Manuscript prepared for Nonlin. Processes Geophys.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 16 September 2015

Response to the Reviewers and the Editor regarding "Hybrid Levenberg–Marquardt and weak constraint ensemble Kalman smoother method"

J. Mandel[1], E. Bergou[2,3], S. Gürol[3], and S. Gratton[3]
[1]University of Colorado Denver, Denver, CO, USA
[2]Institut de Mathématiques de Toulouse, Toulouse, France
[3]INP-ENSEEIHT and CERFACS, Toulouse, France
Correspondence to: J. Mandel (jan.mandel@gmail.com)

1 Introduction

We would like to thank the editor and the reviewers for their helpful comments, which will contribute to improving the paper.

2 Reviewer 1

1. About the numerical experiments and results

1.1 For the proposed method to be considered as a serious alternative to others, its performance must be compared to others, like the standard EnKF (or ETKF), the SIR filter, on a systematic basis. See for example the works of Oke, Sakov, Bocquet, to cite only a few names.

The present method consists of introducing a stochastic solver inside 4DVAR, which is defined as minimizing an objective function, so we evaluate how close its performance is to 4DVAR in terms of that objective function. We also verify that the approximate solver still results in good performance overall. It was not our objective to repeat the studies comparing variational methods with sequential filters. We will clarify the objective of the computational tests. We will also run a similar comparison in terms of RMSE as in Bocquet and Sakov (2013, Fig. 6) and Goodliff et al. (2015) for Lorenz 63 to assess the overall performance of the method.

1.2 The diagnostics must be statistically robust. According to the authors cited previously, and also based on my personal experience (Metref et al., 2014), the diagnostics of DA with the Lorenz 63 system are robust if they are based on 100,000 assimilation steps at least, excluding spin-up.

We will run the assimilation on Lorenz 63 longer as indicated.

1.3 For the QG experiments, I find it quite reductive to limit the diagnostics to the value of the objective function.

The purpose of measuring the objective function was to assess how close the method is to 4DVAR. We will measure also RMSE to assess the overall performance.

1.4 Several statements are not motivated, unclear, or inaccurate: p.882: why choosing 8 iterations? p. 882, lines 23-25: "the objective function decreases with iterations". Not for $\tau = 0.1$ and 0.01 where the function increases at the last iteration. What happens with more iterations? This requires further investigations. p.883, line 8: "for the first iteration, the best decrease in objective function is obtained when τ=1": actually, it increases from 1 to 2. Is the first iteration before that? What is the initial value of the objective function?

p. 882 lines 23-25 refer to Table 2. After a small number of iterations (about 6) the value of the objective function stopped decreasing (essentially, convergence was reached) and its values varied randomly around a limit value forever (this is a randomized method). This is similar to what happens in iterative methods when the rounding precision is reached: the error stops decreasing and then it only varies due to rounding.

883 line 8 is a conclusion drawn from Fig. 4 Larger τ was somewhat better early in the iterations but mainly it prevented the minimum from being reached in later iterations. $\tau = 1$ should be $\tau = 10^{-2}$.

We will clarify what happens for a larger number of iterations as stated above. Fig. 4 will be only used to draw the conclusion that τ needs to be small enough.

2. About the presentation of the method

2.1 More must be said about the computational complexity and the implementation complexity. Of particular importance is the increment of work from an EnKF, for example. And what are the assets of the method, compared with others?

The cost is $(\text{ensemble size} + 1) * (\text{times the number of iterations}) * (\text{lag}) + (\text{ensemble size}) * (\text{lag})$ *in each analysis cycle. For comparison, the cost of the EnKF is* ensemble size *evaluations of the model in each analysis cycle.*

The main advantages of the new method compared with the literature are as follows:

- *The method asymptotically approaches 4DVAR for small τ and large ensemble size, which can be proved rigorously (Bergou et al., 2014), thus it inherits (in the limit) the advantages of 4DVAR.*
- *The implementation of the Levenberg-Marquard reqularization as additional observation is statistically correct, because the EnKS is used to solve only the linear least squares.*
- *The ensemble of increments is generated fresh in every iteration, which prevents the minimization from being restricted to the span of a single fixed ensemble, as in, e.g., Bocquet and Sakov (2012, Algorithm 3).*

2.2 Algorithm 3: If I understand it well, the algorithm consists in: - compute the full model trajectory from x_0; - Apply an EnKS on z, which dynamics are governed by the model linearized at the previously computed trajectory; - update the trajectory, re-linearize the model (and H), and iterate. I understand that the method is proposed as an improvement (or another way to solve the inner loop) of the incremental 4DVar. But I do not see where the "variational" part of the algorithm is, other than in the perturbations z used for the EnKS. Meteorologists and oceanographers are used to speak about variational methods when the objective function is explicitly minimized to reach the solution. Here, it seems that the calculation of the objective function is not essential to solve the problem, but is only used as a diagnostic. To me, it looks more like a hybrid of a "two-step" smoother and an EnKS (see for example Cosme et al. (2011), but I do not request you to cite my work). Could you clarify that?

