

Changes in Response to the Reviewers and the Editor regarding “Hybrid Levenberg–Marquardt and weak constraint ensemble Kalman smoother method”

J. Mandel^{1,5}, E. Bergou², S. Gürol³, S. Gratton^{3,4}, and I. Kasanický⁵

¹University of Colorado Denver, Denver, CO, USA

²INRA, MaIAGE, Domaine de Vilvert, Jouy-en-Josas, France

³CERFACS, Toulouse, France

⁴INP-ENSEEIH, Toulouse, France

⁵Institute of Computer Science, Czech Academy of Sciences, Prague, Czech Republic

Correspondence to: J. Mandel (jan.mandel@gmail.com)

1 Introduction

We would like to thank the editor and the reviewers for their comments, which contributed to improvement of this paper.

2 Reviewer 1

5 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 main intent of the computational tests was to assess how close the method is to incremental
10 *4DVAR in terms of minimizing the objective function. To assess the overall performance, we have now rerun the experiments with the quasi geostrophic model with RMSE, and we have run a long-term comparison with RMSE for the Lorenz 63 model, see the next item.*

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
15 63 system are robust if they are based on 100,000 assimilation steps at least, excluding spin-up.

We have run the assimilation on Lorenz 63 for 100,000 assimilation steps. The results are in new Fig. 6 and described in new Section 6.1.3.

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

20 *The purpose of measuring the objective function was to assess how close the method was to 4DVAR. We have rerun the experiments adding also RMSE, added as new Table 3.*

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
25 investigations. p.883, line 8: "for the first iteration, the best decrease in objective function is obtained
when $\tau=1$ ": actually, it increases from 1 to 2. Is the first iteration before that?

*We have clarified that by 6 iterations the method achieved the decrease in the objective function
and then it just oscillates randomly, that's why lines beyond 8 are not shown (Section 5.1.2, around
line 363).*

30 What is the initial value of the objective function?

*We have added line with the initial value of the objective function to Table 2 for the Lorenz 63
model. For the QG model, the figures now start with iteration 0, which is the initial value.*

2. About the presentation of the method

2.1 More must be said about the computational complexity and the implementation complexity.

35 Of particular importance is the increment of work from an EnKF, for example.

*We have added a summary of the cost in terms of the number of evaluations of the model and the
observation operator (Section 4, around line 270)*

And what are the assets of the method, compared with others?

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

40 – *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 regularization as additional observation is
statistically correct, because the EnKS is used to solve only the linear least squares.*

45 – *The ensemble of increments is generated fresh in every iteration, which prevents the
minimization from being restricted to combinations of a single fixed ensemble.*

This is described in the introduction, around line 67.

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
50 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
55 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. (2012), but I do not request you to cite my work). Could you clarify that?

60 *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 asymptotically the same as incremental 4DVAR, as proved in Bergou et al. (2014). That is, the objective function is minimized, which was the reason for the original choice of the numerical experiments. This is in fact an important motivation of this method, and it is now noted in the introduction, around line 68.*

65 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".

70 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 have added reference to Evensen (2009) and Kalnay (2003) in the introduction.

75 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.

80 *We rewrote the section consistently in terms of δx instead of z . The EnKF and EnKS are formulated with the variable x .*

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.

Left it as is.

90 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.

Left as is. "Kalman smoother" is the commonly used term in the references, even if, as the reviewer points out, it may not be historically entirely right..

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

No longer relevant, the section was rewritten with x.

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

No longer relevant, the section was rewritten.

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

Reference to these figures was inadvertently omitted. They provide more detailed statistics for the convergence of the iterations, as is now noted around line 346.

100 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).

105 *In the new Section 5.1.3, we have now used observations of the Lorenz 63 model from $\Delta t = 0.05$ to 0.55 (strongly nonlinear), cf., Bocquet and Sakov (2013, Fig. 6). Bocquet (2011), and Metref et al. (2014). For the QG model, we run the experiments for 10 days when nonlinearity is increasing (Fisher et al., 2011, Fig. 2), as now noted around line 477.*

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

We have changed to "may be better" (around line 530).

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.

115 *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," at the beginning of Section 5.2.1, around line 413.*

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.

125 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.

130 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, 135 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..." :

140 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 145 (fully cycled) 4D-Var.

We no longer mention the preprint (Mandel et al., 2013) of the original version of this paper. We have added a discussion in the introduction along the lines suggested by the referee around line 103.

(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 150 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.

Added the following note before Theorem 1, around line 280: "It is interesting that the ensemble transform approach in Sakov et al. (2012); Bocquet and Sakov (2012, 2013, 2014) corresponds to 155 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 160 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.

165 *We have added a discussion in the introduction, see lines 88-102.*

(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.

175 (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 have added a discussion of localization at the end of the introduction, starting at line 104.

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...

We have clarified that the regularization was meant to be Levenberg-Marquardt, around line 93.

Please mention those references, and make it clear. Bocquet and Sakov (2012); Chen and Oliver (2013) 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.

We have paraphrased this note around line 95.

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

200 (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
205 algorithm that guarantees convergence of the minimization. These alternatives have been suggested
and successfully tested in Bocquet and Sakov (2012).”

3. page 868, 1.23-26. “Gradient methods in the span of the ensemble for one analysis cycle (i.e.,
3DVAR) include Zupanski (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
210 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
215 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 .

*We have noted that IEnKF is similar to the finite differences and the use of EnKS in this paper,
around line 90. There are of course some differences, but we do not go into further details there. We
have added a more comprehensive discussion reflecting the above items and our best understanding
220 based on fresh reading of the mentioned literature, starting around line 70.*

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
225 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 referred to the reduced rank
Kalman filter approach (RRSQRT) versus the EnKF which differ by the use of the tangent linear or
230 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. (2012)). This should be stated clearly.

235 *We meant the standard EnKS and used “nonlinear” only to distinguish it from the separate presentation in the linear case. We have restructured the presentation now, and the EnKS is stated for reference only once, starting around line 190.*

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 to the 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, 245 explain precisely how you solve Eq. (23) and give us the complete algorithm. This is critical for the paper.

We have restructured the presentation. We minimize the statement of known methods and state the final algorithm from around line 237 linearly in full without references to elsewhere, even if it creates some duplication and an implementation would normally use calls to subprograms The 250 second analysis step with the regularization is added as requested and written out for this particular case. In practice, it might be implemented simply as a second call to a subprogram that performs the analysis step.

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 255 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 state that this is only a partial assessment and localization and cycling in the QG model are a 260 project in itself beyond the scope of this paper, around line 481.

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.

265 *The QG problem diverges without regularization, which shows that divergence of the Gauss-Newton method can appear outside of artificially contrived examples. Perhaps the absence of divergence is one of the attributes that define what a well designed system is. We have added a note that “a careful design of an application system may avoid divergence in practice” in the introduction, around line 27.*

270 2. page 868, 1.7 "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 have noted the storage requirements in the introduction, around line 56.

3. page 867, 1.17-18: "It is well known that weak constraint 4DVAR is equivalent to the Kalman
275 smoother in the linear case." This is only true for the analysis within the data assimilation window.

Added "when all observations are in the assimilation window", around line 38.

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
280 propagate later the ensemble (hence the errors) through the window.

5. page 887, 1.17: "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.

285 *We have added long-cycling experiments including in the strongly nonlinear regime for the Lorenz
63 model, in new Section 5.1.3.*

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
290 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
295 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.
300 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,

305 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
 310 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.

315 *The algorithms typeset in the algorithm environment used by Copernicus are \LaTeX floats, therefore
 our control over their placement is limited. We have now formulated the algorithms as inline text
 with numbered equations so that we can be sure that they are in the right place.*

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
 320 (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 ?).

*We corrected this item, around line 473: “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 for
 Q_i and 0.2 for B .”*

325 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, starting around
 line 424: “The non-dimensional equations (Fandry and Leslie, 1984; Pedlosky, 1979) can be derived
 330 as follows:*

$$t = \frac{\tilde{t}}{\bar{U}}, \quad x = \frac{\tilde{x}}{L}, \quad y = \frac{\tilde{y}}{L},$$

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

*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
 335 derivative, and the tilde notation refers to the dimensionalized parameters . . . For the experiments
 in this paper, we choose $L = 10^6$ m, $\bar{U} = 10$ m s⁻¹, $H_1 = 6000$ m, $H_2 = 4000$ m, $f_0 = 10^{-4}$ s⁻¹,
 $\beta_0 = 1.5 \times 10^{-11}$ 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
 340 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.

345 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

350 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 to SIAM/ASA Journal for Uncertainty Quantification, 2014.
- Bocquet, M.: Ensemble Kalman filtering without the intrinsic need for inflation, *Nonlinear Processes in Geophysics*, 18, 735–750, doi:10.5194/npg-18-735-2011, 2011.
- 355 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.
- 360 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.: 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.
- 365 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, 2012.
- 370 Desroziers, G., Camino, J.-T., and Berre, L.: 4DVar: 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.
- 375 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.
- 380 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-4DVar and Hybrid-4DVar Data Assimilation Methods for Global NWP, *Monthly Weather Review*, 143, 212–229, doi:10.1175/MWR-D-14-00195.1, 2014.
- 385 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
390 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.
- 395 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,
400 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*
405 *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,
410 1710–1726, doi:10.1175/MWR2946.1, 2005.

Summary

11/30/15, 12:34:34 PM

Differences exist between documents.

New Document:

[npg-2015-26-R1](#)

29 pages (644 KB)

11/30/15, 12:33:42 PM

Used to display results.

Old Document:

[npg-2015-26-R0](#)

24 pages (508 KB)

11/30/15, 12:33:41 PM

[Get started: first change is on page 1.](#)


No pages were deleted

How to read this report

Highlight indicates a change.

Deleted indicates deleted content.

 indicates pages were changed.

 indicates pages were moved.

Hybrid Levenberg–Marquardt and weak constraint ensemble Kalman smoother method

J. Mandel^{1,5}, E. Bergou², S. Gürol³, S. Gratton^{3,4}, and I. Kasanický⁵

¹University of Colorado Denver, Denver, CO, USA

²INRA, MaIAGE, Domaine de Vilvert, Jouy-en-Josas, France

³CERFACS, Toulouse, France

⁴INP-ENSEEIH, Toulouse, France

⁵Institute of Computer Science, Czech Academy of Sciences, Prague, Czech Republic

Abstract. The ensemble Kalman smoother (EnKS) is used as a linear least squares solver in the Gauss–Newton method for the large nonlinear least squares system in incremental 4DVAR. The ensemble approach is naturally parallel over the ensemble members and no tangent or adjoint operators are needed. Further, adding a regularization term results in replacing the Gauss–Newton method, which may diverge, by the Levenberg–Marquardt method, which is known to be convergent. The regularization is implemented efficiently as an additional observation in the EnKS. The method is illustrated on the Lorenz 63 model and a two-level quasi-geostrophic model.

1 Introduction

Four dimensional variational data assimilation (4DVAR) is a dominant data assimilation method used in weather forecasting centers worldwide. 4DVAR attempts to reconcile model and data variationally, by solving a large weighted nonlinear least squares problem. The unknown is a vector of system states over discrete points in time, when the data are given. The objective function minimized is the sum of the squares of the differences of the initial state from a known background state at the initial time and the differences of the values of observation operator and the data at every given time point. In the weak-constraint 4DVAR (Trémolet, 2007), considered here, the model error is accounted for by allowing the ending and starting states of the model at every given time point to be different, and adding to the objective function also the sums of the squares of those differences. The sums of the squares are weighted by the inverses of the appropriate error covariance matrices, and much of the work in the applications of 4DVAR goes into modeling those covariance matrices.

In the incremental approach (Courtier et al., 1994), the nonlinear least squares problem is solved iteratively by solving a succession of linearized least square problems. The major cost in 4DVAR iterations is in evaluating the model, tangent and adjoint operators, and solving the large linear least squares. A significant software development effort is needed for the additional code to implement the tangent and adjoint operators to the model and the observation operators. Straightforward lineariza-

25 tion leads to the Gauss–Newton method for nonlinear least squares (Bell, 1994; Tshimanga et al.,
2008). Gauss–Newton iterations are not guaranteed to converge, not even locally, though a careful
design of an application system may avoid divergence in practice. Finally, while the evaluation of the
model operator is typically parallelized on modern computer architectures, there is a need to further
parallelize the 4DVAR process itself.

30 The Kalman filter is a sequential Bayesian estimation of the gaussian state of a linear system at
a sequence of discrete time points. At each of the time points, the use of the Bayes theorem results in
an update of the state, represented by its mean and covariance. The Kalman smoother considers all
states within an assimilation time window to be a large composite state. Consequently, the Kalman
smoother can be obtained from the Kalman filter by simply applying the same update as in the filter
35 to the past states as well. However, historically, the focus was on efficient short recursions (Rauch
et al., 1965; Strang and Borre, 1997), similarly as in the Kalman filter.

