Manuscript prepared for Nonlin. Processes Geophys.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 31 January 2016

Response to reviewers comments - "Hybrid Levenberg–Marquardt and weak-constraint ensemble Kalman smoother method"

J. Mandel[1,5], E. Bergou[2], S. Gürol[3], S. Gratton[3,4], and I. Kasanický[5]

[1]University of Colorado Denver, Denver, CO, USA
[2]INRA, MaIAGE, Domaine de Vilvert, Jouy-en-Josas, France
[3]CERFACS, Toulouse, France
[4]INP-ENSEEIHT, Toulouse, France
[5]Institute 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.

In addition to the specific changes in response to the reviewers, listed below, we have done some proofreading and minor corrections all over.

Reviewer 1

I apologize to the authors for the late review. I found the new manuscript well improved and still easy to read. I am ready to propose its publication to the editor after another few minor improvements:

- The fact that the regularization term favors convergence but limits agreement to the observations (Table 3) is an objective and useful conclusion (perhaps expected, too) and should be stated in the abstract.

  The best results in Table 3 are obtained for $\gamma = 1$, both in terms of RMSE and the (original, non-penalized) objective function. The penalization affects the progress of the iterations by modifying the search direction and the step size, but, at least in the limit for large ensembles, it does not change the solution (the minimum of the 4DVAR objective function) itself and therefore its agreement with the observations. The behavior of the method for finite ensembles and small number of iterations, or sensitivity to local minima, can of course differ, and we appreciate the observation of the reviewer.

  No change.

- line 180: at this point, it is worth giving the expression of T matrix.

  Added.

- Eq. 15: is it x?

  Yes, fixed.

- line 228: correct references in brackets

  Removed.

- Figure 2 is redundant with Fig. 3 and not very informative. You should emphasize what this figure says in addition to Fig 3, or remove it.

  Deleted Fig. 2.

- line 349: the sentence seems incomplete

  Removed unwanted "of".

- line 364 also

  Removed repeated "randomly".

- section 5.1.3 deserves proofreading for punctuation, tenses, cases...

  Done.

- Figure 6: any idea why this discontinuity at 0.35?

  We can only speculate that one possible reason for such behavior of the EnKF could be covariance inflation. As mentioned in Bocquet and Sakov (2012), properly tuned covariance inflation is needed to guarantee, that EnKF perform well over multiple cycles. We did not perform the tuning of covariance inflation parameters. Also, (Bocquet and Sakov, 2012, Figure 7) also shows a small kink in the error of EnKF around 0.35-0.40. Still, we do not have a plausible reason.

  No changes made.

- Figures 8 and 9, independently: the y axis scales should be at least uniform. I actually think that both figures (8 graphs) could be compressed in one graph with 8 lines.

  Done.

Reviewer 2

Minor comments

1. l.30: “gaussian” $\longrightarrow$ “Gaussian”

   Fixed.

2. l.50: the “e.g.” should be in the parentheses.

   Fixed.

3. l.118: “For background...(2003).” I don’t understand what the purpose of this sentence is.

   This sentence had been added because Reviewer 1 asked in the report on the previous version: “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).”

   Deleted.

4. l.123: please explain that L refers to a time index.

   We have added that that L is the number of cycles in an assimilation window. We have unified the nomenclature and replaced “time step” by “cycle” where appropriate.

5. l.151: “$N(m, A)$”: explain - at least once - that N is the Gaussian distribution.

   Added.

6. l.174-175: “This technique is commonly used...(2009).” This is indeed a very well-known trick at the heart of the EnKF. Citing Chen and Snyder (2007) and Mandel et al. (2009) is really inappropriate. I believe that for such a very well known property, you either don’t cite anyone (implicitly referring to Evensen (2009)), or you cite one of the earlier papers by Houtekamer and Mitchell who introduced/emphasized it.

   Citations removed.

7. l.235: “gaussian” $\longrightarrow$ “Gaussian”

   Fixed.

8. l.325: You should mention the time-step while referring to the Runge-Kutta method.

   Noted that “The system is discretized using the fourth-order Runge–Kutta method with time step 0.1 time units in Sect. 5.1.1 and 5.1.2, and with time step 0.01 time units in Sect. 5.1.3.”

9. l.337: 0.1 as a time-step seems too large for me. This is definitely too large for the 40-variable Lorenz model. The Lorenz-63 model can be even more sensitive. Did you perform a sensitivity analysis?

