Ocean Science
Discussions

Open Access

OSD

**Interactive
comment**

Interactive comment on "A hybrid data assimilation method and its comparison with an Ensemble Optimal Interpolation scheme in conjunction with the numerical ocean model using altimetry data" *by* Konstantin Belyaev et al.

Anonymous Referee #1

Received and published: 25 July 2019

The manuscript presents an application of a recently formulated method, the Generalized Kalman Filter (GKF), to the assimilation of altimeter data into an eddy permitting model of the Atlantic Ocean. The results in terms of analysis and forecast error variance are compared against a simulation and an assimilation based on an Ensemble Optimal Interpolation method (EnOI). When evaluated with the assimilated altimeter data, GKF is shown to perform better for both metrics than the EnOI. The comparison is kept to the minimum and only the evolution of the globally averaged metrics and fields of one time instance are shown. The additionally impact on SST provides very little in-

formation except that GKF is able to produce larger changes in SST on eddy-scales than EnOI. Whether they go into the right direction is not possible to judge. Overall the amount of verification would be appropriate for an illustration in a paper that presents the method, which however has already been published. The presentation is very poor particularly due to the abundance of grammatical errors. I stopped to list the necessary changes on page two, since I felt that the manuscript would require too much extra work. The method remains obscure particularly because of deficiencies in the presentation. There are also problems in the application (potentially also in the formulation) of the method noted. Notably the application of the Birkhoff-Khinchin theorem on the drifts does not make sense to me as substantiated below. Additionally, the application of the EnOI seems to be problematic since it depends on two parameters whose effects have not been explored and their choices have neither been stated nor the selection criteria explained. One parameter, the scaling factor for the error covariance, seems to be incorrectly implemented, the other parameter, the error covariance of the observations, does not appear to be relevant in GKF, which hardly can be correct unless it is assumed to be zero there. The application of EnOI actually demonstrates that EnOI as almost a failure, not being able to adjust the SSS on eddy-scales. This is surprising since this method has been proven to work well with this model before in the Gulf of Mexico. I would expect that the analysis error could be almost arbitrarily reduced with decreasing observational error R, which means that the performance of GKF relative to EnOI can be adjusted by varying R in a way that desired result is obtained. This could be for instance be the one shown in the manuscript: EnOI reduces the analzsis error but less than GKF. Last, it remains unclear in what way GKF is a hybrid method, since this term is only mentioned in the Title and the Abstract and no further motivation or explanation are given.

Details (improvements to the Grammar only for page 1 and 2):

P1

L14-15: How is the ability to assimilate data being judged? I guess by producing a

result that this closer to the observation after assimilation. I suggest to connect the two parts into one and make clear relative to what state the analysis is "closer"

L20: remove "a" from "A data assimilation" and "a great"

L22 "the" in front of "Australian" and "American"

L24 "altogether" instead of "at all"

L25 remove "on how"

L31 "Ghil et al (1991)"

L34 "led"

L35 remove "the" before "data

L36 replace "the" before essential by "an"

P2

L1 replace "indicate" with "name"

L2 add "the" in front of "Australian" and "American"

L5 "economic" instead of "economy"

L6 "medium" instead of "media"

L7 add "the" before "used"

L9 add "a" before "given" and change "with respect to" to "represented by"

L12 replace "seeking" with "sought" and replace "in a" by "depending on"

L13 research is uncountable and cannot be used with several replace "be found in (Marchuk" with " be found (e.g. Marchuk"

L16 not clear what the sought field is in this context. It was the initial state in the

Printer-friendly version

Discussion paper

previous paragraph but minimizing is variance would not make sense to me. Do you mean variance or variation, it the latter case it seems to be just the variational method.

L17 I think it should be "the observed variables are the sum of the true signal, which the model is supposed to represent, and stochastic ...."

L18 remove "the" before "known"

L19 should be "Pendruff et al. (2002)" this is one of the modern development but not the only one.

L20 sloppy formulation, should read "to the papers of Evensen (2009) and Xie et al (2010)"

L22 I don't understand what that means. Change to "Van Leeuwen (2015)" and "Van Leeuwen (2011)", respectively

L25 remove "the" before "hybrid and "both" and replace "are" with "have"

L26 add "a" before "functional"

L28 should be " found (e.g. Lorenc et. al., 2015; Tanajura et. al., 2013). We may also refer to Tanajura et. al. (2009),"

L34 In the EnKF you have and ensemble of model runs but only one representation of the observations. Ensembles of observations are the same as just observations.