It is well known that weak constraint 4DVAR is equivalent to the Kalman smoother in the linear
case and when all observations are in the assimilation window. Use of the Kalman smoother to solve
the linear least squares in the Gauss–Newton method is known as the iterated Kalman smoother, and
40 considerable improvements can be obtained against running the Kalman smoother only once (Bell,
1994; Fisher et al., 2005).

The Kalman filter and smoother require maintaining the covariance of the state, which is not
feasible for large systems, such as in numerical weather prediction. Hence, the ensemble Kalman
filter (EnKF) and ensemble Kalman smoother (EnKS) (Evensen, 2009) use a Monte-Carlo approach
45 for large systems, representing the state by an ensemble of simulations, and estimating the state
covariance from the ensemble. The implementation of the EnKS in Stroud et al. (2010) uses the
adjoint model explicitly, with the short recursions and a forward and a backward pass, as in the
Kalman smoother. However, the implementations in Khare et al. (2008); Evensen (2009) do not
depend on the adjoint model and simply apply EnKF algorithms to the composite state over multiple
50 time points. Such composite variables are also called 4D vectors, e.g., (Desroziers et al., 2014). We
use the latter approach in the computations reported here.

In this paper, we use the EnKS as a linear least squares solver in 4DVAR. The EnKS is implemented
in the physical space and with randomization. The ensemble approach is naturally parallel over the
ensemble members. The rest of the computational work is relatively cheap compared to the ensemble
55 of simulations, and parallel dense linear algebra libraries can be used; however, in high-dimensional
systems or for a large lag, the storage requirements can be prohibitive (e.g., Cosme et al., 2010).
The proposed approach uses finite differences from the ensemble, and no tangent or adjoint opera-
tors are needed. To stabilize the method and assure convergence, a Tikhonov regularization term is
added to the linear least squares, and the Gauss–Newton method becomes the Levenberg–Marquardt
60 method (Levenberg, 1944; Marquardt, 1963). The Tikhonov regularization is implemented within
EnKS as an independent observation following Johns and Mandel (2008) in a computationally cheap

additional analysis step, which is statistically correct because the smoother operates only on the linearized problem. A new probabilistic ensemble is generated in every iteration, so the minimization is not restricted to the combinations of a single ensemble. We use finite differences from ensemble mean towards the ensemble members to linearize the model and observation operators. The iterations can be proved to converge to incremental 4DVAR iterations for small finite difference step and large ensemble size (Bergou et al., 2014). Thus, in the limit, the method performs actual minimization of the weak-constraint objective function and inherits the advantages of 4DVAR in handling nonlinear problems. We call the resulting method EnKS-4DVAR.

Combinations of ensemble and variational approaches have been of considerable recent interest. Estimating the background covariance for 4DVAR from an ensemble was one of the first connections (Hamill and Snyder, 2000b). It is now standard and became operational (Wang, 2010). Zhang et al. (2009) use a two-way connection between EnKF and 4DVAR to obtain the covariance for 4DVAR, and 4DVAR to feed the mean analysis into EnKF. EnKF is operational at the National Centers for Environmental Prediction (NCEP) as part of its Global Forecast System Hybrid Variational Ensemble Data Assimilation System (GDAS), together with the Gridpoint Statistical Interpolation (GSI) variational data assimilation system (Developmental Testbed Center, 2015).

The first methods that use ensembles for more than computing the covariance minimized the 3DVAR objective function in the analysis step. The MLEF method by Zupanski (2005) works in the ensemble space, i.e., minimizing in the span of the ensemble members, with the control variables being the coefficients of a linear combination of the ensemble members. Gu and Oliver (2007) use iterated ensemble Kalman filter (with randomization) in the state space, with a linearization of the observation operator obtained by a regression on the increments given by the ensemble. This approach was extended by Chen and Oliver (2013) to a Levenberg-Marquardt method, with the regularization done by a multiplicative inflation of the covariance in the linearized problem rather than adding a Tikhonov regularization term. Liu et al. (2008, 2009); Liu and Xiao (2013) minimize the (strong constraint) 4DVAR objective function over linear combinations of the ensemble by computations in the observation space.

The IEnKF method by Sakov et al. (2012) minimizes the lag-one 4DVAR objective function in the ensemble space, using the square root EnKF as a linear solver in Newton-Gauss method, and rescaling the ensemble to approximate the tangent operators, which is similar to the use of finite differences and EnKS here. Bocquet and Sakov (2012) combined the IEnKF method of Sakov et al. (2012) with an inflation-free approach to obtain a 4D ensemble variational method, and with the Levenberg-Marquardt method by adding a diagonal regularization to the Hessian. Bocquet and Sakov (2012); Chen and Oliver (2013) used Levenberg-Marquardt for faster convergence, as an adaptive method between steepest descent and Gauss-Newton method rather than to overcome divergence. Bocquet and Sakov (2012) also considered scaling the ensemble to approximate the tangent operators (“bundle variant”) as in Sakov et al. (2012). Bocquet and Sakov (2013) extended IEnKF to

smoother (IEnKS) with fixed-lag and moving window and noted that Gauss-Newton can be replaced
 100 by Levenberg-Marquard. The method is formulated in terms of the composite model operator, i.e.,
 with strong constraints. Bocquet and Sakov (2014) developed the method further, including cycling.
 (Bocquet and Sakov, 2012, 2013, 2014) note that various optimizers could be used in IEnKF/IEnKS;
 the present method can be understood as EnKS used as such optimizer.

It is well known that for good practical performance, ensemble methods need to be modified
 105 by localization to improve the sampling error. Ensemble methods can be localized in multiple
 ways (Sakov and Bertino, 2011). For methods operating in the physical space, localization can be
 achieved, e.g., by tapering of the covariance matrix (Furrer and Bengtsson, 2007) or by replacing the
 sample covariance by its diagonal in a spectral space (Kasanický et al., 2015). This is not completely
 straightforward for the EnKS, but implementations of the EnKS based on the Bryson–Frazier ver-
 110 sion of the classical formulation of the Kalman smoother, with a forward and a backward pass, are
 more flexible (Butala, 2012). Methods in the ensemble space can be modified to update only nodes
 in a neighborhood of the observation (e.g., Ott et al., 2004). The 4DEnVAR method of Desroziers
 et al. (2014) uses ensemble-derived background covariance and the authors propose several methods
 to solve the linearized problem in each iteration by combinations of ensemble members with the
 115 weights allowed to vary spatially. Lorenc et al. (2014) compares 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.

For background in data assimilation, see, e.g., Evensen (2009) and Kalnay (2003).

The paper is organized as follows. In Sect. 2, we review the formulation of 4DVAR. The EnKF and
 120 the EnKS are reviewed in Sect. 3. The proposed method is described in Sect. 4. Section 5 contains
 the results of the computational experiments, and Sect. 6 is the conclusion.

2 Incremental 4DVAR

For vectors \mathbf{u}_i , $i = 1, \dots, L$, denote the composite (column) 4D vector

$$\mathbf{u}_{0:L} = \begin{bmatrix} \mathbf{u}_0 \\ \vdots \\ \mathbf{u}_L \end{bmatrix}.$$

125 We want to estimate $\mathbf{x}_0, \dots, \mathbf{x}_L$, where \mathbf{x}_i is the state at time i , from the background state, $\mathbf{x}_0 \approx \mathbf{x}_b$,
 the model, $\mathbf{x}_i \approx \mathcal{M}_i(\mathbf{x}_{i-1})$, and the observations $\mathcal{H}_i(\mathbf{x}_i) \approx \mathbf{y}_i$, where \mathcal{M}_i is the model operator,
 and \mathcal{H}_i is the observation operator. Quantifying the uncertainty by covariances, with $\mathbf{x}_0 \approx \mathbf{x}_b$ taken
 as $(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) \approx 0$, etc., we get the nonlinear least squares problem

$$\|\mathbf{x}_0 - \mathbf{x}_b\|_{\mathbf{B}^{-1}}^2 + \sum_{i=1}^L \|\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1})\|_{\mathbf{Q}_i^{-1}}^2 + \sum_{i=1}^L \|\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)\|_{\mathbf{R}_i^{-1}}^2 \rightarrow \min_{\mathbf{x}_{0:L}} \quad (1)$$

130 called weak-constraint 4DVAR (Trémolet, 2007). Originally, in 4DVAR, $\mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1})$; the weak constraint $\mathbf{x}_i \approx \mathcal{M}_i(\mathbf{x}_{i-1})$ accounts for model error.

The least squares problem (Eq. 1) is solved iteratively by linearization,

$$\begin{aligned} \mathcal{M}_i(\mathbf{x}_{i-1} + \delta\mathbf{x}_{i-1}) &\approx \mathcal{M}_i(\mathbf{x}_{i-1}) + \mathcal{M}'_i(\mathbf{x}_{i-1})\delta\mathbf{x}_{i-1}, \\ \mathcal{H}_i(\mathbf{x}_i + \delta\mathbf{x}_i) &\approx \mathcal{H}_i(\mathbf{x}_i) + \mathcal{H}'_i(\mathbf{x}_i)\delta\mathbf{x}_i. \end{aligned}$$

135 In each iteration $\mathbf{x}_{0:L} \leftarrow \mathbf{x}_{0:L} + \delta\mathbf{x}_{0:L}$, one solves the auxiliary linear least squares problem for the increments $\delta\mathbf{x}_{0:L}$,

$$\begin{aligned} &\|\mathbf{x}_0 + \delta\mathbf{x}_0 - \mathbf{x}_b\|_{\mathbf{B}^{-1}}^2 + \sum_{i=1}^L \|\mathbf{x}_i + \delta\mathbf{x}_i - (\mathcal{M}_i(\mathbf{x}_{i-1}) + \mathcal{M}'_i(\mathbf{x}_{i-1})\delta\mathbf{x}_{i-1})\|_{\mathbf{Q}_i^{-1}}^2 \\ &+ \sum_{i=1}^L \|\mathbf{y}_i - (\mathcal{H}_i(\mathbf{x}_i) + \mathcal{H}'_i(\mathbf{x}_i)\delta\mathbf{x}_i)\|_{\mathbf{R}_i^{-1}}^2 \rightarrow \min_{\delta\mathbf{x}_{0:L}}. \end{aligned} \quad (2)$$

This is the Gauss–Newton method (Bell, 1994; Tshimanga et al., 2008) for nonlinear squares, known

140 in 4DVAR as the incremental approach (Courtier et al., 1994). Write the auxiliary linear least squares problem (Eq. 2) for $\delta\mathbf{x}_{0:L}$ as

$$\|\delta\mathbf{x} - \delta\mathbf{x}_b\|_{\mathbf{B}^{-1}}^2 + \sum_{i=1}^L \|\delta\mathbf{x}_i - (\mathbf{M}_i\delta\mathbf{x}_{i-1} + \mathbf{m}_i)\|_{\mathbf{Q}_i^{-1}}^2 + \sum_{i=1}^L \|\mathbf{d}_i - \mathbf{H}_i\delta\mathbf{x}_i\|_{\mathbf{R}_i^{-1}}^2 \rightarrow \min_{\delta\mathbf{x}_{0:L}} \quad (3)$$

where

$$\delta\mathbf{x}_b = \mathbf{x}_b - \mathbf{x}_0, \quad \mathbf{m}_i = \mathcal{M}_i(\mathbf{x}_{i-1}) - \mathbf{x}_i, \quad \mathbf{d}_i = \mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i), \quad (4)$$

145 $\mathbf{M}_i = \mathcal{M}'_i(\mathbf{x}_{i-1})$, $\mathbf{H}_i = \mathcal{H}'_i(\mathbf{x}_i)$.

The function minimized in Eq. (3) is the same as the one minimized in the Kalman smoother (Bell, 1994).

3 Ensemble Kalman filter and smoother

We present the EnKF and EnKS algorithms, essentially following Evensen (2009), in a form suitable

150 for our purposes. We start with a formulation of the EnKF, in a notation useful for the extension to EnKS. The notation $\mathbf{v}^\ell \sim N(\mathbf{m}, \mathbf{A})$ means that \mathbf{v}^ℓ is sampled from $N(\mathbf{m}, \mathbf{A})$ independently of anything else. The ensemble of states of the linearized model at time i , conditioned on data up to time j (that is, with the data up to time j already ingested), is denoted by $\mathbf{X}_{i|j}^N = [\mathbf{x}_{i|j}^1, \dots, \mathbf{x}_{i|j}^N] = [\mathbf{x}_{i|j}^\ell]$, where the ensemble member index ℓ always runs over $\ell = 1, \dots, N$, and similarly for other

155 ensembles. Assume for the moment that the observation operator \mathcal{H}_i is linear, that is, $\mathcal{H}_i(\mathbf{u}) = \mathbf{H}_i\mathbf{u}$.

