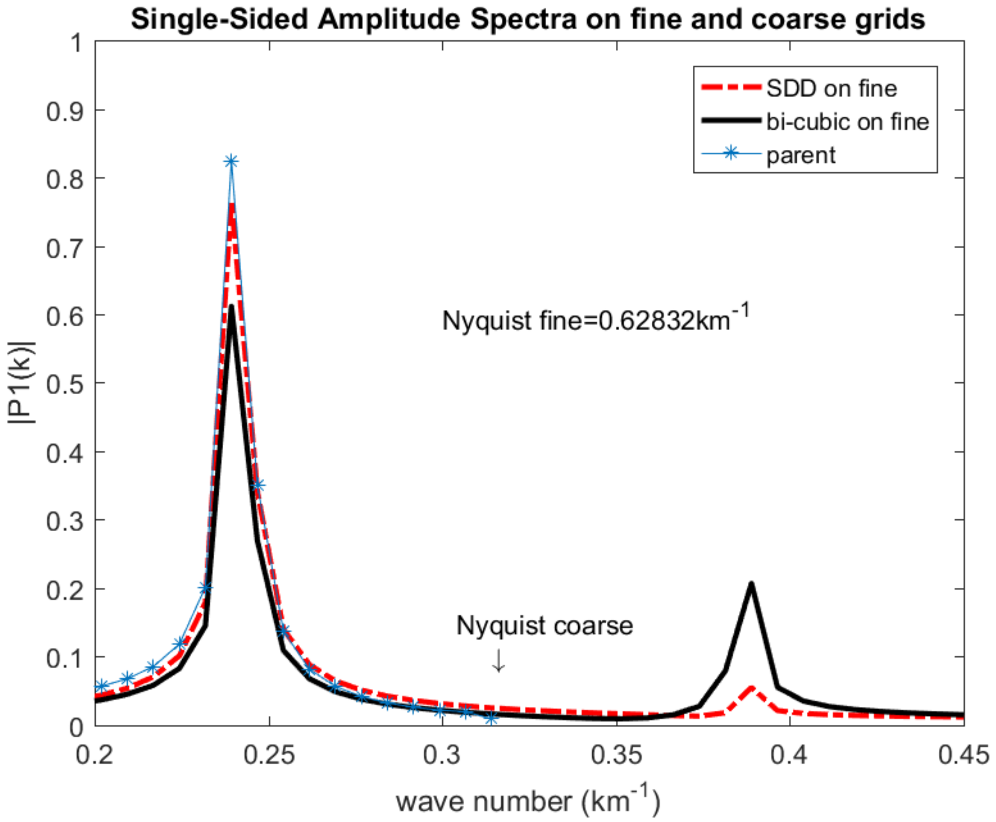
New comments.

Comment. *Lines 18-19 of the abstract state: "Then the method is applied to create an operational eddy resolving Stochastic Model of the Red Sea (SMORS) with the parent model being the eddy permitting Mercator Global Ocean Analysis and Forecast System."*
This claim that an eddy-permitting model is transformed into an eddy-resolving model is a significant exaggeration in my view.

Response. We have modified the text as advised. **The sentence is rephrased** avoiding the terms of 'eddy-resolving' and 'eddy-permitting' as follows : 'Then the method is applied to create an operational Stochastic Model of the Red Sea (SMORS) with the parent model being the Mercator Global Ocean Analysis and Forecast System at 1/12th degree resolution'.

Comment.. *I have tried to work out whether the proposed method truly provides some down-scaling. The following argument suggests that it does and that one would expect some form of down-scaling penalty to attach to it. It seems to me that the Fourier spectrum for the fine grid fields will be very close to that of the coarse grid fields down to the Nyquist wavenumber of the coarse grid. The fine grid will then have (probably small) non-zero amplitudes in the Fourier spectrum down to its Nyquist wavenumber. These intermediate Fourier amplitudes will be guided by the form of the correlation function. This additional power seems to me to be a modest form of down-scaling that could be based on the estimates of the statistics of the ocean fields.*

Response. We agree that the SDD method is prone to some form of down-scaling penalty. However, such penalty is significantly smaller than when using deterministic interpolation methods. We tried bi-linear and bi-cubic, however we think that a similar behaviour will be with other deterministic interpolators such as inverse distance, Sheppard's method, moving average etc. The actual values of downscaling errors, i. e. the difference between the downscaled or interpolated field and the true field are given in Figure 3 for the true field sampled with zero errors, and in Table 1 for a noisy true field. We attribute the better skill of the SDD method by the fact that it calculates the weighting coefficients based on the properties of surrounding data while the usual interpolation methods use weighting coefficients based only on the distance between the points no matter what the data are.



It is true that the Fourier spectrum of the field produced by SDD is close to the spectrum of the true field on the coarse grid up to the Nyquist wavelength of the coarse grid as can be seen on the figure below (also added to the revised MS). This figure shows the spectra of the following fields (i) true field on coarse grid, (ii) downscaled with SDD on fine grid, (iii) bi-cubic interpolated onto the same fine grid. The main peak on SDD spectrum is closer to the true peak than that produced by bi-linear interpolation. In the spectral region between the coarse and fine grids Nyquist wavenumbers, there is a parasitic peak which is an artefact caused by distortion of the fields downscaled by SDD as well as by bi-cubic interpolation from the coarse grid. However, this artefact is much smaller in the case of SDD which demonstrates its better skill of recovering the true field.

The idealised case where we know the true field gives us some confidence that the additional powers at high wavenumbers in the real world situation are mainly a representation of the true field, not artefacts.

We agree that the better representation of smaller scales by the SDD is due to resolving additional range of wavenumbers between the Nyquist values for coarse and fine grids.

We have amended the text to reflect the above (lines 255-265 in the revised MS) and added a new figure (now Fig 4)

Comment. *The enstrophy field clearly has larger values on the fine than the coarse grid. I am not sure whether this relates to the additional power in the Fourier spectrum or the fact that the derivatives for the fine grid are calculated using smaller grid spacing than on the coarse grid.*

Response. We think that the two effects are closely related. The additional range in the Fourier power spectrum is a reflection of smaller grid spacing and hence a potential of having sharper gradients. **We have clarified this is the text, lines 579-581.**