The present method is set as Gauss-Newton (or Levenberg-Marquardt, when regularization is added) for the minimization problem in 4DVAR. With a large enough ensemble, and small enough τ, it becomes asyptotically incremental 4DVAR, as proved in Bergou et al. (2014). That is, the objective function is explicitly minimized – even if the objective function or its gradient are not actually used in the algorithm. This is in fact an important motivation of this method.

2.3 Algorithm 3: Following the previous point, if I am right and if you agree, the name “EnKS-4DVar” should be modified.

We would prefer to keep the same name. Also, other related methods use names with various combinations of "En" and "Var", so it is an established nomenclature.

2.4 I wrote earlier that the presentation was clear and concise, and I like it. But it requires a significant amount of background in data assimilation to understand the paper. Perhaps the authors could guide the reader toward some appropriate references for his/her self-education if necessary (more than in the present version).

We would like to thank the reviewer for kind words. We will recommend Evensen (2009) and Kalnay (2003) for background material.

3. Minor comments, typos, etc p.872, Beginning of section 3: the notation z for the states (line 5) is confusing after notation x in the previous section. Perhaps you can anticipate this by stating shortly why you adopt this notation (a short statement in brackets should be enough). When we understand the rationale of this notation (later in the text) Equation 5 becomes all the more confusing because it involves a nonlinear model. I am sure the authors will find a smart way to make things a bit clearer.

We went back and forth few times on this before. At one point we tried δx instead of z. Algorithm 1, where Eq. 5 resides, serves double duty: first it is the linear tangent equation for the increments z, then it is the reference statement for the nonlinear EnKF which is further developed in the smoother in Algorithm 4. The reference statement is needed so that one can formulate Theorem 1, namely, the method with $\tau = 1$ becomes the nonlinear EnKS. We had tried to duplicate the algorithms in a previous version but since then we have streamlined the paper. We will explain the reason for the nonlinear M in eq. 5 and the notation z as above.

Equation 6, 19, and elsewhere: I know it is common to perturb the observations y, but the usual observation equation ($y = h(x) + \epsilon$) says that $h(x)$ should be perturbed instead, as in the model equation. This reverses the sign of the perturbation. Of course, it is equivalent with centered and symmetric noise as it is here. Since it is (unfortunately) common to present things that way, this changes of sign is not a strong requirement.

We would like to leave it as is.

Introduction and almost everywhere: Although the Kalman filter is indeed due to Kalman (1960), optimal linear smoothers are not. "Kalman smoothers" should be replaced by "Smoothers based on Kalman's hypotheses". But I agree this is quite cumbersome. If the authors does not find an easy way around this, I do not make it a strong requirement.

We would like to leave it as is. "Kalman smoother" is the commonly used term, even if, as the reviewer points out, it may not be quite right..

Equations 14 and before 11: should the first x be replaced by a z?

Yes, replaced.

p.875, first line: perhaps a reference to Eq. 4 rather than Eq. 2 would be more appropriate.

Agreed, replaced.

Figures 4 and 5 are not used in the discussion. They could be removed.

Reference to these figures was inadvertently omitted in the discussion on p. 883, see the response to 1.4 above.

p.887, lines 16-17: "it is capable of handling strongly nonlinear problems". I tend to disagree with this statement. The Lorenz 63 system fully observed every 25 steps is considered "weakly nonlinear" (Sakov et al., 2012; Verlaan and Heemink, 2001; Metref et al., 2014) and the QG model with grid meshes of 300 km and observed with a ratio 1/32 is probably not very nonlinear (I do not have a reference for this).

We will use observations of Lorenz 63 as in Bocquet and Sakov (2013, Fig. 6) and Metref et al. (2014). For QG model we run the experiments for 10 days where we can see that nonlinearity is increasing (Fisher et al., 2011, Fig. 2). We will mention this more clearly.

p.888, lines 2-5: the advantages of using varying τ's are only speculative from the results presented. They are not shown.