The EnKF algorithm consists of the following steps:

1. Initialize

$$\mathbf{x}_{0|0}^\ell \sim N(\mathbf{x}_b, \mathbf{B}), \quad \ell = 1, \dots, N. \quad (5)$$

2. For $i = 1, 2, \dots$,

160 (a) advance in time

$$\mathbf{x}_{i|i-1}^\ell = \mathcal{M}_i(\mathbf{x}_{i-1|i-1}^\ell) + \mathbf{v}_i^\ell, \quad \mathbf{v}_i^\ell \sim N(\mathbf{0}, \mathbf{Q}_i), \quad (6)$$

(b) The analysis step

$$\begin{aligned} \mathbf{x}_{i|i}^\ell &= \mathbf{x}_{i|i-1}^\ell - \mathbf{P}_{i,i}^N \mathbf{H}_i^T (\mathbf{H}_i \mathbf{P}_{i,i}^N \mathbf{H}_i^T + \mathbf{R}_i)^{-1} (\mathbf{H}_i(\mathbf{x}_{i|i-1}^\ell) - \mathbf{d}_i - \mathbf{w}_i^\ell), \\ \mathbf{w}_i^\ell &\sim N(0, \mathbf{R}_i), \end{aligned} \quad (7)$$

165 where $\mathbf{P}_{i,i}^N$ is the sample covariance computed from the the ensemble $\mathbf{X}_{i|i-1}^N$.

Denote by \mathbf{A}_i^N the matrix of anomalies of the ensemble $\mathbf{Z}_{i|i-1}^N$,

$$\mathbf{A}_i^N = [\mathbf{a}_i^1, \dots, \mathbf{a}_i^N] = \left[\mathbf{x}_{i|i-1}^1 - \bar{\mathbf{x}}_{i|i-1}, \dots, \mathbf{x}_{i|i-1}^N - \bar{\mathbf{x}}_{i|i-1} \right], \quad \bar{\mathbf{x}}_{i|i-1} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_{i|i-1}^j. \quad (8)$$

Then $\mathbf{P}_{i,i}^N = \frac{1}{N-1} \mathbf{A}_i^N (\mathbf{A}_i^N)^T$, and we can write the matrices in Eq. (7) as

$$\mathbf{P}_{i,i}^N \mathbf{H}_i^T = \frac{1}{N-1} \mathbf{A}_i^N (\mathbf{H}_i \mathbf{A}_i^N)^T, \quad \mathbf{H}_i \mathbf{P}_{i,i}^N \mathbf{H}_i^T = \frac{1}{N-1} \mathbf{H}_i \mathbf{A}_i^N (\mathbf{H}_i \mathbf{A}_i^N)^T. \quad (9)$$

170 In particular, the matrix \mathbf{H}_i is used here only in the matrix-vector multiplications

$$\mathbf{g}_i^\ell = \mathbf{H}_i \mathbf{a}_i^\ell = \mathbf{H}_i (\mathbf{x}_{i|i-1}^\ell - \bar{\mathbf{x}}_{i|i-1}) = \mathbf{H}_i \mathbf{x}_{i|i-1}^\ell - \frac{1}{N} \sum_{j=1}^N \mathbf{H}_i \mathbf{x}_{i|i-1}^j, \quad (10)$$

which allows the matrix-vector multiplication to be replaced by the use of a possibly nonlinear observation operator \mathcal{H}_i evaluated on the ensemble members only (Eq. 18 below). This technique is commonly used for nonlinear observation operators, e.g., Chen and Snyder (2007); Mandel et al.

175 (2009). With $\mathbf{H}_i \mathbf{A}_i^N = \mathbf{G}_i^N = [\mathbf{g}_i^1, \dots, \mathbf{g}_i^N]$, Eq. (9) becomes

$$\mathbf{P}_{i,i}^N \mathbf{H}_i^T = \frac{1}{N-1} \mathbf{A}_i^N (\mathbf{G}_i^N)^T, \quad \mathbf{H}_i \mathbf{P}_{i,i}^N \mathbf{H}_i^T = \frac{1}{N-1} \mathbf{G}_i^N (\mathbf{G}_i^N)^T, \quad (11)$$

Also, from Eqs. (9,8), it follows that the analysis ensemble $\mathbf{X}_{i|i}^N$ consists of linear combinations of the forecast ensemble, hence it can be written as multiplying the forecast ensemble by a suitable transformation matrix \mathbf{T}_i^N ,

$$180 \mathbf{X}_{i|i}^N = \mathbf{X}_{i|i-1}^N \mathbf{T}_i^N, \quad \mathbf{T}_i^N \in \mathbb{R}^{N \times N}. \quad (12)$$

The EnKS is obtained by applying the same analysis step as in EnKF (Eq. 7) to the ensemble $\mathbf{X}_{0:i|i-1}$ of 4D composite states from time 0 to i , conditioned on data up to time $i-1$,

$$\mathbf{X}_{0:i|i-1}^N = \begin{bmatrix} \mathbf{X}_{0|i-1}^N \\ \vdots \\ \mathbf{X}_{i|i-1}^N \end{bmatrix},$$

in the place of $\mathbf{X}_{0:i|i-1}$, with the observation matrix $\tilde{\mathbf{H}}_{0:i} = [0, \dots, \mathbf{H}_i]$. Then, Eq. (7) becomes

$$185 \quad \mathbf{x}_{0:i|i}^\ell = \mathbf{x}_{0:i|i-1}^\ell - \mathbf{P}_{0:i,0:i}^N \tilde{\mathbf{H}}_{0:i}^\top (\tilde{\mathbf{H}}_{0:i} \mathbf{P}_{0:i,0:i}^N \tilde{\mathbf{H}}_{0:i}^\top + \mathbf{R}_i)^{-1} (\tilde{\mathbf{H}}_{0:i} \mathbf{x}_{0:i|i-1}^\ell - \mathbf{d}_i^\ell - \mathbf{w}_i^\ell),$$

where $\mathbf{P}_{0:i,0:i}^N$ is the sample covariance matrix of $\mathbf{X}_{0:i|i-1}^N$. Fortunately, the matrix–vector and matrix–matrix products can be simplified,

$$\tilde{\mathbf{H}}_{0:i} \mathbf{x}_{0:i|i-1}^\ell = [0, \dots, 0, \mathbf{H}_i] \mathbf{x}_{0:i|i-1}^\ell = \mathbf{H}_i \mathbf{x}_{i|i-1}^\ell \quad (13)$$

$$\mathbf{P}_{0:i,0:i}^N \tilde{\mathbf{H}}_{0:i}^\top = \mathbf{P}_{0:i,i}^N \mathbf{H}_i^\top, \quad \tilde{\mathbf{H}}_{0:i} \mathbf{P}_{0:i,0:i}^N \tilde{\mathbf{H}}_{0:i}^\top = \mathbf{H}_i \mathbf{P}_{i,i}^N \mathbf{H}_i^\top, \quad (14)$$

190 which is the same expression as in Eq. (9). Using also Eq. (11), we obtain the **EnKS algorithm**:

1. Initialize

$$\mathbf{z}_{0|0}^\ell \sim N(\mathbf{z}_b, \mathbf{B}), \quad \ell = 1, \dots, N. \quad (15)$$

2. For $i = 1, \dots, L$:

(a) Advance in time:

$$195 \quad \mathbf{x}_{i|i-1}^\ell = \mathcal{M}_i(\mathbf{x}_{i-1|i-1}^\ell) + \mathbf{v}_i^\ell, \quad \mathbf{v}_i^\ell \sim N(\mathbf{0}, \mathbf{Q}_i), \quad \ell = 1, \dots, N \quad (16)$$

(b) Compute the anomalies of the ensemble in the state space and in the observation space:

$$\mathbf{A}_{0:i} = [\mathbf{a}_{0:i}^1, \dots, \mathbf{a}_{0:i}^N], \quad \mathbf{a}_{0:i}^\ell = \mathbf{x}_{0:i|i-1}^\ell - \frac{1}{N} \sum_{j=1}^N \mathbf{x}_{0:i|i-1}^j \quad (17)$$

$$\mathbf{G}_i^N = [\mathbf{g}_i^1, \dots, \mathbf{g}_i^N], \quad \mathbf{g}_i^\ell = \mathcal{H}_i(\mathbf{x}_{i|i-1}^\ell) - \frac{1}{N} \sum_{j=1}^N \mathcal{H}_i(\mathbf{x}_{i|i-1}^j) \quad (18)$$

(c) The analysis step:

$$200 \quad \mathbf{x}_{0:i|i}^\ell = \mathbf{x}_{0:i|i-1}^\ell - \frac{1}{N-1} \mathbf{A}_{0:i}^N (\mathbf{G}_i^N)^\top \left(\frac{1}{N-1} \mathbf{G}_i^N (\mathbf{G}_i^N)^\top + \mathbf{R}_i \right)^{-1} \cdot \left(\mathcal{H}_i(\mathbf{x}_{i|i-1}^\ell) - \mathbf{y}_i - \mathbf{w}_i^\ell \right), \quad \mathbf{w}_i^\ell \sim N(\mathbf{0}, \mathbf{R}_i), \quad \ell = 1, \dots, N. \quad (19)$$

Comparing Eq. (7) and Eq. (19), we see that the EnKS can be implemented in a straightforward manner by applying the same transformation as in the EnKF to the composite 4D state vector from times 0 to i , $\mathbf{X}_{0:i|i}^N = \mathbf{X}_{0:i|i-1}^N \mathbf{T}_i^N$, where \mathbf{T}_i^N is the transformation matrix in Eq. (12) (Brusdal et al., 2003, Eq. 20).

4 EnKS-4DVAR

We apply the EnKS algorithm (Eqs. 15–18) with the increments $\delta \mathbf{x}$ in place of \mathbf{x} to solve the linearized auxiliary least squares problem (Eq. 3). Approximating by finite differences based at \mathbf{x}_{i-1} with step $\tau > 0$, we get the action of the linearized model operator

$$210 \quad \mathbf{M}_i \delta \mathbf{x}_{i-1}^\ell + m_i \approx \frac{\mathcal{M}_i(\mathbf{x}_{i-1} + \tau \delta \mathbf{x}_{i-1}^\ell) - \mathcal{M}_i(\mathbf{x}_{i-1})}{\tau} + \mathcal{M}_i(\mathbf{x}_{i-1}) - \mathbf{x}_i, \quad (20)$$

and the linearized observation operator

$$\mathbf{H}_i \delta \mathbf{x}_i^\ell \approx \frac{\mathcal{H}_i(\mathbf{x}_i + \tau \delta \mathbf{x}_i^\ell) - \mathcal{H}_i(\mathbf{x}_i)}{\tau}. \quad (21)$$

The Gauss–Newton method may diverge, but convergence to a stationary point of (Eq. 1) can be recovered by a control of the step $\delta \mathbf{x}$. Adding a constraint of the form $\|\delta \mathbf{x}_i\| \leq \varepsilon$ leads to globally convergent trust region methods (Gratton et al., 2013). Here, we add to (Eq. 3) a Tikhonov regularization term of the form $\gamma \|\delta \mathbf{x}_i\|_{\mathbf{S}_i}^2$, which controls the step size as well as rotates the step direction towards the steepest descent, and obtain the Levenberg–Marquardt method (Levenberg, 1944; Marquardt, 1963) $\mathbf{x}_{0:L} \leftarrow \mathbf{x}_{0:L} + \delta \mathbf{x}_{0:L}$, where

$$\|\delta \mathbf{x}_0 - \delta \mathbf{x}_b\|_{\mathbf{B}^{-1}}^2 + \sum_{i=1}^L \|\delta \mathbf{x}_i - \mathbf{M}_i \delta \mathbf{x}_{i-1} - \mathbf{m}_i\|_{\mathbf{Q}_i^{-1}}^2 + \sum_{i=1}^L \|\mathbf{d}_i - \mathbf{H}_i \delta \mathbf{x}_i\|_{\mathbf{R}_i^{-1}}^2 + \gamma \sum_{i=0}^L \|\delta \mathbf{x}_i\|_{\mathbf{S}_i^{-1}}^2 \rightarrow \min_{\delta \mathbf{x}_{0:L}}. \quad (22)$$

Under suitable technical assumptions, the Levenberg–Marquardt method is guaranteed to converge globally if the regularization parameter $\gamma \geq 0$ is large enough (Gill and Murray, 1978; Osborne, 1976). Estimates for the convergence of the Levenberg–Marquardt method in the case when the linear system is solved only approximately exist (Wright and Holt, 1985).