   The statement about the time step was left inadvertently from an earlier version of the paper. The current computations are using adaptive RK method of order 4 and 5 within each cycle.

   Corrected.

10. l.369-373: You reach the same conclusion as in Bocquet and Sakov (2014) (τ is really the same parameter as ε) which shows there is a comfortable range of safe values. This is worth mentioning.

    Done, reformulated.

11. l.387: "6 observations" $\longrightarrow$ "6 observation vectors"

    Changed.

12. l.405: "into Object" $\longrightarrow$ "into the Object"

    Done.

13. "developed by European" $\longrightarrow$ "developed by the European"

    Done.

14. l.480: "$N = 30000$" that is very, very large. Especially for 1600 state variables. Why such a choice?

    Because no localization was used in the experiments, the ensemble size had to be large enough, as already noted in the text.

15. l.505: "For smaller...small": Did you intend to split the sentence?

    Yes. Replaced by "for smaller...slow" for more clarity. Also corrected the wording.

16. l.530: "those of obtained" $\longrightarrow$ "those obtained"

    Fixed.

17. l.550: "there is nothing to prevent the use" $\longrightarrow$ "there is -a priori- nothing to prevent the use" would be so much closer to the truth...

    Changed.

References

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

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.

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

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.

Manuscript prepared for Nonlin. Processes Geophys.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 31 January 2016

Hybrid Levenberg–Marquardt and ~~weak-constraint~~ weak-constraint ensemble Kalman smoother method

J. Mandel[1,5], E. Bergou[2], S. Gürol[3], S. Gratton[3,4], and I. Kasanický[5]

[1]University of Colorado Denver, Denver, CO 80217-3364, USA
[2]INRA, MaIAGE, Université Paris-Saclay, 78350 Jouy-en-Josas, France
[3]CERFACS, 31100 Toulouse, France
[4]INP-ENSEEIHT, 31071 Toulouse, France
[5]Institute of Computer Science, Academy of Sciences of the Czech Republic, 182 07 Prague, Czech Republic

Abstract. The ensemble Kalman smoother (EnKS) is used as a linear ~~least-squares~~ least-squares solver in the Gauss–Newton method for the large nonlinear ~~least-squares~~ 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~~ 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~~ least-squares problem is solved iteratively by solving a succession of linearized ~~least-square~~ least-squares problems. The major cost in 4DVAR iterations is ~~in~~ evaluating the model, the 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 linearization 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.

The Kalman filter is a sequential Bayesian estimation of the ~~gaussian~~ 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 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~~ 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 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 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 time points. Such composite variables are also called 4D vectors, ~~e.g., (Desroziers et al., 2014)~~(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 lest 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 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 operators 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 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 probabilitistic 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-Marquard 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 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 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 version 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 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 ?.~~

The paper is organized as follows. In Sect. 2, we review the formulation of 4DVAR. The EnKF and 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 $\boldsymbol{u}_i$, $i = 1, \ldots, L$, denote the composite (column) 4D vector

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

where L is the number of cycles in the assimilation window. We want to estimate $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_L$, where $\boldsymbol{x}_i$ is the state at time i, from the background state, ~~$\boldsymbol{x}_0 \approx \boldsymbol{x}_\mathrm{b},$~~ $\boldsymbol{x}_0 \approx \boldsymbol{x}_\mathrm{b}$, the model, $\boldsymbol{x}_i \approx \mathcal{M}_i(\boldsymbol{x}_{i-1})$, and the observations ~~$\mathcal{H}_i(\boldsymbol{x}_i) \approx \boldsymbol{y}_i,$~~ $\mathcal{H}_i(\boldsymbol{x}_i) \approx \boldsymbol{y}_i$, where $\mathcal{M}_i$ is the model operator, and $\mathcal{H}_i$ is the observation operator. Quantifying the uncertainty by covariances, with $\boldsymbol{x}_0 \approx \boldsymbol{x}_\mathrm{b}$ taken

as ~~$(\boldsymbol{x}_0 - \boldsymbol{x}_\mathrm{b})^\mathrm{T} \mathbf{B}^{-1} (\boldsymbol{x}_0 - \boldsymbol{x}_\mathrm{b}) \approx 0$~~$(\boldsymbol{x}_0 - \boldsymbol{x}_\mathrm{b})^\mathrm{T} \mathbf{B}^{-1} (\boldsymbol{x}_0 - \boldsymbol{x}_\mathrm{b}) \approx 0$, etc., we get the nonlinear ~~least squares~~ least-squares problem