We will make clear that we are only suggesting it as a possibility in the discussion, rather than actually doing it.

p.888, line 7: the QG model is one of the simplest models of the atmospheric circulation. It cannot be considered standard, because it is rarely used for meteorological applications.

We have rephrased as "one of the widely used model in theoretical atmospheric studies, since it is simple enough for numerical calculations and it adequately captures an important aspect of large-scale dynamics in the atmosphere."

3 Anonymous Reviewer 2

1 General impression

I believe the paper is interesting. In particular, the use of the EnKS to solve the inner loop problem is the real novelty of the paper worth investigating. I am less pleased with the treatment of the literature. Some contributions need to be mentioned. Others are discussed and mentioned but not properly described, or part of the results relevant to this paper omitted. Grey literature is mentioned. In theory it should not. I personally don't mind but then you should also mention other non peer-reviewed contributions of other colleagues.

Moreover, there are a few unjustified statements. For instance, the standard EnKS as presented as if it was a novelty. Also, the paper does not truly deliver on the promise, especially at the end of Section 5. The numerics is technically fine, but not entirely convincing. Overall I would ultimately recommend the publication of this paper, but on the condition that the following remarks are properly addressed.

Main comments

1. (a) page 869, l.11-41: This passage has wrong statements, and uses gray and peer-reviewed literature in a biased way. First of all, let me say that the IEnKF/IEnKS is quite complementary to your idea of using the EnKS to solve the inner loop problem. It has always been claimed (Bocquet and Sakov, 2012, 2013, 2014) that the IEnKS/IEnKS could use a different optimizer (on the shelf, Quasi-Newton, Levenberg-Marquart, etc.). Quasi-Newton and Levenberg-Marquart methods have indeed also been used in those papers. The IEnKS could easily incorporate your idea and use the EnKS to to solve the inner loop problem, which would make a nice blending!

(b) "Additional work appeared after the first version of this paper was written (Mandel et al., 2013). Bocquet and Sakov (2014) extend the method of Bocquet and Sakov (2012) to 4DVAR..." : This chronology is biased and incorrect for these reasons:

- If you use gray literature then you should mention: http://www.meteo.fr/cic/meetings/2012/ensemble.conference/presentations/session04/1.pdf
- Bocquet and Sakov (2014) appeared online in final form with a doi number in 2013.
- Please also cite Bocquet and Sakov (2013), which additionally offers a comparison with a (fully cycled) 4D-Var.

We will not mention the preprint of this paper, but we still need to cite Bergou et al. (2014), which is submitted and in review, and important for the argument here.

(c) Sakov et al. (2012); Bocquet and Sakov (2012, 2013, 2014) not only use finite-differences but also an ensemble transform approach without rescaling which proved to lead to very similar performances. Finite-difference/bundle is interesting in that it mimics the tangent linear, although the ensemble transform is more elegant. This is of direct relevance to your discussion of τ in Section 4. Please mention it.

We will add something like: It is interesting that the ensemble transform approach in Sakov et al. (2012); Bocquet and Sakov (2012, 2013, 2014) corresponds to our $\tau = 1$, but it does not seem to reduce to the standard EnKS.

(d) "However, Bocquet and Sakov (2014) nest the minimization loop for the 4DVAR objective function inside a square root version of the EnKS and minimize over the span of the ensemble, rather than nesting EnKS as a linear solver inside the 4DVAR minimization loop over the full state space as here." This sentence seems nice but it is partially misleading in at least two ways: (i) the IEnKS is more than what is implicit here as it incorporates cycling which is one of the main results of Bocquet and Sakov (2014). So the sentence should start with something like "Focusing only on the variational analysis..." (ii) Bocquet and Sakov (2012, 2013, 2014) emphasized that the minimization can be performed differently opening the way to many consistent variant in the variational analysis. Using your idea of the EnKS for solving would actually be a nice addition to the IEnKS.

We provide the response to (d)-(3) after (3) below.