L35 "As a consequence..." is this a problem or the goal of the method? In the latter case it should be formulated differently.

l36 Not clear why it becomes better if no assimilation occurs. Better in the sense of what? Numerical forecast does not make sense to me because this might be ultimately happening but I don't see how this as a direct consequence. It could be better in the sense of obeying the dynamical equations

L37-40 Minimum energy is characteristic for equilibrium states. For transitions I would

apply conservation principles, e.g. those of mass or momentum. This would be equivalent to minimum energy change. Maybe this is meant here?

P3

L4-5 Please motivate why this method is used for comparison and not the EnKF as a more advanced method. Is this is more similar to the GKF?

L7 "data" instead of archive

L8 Strange formulation, the following is not about the possibility.

l14 I don't get where the Fokker-Planck equation may play a role in this comparison. Eq. (3) eq (3) doesn't make sense from the dimensional point of view. The first () is a scalar, the second () is a vector of the model-dimension while the third () is a vector of observational dimension and Q a square matrix of observational dimensions.

P4

L6 I suggest to use "the model variables that are observed". "the observed variables" are understood as the observations.

l9 As the observational trend I would expect the difference $Y_{n+1} - Y_n$ rather than the difference between model and data

l10-11 should read "coincides". The operator for the "prolongation" needs to be specified. Prolongation is as term that refers to time, while I have the impression that actually interpolation/extrapolation in (phase) space is more relevant here.

l11-12 I don't understand the ensemble average. This requires that something is observed in multiple ways which is rarely possible, except for if two satellite tracks cross. I think it should be made clear from the beginning that your ensemble is generated from several time instances before the assimilation step

L13 this Y needs a different symbol than the one above on line 10 because they have

different dimensions. Also Yn+1 depends on Cn+1 while Cn+1 depends on Yn+1, so it is not clear how this could work Eq.5 Why is the formula for Cn+1 different here from line 10 ?

L32-33 "In the expanded form": If different forms for the equations exist this should be made clear from the beginning and transition operators defined.

L37 you mean Eq. (5)

l36-37 I don't think this theorem applies for the cases that you have in mind. If the system is stationary there should be no trend. Eq. (5) provides an estimate of the tendency at a particular time. This is supposed to change with time. While Eq.(7) will change less and less with progressing time. For a constant tendency (that means slowly varying changes) (7) will even have a different sign because most of the Xai will soon be in the past relative to n.

P5

L9 The layers should have different but uniform densities.

L17 innovative use of language, but for normal people better remove relatively and just say "the temporal average ... has been removed"

L19 Previously..." If you did this you have to describe what has been done otherwise a suitable reference should be provided.

L22-23 I don't understand what this has to do with parallelization. After parallelization you should be able to handle larger problems. And what does "reduced" mean - what has been done to reduce the size.

L25 why not say what they use?

L26-27 Particularly here you need to provide details of the scheme that was used to create the data for comparison. It is not even clear whether the respective results are discussed by Tanajura et al (2015) or elsewhere.

L33-34 "The high parallelization..." I don't understand what this means?

P6

L3-4 no freshwater fluxes? Is there at least surface SSS relaxation?

L6 what does real mean. Are the without error.

L14 I don't understand how you get the number 50, do I need to know the time step for this? Eq.(8) in the ensemble OI the factor alpha also appears inside the () in front of H thereby acknowledging the fact that the variance over time typically is too large.

L24 what are these considerations and how are the actual values for R and alpha? Why is there no error covariance matrix R necessary in GKF? Does it mean GKF uses zero error?

P7

Figure 1 which day is this and which model run?

L15 better: initialized from the analysis at the past time n-1

L16 simply: SLAa is the analysis at the time n Figure 2/3 what are the units on both axes? Are these days, which means the forecast period is one day? The forecast period should explicitly stated somewhere

P8

L5 How do the results depend on the heuristically defined values alpha and R in the EnOI? Can EnOI get closer to the data with a larger alpha and/or a smaller R, particularly since R seems to be zero in GKF?

L14-15 My interpretation of Figs.2/3 is that the curves are displayed only until day 27 therefore I can not verify this based on Figure 2 and 3

Figure 4:please provide units and a colorbar the I can only read 0.3.

P 11

L4-5 What is the argument here, the states are too different therefore B too large? Maybe the reason is the omission of alpha inside of the brackets of Eq. () or an inappropriate value for alpha. EnOI hardy improves the state in regions where the variability is large.

P13

l10 Why not the RMS differences instead of the mean bias, the mean bias across all moorings is not so interesting.

l 14-15 move or copy to the captions.

P14

l1-6 This comparison is too superficial to be helpful since SST is only shown to demonstrates only the existence of corrections on eddy-scales.

l9-10 I would have hoped that the study provides evidence rather than just asserting something.

———————————————————

Interactive comment on Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-56, 2019.