Similarly as in Johns and Mandel (2008), we interpret the regularization term $\gamma \|\delta \mathbf{x}_i\|_{\mathbf{S}_i}^2$ in Eq. (22) as arising from additional independent observations $\delta \mathbf{x}_i \approx \mathbf{0}$ with covariance $\gamma^{-1} \mathbf{S}_i$. The independent observation can be assimilated separately, resulting in a mathematically equivalent but often more efficient two-stage method – simply run the EnKF analysis (Eqs. 25, 26) twice. With the choice of \mathbf{S}_i as identity or, more generally a diagonal matrix, the implementation of these large observations can be made efficient (Mandel et al., 2009). We use the notation $\delta \mathbf{x}_{0:i|i-1/2}^\ell$ for the increments after the first half-step, conditioned on the original observations only, and $\delta \mathbf{x}_{0:i|i}^\ell$ for the increments conditioned also on the regularization $\delta \mathbf{x}_i \approx \mathbf{0}$. Note that unlike in Johns and Mandel (2008), where the regularization was applied to a nonlinear problem and thus the sequential data assimilation was only approximate, here the EnKS is run on the auxiliary linearized problem, so all distributions are gaussian and the equivalence of assimilating the observations at the same time and sequentially is statistically exact.

We obtain the following algorithm **EnKS-4DVAR** for Eq. (1).

1. Initialize

$$\mathbf{x}_0 = \mathbf{x}_b, \quad \mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1}), i = 1, \dots, L,$$

if not given already.

2. Incremental 4DVAR (Eq. 2): Given $\mathbf{x}_0, \dots, \mathbf{x}_L$, initialize the ensemble of increments

$$\delta \mathbf{x}_{0|0}^\ell \sim N(\delta \mathbf{x}_b, \mathbf{B}), \quad \ell = 1, \dots, N, \quad \delta \mathbf{x}_b = \mathbf{x}_b - \mathbf{x}_0. \quad (23)$$

(a) For $i = 1, \dots, L$:

- i. Advance the ensemble of increments $\delta \mathbf{x}^\ell$ in time following Eq. (16), with the linearized operator approximated from Eq. (20),

$$\delta \mathbf{x}_{i|i-1}^\ell = \frac{\mathcal{M}_i(\mathbf{x}_{i-1} + \tau \delta \mathbf{x}_{i-1|i-1}^\ell) - \mathcal{M}_i(\mathbf{x}_{i-1})}{\tau} + \mathcal{M}_i(\mathbf{x}_{i-1}) - \mathbf{x}_i + \mathbf{v}_i^\ell, \quad (24)$$

$$\mathbf{v}_i^\ell \sim N(\mathbf{0}, \mathbf{Q}_i), \quad \ell = 1, \dots, N.$$

- ii. Compute the anomalies of the ensemble in the 4D state space and in the observation space:

$$\mathbf{A}_{0:i} = [\mathbf{a}_{0:i}^1, \dots, \mathbf{a}_{0:i}^N], \quad \mathbf{a}_{0:i}^\ell = \delta \mathbf{x}_{i|i-1}^\ell - \frac{1}{N} \sum_{j=1}^N \delta \mathbf{x}_{i|i-1}^j$$

$$\mathbf{G}_i^N = [\mathbf{g}_i^1, \dots, \mathbf{g}_i^N], \quad \mathbf{g}_i^\ell = \frac{1}{\tau} \left(\mathcal{H}_i(\mathbf{x}_i + \tau \delta \mathbf{x}_{i|i-1}^\ell) - \frac{1}{N} \sum_{j=1}^N \mathcal{H}_i(\mathbf{x}_i + \tau \delta \mathbf{x}_{i|i-1}^j) \right)$$
(25)

- iii. The first analysis step:

$$\delta \mathbf{x}_{0:i|i-1/2}^\ell = \delta \mathbf{x}_{0:i|i-1}^\ell - \frac{1}{N-1} \mathbf{A}_{0:i}^N (\mathbf{G}_i^N)^\top \left(\frac{1}{N-1} \mathbf{G}_i^N (\mathbf{G}_i^N)^\top + \mathbf{R}_i \right)^{-1} \cdot$$

$$\left(\mathcal{H}_i(\mathbf{x}_i) + \frac{\mathcal{H}_i(\mathbf{x}_i + \tau \delta \mathbf{x}_{i|i-1}^\ell) - \mathcal{H}_i(\mathbf{x}_i)}{\tau} - \mathbf{y}_i - \mathbf{w}_i^\ell \right), \quad (26)$$

$$\mathbf{w}_i^\ell \sim N(\mathbf{0}, \mathbf{R}_i), \quad \ell = 1, \dots, N.$$

- iv. If $\gamma > 0$, compute the anomalies of the ensemble in the 4D state space:

$$\mathbf{Z}_{0:i}^N = [\mathbf{z}_{0:i}^1, \dots, \mathbf{z}_{0:i}^N], \quad \mathbf{z}_{0:i}^\ell = \delta \mathbf{x}_{i|i-1/2}^\ell - \frac{1}{N} \sum_{j=1}^N \delta \mathbf{x}_{i|i-1/2}^j \quad (27)$$

Observation operator for the regularization is the identity, so the anomalies in the observation space are simply \mathbf{Z}_i^N .

- v. If $\gamma > 0$, regularization as the second analysis step with zero data and data covariance $\gamma^{-1} \mathbf{S}_i$:

$$\delta \mathbf{x}_{0:i|i}^\ell = \delta \mathbf{x}_{0:i|i-1/2}^\ell - \frac{1}{N-1} \mathbf{Z}_{0:i}^N (\mathbf{Z}_i^N)^\top \left(\frac{1}{N-1} \mathbf{Z}_i^N (\mathbf{Z}_i^N)^\top + \frac{1}{\gamma} \mathbf{S}_i \right)^{-1} \cdot \quad (28)$$

$$\left(\delta \mathbf{x}_{i|i-1/2}^\ell - \mathbf{v}_i^\ell \right), \quad \mathbf{v}_i^\ell \sim N(\mathbf{0}, \mathbf{S}_i), \quad \ell = 1, \dots, N,$$

otherwise $\delta \mathbf{x}_{0:i|i}^\ell = \delta \mathbf{x}_{0:i|i-1/2}^\ell, \ell = 1, \dots, N.$

- (b) Complete the approximate incremental 4DVAR iteration: update

$$\mathbf{x}_{0:L} \leftarrow \mathbf{x}_{0:L} + \frac{1}{N} \sum_{\ell=1}^N \delta \mathbf{x}_{0:L|L}^\ell \quad (29)$$

Note that for small $\gamma \rightarrow 0$, (Eq. 28) has asymptotically no effect, $\delta \mathbf{x}_{0:i|i}^\ell \Rightarrow \delta \mathbf{x}_{0:i|i-1/2}^\ell$. The computational cost of EnKS-4DVAR is one evaluations of the model \mathcal{M}_i for the initialization (Eq. 23), $N+1$ evaluations of the model \mathcal{M}_i , and N evaluations of the observation operator \mathcal{H}_i in each incremental 4DVAR iteration, in each of the L observation periods. In comparison, the cost of EnKF is N evaluation of the model \mathcal{M}_i and of the observation operator \mathcal{H}_i in each observation period. Running the model and evaluating the observation operator are the major cost in practical problems such as weather models, rather than the linear algebra of the EnKS itself, in a reasonably efficient EnKF/EnKS implementation.

It can be proved that for small τ and large N , the iterates $\mathbf{x}_{0:k}$ converge to those of incremental 4DVAR (Bergou et al., 2014). Surprisingly, it turns out that in the case when $\tau = 1$, we recover the standard EnKS applied directly to the nonlinear problem (Eq. 1), as shown by the following theorem. In particular, EnKS-4DVAR does not converge when $\tau = 1$ for nonlinear problems, because the result of each iteration is determined only by the starting value \mathbf{x}_0 . 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.

Theorem 1 If $\tau = 1$, then one step of EnKS-4DVAR (Eqs. 23-26) becomes the EnKS (Eqs. 15-19) (modified by including the additional regularization observation if $\gamma > 0$). In particular, in that case, the values of $\mathbf{x}_{0:L} + \delta \mathbf{x}_{0:L}^\ell$ do not depend on the previous values of $\mathbf{x}_{1:L}$.

Proof. Indeed, Eq. (24) becomes

$$\begin{aligned} \delta \mathbf{x}_{i|i-1}^\ell &= \frac{\mathcal{M}_i(\mathbf{x}_{i-1} + \delta \mathbf{x}_{i-1|i-1}^\ell) - \mathcal{M}_i(\mathbf{x}_{i-1})}{1} + \mathcal{M}_i(\mathbf{x}_{i-1}) - \mathbf{x}_i + \mathbf{v}_i^\ell \\ &= \mathcal{M}_i(\mathbf{x}_{i-1} + \delta \mathbf{x}_{i-1|i-1}^\ell) - \mathbf{x}_i + \mathbf{v}_i^\ell, \end{aligned}$$

hence

$$\mathbf{x}_i + \delta \mathbf{x}_{i|i-1}^\ell = \mathcal{M}_i(\mathbf{x}_{i-1} + \delta \mathbf{x}_{i-1|i-1}^\ell) + \mathbf{v}_i^\ell$$

which is the same as Eq. (16) for $\mathbf{x}_{i-1} + \delta \mathbf{x}_{i-1|i-1}^\ell$ in place of $\mathbf{x}_{i-1|i-1}$. Similarly, Eq. (25) becomes with $\tau = 1$,

$$\mathbf{g}_i^\ell = \frac{\mathcal{H}_i(\mathbf{x}_i + \delta \mathbf{x}_{i|i-1}^\ell) - \mathcal{H}_i(\mathbf{x}_i)}{1} - \frac{1}{N} \sum_{j=1}^N \frac{\mathcal{H}_i(\mathbf{x}_i + \delta \mathbf{x}_{i|i-1}^j) - \mathcal{H}_i(\mathbf{x}_i)}{1} \quad (30)$$

$$= \mathcal{H}_i(\mathbf{x}_i + \delta \mathbf{x}_{i|i-1}^\ell) - \frac{1}{N} \sum_{j=1}^N \mathcal{H}_i(\mathbf{x}_i + \delta \mathbf{x}_{i|i-1}^j), \quad (31)$$

which is again the same as Eq. (18) for $\mathbf{x}_i + \delta \mathbf{x}_{i|i-1}^\ell$ in place of $\mathbf{x}_{i|i-1}$. Finally, the innovation term in Eq. (26) becomes using Eq. (4),

$$\mathcal{H}_i(\mathbf{x}_i) + \frac{\mathcal{H}_i(\mathbf{x}_i + \tau \delta \mathbf{x}_{i|i-1}^\ell) - \mathcal{H}_i(\mathbf{x}_i)}{1} - \mathbf{y}_i = \mathcal{H}_i(\mathbf{x}_i + \delta \mathbf{x}_{i|i-1}^\ell) - \mathbf{y}_i,$$

which is again the same as in Eq. (19) $\mathbf{x}_i + \delta \mathbf{x}_{i|i-1}^\ell$ in place of $\mathbf{x}_{i|i-1}$. \square

5 Computational results

In this section, we investigate the performance of EnKS-4DVAR method, described in this paper, by solving the nonlinear least-squares problem (Eq. 1) in which the dynamical models are chosen either the Lorenz 63 system (Lorenz, 1963) or the two-level quasi-geostrophic model (Fandry and Leslie, 1984). Most of the experiments assess the convergence of the incremental 4DVAR iterations with EnKS as the linear solver in a single assimilation cycle (Sections 5.1.1, 5.1.2). We also demonstrate the overall long-term performance on a large number of assimilation cycles on the Lorenz 63 model in Section 5.1.3.

We first consider experiments where the regularisation is not necessary to guarantee the convergence (i.e., $\gamma = 0$). Lorenz 63 equations are used as a forecast model for these experiments. Section 5.1 describes the Lorenz 63 model and presents numerical results on the convergence. Using the same model, in Sect. 5.1.2, we investigate the impact of the finite differences parameter τ , used to approximate the derivatives of the model and observation operators, along the iterations.

Experiments where the regularisation is necessary to guarantee the convergence are shown in Sect. 5.2, and we analyse the impact of the regularisation parameter γ on the application to the two-level quasi-geostrophic model.