(e) "Their method is tied to the use of the sample covariance matrix of the state without localization of the covariance and to strong-constraint 4DVAR": This is partially incorrect for the second statement and plain wrong for the first. Please remove entirely this sentence. I agree that (Bocquet and Sakov, 2014) strongly rely on the strong-constraint hypothesis (which is not the case for Bocquet and Sakov (2013)). As for localization, it seems that it was not used in Bocquet and Sakov (2014) on purpose. But it was not claimed it is not possible to use it, only that this is not as simple as with the EnKF. Actually, localization can be used in the IEnKS. Preliminary results were reported early in 2013 http://das6.umd.edu/program/das6_program.html in the largest international data assimilation conference. Please mention clearly that localization has been shown to be possible with the IEnKS.

(f) "However, limiting the EnKF to linear combinations only does not allow common approaches to localization (Sakov and Bertino, 2011)." This is wrong. Please remove the sentence. Local analysis/domain analysis which limits the EnKF to local linear combinations, is extensively used in data assimilation, notably, but not only, via the popular LETKF (Ott et al., 2004). Please read Sakov and Bertino (2011); Nerger et al. (2012). That is why it is rather straightforward to implement localization in the IEnKS. It seems to me that you try to create an opposition that does not exist.

(g) "Ensemble methods for the solution of the 4DVAR nonlinear least squares problem in the weak constraint 4DVAR, or ensemble methods for this problem which allow localization, do not seem to have been developed before.": I disagree. There are published papers (not to mention gray literature) that already discuss the issue in an ensemble variational context, some of them being difficult to ignore for the readership of Nonlinear Processes in Geophysics. For instance: Chen and Oliver (2013); Desroziers et al. (2014); Lorenc et al. (2014) to quote just a few.

We wil add something like: In principle, methods that work in the span of the ensemble can be developed into localized method by the use of local linear combinations (e.g., Ott et al., 2004), or by the Schur product, and methods that rely on Gauss-Newton can be regularized to become Levenberg-Marquardt. The impact on computational cost can vary. For related methods with localization, see, e.g., Chen and Oliver (2013); Desroziers et al. (2014); Lorenc et al. (2014).

2. Implementing Levenberg-Marquardt in the solution of an EnVar problem has been considered first, tested and validated in Bocquet and Sakov (2012) and Chen and Oliver (2013). Surprisingly the authors mentioned "and Bocquet and Sakov (2012), who added regularization" but not the fact that this regularized is based on the Levenberg-Marquardt scheme... Please mention those references, and make it clear. Bocquet and Sakov (2012) did not find any convergence problem with their application, but rather use it as a faster convergence method, as an adaptive method between steepest descent and Gauss-Newton.

Here are quotations from (Bocquet and Sakov, 2013, 2014):

“One has a choice of minimization scheme: for instance, Sakov et al. (2012) used a Gauss-Newton scheme whereas Bocquet and Sakov (2012) advocated the use of the Levenberg-Marquardt scheme (Levenberg, 1944; Marquardt, 1963) for strongly nonlinear systems. In this article we shall use a Gauss-Newton scheme, because the emphasis is not specifically on strongly nonlinear systems and the number of iterations for convergence in the experiments below is rather limited for most experiments.” “The Gauss-Newton minimization scheme shown in Eq. (2) can easily be replaced by a quasi-Newton scheme that avoids the computation of the Hessian, or by a Levenberg-Marquardt algorithm that guarantees convergence of the minimization. These alternatives have been suggested and successfully tested in Bocquet and Sakov (2012).”

3. page 868, l.23-26. “Gradient methods in the span of the ensemble for one analysis cycle (i.e., 3DVAR) includeZupanski (2005); Sakov et al. (2012) (with square root EnKF as a linear solver in Newton method), and Bocquet and Sakov (2012)” This is wrong. The iterative ensemble Kalman filter in Sakov et al. (2012) and Bocquet and Sakov (2012) is already a 4D ensemble variational method as it has a temporal variational analysis. It coincides with the iterative ensemble Kalman smoother with only one batch of observations. It can be seen as a one-lag smoother. Actually your method essentially coincides with the IEnKF in the lag-one case (modulo some irrelevant details such as the use of stochastic perturbations or not)! Note that Sakov et al. (2012) actually compared two variants of the IEnKF (lag-one smoother): one with the tangent linear model and one with the nonlinear model, which is of direct relevance to your discussion of τ.

The reviewer’s understanding of the related work is clearly much deeper than ours and we would like to thank the reviewer for valuable insights. With the reviewer’s permission, we would like to use some of them in the paper. We will provide brief synopses like the following, based on our fresh reading of those papers, and the reviewers’ comments. They are in separate paragraphs for clarity of this response, but they may be more in flowing text in the revised paper.

