

This version of the manuscript has greatly improved the previous one. The reorganization of the sections and the rewriting of a substantial part of the text has brought to the document a well defined structure and a language that is clear and easy to follow. I have particularly enjoyed the didactic style of some of the parts, which is very appropriate for a review document where the most likely readers will be newcomers to the topic.

I would like to thank the authors for the effort spent correcting what was suggested by the reviewers, as well as other things.

The paper is now close to be accepted for publication, only some minor issues and inconsistencies should be corrected.

I understand that this version of the manuscript has been submitted without highlighting the changes because it has been extensively re-written. I encourage the authors to provide a manuscript highlighting the changes for the next version.

Minor revisions

There are some issues in the description of the data assimilation methods

Section 4

P24L25 “by partially correcting surface wind forcing”. This phrase suggests that DA is correcting the forcing data and DA usually only corrects the model state. Can this be clarified a bit further?

Section 4.2

It would be worth mentioning that in EnKF applications the error covariance matrix P_f is almost never calculated explicitly using the formula indicated in (55). Instead, the model is first converted to the observation space and the terms $P_f H_t$ and $H P_f H_t$ are then calculated as summations. This substantially reduces the computational cost and removes the need for a H_t operator. Please see eqs 6 and 7 in Houtekamer and Zhang (2017), for a clarification on this.

Section 4.3

It should be noted that Hoteit (2009) description of 4DVAR differs from the more common formulation used in several operational atmospheric and ocean implementations. The difference lies in the second term of (59) that is commonly defined at the analysis time (t_a) and is not time dependant. This formulation ensures that the analysis minimizes the error with the background field at $t=0$ and with the observations across the time window, and is the reason why only the observation term is time dependant. For a reference on this, please check, Mike Fisher ECMWF lecture notes, eq 4, and Mogensen and Alonso-Balmaseda ECMWF technical note 668, eq (1) (see at the bottom for links to these references). These documents contain descriptions of 4DVAR from atmospheric and ocean implementations as they are used in several operational centres. A more theoretical discussion can be found at LeDimet & Talagrand (1986), which is a 4DVAR classical paper and it would be a nice addition to the paper references. Authors might want to reconsider the proposed formulation for 4DVAR.

The matrix B is named inconsistently as “control variance matrix” (P32L15, P32L18), “covariance matrix” (P32L25, P32L29) and “error covariance matrix” (P33L19). I mentioned in my first review that the B matrix can only be interpreted as the “error covariance matrix”. Please check the paper

Hoteit et al (2009), cited at the manuscript and used as a guideline for the section, for a clarification on this. In Hoteit (eq 2, equivalent to eq 59 in the manuscript), the R and B matrices are noted as R and Q and described as: “R(t) and Q(t) and are the covariance matrices of observational and first-guess control uncertainties, respectively”. In the context of Hoteit et al (2009), “uncertainties” is equivalent to “errors” or “unkowns” in other formulations. Please, review the manuscript to ensure that the B matrix is noted in the appropriate way.

P32L15, “R and B ... are a function of time”. While at Hoteit et al (2009) these matrices are defined as time dependant, this is not always the case. In many applications (i. e. operational implementations), R and B are commonly formulated as climatic (static) matrices. I suggest to re-phrase as “R and B ... might be time dependant”. Also, this is consistent with the way R and B are noted in (59), where their time dependency is not explicitly indicated.

Comments/ Other

P1L8 fie ld > fields

P1L10 models > models

P1L19 f ew > few

P4L17 “the number of HF has increased...” this phrase is then repeated at P2L21. Authors might consider rewriting.

P4L27 satellite > satellites

P4L31 “Improvements of a better understanding” sounds redundant. May be “A better understanding” or “Progress in the understanding”?

P5L10 I would add a description of section 2 before describing each of the sub-sections. i.e “Section 2 reviews retrievals from satellite observations...”

P5L10 sea level > sea level measurements

P6L24 directions > direction (or is it “perpendicular directions”?)

P7L8 level > the level

P7L11 strung > strong

P8L12 “this example” repeated in P8L13, please re-phrase.

P8L23 “SST of” > “SST at”?

P8L31 “the residuals to respect along track data”, not sure if this part of the phrase is correctly worded.

P9L26 “allowed” followed by “allowing”. I am being picky here, probably fine...

P10L5 “which is the Lagrangian mean velocities due to waves”. Wrong number. “which are” or “which is ... velocity”.

P14L27 wave > waves

P15L24 Ocean > ocean

P19L3 remove “does”

P19L19 “an exponential stratifications”. Wrong number “exponential stratifications” or “an exponential stratification”

P20L14 “SQG” or “eSQG”? Please review.

P22L13 Something is missing in the phrase starting with “An alternative to...”. May be “An alternative is necessary to...” or “There is an alternative to...”

P22L14 “to” duplicated.

P24L5 Wrongly formatted reference.

P25L15 wrong section reference

P27L22 “Gain”, wrong format

P27L24 “... assimilation increment is used...” > “... assimilation increment and is used...”

P27L26 “defined given”, choose one

P27L30 describes > describe

P30L10 Figure, wrong format

P33L3 o f > of

P34L10 After minimization > After the minimization

P34L13 t he > the

P34L17 The phrase “The control ... conditions” is incomplete. Please review.

P35L7-8 What is “Lagrangian predictability”? Please review and clarify.

P35L11 “together with a large spatial coverage”. Why is this bad? Is it because is impacting a large area of the model? Please, clarify.

P35L21-P36L2 Phrase incomplete, please review.

P36L12 this > This

P36L17 open > opens

P36L17 to > the

P37L6 method ologies > methodologies

Caption Fig 5: “paremeters” > “parameters”

References

M Fisher. *Assimilation techniques (4): 4dVar*. ECMWF Lecture Notes.

(<https://www.ecmwf.int/sites/default/files/elibrary/2002/16933-assimilation-techniques-4-4dvar.pdf>)

P. L. Houtekamer, F. Zhang. *Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation*. Monthly Weather Review. December 2016.

F.X. LeDimet, O. Talagrand. *Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects*. Tellus. 1986

K. Mogensen; W. M. Alonso Balmaseda. *The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4*. ECMWF Technnical Note 668.

(<https://www.ecmwf.int/sites/default/files/elibrary/2012/11174-nemovar-ocean-data-assimilation-system-implemented-ecmwf-ocean-analysis-system-4.pdf>)