Note that for the experiments presented here, we do not use localization, hence we choose large ensemble sizes. In all experiments, the regularization covariance $\mathbf{S}_i = \mathbf{I}$.

5.1 Numerical tests using the Lorenz 63 model

The Lorenz 63 equations (Lorenz, 1963) are given by the nonlinear system

$$\frac{dx}{dt} = -\sigma(x - y), \quad \frac{dy}{dt} = \rho x - y - xz, \quad \frac{dz}{dt} = xy - \beta z, \quad (32)$$

where $x = x(t)$, $y = y(t)$, $z = z(t)$ and σ , ρ , β are parameters, whose values are chosen as 10, 28 and $8/3$ respectively for the experiments described in this paper. These values result in a chaotic behaviour with two regimes as illustrated in Fig. 1. This figure shows the Lorenz attractor, which has two lobes connected near the origin, and the trajectories of the system in this saddle region are particularly sensitive to perturbations. Hence, slight perturbations can alter the subsequent path from one lobe to the other.

The system is discretized using the fourth-order Runge–Kutta method. The state at time t is denoted by $\mathbf{X}_t = [x(t), y(t), z(t)]^\top$, $\mathbf{X}_t \in \mathbb{R}^3$.

To evaluate the performance of EnKS-4DVAR method, we will test it using the classical twin experiment technique, which consists on fixing an initial true state, denoted by truth_0 , and then integrating the initial truth in time using the model to obtain the true state $\text{truth}_i = \mathcal{M}(\text{truth}_{i-1})$ at each time i . We then build the data \mathbf{y}_i by applying the observation operator \mathcal{H}_i to the truth at time i and by adding a Gaussian perturbation $\mathcal{N}(0, \mathbf{R}_i)$. Similarly, the background \mathbf{x}_b is sampled from the

Gaussian distribution with the mean truth_0 and the covariance matrix \mathbf{B} . Then, we try to recover the truth using the observations and the background.

5.1.1 Convergence of the iterations

335 We perform numerical experiments without model error. The initial truth is set to $\text{truth}_0 = [1, 1, 1]^\top$ and the background covariance is chosen as the identity matrix of order three, i.e. $\mathbf{B} = \mathbf{I}_3$. The model is advanced in time by using the Runge–Kutta method with a time step of 0.1 time unit. The time window length is $L = 50$ time steps (5 time units). The observation operator is defined as $\mathcal{H}_i(x, y, z) = (x^2, y^2, z^2)$. At each time i , the observations are constructed as follows:
 340 $\mathbf{y}_i = \mathcal{H}_i(\text{truth}_i) + \mathbf{v}_i$, where \mathbf{v}_i is sampled from $N(0, \mathbf{R})$ with $\mathbf{R} = \mathbf{I}_3$. Observations are taken for each time step ($i = 1, \dots, 50$). The ensemble size is fixed to $N = 100$.

Figure 2 shows the estimator of the state vector \mathbf{X}_i , $i = 1, \dots, 10$, for the first five iterations. Figure 3 shows the root square error (RSE) for the same iterates shown in Fig. 2. RSE is defined as

$$\text{RSE}_i^{(j)} = \sqrt{\frac{1}{n}(\text{truth}_i - \mathbf{x}_i^{(j)})^\top (\text{truth}_i - \mathbf{x}_i^{(j)})}, \quad j = 1, \dots, 5, \quad (33)$$

345 where truth_i is the true vector state at time i , $\mathbf{x}_i^{(j)}$ is the j th iterate at time i and n is the length of \mathbf{x}_i . Table 1 shows the root mean square error (RMSE) for each iterate given by

$$\text{RMSE}^{(j)} = \frac{1}{k} \sum_{i=0}^k \text{RSE}_i^{(j)}, \quad j = 1, \dots, 5, \quad (34)$$

where k is the number of time steps.

350 From Table 1 and Figs. 2 and 3, it can be seen that the iterates of converge to the solution (without using a regularization). For these experiments, we observe that RMSE is reduced significantly in five iterations. Note that the error does not converge to zero, because of the approximation and variability inherent in the ensemble approach.

5.1.2 The impact of the finite difference parameter

355 Now we investigate the influence of the finite differences parameter τ used to approximate the derivatives of the model and observation operators. We use the same experimental set-up as described in the previous section. The numerical results are based on 30 runs with eight iterations for Lorenz 63 problem, with the following choices for the parameter τ : 1, 10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} and 10^{-6} .

360 Table 2 shows the mean of the objective function value as a function of the finite difference step τ and the number of iterations. When $\tau = 1$, the iterations after the first one do not improve the objective function. However, when $\tau \leq 10^{-1}$, the objective function was overall decreasing along the iterations, after a large initial increase. Because of the stochastic nature of the algorithm, the objective function does not necessarily decrease every iteration and its values eventually fluctuate

around a limit value randomly randomly. This stage was achieved after at most 6 iterations, so only
365 8 iterations are shown; further lines (not shown) exhibit the same fluctuating pattern in all columns.
This limit value of the objective function decreases with smaller τ , until it stabilizes for $\tau \leq 10^{-3}$.
Figs. 4 and 5 show more details of the statistics as boxplots of the objective function values. Each
panel corresponds to one line of Table 2.

We can conclude that for this toy test case at least, the method was insensitive to the choice of
370 $\tau \leq 10^{-3}$, except that small τ , when the problem solved by the smoother is essentially the tangent
problem, resulted in a large increase of the value of the objective function in the first iteration. This is
not uncommon in Newton type methods and highly nonlinear problems. Hence, an adaptive method,
which decreases τ adaptively, may be of interest. This issue will be studied elsewhere.

5.1.3 Cycling

375 So far, we have studied the impact of the use of the stochastic solver for a single assimilation window
only. Now we test the overall long-term performance. Consider again the Lorenz 63 model (32) with
the parameters $\sigma = 10$, $\rho = 28$, $\beta = 8/3$, and integration step set to 0.01 time unit. This is the same
parameters setup as the one used in Bocquet and Sakov (2012). We then proceed with similar testing
as in Metref et al. (2014). we perform usual twin model experiment. The initial truth state Y_0 is gen-
380 erated from $N(0, \mathbf{I}_3)$ distribution and the initial forecast state is then simulated by sampling from
 $N(Y_0, \mathbf{I}_3)$. Both states are advanced for 50,000 model time-steps burn-in period We used nonlinear
observational operator $h(x, y, z) = (x^3, y^3, z^3)$ with observational error generated from $N(0, \sigma^2 \mathbf{I}_3)$
with $\sigma^2 = 8$, and $\tau = 10^{-4}$. The time between two available observations Δt varies from 0.05 time
units, when the model is nearly linear, to 0.55, when the model is strongly nonlinear. We use ensem-
385 ble of size 10. After running multiple simulations we have found suitable values of parameters of the
method: the number of iterations is 25, and the parameter $\gamma = 10^{-9}$ when $\Delta t = 0.05$ and $\gamma = 1000$
otherwise. The length of assimilation window is set to $L = 6$, in other words we assimilate 6 ob-
servations at once. We assimilate each observation only once, i.e., the assimilation windows do not
overlap. To create initial ensemble at the beginning of each iteration, we use background covariance
390 created as a weighted average of the sample covariance from the last iteration in the previous assim-
ilation window and the identity matrix, similarly as in Hamill and Snyder (2000a). The weights are
0.99 for sample covariance and 0.01 for identity. The model error covariance in each time-step set
to $Q = 0.01 \mathbf{I}_3$. The experiment is run for 100,000 observation cycles.

We compare the proposed method with the standard EnKF with 10 ensemble members, where
395 the initial ensemble is created after burn-in period by adding $N(0, \mathbf{I}_3)$ perturbations. For stability
reason and to preserve covariance between ensemble members we add $N(0, 0.01 \mathbf{I}_3)$ noise after each
advance of each ensemble member. The necessity of related covariance inflation was identified also
in Bocquet and Sakov (2012). The EnKF algorithm is run every time when new observations are

available. The initial ensemble is created by adding white noise perturbation to forecasted state
 400 directly after burn-in period.

Fig. 6 shows that the proposed method has significantly smaller RMSE in the case when the
 model is nonlinear. Only in the situation, when the time between observation is 0.05 time unit, i.e.,
 the model is nearly linear EnKF gives a comparable result as the proposed method.

5.2 Numerical tests using a two-layer Quasi Geostrophic model (QG)

405 The EnKS-4DVAR algorithm has been implemented into Object Oriented Prediction System
 (OOPS) (Trémolet, 2013), which is a data assimilation framework developed by European Centre
 for Medium-Range Weather Forecasts (ECMWF). Numerical experiments are performed by using
 the simple two-layer quasi-geostrophic model of OOPS platform. The details for the model and the
 data assimilation system are given in Sects. 5.2.1 and 5.2.2 respectively. Numerical experiments are
 410 performed to solve the weak-constraint data assimilation problem (Eq. 1) by using EnKS-4DVAR
 with regularization. Numerical results are presented in Sect. 5.2.3.

5.2.1 A two-layer quasi-geostrophic model

The two-layer quasi-geostrophic channel model is widely used in theoretical atmospheric studies,
 since it is simple enough for numerical calculations and it adequately captures an important aspect
 415 of large-scale dynamics in the atmosphere.

The two-layer quasi-geostrophic model equations are based on the non-dimensional quasi-
 geostrophic potential vorticity, whose evolution represents large scale circulations of the atmosphere.
 The quasi-geostrophic potential vorticity on the first (upper) and second (lower) layers can be written
 respectively as

$$420 \quad q_1 = \nabla^2 \psi_1 - \frac{f_0^2 L^2}{g' H_1} (\psi_1 - \psi_2) + \beta y, \quad q_2 = \nabla^2 \psi_2 - \frac{f_0^2 L^2}{g' H_2} (\psi_2 - \psi_1) + \beta y + R_s, \quad (35)$$

where ψ is the stream function, ∇^2 is the two-dimensional Laplacian, R_s represents orography
 or heating, β is the (non-dimensionalised) northward variation of the Coriolis parameter at the fixed
 latitude y , f_0 is the Coriolis parameter at the southern boundary of the domain. L is the typical length
 425 scale of the motion we wish to describe, H_1 and H_2 are the depths of the two layers, $g' = g\Delta\theta/\bar{\theta}$ is
 the reduced gravity where $\bar{\theta}$ is the mean potential temperature, and $\Delta\theta$ is the difference in potential
 temperature across the layer interface. The non-dimensional equations (Fandry and Leslie, 1984;
 Pedlosky, 1979) can be derived as follows:

$$\begin{aligned} t &= \tilde{t} \frac{\bar{U}}{L}, & x &= \frac{\tilde{x}}{L}, & y &= \frac{\tilde{y}}{L}, \\ u &= \frac{\tilde{u}}{\bar{U}}, & v &= \frac{\tilde{v}}{\bar{U}}, & \beta &= \beta_0 \frac{L^2}{\bar{U}}. \end{aligned}$$

430 where t denotes time, \bar{U} is a typical velocity scale, x and y are the eastward and northward coordi-
 nates respectively, u and v are the horizontal velocity components, β_0 is the northward derivative,
 and the tilde notation refers to the dimensionalized parameters.

Potential vorticity in each layer is conserved and thus is described by

$$\frac{D_i q_i}{Dt} = 0, \quad i = 1, 2. \quad (36)$$

435 where D_i/Dt , is the total derivative, defined by

$$\frac{D_i}{Dt} = \frac{\partial}{\partial t} + u_i \frac{\partial}{\partial x} + v_i \frac{\partial}{\partial y} \quad (37)$$

and

$$u_i = -\frac{\partial \psi_i}{\partial y}, \quad v_i = \frac{\partial \psi_i}{\partial x}, \quad (38)$$

are the horizontal velocity components in each layer. Therefore, the potential vorticity at each time
 440 step is determined by using the conservation of potential vorticity given by Eq. (36). In this process,
 time stepping consists of a simple first order semi-Lagrangian advection of potential vorticity.

Given the potential vorticity at a fixed time, Eq. (35) can be solved for the stream function at
 each gridpoint and then the velocity fields obtained through Eq. (38). The equations are solved by
 using periodic boundary conditions in the west–east direction and Dirichlet boundary condition in
 445 the north–south direction. For the experiments in this paper, we choose $L = 10^6$ m, $\bar{U} = 10$ m s⁻¹,
 $H_1 = 6000$ m, $H_2 = 4000$ m, $f_0 = 10^{-4}$ s⁻¹, $\beta_0 = 1.5 \times 10^{-11}$ s⁻¹ m⁻¹. For more details on the
 model and its solution, we refer to Fisher et al. (2011).