$$\|\boldsymbol{x}_0 - \boldsymbol{x}_\mathrm{b}\|^2_{\mathbf{B}^{-1}} + \sum_{i=1}^{L} \|\boldsymbol{x}_i - \mathcal{M}_i(\boldsymbol{x}_{i-1})\|^2_{\mathbf{Q}_i^{-1}} + \sum_{i=1}^{L} \|\boldsymbol{y}_i - \mathcal{H}_i(\boldsymbol{x}_i)\|^2_{\mathbf{R}_i^{-1}} \to \min_{\boldsymbol{x}_{0:L}}, \tag{1}$$

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

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

$$\mathcal{M}_i(\boldsymbol{x}_{i-1} + \boldsymbol{\delta x}_{i-1}) \approx \mathcal{M}_i(\boldsymbol{x}_{i-1}) + \mathcal{M}'_i(\boldsymbol{x}_{i-1})\,\boldsymbol{\delta x}_{i-1},$$
$$\mathcal{H}_i(\boldsymbol{x}_i + \boldsymbol{\delta x}_i) \approx \mathcal{H}_i(\boldsymbol{x}_i) + \mathcal{H}'_i(\boldsymbol{x}_i)\,\boldsymbol{\delta x}_i.$$

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

$$\|\boldsymbol{x}_0 + \boldsymbol{\delta x}_0 - \boldsymbol{x}_\mathrm{b}\|^2_{\mathbf{B}^{-1}} + \sum_{i=1}^{L} \|\boldsymbol{x}_i + \boldsymbol{\delta x}_i - (\mathcal{M}_i(\boldsymbol{x}_{i-1}) + \mathcal{M}'_i(\boldsymbol{x}_{i-1})\,\boldsymbol{\delta x}_{i-1})\|^2_{\mathbf{Q}_i^{-1}}$$
$$+ \sum_{i=1}^{L} \|\boldsymbol{y}_i - (\mathcal{H}_i(\boldsymbol{x}_i) + \mathcal{H}'_i(\boldsymbol{x}_i)\,\boldsymbol{\delta x}_i)\|^2_{\mathbf{R}_i^{-1}} \to \min_{\boldsymbol{\delta x}_{0:k}\,\boldsymbol{\delta x}_{0:L}}. \tag{2}$$

This is the Gauss–Newton method (Bell, 1994; Tshimanga et al., 2008) for nonlinear least squares, known in 4DVAR as the incremental approach (Courtier et al., 1994). Write the auxiliary linear ~~least squares~~ least-squares problem (Eq. 2) for $\boldsymbol{\delta x}_{0:L}$ as

$$\left\|\boldsymbol{\delta x}_0 - \boldsymbol{\delta x}_\mathrm{b}\right\|^2_{\mathbf{B}^{-1}} + \sum_{i=1}^{L} \|\boldsymbol{\delta x}_i - (\mathbf{M}_i \boldsymbol{\delta x}_{i-1} + \boldsymbol{m}_i)\|^2_{\mathbf{Q}_i^{-1}} + \sum_{i=1}^{L} \|\boldsymbol{d}_i - \mathbf{H}_i \boldsymbol{\delta x}_i\|^2_{\mathbf{R}_i^{-1}} \to \min_{\boldsymbol{\delta x}_{0:L}} \tag{3}$$

where

$$\boldsymbol{\delta x}_\mathrm{b} = \boldsymbol{x}_\mathrm{b} - \boldsymbol{x}_0, \quad \boldsymbol{m}_i = \mathcal{M}_i(\boldsymbol{x}_{i-1}) - \boldsymbol{x}_i, \quad \boldsymbol{d}_i = \boldsymbol{y}_i - \mathcal{H}_i(\boldsymbol{x}_i), \tag{4}$$
$$\mathbf{M}_i = \mathcal{M}'_i(\boldsymbol{x}_{i-1}), \quad \mathbf{H}_i = \mathcal{H}'_i(\boldsymbol{x}_i).$$