– *Zupanski (2005), Maximum Likelihood Ensemble Filter (MLEF): iterative minimization of the 3DVAR cost function in the span of the ensemble in eache analysis cycle, with preconditioning by approximate Hessian computed as in ensemble square root filter (ESRF), the ETKF variant.*

– *Sakov et al. (2012), IEnKF: minimization the 3DVAR cost function in each assimilation cycle in the span of the ensemble by Newton's method with ETKF as linear solver. Rescaling of the ensemble spread to approximate the tangent is used, similar to the the τ here, rather than a finite-spread approximation.*

– *Bocquet and Sakov (2012) combined the IEnKF method of Sakov et al. (2012) with an inflation-free approach, and Levenberg-Marquard method by adding a diagonal regularization to the Hessian. The present method essentially coincides with the IEnKF in the lag-one case, except for the use of random perturbations and implementation of the regularization as additional observation.*

– *Bocquet and Sakov (2013), extended IEnKF to smoother (IEnKS) with fixed-lag and moving window and noted that Gauss-Newton can be replaced by Levenberg-Marquard. The method is formulated in terms of the composite model operator $M_{k\leftarrow 0}$, i.e., with strong constraints.*

– *Bocquet and Sakov (2014) incorporate cycling and minimize over the span of the ensemble, nesting the minimization loop for the 4DVAR objective function inside a square root version of the EnKS.*

– *Chen and Oliver (2013): Levenberg-Marquardt-ensemble randomized maximum likelihood (LM-EnRML) is an incremental method variational methos with square-root smoother as linear solver, with localization of the covariance by a Schur product (term-by-term multiplication).*

– *Lorenc et al. (2014) provides a comparison of the hybrid 4DEnVAR and hybrid 4DVAR for operational weather forecasts. "Hybrid" refers to a combination of a fixed climatological model of the background error covariances and localised covariances obtained from ensembles. 4DVAR is the traditional variational method whereas 4DEnVAR is a variational-ensemble method in which the localised linear combination of an ensemble of nonlinear forecasts is used for the minimization. This is similar to the case $\tau = 1$, plus the localization.*

– *Desroziers et al. (2014) use the Gauss-Newton method for the solution of the weak-constraint 4DVAR minimization problem, with the inner loop performed by ensemble Kalman smoother. For adjoint computations they also use $\tau = 1$.*

4. What you called the nonlinear EnKS (Algorithm 4) is actually the standard EnKS as implemented by the geophysical data assimilation community! You will find many variants (depending on the flavor of the EnKF, perturbed observations or not, with or without model error,

with or without localization), but they strictly follow the same smoothing principle: an EnKF pass operated with the nonlinear model, and a backward smoothing pass.

As as far as the EnKS is concerned (the question is richer in the IEnKS context, and could be in your section 5), the question of using the tangent linear model or not only appears in the EnKF pass and it has been discussed over 20 years. This is what is commonly refereed to the reduced rank Kalman filter approach (RRSQRT) versus the EnKF which differ by the use of the tangent linear or the full model in the propagation. The reason why the nonlinear model is preferred is because it is simpler and natural and capture some nonlinear effects (which turns out to be often more precise). Hence, what you call the nonlinear EnKS (which in light of the previous comment is a pleonasm) is what is actually used in Evensen (2009); Cosme et al. (2010); Nerger et al. (2014); Bocquet and Sakov (2012, 2013) and several others (see also Cosme et al. (2011)). This should be stated clearly.

On page 877, we say "we recover the standard EnKS applied directly to the nonlinear problems... Algorithm 4," which is labeled "Nonlinear EnKS". We'll clarify this and label the algorithm consistently "Standard EnKS."

5. As mentioned earlier the novel and appealing idea of this manuscript is the use of the EnKS to solve the inner loop problem of a nonlinear problem. Almost up the to end of section 5, the discussion is on the reformulation of known methods and techniques, and the expectation of the reader is great at this point. But, the final theoretical piece of the study does not seem to be given. Where do you describe the full algorithm with the regularization? It is necessary that you give it, because this should stand as the essential piece of the paper and one might think that there is nothing essentially new without it. Besides, this is where nonlinear ensemble variational methods gets trickier. Please, explain precisely how you solve Eq. (23) and give us the complete algorithm. This is critical for the paper.