The domain for the experiments is 12000 km by 6300 km for both layers. The horizontal dis-
 cretization consists of 40×20 points, so that the east–west and the north–south resolution is approx-
 450 imately 300 km. The dimension of the state vector of the model is then 1600. Note that the state
 vector is defined only in terms of the stream function.

5.2.2 Experimental setup

The performance of EnKS-4DVAR with regularization is analyzed by using twin experiments
 (Sect. 5.1).

455 The truth is generated from a model with layer depths of $D_1 = 6000$ m and $D_2 = 4000$ m, and
 the time step is set to 300 s, whereas the assimilating model has layer depths of $D_1 = 5500$ m and
 $D_2 = 4500$ m, and the time step is set to 3600 s. These differences in the layer depths and the time
 step provide a source of model error.

For all the experiments presented here, observations of non-dimensional stream function, vector
 460 wind and wind speed were taken from a truth of the model at 100 points randomly distributed over
 both levels. Observations were taken every 12 hours. We note that the number of observations is
 much smaller than the dimension of the state vector. Observation errors were assumed to be inde-
 pendent from each others and uncorrelated in time. The standard deviations (SD) were chosen to

be equal to 0.4 for stream function observation error, 0.6 for vector wind and 1.2 for wind speed.

465 The observation operator is the bi-linear interpolation of the model fields to horizontal observation locations.

470 The background error covariance matrix (matrix \mathbf{B}) and the model error covariances (matrices \mathbf{Q}_i) used in these experiments correspond to vertical and horizontal correlations. The vertical and horizontal structures are assumed to be separable. In the horizontal plane covariance matrices correspond to isotropic, homogeneous correlations of stream function with Gaussian spatial structure obtained from a Fast Fourier Transform approach (Dietrich and Newsam, 1997; Nowak et al., 2003). For the background covariance matrix \mathbf{B} , the SD and the horizontal correlation length scale in this experiments was set to 0.8 and 10^6 m respectively. For the model error covariance matrices \mathbf{Q}_i , the SD and the horizontal correlation length scale was set to 0.2 and 2×10^6 m respectively. The vertical correlation is assumed to be constant over the horizontal grid and the correlation coefficient value 475 between the two layers was taken as 0.5 for \mathbf{Q}_i and 0.2 for \mathbf{B} .

5.2.3 Numerical results

We perform one cycle for the experiments. The window length is set to 10 days when nonlinearity is increasing (Fisher et al., 2011, Fig. 2), with two sub-windows of 5 days ($L = 2$). No localization is 480 used in the experiments, as a result the ensemble size is chosen to be large enough, i.e. $N = 30000$. Therefore, this test is only a partial assessment. Localization and cycling in the QG model are beyond the scope of this paper. For the finite difference approximation, the parameter τ is set to 10^{-4} for all experiments. We have performed experiments for incremental 4DVAR and EnKS-4DVAR. The incremental 4DVAR method used conjugate gradients to solve the linearized problem with exact 485 tangent and adjoint models in each iteration, with no ensembles involved. The numerical results are presented as follows.

Figure 7 shows the objective function values along iterations of the incremental 4DVAR method. The objective function is oscillating with the iteration number, therefore incremental 4DVAR method without regularization is diverging. This divergence is due to the highly nonlinear behaviour of the 490 model for a long window (10 days). In such a case, as explained in Sect. 4, a convergence to a stationary point can be recovered by controlling the step which is done by introducing an additional regularization term in this study. In order to see the affect of this regularization, we performed EnKS-4DVAR with different values of the regularization parameter γ . Figures 8 and 9 show the objective function values along iterations for eight different choices of γ . RMSE values along the iterations 495 for the same experiments performed with 4DVAR and EnKS-4DVAR are presented in Table 3.

It can be seen from Figures 8 and 9 that when $\gamma = 0$, the iterations diverging as expected (since we do not use regularization and we only approximate the linearized subproblem using ensembles). For small values of γ (for instance $\gamma \leq 10^{-1}$), the objective function is not monotonically decreasing, hence the iterations are still diverging even if we use the regularization. Therefore, small values of γ

500 can not guarantee the convergence. For large values of γ (for instance $\gamma \geq 10$), we can observe the decrease on the objective function along iterations. Moreover, the fastest decrease on the objective function is obtained for $\gamma = 10$.

If we look at the RMSE values from Table 3, we can see that increasing γ values result in higher RMSE values. For large values of γ , for example $\gamma \geq 10$, the reduction in RMSE values is very small. For smaller values of γ the reduction is faster however RMSE values oscillates along the iterations. We want to note also that all RMSE values are lower than the initial RMSE value.


In conclusion, when the regularization is used, the choice of the regularization parameter γ is crucial to ensure the convergence. For instance, for small values of γ , the method can still diverge, and for large values of γ , the objective function decreases, but slowly (and many iterations may be needed to attain some predefined decrease). On the other hand, small γ values results in small RMSE values with oscillation along the iterations and RMSE values decrease slowly for the larger values of γ . Therefore the regularization parameter should be neither “very small” nor “very large”. An adaptive γ over iterations can be a better compromise, which will be explored in future studies.

6 Conclusions

515 We have proposed a stochastic solver for the incremental 4DVAR weak constraint method. The regularization term added to the Gauss–Newton method, resulting in a globally convergent Levenberg–Marquardt method, maintains the structure of the linearized least squares subproblem, enabling us to use ensemble Kalman smoother as linear solver while simultaneously controlling the convergence. We have formulated the EnKS-4DVAR method and have shown that it is capable of handling strongly nonlinear problems. We have demonstrated that the randomness of the EnKS version used (with perturbed data) eventually limits the convergence to a minimum, but a sufficiently large decrease of the objective function can be achieved for successful data assimilation. On the contrary, we suspect that the randomization may help to increase the supply of the search directions over the iterations, as opposed to deterministic methods locked into one low-dimensional subspace, such as the span of one given ensemble.


We have numerically illustrated the new method on the Lorenz 63 model and the two-level quasi-geostrophic model. We have analyzed the impact of the finite differences parameter τ used to approximate the derivatives of the model and observation operators. We have shown that for $\tau = 1$, the iterates obtained from EnKS-4DVAR are equivalent to those of obtained from the standard EnKS. Based on computational experiments, it may be better to start with the EnKS (i.e., $\tau = 1$) and then to decrease τ in further the iterations.

We have demonstrated long-term stability of the method on the Lorenz 63 method and shown that it achieves lower RMSE than standard EnKF for a highly nonlinear problem. This, however, took some parameter turning, in particular the data error variance.

535  For the second part of the experiments, we have shown the performance of the EnKS-4DVAR method with regularization on the two-level quasi-geostrophic problem, 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. We have observed that the incremental 4DVAR method is not converging for a long time window length, and that the 540 regularization is necessary to guarantee convergence. We have concluded that the choice of the regularization parameter is crucial to ensure the convergence and different choices of this parameter can change the rate of decrease in the objective function. As a summary, an adaptive regularization parameter can be a better compromise to achieve the approximate solution in a reasonable number of iterations.

545 The choice of the parameters used in our approach is of crucial importance for the computational cost of the algorithm, for instance the number of iterations to obtain some desired reduction. The exploration in more detail of the best strategies to adapt these parameters course of the iterations will be studied elsewhere.

The base method, used in the computational experiments here, is using sample covariance. How- 550 ever, there is nothing to prevent the use of more sophisticated variants of EnKS with localization and the covariance inflation, and square root filters instead of EnKS with data perturbation, as is done in related methods in the literature. These issues, as well as, the performance on larger and realistic problems, will be studied elsewhere.

Acknowledgements. This research was partially supported by the Fondation STAE project ADTAO, the Czech 555 Science Foundation under the grant GA13-34856S, and the US National Science Foundation under the grant DMS-1216481. A part of this work was done when Jan Mandel was visiting INP-ENSEEIH and CERFACS, and when Elhoucine Bergou, Serge Gratton, and Ivan Kasanicky were visiting the University of Colorado Denver. The authors would like to thank the editor, Prof. Olivier Talagrand, reviewer Dr. Emmanuel Cosme, and an anonymous reviewer for their comments, which contributed to improvement of this paper. 

560 References

- Bell, B.: The Iterated Kalman Smoother as a Gauss-Newton Method, *SIAM Journal on Optimization*, 4, 626–636, doi:10.1137/0804035, 1994.
- Bergou, E., Gratton, S., and Mandel, J.: On the Convergence of a Non-linear Ensemble Kalman Smoother, arXiv:1411.4608, submitted to *SIAM/ASA Journal for Uncertainty Quantification*, 2014.
- 565 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.
- 570 Brusdal, K., Brankart, J. M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P. J., Dombrowsky, E., and Verron, J.: A demonstration of ensemble based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems, *Journal of Marine Systems*, 40–41, 253–289, doi:10.1016/S0924-7963(03)00021-6, 2003.
- 575 Butala, M. D.: A localized ensemble Kalman smoother, in: 2012 IEEE Statistical Signal Processing Workshop (SSP), pp. 21–24, IEEE, doi:10.1109/SSP.2012.6319665, 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.
- 580 Chen, Y. and Snyder, C.: Assimilating Vortex Position with an Ensemble Kalman Filter, *Monthly Weather Review*, 135, 1828–1845, doi:10.1175/MWR3351.1, 2007.
- 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.
- 585 Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, *Quarterly Journal of the Royal Meteorological Society*, 120, 1367–1387, doi:10.1002/qj.49712051912, 1994.
- Desroziers, G., Camino, J.-T., and Berre, L.: 4D-EnVar: 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.
- 590 Developmental Testbed Center: NOAA Ensemble Kalman Filter Beta Release v1.0, <http://www.dtcenter.org/com-GSI/users/docs>, retrieved March 2015, 2015.
- Dietrich, C. R. and Newsam, G. N.: Fast and Exact Simulation of Stationary Gaussian Processes through Circulant Embedding of the Covariance Matrix, *SIAM Journal on Scientific Computing*, 18, 1088–1107, 1997.
- 595 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., Trémolet, 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.
- 600
- Fisher, M., Leutbecher, M., and Kelly, G. A.: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 131, 3235–3246, doi:10.1256/qj.04.142, 2005.
- Furrer, R. and Bengtsson, T.: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants, *J. Multivariate Anal.*, 98, 227–255, doi:10.1016/j.jmva.2006.08.003, 2007.
- 605
- Gill, P. E. and Murray, W.: Algorithms for the solution of the nonlinear least-squares problem, *SIAM J. Numer. Anal.*, 15, 977–992, doi:10.1137/0715063, 1978.
- Gratton, S., Gürol, S., and Toint, P.: Preconditioning and globalizing conjugate gradients in dual space for quadratically penalized nonlinear-least squares problems, *Computational Optimization and Applications*, 610 54, 1–25, doi:10.1007/s10589-012-9478-7, 2013.
- Gu, Y. and Oliver, D.: An iterative ensemble Kalman filter for multiphase fluid flow data assimilation, *SPE Journal*, 12, 438–446, doi:10.2118/108438-PA, 2007.
- Hamill, T. M. and Snyder, C.: A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme, *Monthly Weather Review*, 128, 2905–2919, doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2, 2000a.
- 615 Hamill, T. M. and Snyder, C.: A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme, *Monthly Weather Review*, 128, 2905–2919, doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2, 2000b.
- 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.
- 620 Kasanický, I., Mandel, J., and Vejmelka, M.: Spectral diagonal ensemble Kalman filters, *Nonlinear Processes in Geophysics*, 22, 485 – 497, doi:10.5194/npg-22-485-2015, 2015.
- Khare, S. P., Anderson, J. L., Hoar, T. J., and Nychka, D.: An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems, *Tellus A*, 60, 97–112, doi:10.1111/j.1600-0870.2007.00281.x, 2008.
- 625 Levenberg, K.: A method for the solution of certain non-linear problems in least squares, *Quarterly of Applied Mathematics*, 2, 164–168, 1944.
- Liu, C. and Xiao, Q.: An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part III: Antarctic Applications with Advanced Research WRF Using Real Data, *Monthly Weather Review*, 141, 2721–2739, doi:10.1175/MWR-D-12-00130.1, 2013.
- 630 Liu, C., Xiao, Q., and Wang, B.: An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test, *Monthly Weather Review*, 136, 3363–3373, doi:10.1175/2008MWR2312.1, 2008.
- Liu, C., Xiao, Q., and Wang, B.: An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part II: Observing System Simulation Experiments with Advanced Research WRF (ARW), *Monthly Weather Review*, 635 137, 1687–1704, doi:10.1175/2008MWR2699.1, 2009.
- Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R., and Fairbairn, D.: Comparison of Hybrid-4DVar and Hybrid-4DVar Data Assimilation Methods for Global NWP, *Monthly Weather Review*, 143, 212–229, doi:10.1175/MWR-D-14-00195.1, 2014.