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 for our purposes. We start with a formulation of the EnKF, in a notation useful for the extension to EnKS. The notation $\boldsymbol{v}^\ell \sim N(\boldsymbol{m}, \mathbf{A})$ means that $\boldsymbol{v}^\ell$ is sampled from ~~$N(\boldsymbol{m}, \mathbf{A})$~~ the Gaussian distribution $N(\boldsymbol{m}, \mathbf{A})$ with mean $\boldsymbol{m}$ and covariance $\mathbf{A}$, 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 $\mathbf{X}_{i|j}^N = \left[\boldsymbol{x}_{i|j}^1, \ldots, \boldsymbol{x}_{i|j}^N\right] = \left[\boldsymbol{x}_{i|j}^\ell\right]$, where the ensemble member index ℓ always runs over $\ell = 1, \ldots, N$, and similarly for other ensembles. Assume for the moment that the observation operator $\mathcal{H}_i$ is linear, that is, $\mathcal{H}_i(\boldsymbol{u}) = \mathbf{H}_i \boldsymbol{u}$. The **EnKF algorithm** consists of the following steps:

1. Initialize

$$\boldsymbol{x}_{0|0}^\ell \sim N(\boldsymbol{x}_\mathrm{b}, \mathbf{B}), \quad \ell = 1, \ldots, N. \tag{5}$$

2. For ~~$i = 1, 2, \ldots$~~ $i = 1, 2, \ldots, L$,

   (a) advance in time

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

   (b) The analysis step

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

   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 ~~$\mathbf{Z}_{i|i-1}^N$~~ $\mathbf{X}_{i|i-1}^N$,

$$\mathbf{A}_i^N = \left[\boldsymbol{a}_i^1, \ldots, \boldsymbol{a}_i^N\right] = \left[\boldsymbol{x}_{i|i-1}^1 - \overline{\boldsymbol{x}}_{i|i-1}, \ldots, \boldsymbol{x}_{i|i-1}^N - \overline{\boldsymbol{x}}_{i|i-1}\right], \quad \overline{\boldsymbol{x}}_{i|i-1} = \frac{1}{N}\sum_{j=1}^N \boldsymbol{x}_{i|i-1}. \tag{8}$$

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

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

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

$$\boldsymbol{g}_i^\ell = \mathbf{H}_i \boldsymbol{a}_i^\ell = \mathbf{H}_i\left(\boldsymbol{x}_{i|i-1}^\ell - \overline{\boldsymbol{x}}_{i|i-1}\right) = \mathbf{H}_i \boldsymbol{x}_{i|i-1}^\ell - \frac{1}{N}\sum_{j=1}^N \mathbf{H}_i \boldsymbol{x}_{i|i-1}^j, \tag{10}$$

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., ?Mandel et al. (2009). With~~ ~~$\mathbf{H}_i \mathbf{A}_i^N = \mathbf{G}_i^N = [\boldsymbol{g}_i^1, \ldots, \boldsymbol{g}_i^N]$~~. With $\mathbf{H}_i \mathbf{A}_i^N = \mathbf{G}_i^N = [\boldsymbol{g}_i^1, \ldots, \boldsymbol{g}_i^N]$, Eq. (9) becomes

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

Also, from Eqs. (~~9,8~~) 7,9) and writing the matrix of anomalies in the form

$$\mathbf{A}_i^N = \mathbf{X}_{i|j}^N \left(\boldsymbol{I} - \frac{\mathbf{1}\mathbf{1}^\mathrm{T}}{N}\right),$$

where $\mathbf{1}$ is the column vector of all ones of length N, it follows that the analysis ensemble $\mathbf{X}^N_{i|i}$ consists of linear combinations of the forecast ensemble, hence it can be written as multiplying the forecast ensemble by a suitable transformation matrix ~~$\mathbf{T}^N_i$,~~

~~$\mathbf{X}^N_{i|i} = \mathbf{X}^N_{i|i-1}\mathbf{T}^N_i, \quad \mathbf{T}^N_i \in \mathbb{R}^{N\times N}.$~~