The final algorithm is Algorithm 3 EnKS-4DVAR on p. 876, and it incorporates the previously specified components that build up to it, Algorithms 1 and 2 by reference, much like calling a function in code. Algorithm 4 Nonlinear EnKS is called standard EnKS in the text on the same page. It is needed only to show what happens when $\tau = 1$. *We will remove the "nonlinear". The regularized eq. (23) was solved also by EnKS using the penalty as additional observation following Johns and Mandel (2008), as shown in the text. The description is informal rather than stated as an algorithm, but we think in sufficient detail to program. We did not want to repeat significant parts of Johns and Mandel (2008) and burden the paper with more complicated notation. We will provide the final algorithm in a complete form now, withough relying on references to equations and verbal descriptions.*

6. The numeric is technically fine and using the OOPS QG model offers a nice illustration. But it is not entirely convincing. This seems a mere check of consistency. Some of the early claims of the paper are not supported, because, for instance, of the absence of localization and cycling (the latter being critical in ensemble methods). The use of localization could have made this paper a

bit different from other contributions. I would suggest you to be more caution and state that these experiments offer a partial assessment of the scheme.

We will state that this is only a partial assessment. Localization in the QG model is a project in itself beyond the scope of this paper. We leave it to future work.

Minor points or comments related to the major points

1. page 867, l.5-7: "However, Gauss-Newton iterations may not converge, not even locally." Yes, it is important that you mention it. However, in practice (which is also important for this journal), for a well designed system failures to converge are rare.

The QG problem in 6.3.3 diverges without regularization, which shows that divergence of the Gauss-Newton method can appear outside of artificially contrived examples. We will note that divergence of Gauss-Newton method is in practice rare. Perhaps the absence of divergence is one of the attributes that define what a well designed system is.

2. page 868, l.7 "work is relatively cheap": The EnKS is wonderful as it is computationally cheap. But in high-dimensional systems, it has a huge storage requirement which has been warned against (Cosme et al. (2010) and earlier references).

We will be more specific, esp. about the storage of the ensemble over a large time lag.

We will rephrase and mention storage requirement for high-dimensional systems

3. page 867, l.17-18: "It is well known that weak constraint 4DVAR is equivalent to the Kalman smoother in the linear case." This is only true for the analysis within the data assimilation window.

We will note that.

4. page 880-886: I believe the discussion on the impact of the hyper-parameters should also depend on the outcome of a long cycling of the experiment. You may not have to achieve a high precision minimization to address properly the nonlinear effects within the data assimilation window and propagate later the ensemble (hence the errors) through the window.

5. page 887, l.17: “and have shown that it is capable of handling strongly nonlinear problems”: in the absence of cycling, it is difficult to really conclude. Cycling is important for the L63 and the QG model. That said, the numerical experiments are convincing enough for the case of a single nonlinear minimization. Please mitigate your statements.

We will run some more experiments with cycling for Lorenz 63, as also requested by Reviewer 1. We will state that this is only a partial assessment.

4 Letter from the editor

Dear Prof. Mandel,

You must have seen the reports of the two referees of your paper. You must also have received a message from Copernicus Publications asking you to send your own response to the referees’ reports by 2 September next. That same message must also mention the possibility of your submitting a new version of the paper after you have responded to the referees. I as Editor encourage you (if you have not already done so) to start preparing without delay a new version of your paper. And, in order to save time, I want to send you now my comments on the referees' reports, as well as my suggestions and requests for the new version. The two referees are qualified experts on assimilation of observations, and especially on Kalman filters and smoothers. Referee 1, who has let his name known, is E. Cosme from Grenoble University. Both referees consider your paper contains material that deserves publication, but both also consider that it requires major revisions. Referee 1 has comments on both the general presentation of the method you use, and on your numerical results. Concerning the method, he questions in particular the use of the word 'variational' for qualifying it (his comments 2.2 and 2.3). Concerning your results, he asks for comparison with other assimilation algorithms (his comment 1.1). He also considers that it is insufficient to use only the value of the objective function as diagnostic for the quality of the assimilation performed with the QG model (his comment 1.3). Concerning this last point, you know the 'true' field at all gridpoints and timesteps, and there is fundamentally a circular argument in evaluating the accuracy of the reconstructed fields by their fit to the observations that have been used in the assimilation. Referee 2 strongly stresses that you have not in his/her opinion given proper credit to recent works on ensemble Kalman filtering and smoothing (main comments 1 to 3). He/she also asks for a more detailed description of your implementation of algorithm 3 (main comment 5) and, as Referee 1, says he/she is not convinced by your numerical results concerning the QG model (main comment 6).