- Lorenz, E. N.: Deterministic Nonperiodic Flow, *Journal of the Atmospheric Sciences*, 20, 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2, 1963.
- Mandel, J., Beezley, J. D., Coen, J. L., and Kim, M.: Data Assimilation for Wildland Fires: Ensemble Kalman filters in coupled atmosphere-surface models, *IEEE Control Systems Magazine*, 29, 47–65, doi:10.1109/MCS.2009.932224, 2009.
- 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.
- Nowak, W., Tenkleve, S., and Cirpka, O.: Efficient Computation of Linearized Cross-Covariance and Auto-Covariance Matrices of Interdependent Quantities, *Mathematical Geology*, 35, 53–66, 2003.
- Osborne, M. R.: Nonlinear least squares – the Levenberg algorithm revisited, *Journal of the Australian Mathematical Society Series B*, 19, 343–357, doi:10.1017/S03342700000120X, 1976.
- 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.
- Rauch, H. E., Tung, F., and Striebel, C. T.: Maximum likelihood estimates of linear dynamic systems, *AIAA Journal*, 3, 1445–1450, 1965.
- 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.
- Strang, G. and Borre, K.: *Linear Algebra, Geodesy, and GPS*, Wellesley-Cambridge Press, 1997.
- Stroud, J. R., Stein, M. L., Lesht, B. M., Schwab, D. J., and Beletsky, D.: An Ensemble Kalman Filter and Smoother for Satellite Data Assimilation, *Journal of the American Statistical Association*, 105, 978–990, doi:10.1198/jasa.2010.ap07636, 2010.
- Trémolet, Y.: Model-error estimation in 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, 133, 1267–1280, doi:10.1002/qj.94, 2007.
- Trémolet, Y.: Object-Oriented Prediction System, <http://www.data-assimilation.net/Events/Year3/OOPS.pdf>, 2013.
- Tshimanga, J., Gratton, S., Weaver, A. T., and Sartenaer, A.: Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 134, 751–769, doi:10.1002/qj.228, 2008.
- Wang, X.: Incorporating Ensemble Covariance in the Gridpoint Statistical Interpolation Variational Minimization: A Mathematical Framework, *Monthly Weather Review*, 138, 2990–2995, doi:10.1175/2010MWR3245.1, 2010.
- Wright, S. J. and Holt, J. N.: An inexact Levenberg-Marquardt method for large sparse nonlinear least squares, *J. Austral. Math. Soc. Ser. B*, 26, 387–403, doi:10.1017/S0334270000004604, 1985.

- Zhang, F., Zhang, M., and Hansen, J.: Coupling ensemble Kalman filter with four-dimensional variational data assimilation, *Advances in Atmospheric Sciences*, 26, 1–8, doi:10.1007/s00376-009-0001-8, 2009.
- 680 Zupanski, M.: Maximum Likelihood Ensemble Filter: Theoretical Aspects, *Monthly Weather Review*, 133, 1710–1726, doi:10.1175/MWR2946.1, 2005.

Table 1. The root mean square error given by Eq. (34) for the first six Gauss–Newton iterations, for Lorenz 63 problem. The whole state is observed. Ensemble size is 100. The time window length is 50 time steps. Finite differences parameter is 10^{-3} .

Iteration	1	2	3	4	5	6
RMSE	20.16	15.37	3.73	2.53	0.09	0.09

Table 2. Mean of the objective function from 30 runs of the EnKS-4DVAR algorithm for the Lorenz 63 problem and for different values of τ (finite differences parameter). The whole state is observed. Ensemble size is 50. The time window length is 50 time steps.

Iter.	$\tau = 1$	$\tau = 10^{-1}$	$\tau = 10^{-2}$	$\tau = 10^{-3}$	$\tau = 10^{-4}$	$\tau = 10^{-5}$	$\tau = 10^{-6}$
Init	$5.61e+6$	$5.61e+6$	$5.61e+6$	$5.61e+6$	$5.61e+6$	$5.61e+6$	$5.61e+6$
1	$1.02e+6$	$1.39e+9$	$3.21e+9$	$3.54e+9$	$3.58e+9$	$3.58e+9$	$3.58e+9$
2	$1.39e+6$	$5.27e+7$	$1.70e+8$	$1.93e+8$	$1.96e+8$	$1.96e+8$	$1.96e+8$
3	$1.32e+6$	$4.14e+6$	$2.99e+6$	$3.69e+6$	$3.76e+6$	$3.77e+6$	$3.77e+6$
4	$1.38e+6$	5699	3266	4431	4581.31	4594	4598
5	$1.55e+6$	1299	89.22	65.69	65.4442	65.41	65.26
6	$1.34e+6$	830.1	17.08	6.933	6.844	6.856	6.923
7	$2.05e+6$	826.8	10.75	1.885	1.89082	1.8	1.721
8	$1.47e+6$	847.4	10.82	1.68	1.63813	1.547	1.641

Table 3. RMSE values calculated by Eq. (34) along the incremental 4DVAR and EnKS-4DVAR iterations for different values of the regularization parameter γ , for the two-level quasi-geostrophic model (Sect. 5.2.2).

Iter.	4DVAR	$\gamma = 0$	$\gamma = 10^{-3}$	$\gamma = 0.1$	$\gamma = 1$	$\gamma = 10$	$\gamma = 100$	$\gamma = 500$	$\gamma = 10^3$
Init	5.3026	5.3026	5.3026	5.3026	5.3026	5.3026	5.3026	5.3026	5.3026
1	3.9666	3.9713	3.9716	4.0274	4.4051	4.7046	4.8194	4.8774	4.9028
2	3.8167	3.8879	3.8903	3.8388	4.1949	4.3618	4.7136	4.8233	4.8514
3	3.8394	3.9703	3.9539	4.0927	4.1092	4.4898	4.6993	4.8093	4.8222
4	4.3390	4.1093	4.1891	3.9588	4.0232	4.4697	4.7348	4.7781	4.7771
5	3.9726	3.7723	3.7337	3.9000	3.9490	4.3866	4.7104	4.7802	4.7729
6	3.8984	3.8202	3.7302	3.8222	3.8045	4.3587	4.6785	4.7800	4.7624
7	3.7553	3.8873	3.8004	3.8619	4.0068	4.3369	4.6562	4.7742	4.7533
8	4.005	3.8183	4.1342	4.0614	3.7866	4.3147	4.6521	4.7578	4.7514
9	3.8429	3.7907	4.0450	3.7049	3.7159	4.2962	4.6358	4.7436	4.7409
10	3.8759	3.7177	4.0983	3.7242	3.6996	4.2805	4.6280	4.7239	4.7327

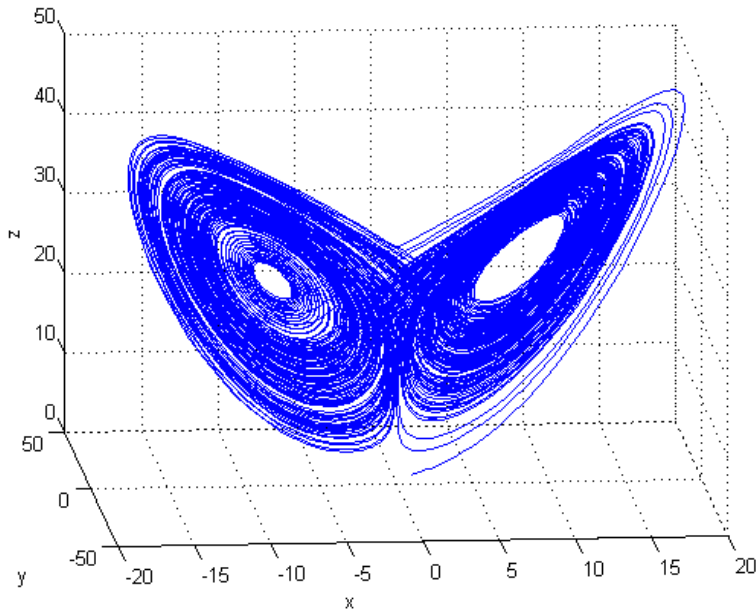


Figure 1. The Lorenz attractor, initial values $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, discretization time step is $dt = 0.1$ time unit.

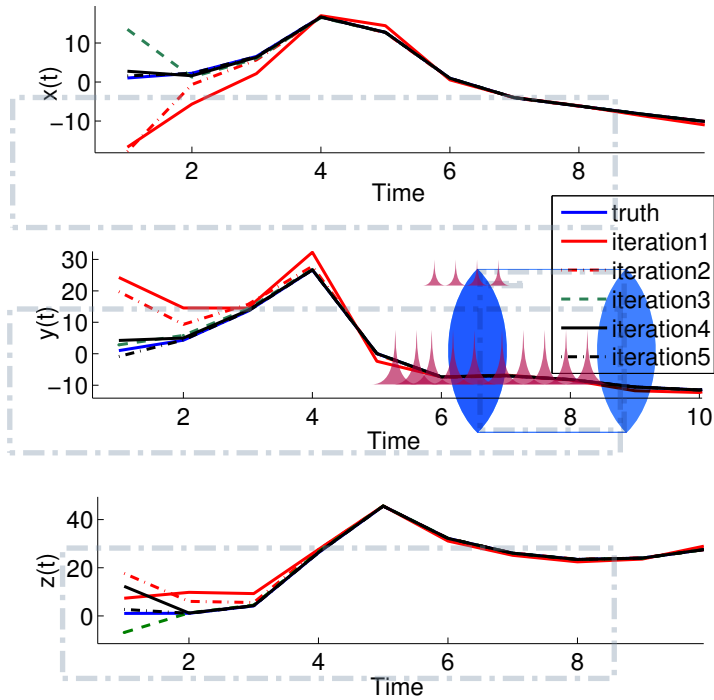


Figure 2. The three components x , y , z of the truth and the first five Gauss–Newton iterations from Lorenz 63 problem, for the first 10 time steps. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$. Time step is $dt = 0.1$ time unit. Observations are the full state at each time step. Ensemble size is 100. The time window length is 50 time steps. Finite differences parameter is 10^{-3} .

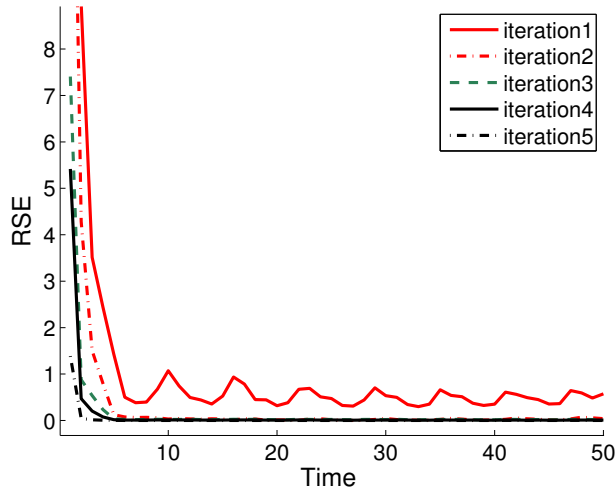


Figure 3. Root square error given by Eq. (33) for the first five Gauss–Newton iterations from Lorenz 63 problem. The problem setting is the same as in Fig. 2.

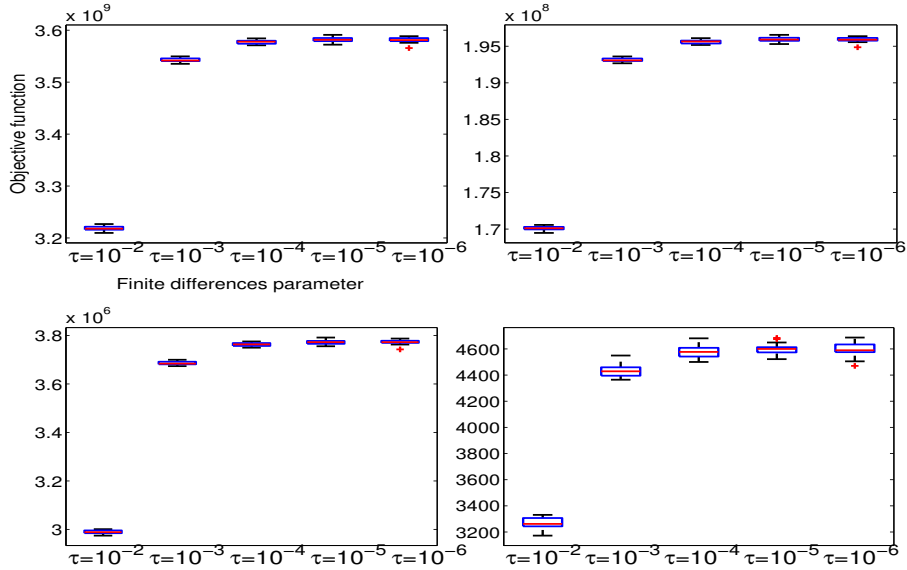


Figure 4. Box plots of objective function values for Lorenz 63 problem. From the left to the right and from the top to the bottom, the figures correspond to the results of the first, the second, the third and the fourth iteration respectively. The whole state is observed. Ensemble size is 50. The time window length is 50 time steps. In each box, the central line presents the median (red line), the edges are the 25th and 75th percentiles (blue line), the whiskers extend to the most extreme data points the plot algorithm considers to be not outliers (black line), and the outliers are plotted individually (red dots).