$\mathbf{T}^N_i$,

$\mathbf{X}^N_{i|i} = \mathbf{X}^N_{i|i-1}\mathbf{T}^N_i, \quad \mathbf{T}^N_i \in \mathbb{R}^{N\times N},$

where

$$\mathbf{T}^N_i = \mathbf{I} - \frac{1}{N-1}\left(\boldsymbol{I} - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}}{N}\right)\left(\mathbf{A}^N_i\right)^{\mathrm{T}}\left(\frac{1}{N-1}\mathbf{A}^N_i\left(\mathbf{A}^N_i\right)^{\mathrm{T}} + \boldsymbol{R}_i\right)^{-1} \cdot \left[\boldsymbol{H}_i\boldsymbol{x}^\ell_{i|i-1} - \boldsymbol{d}_i + \boldsymbol{w}^\ell_i\right]_{\ell=1,N}. \tag{12}$$

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 $i-1$,

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

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

$$\boldsymbol{x}^\ell_{0:i|i} = \boldsymbol{x}^N_{0:i|i-1} - \mathbf{P}^N_{0:i,0:i}\widetilde{\mathbf{H}}^{\mathrm{T}}_{0:i}(\widetilde{\mathbf{H}}_{0:i}\mathbf{P}_{0:i,0:i}\widetilde{\mathbf{H}}^{\mathrm{T}}_{0:i} + \mathbf{R}_i)^{-1}(\widetilde{\mathbf{H}}_{0:i}\boldsymbol{x}^\ell_{0:i|i-1} - \boldsymbol{d}^\ell_i - \boldsymbol{w}^\ell_i),$$

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

$$\widetilde{\mathbf{H}}_{0:i}\boldsymbol{x}^\ell_{0:i|i-1} = [0, \ldots, 0, \mathbf{H}_i]\,\boldsymbol{x}^\ell_{0:i|i-1} = \mathbf{H}_i\boldsymbol{x}^\ell_{i|i-1} \tag{13}$$

$$\mathbf{P}^N_{0:i,0:i}\widetilde{\mathbf{H}}^{\mathrm{T}}_{0:i} = \mathbf{P}^N_{0:i,i}\mathbf{H}^{\mathrm{T}}_i, \quad \widetilde{\mathbf{H}}_{0:i}\mathbf{P}_{0:i,0:i}\widetilde{\mathbf{H}}^{\mathrm{T}}_{0:i} = \mathbf{H}_i\mathbf{P}^N_{i,i}\mathbf{H}^{\mathrm{T}}_i, \tag{14}$$

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

1. Initialize

$$\boldsymbol{x}^\ell_{0|0} \sim N(\boldsymbol{x}_{\mathrm{b}}, \mathbf{B}), \quad \ell = 1, \ldots, N. \tag{15}$$

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

   (a) Advance in time:

$$\boldsymbol{x}^\ell_{i|i-1} = \mathcal{M}_i(\boldsymbol{x}^\ell_{i-1|i-1}) + \boldsymbol{v}^\ell_i, \quad \boldsymbol{v}^\ell_i \sim N(\mathbf{0}, \mathbf{Q}_i), \quad \ell = 1, \ldots, N \tag{16}$$

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

$$\mathbf{A}_{0:i} = \left[\boldsymbol{a}_{0:i}^1, \ldots, \boldsymbol{a}_{0:i}^N\right], \quad \boldsymbol{a}_{0:i}^\ell = \boldsymbol{x}_{0:i|i-1}^\ell - \frac{1}{N}\sum_{j=1}^N \boldsymbol{x}_{0:i|i-1}^j \tag{17}$$

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

(c) The analysis step:

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

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~~19) with the increments $\boldsymbol{\delta x}$ in place of $\boldsymbol{x}$ to solve the linearized auxiliary ~~least squares~~ least-squares problem (Eq. 3). Approximating by finite differences based at $\boldsymbol{x}_{i-1}$ with step $\tau > 0$, we get the action of the linearized model operator

$$\mathbf{M}_i \boldsymbol{\delta x}_{i-1}^\ell + \boldsymbol{m}_i \approx \frac{\mathcal{M}_i\left(\boldsymbol{x}_{i-1} + \tau \boldsymbol{\delta x}_{i-1}^\ell\right) - \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right)}{\tau} + \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right) - \boldsymbol{x}_i, \tag{20}$$

and the linearized observation operator

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

The Gauss–Newton method may diverge, but convergence to a stationary point of (Eq. 1) can be recovered by a control of the step $\boldsymbol{\delta x}$. Adding a constraint of the form $\|\boldsymbol{\delta x}_i\| \leq \varepsilon$ leads to globally convergent trust region methods (Gratton et al., 2013). Here, we add to (Eq. 3) a Tikhonov regularization term of the form $\gamma \|\boldsymbol{\delta x}_i\|^2_{\mathbf{S}_i^{-1}}$, 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) $\boldsymbol{x}_{0:L} \leftarrow \boldsymbol{x}_{0:L} + \boldsymbol{\delta x}_{0:L}$, where