I as Editor also have a few comments. I mention two at this stage.

1. You refer to Algorithms 3 and 4 (statement of Theorem 1, Section 4) without having described what they are, nor even mentioning the Tables in which the corresponding equations are given. These algorithms must be described in the text before they are discussed.

In the discussion style, pages are short and the algorithm environment became floats. The floats got placed more in the back, since only one float fits per page. This was not a problem in the manuscript with normal-sized pages, and it should not be a problem in the final paper format. We will employ the usual techniques to place floats, or replace them by text.

2. The setting of the QG experiments (independently of their validation) should be described in more detail. For instance, the sentence The vertical correlation function value was taken as 0.2 (subsection 6.3.2, about three lines before end of penultimate paragraph) does not make much sense (in which unit is the value 0.2 expressed ?).

This should have been "The vertical correlation is assumed to be constant over the horizontal grid and the correlation coefficient value between the two layers was taken as 0.5. We will add more detail to make the computations more reproducible."

And how do you 'non-dimensionalise' the parameter β (2 lines after Eq. 26; it is somewhat inconsistent to keep a dimensional Coriolis parameter f_0, and then to non-dimensionalise its spatial derivative) ?

We have added more information to the text regarding the non-dimensionalisation: "The non-dimensional equations (Fandry and Leslie, 1984; Pedlosky, 1979) can be derived as follows:

$$t = \tilde{t}\frac{\bar{U}}{L}, \quad x = \frac{\tilde{x}}{L}, \quad y = \frac{\tilde{y}}{L},$$
$$u = \frac{\tilde{u}}{\bar{U}}, \quad v = \frac{\tilde{v}}{\bar{U}}, \quad \beta = \beta_0 \frac{L^2}{\bar{U}},$$

where t denotes time, $\bar{U}$ is a typical velocity scale, x and y are the eastward and northward coordinates respectively, u and v are the horizontal velocity components, β_0 is the northward derivative, and the tilde notation refers to the dimensionalized parameters . . . For the experiments in this paper, we choose $L = 10^6\,\mathrm{m}$, $\bar{U} = 10\,\mathrm{m\,s^{-1}}$, $H_1 = 6000\,\mathrm{m}$, $H_2 = 4000\,\mathrm{m}$, $f_0 = 10^{-4}\,\mathrm{s^{-1}}$, $\beta_0 = 1.5 \times 10^{-11}\,\mathrm{s^{-1}\,m^{-1}}$."

Please revise your paper according to the comments and suggestions of the two referees, as well as to mine. Concerning the referees' requests, that may require additional diagnostics or even numerical experiments. As requested by Copernicus Publications, give a point-by-point answer to all these comments and suggestions (including mine). Should you disagree with one particular comment, or decide not to follow one particular suggestion, please state precisely your reasons for that. As far as I am concerned, your response can be submitted in the open discussion, or in a letter attached to your revised version.

Both referees have stated they would be willing to review your paper again, and I will send your revised version to both of them.

I thank you for having submitted your paper to Nonlinear Processes in Geophysics, and look forward to receiving a new version.

Olivier Talagrand

Editor, Nonlinear Processes in Geophysics

References

Bergou, E., Gratton, S., and Mandel, J.: On the Convergence of a Non-linear Ensemble Kalman Smoother, arXiv:1411.4608, submitted, 2014.

Bocquet, M. and Sakov, P.: Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems, Nonlinear Processes in Geophysics, 19, 383–399, doi:10.5194/npg-19-383-2012, 2012.

Bocquet, M. and Sakov, P.: Joint state and parameter estimation with an iterative ensemble Kalman smoother, Nonlinear Processes in Geophysics, 20, 803–818, doi:10.5194/npg-20-803-2013, 2013.

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Quarterly Journal of the Royal Meteorological Society, 140, 1521–1535, doi:10.1002/qj.2236, 2014.

Chen, Y. and Oliver, D.: Ensemble Randomized Maximum Likelihood Method as an Iterative Ensemble Smoother, Mathematical Geosciences, 44, 1–26, doi:10.1007/s11004-011-9376-z, 2012.

Chen, Y. and Oliver, D.: Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification, Computational Geosciences, 17, 689–703, doi:10.1007/s10596-013-9351-5, http://dx.doi.org/10.1007/s10596-013-9351-5, 2013.