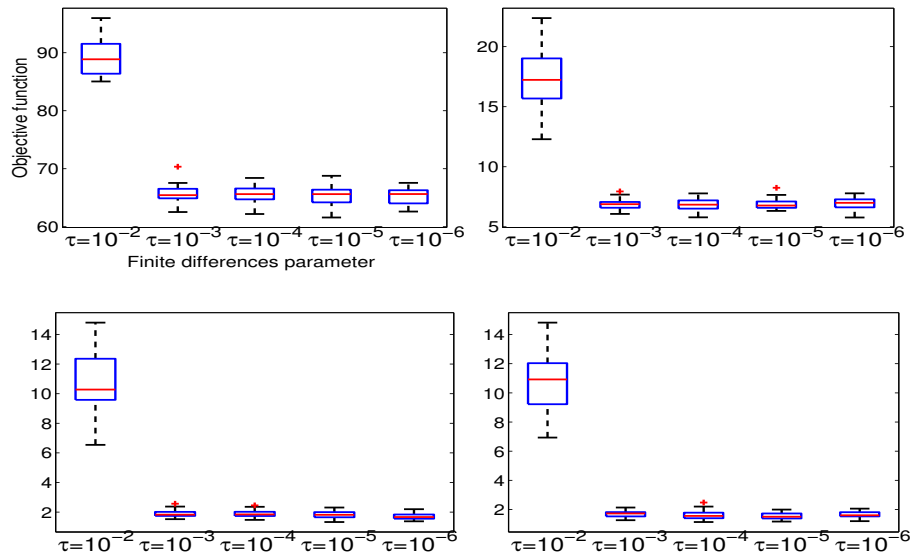


Figure 5. Same as Fig. 4, but for the fifth, the sixth, the seventh and the eighth iteration respectively.

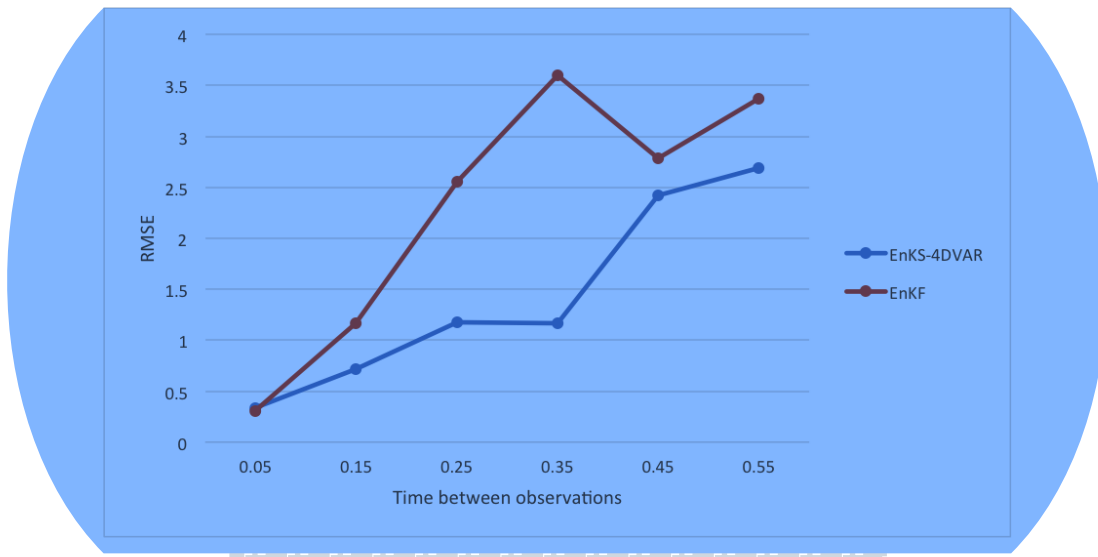


Figure 6. Comparison of RMSE between EnKF and EnKS-4DVAR from twin experiment for the Lorenz 63 model. EnKS-4DVAR has better performance for larger time interval between the observations as the model become more nonlinear. See Section 5.1.3 for further details.

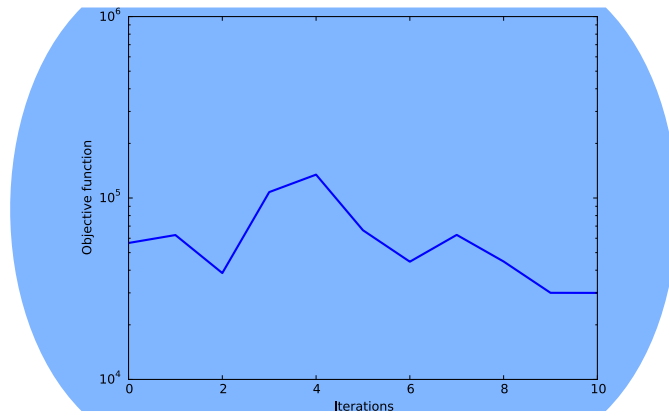


Figure 7. Objective function values along incremental 4DVAR iterations, for two-level quasi-geostrophic problem from Sect. 5.2.2.

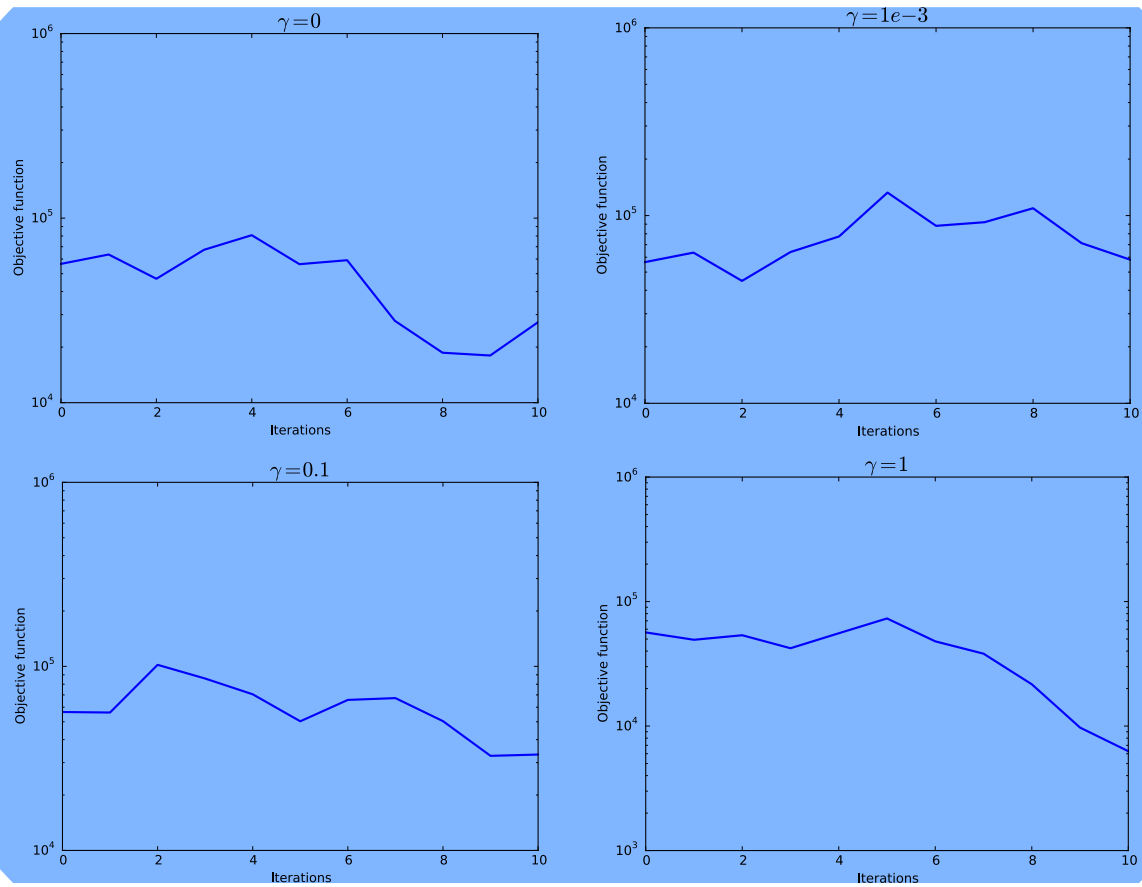


Figure 8. Objective function values along EnKS-4DVAR with regularization iterations for two-level quasi-geostrophic problem (Sect. 5.2.2). From the left to the right and from the top to the bottom: $\gamma = 0$, $\gamma = 0.001$, $\gamma = 0.1$, $\gamma = 1$.

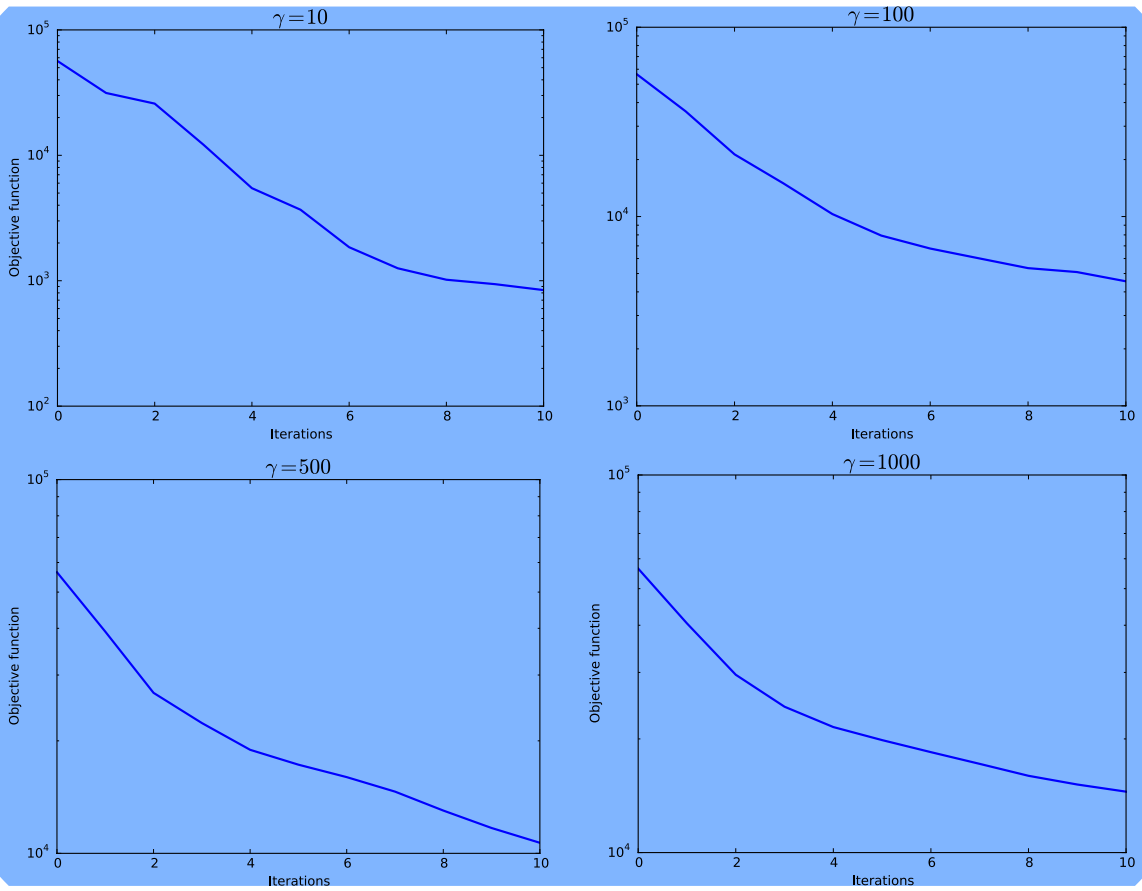


Figure 9. As Fig. 8, but for $\gamma = 10, \gamma = 100, \gamma = 500, \gamma = 1000$, respectively.