$$\|\boldsymbol{\delta x}_0 - \boldsymbol{\delta x}_\mathrm{b}\|^2_{\mathbf{B}^{-1}} + \sum_{i=1}^L \|\boldsymbol{\delta x}_i - \mathbf{M}_i \boldsymbol{\delta x}_{i-1} - \boldsymbol{m}_i\|^2_{\mathbf{Q}_i^{-1}} + \sum_{i=1}^L \|\boldsymbol{d}_i - \mathbf{H}_i \boldsymbol{\delta x}_i\|^2_{\mathbf{R}_i^{-1}} + \gamma \sum_{i=0}^L \|\boldsymbol{\delta x}_i\|^2_{\mathbf{S}_i^{-1}} \to \min_{\boldsymbol{\delta x}_{0:L}}. \tag{22}$$

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 $\gamma \|\boldsymbol{\delta x}_i\|^2_{\mathbf{S}_i^{-1}}$ in Eq. (22) as arising from additional independent observations $\delta \boldsymbol{x}_i \approx \mathbf{0}$ 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 $\boldsymbol{\delta x}^\ell_{0:i|i-1/2}$ for the increments after the first half-step, conditioned on the original observations only, and $\boldsymbol{\delta x}^\ell_{0:i|i}$ for the increments conditioned also on the regularization $\delta \boldsymbol{x}_i \approx \mathbf{0}$. 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~~ 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

$$\boldsymbol{x}_0 = \boldsymbol{x}_\mathrm{b}, \quad \boldsymbol{x}_i = \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right), i = 1, \ldots, L,$$

if not given already.

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

$$\boldsymbol{\delta x}^\ell_{0|0} \sim N\left(\mathrm{b}\mathbf{0}, \mathbf{B}\right), \quad \ell = 1, \ldots, N, \mathrm{b} = \mathrm{b} - 0. \tag{23}$$

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

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

$$\boldsymbol{\delta x}^\ell_{i|i-1} = \frac{\mathcal{M}_i\left(\boldsymbol{x}_{i-1} + \tau \boldsymbol{\delta x}^\ell_{i-1|i-1}\right) - \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right)}{\tau} + \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right) - \boldsymbol{x}_i + \boldsymbol{v}^\ell_i, \tag{24}$$

$$\boldsymbol{v}^\ell_i \sim N\left(\mathbf{0}, \mathbf{Q}_i\right), \quad \ell = 1, \ldots, N.$$

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

$$\mathbf{A}_{0:i} = \left[\boldsymbol{a}^1_{0:i}, \ldots, \boldsymbol{a}^N_{0:i}\right], \quad \boldsymbol{a}^\ell_{0:i} = \boldsymbol{\delta x}^\ell_{i|i-1} - \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{\delta x}^j_{i|i-1}$$

$$\mathbf{G}^N_i = \left[\boldsymbol{g}^1_i, \ldots, \boldsymbol{g}^N_i\right], \quad \boldsymbol{g}^\ell_i = \frac{1}{\tau}\left(\mathcal{H}_i\left(\boldsymbol{x}_i + \tau\boldsymbol{\delta x}^\ell_{i|i-1}\right) - \frac{1}{N}\sum_{j=1}^{N}\mathcal{H}_i(\boldsymbol{x}_i + \tau\boldsymbol{\delta x}^j_{i|i-1})\right) \tag{25}$$

iii. The first analysis step:

$$\begin{aligned}\boldsymbol{\delta x}^\ell_{0:i|i-1/2} =& \boldsymbol{\delta x}^\ell_{0:i|i-1} - \frac{1}{N-1}\mathbf{A}^N_{0:i}\left(\mathbf{G}^N_i\right)^{\mathrm{T}}\left(\frac{1}{N-1}\mathbf{G}^N_i\left(\mathbf{G}^N_i\right)^{\mathrm{T}} + \mathbf{R}_i\right)^{-1} \cdot \\ & \left(\mathcal{H}_i(\boldsymbol{x}_i) + \frac{\mathcal{H}_i\left(\boldsymbol{x}_i + \tau\boldsymbol{\delta x}^\ell_{i|i-1}\right) - \mathcal{H}_i\left(\boldsymbol{x}_i\right)}{\tau} - \boldsymbol{y}_i - \boldsymbol{w}^\ell_i\right), \qquad (26)\\ \boldsymbol{w}^\ell_i \sim& N\left(\mathbf{0}, \mathbf{R}_i\right), \quad \ell = 1, \ldots, N.\end{aligned}$$