Cosme, E., Brankart, J.-M., Verron, J., Brasseur, P., and Krysta, M.: Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation, Ocean Modelling, 33, 87–100, doi:10.1016/j.ocemod.2009.12.004, 2010.

Cosme, E., Verron, J., Brasseur, P., Blum, J., and Auroux, D.: Smoothing Problems in a Bayesian Framework and Their Linear Gaussian Solutions, Monthly Weather Review, 140, 683–695, doi:10.1175/MWR-D-10-05025.1, 2011.

Desroziers, G., Camino, J.-T., and Berre, L.: 4DEnVar: link with 4D state formulation of variational assimilation and different possible implementations, Quarterly Journal of the Royal Meteorological Society, 140, 2097–2110, doi:10.1002/qj.2325, 2014.

Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, Springer, 2nd edn., doi:10.1007/978-3-642-03711-5, 2009.

Fandry, C. and Leslie, L.: A Two-Layer Quasi-Geostrophic Model of Summer Trough Formation in the Australian Subtropical Easterlies., Journal of the Atmospheric Sciences, 41, 807–817, 1984.

Fisher, M., Tr'emolet, Y., Auvinen, H., Tan, D., and Poli, P.: Weak-Constraint and Long-Window 4D-Var, Tech. rep., European Centre for Medium-Range Weather Forecasts, 2011.

Goodliff, M., Amezcua, J., and Van Leeuwen, P. J.: Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity, Tellus A, 67, doi:10.3402/tellusa.v67.26928, 2015.

Johns, C. J. and Mandel, J.: A Two-Stage Ensemble Kalman Filter for Smooth Data Assimilation, Environmental and Ecological Statistics, 15, 101–110, doi:10.1007/s10651-007-0033-0, 2008.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 2003.

Levenberg, K.: A method for the solution of certain non-linear problems in least squares, Quarterly of Applied Mathematics, 2, 164–168, 1944.

Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R., and Fairbairn, D.: Comparison of Hybrid-4DEnVar and Hybrid-4DVar Data Assimilation Methods for Global NWP, Monthly Weather Review, 143, 212–229, doi:10.1175/MWR-D-14-00195.1, 2014.

Mandel, J., Bergou, E., and Gratton, S.: 4DVAR by Ensemble Kalman Smoother, arxiv:1304.5271, 2013.

Marquardt, D. W.: An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial and Applied Mathematics, 11, 431–441, doi:10.1137/0111030, 1963.

Metref, S., Cosme, E., Snyder, C., and Brasseur, P.: A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation, Nonlinear Processes in Geophysics, 21, 869–885, doi:10.5194/npg-21-869-2014, 2014.

Nerger, L., Janjić, T., Schröter, J., and Hiller, W.: A regulated localization scheme for ensemble-based Kalman filters, Quarterly Journal of the Royal Meteorological Society, 138, 802–812, doi:10.1002/qj.945, http://dx.doi.org/10.1002/qj.945, 2012.

Nerger, L., Schulte, S., and Bunse-Gerstner, A.: On the influence of model nonlinearity and localization on ensemble Kalman smoothing, Quarterly Journal of the Royal Meteorological Society, 140, 2249–2259, doi:10.1002/qj.2293, 2014.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D., and Yorke, J. A.: A Local Ensemble Kalman Filter for Atmospheric Data Assimilation, Tellus, 56A, 415–428, doi:10.1111/j.1600-0870.2004.00076.x, 2004.

Pedlosky, J.: Geophysical Fluid Dynamics, Springer, 1979.

Sakov, P. and Bertino, L.: Relation Between Two Common Localisation Methods for the EnKF, Computational Geosciences, 10, 225–237, doi:10.1007/s10596-010-9202-6, 2011.

Sakov, P., Oliver, D. S., and Bertino, L.: An Iterative EnKF for Strongly Nonlinear Systems, Monthly Weather Review, 140, 1988–2004, doi:10.1175/MWR-D-11-00176.1, 2012.

Verlaan, M. and Heemink, A. W.: Nonlinearity in Data Assimilation Applications: A Practical Method for Analysis, Monthly Weather Review, 129, 1578–1589, doi:10.1175/1520-0493(2001)129<1578:NIDAAA>2.0.CO;2, 2001.

Zupanski, M.: Maximum Likelihood Ensemble Filter: Theoretical Aspects, Monthly Weather Review, 133, 1710–1726, doi:10.1175/MWR2946.1, 2005.