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

$$\mathbf{Z}^N_{0:i} = \left[\boldsymbol{z}^1_{0:i}, \ldots, \boldsymbol{z}^N_{0:i}\right], \quad \boldsymbol{z}^\ell_{0:i} = \boldsymbol{\delta x}^\ell_{i|i-1/2} - \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\delta x}^j_{i|i-1/2} \tag{27}$$

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

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

$$\begin{aligned}\boldsymbol{\delta x}^\ell_{0:i|i} =& \boldsymbol{\delta x}^\ell_{0:i|i-1/2} - \frac{1}{N-1}\mathbf{Z}^N_{0:i}\left(\mathbf{Z}^N_i\right)^{\mathrm{T}}\left(\frac{1}{N-1}\mathbf{Z}^N_i\left(\mathbf{Z}^N_i\right)^{\mathrm{T}} + \frac{1}{\gamma}\mathbf{S}_i\right)^{-1} \cdot \qquad (28)\\ & \left(\boldsymbol{\delta x}^\ell_{i|i-1/2} - \boldsymbol{v}^\ell_i\right), \quad \boldsymbol{v}^\ell_i \sim N\left(\mathbf{0}, \mathbf{S}_i\right), \quad \ell = 1, \ldots, N,\end{aligned}$$

otherwise $\boldsymbol{\delta x}^\ell_{0:i|i} = \boldsymbol{\delta x}^\ell_{0:i|i-1/2}$, $\ell = 1, \ldots, N$.

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

$$\boldsymbol{x}_{0:L} \leftarrow \boldsymbol{x}_{0:L} + \frac{1}{N}\sum_{\ell=1}^{N}\boldsymbol{\delta x}^\ell_{0:L|L}. \tag{29}$$

Note that for small $\gamma \to 0$, (Eq. 28) has asymptotically no effect, $\boldsymbol{\delta x}^\ell_{0:i|i} \to \boldsymbol{\delta x}^\ell_{0:i|i-1/2}$. 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 ~~$\boldsymbol{x}_{0:k}$~~ $\boldsymbol{x}_{0:L}$ 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 $\boldsymbol{x}_0$. 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 $\boldsymbol{x}_{0:L} + \boldsymbol{\delta x}^{\ell}_{0:L}$ do not depend on the previous values of $\boldsymbol{x}_{1:L}$.

Proof. Indeed, Eq. (24) becomes

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

hence

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

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

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

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

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

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

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

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

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\sigma(x-y),\ \frac{\mathrm{d}y}{\mathrm{d}t} = \rho x - y - xz,\ \frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z, \tag{32}$$

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 $\boldsymbol{X}_t = [x(t), y(t), z(t)]^\top$, $\boldsymbol{X}_t \in \mathbb{R}^3$.

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~~ cycle i. We then build the data $\boldsymbol{y}_i$ by applying the observation operator $\mathcal{H}_i$ to the truth at time i and by adding a Gaussian perturbation $N(0, \mathbf{R}_i)$. Similarly, the background $\boldsymbol{x}_b$ 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

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~~ cycles of 0.1 time unit. ~~The~~ Within each cycle, the differential equations are solved by adaptive Runge-Kutta method implemented as MATLAB function `ode45`, with default parameter values. The assimilation time window length is $L = 50$ ~~time steps~~ cycles (5 time units total). The observation operator is defined as ~~$\mathcal{H}_i(x,y,z) = (x^2,y^2,z^2)$~~ $\mathcal{H}_i(x,y,z) = (x^2,y^2,z^2)$. At each time i, the observations are constructed as follows: ~~$\boldsymbol{y}_i = \mathcal{H}_i(\text{truth}_i) + \boldsymbol{v}_i$~~ $\boldsymbol{y}_i = \mathcal{H}_i(\text{truth}_i) + \boldsymbol{v}_i$, where $\boldsymbol{v}_i$ is sampled from $N(0,\mathbf{R})$ with ~~$\mathbf{R} = \mathbf{I}_3$~~ $\mathbf{R} = \mathbf{I}_3$. Observations are taken for each ~~time step~~ cycle ($i = 1,\ldots 50$). The ensemble size is fixed to $N = 100$.

Figure ~~?? shows the estimator of the state vector $\boldsymbol{X}_i$, $i = 1,\ldots,10$, for the first five iterations. Figure~~ 2 shows the root square error (RSE) for the ~~same iterates showm in Fig. ??. RSE is~~ first 5 iterations, defined as

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

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

$$\text{RMSE}^{(j)} = \frac{1}{k}\frac{1}{L}\sum_{i=0}^{kL}\text{RSE}_i^{(j)}, \quad j = 1,\ldots,5,. \tag{34}$$

~~where k is the number of time steps.~~

From Table 1 and ~~Figs. ?? and 2~~Fig. 2, it can be seen that the iterates ~~of~~ converge to the solution~~(without using a regularization)~~, without using 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

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}.

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 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. 3 and 4 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 $\tau \leq 10^{-3}$~~, except that~~. This is a similar conclusion as in Bocquet and Sakov (2014); parameter τ here plays the same role as their ε. It should be noted that very small τ, when the problem solved by the smoother is essentially the tangent problem, ~~resulted~~ results 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

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)~~ Eq. 32), with the parameters $\sigma = 10$, $\rho = 28$, $\beta = 8/3$~~, and integration step set to~~. This time, we use the Runge-Kutta method of order 4 with the time step of 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~~ We perform usual twin model experiment. The initial truth state Y_0 is generated from $N(0, \mathbf{I}_3)$ distribution and the initial forecast state is then simulated by sampling from ~~$N(Y_0, \mathbf{I}_3)$~~$N(Y_0, \mathbf{I}_3)$. Both states are advanced for $50,000$ model ~~time-steps~~ time steps burn-in period~~We used~~. We use the nonlinear observational operator ~~$h(x,y,z) = (x^3, y^3, z^3)$~~ $h(x,y,z) = (x^3, y^3, z^3)$ with observational error generated from $N\left(0, \sigma^2 \mathbf{I}_3\right)$ with ~~$\sigma^2 = 8$~~$\sigma^2 = 8$, and $\tau = 10^{-4}$. The ~~time~~ cycle length Δt between two available observations ~~Δt~~ varies from 0.05 time ~~units~~unit, when the model is nearly linear, to 0.55, when the model is strongly nonlinear. We use ensemble ~~of~~ size 10. After running multiple simulations, we have found suitable values of the parameters of the method ~~:~~ as the number of iterations ~~is~~ 25 ~~, and the parameter~~ and the penalty coefficient $\gamma = 10^{-9}$ when $\Delta t = 0.05$ and $\gamma = 1000$ otherwise. The length of assimilation window is ~~set to~~ $L = 6$, ~~in other words we assimilate~~ i.e., assimilating 6 ~~observations~~ observation vectors at once. ~~We assimilate each observation~~ Each observation vector is assimilated only once, i.e., the assimilation windows do not overlap. To create initial ensemble at the beginning of each iteration, we use the background covariance created as a weighted average of the sample covariance from the last iteration in the previous assimilation window and the identity matrix, similarly as in Hamill and Snyder (2000a). The weights are 0.99 for the sample covariance and 0.01 for the identity. The model

error covariance in each ~~time-step set to~~ cycle is $Q = 0.01\mathbf{I}_3$. The experiment ~~is~~ was run for $100,000$ observation cycles.

We compare the proposed method with the standard EnKF with ~~10 ensemble members~~also ensemble size 10, where the initial ensemble is created after the burn-in period by adding ~~$N(0,\mathbf{I}_3)$ perturbations~~ perturbations sampled from $N(0,\mathbf{I}_3)$. For stability reason and to preserve the covariance between ensemble members~~we add $N(0,0.01\mathbf{I}_3)$noise after each advance of each ensemblemember~~, we add noise sampled from $N(0,0.01\mathbf{I}_3)$ after advancing the ensemble. The necessity of related covariance inflation was ~~identified~~ pointed out 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 directly after burn-in period.~~

Fig. 5 shows that the proposed method has a significantly smaller RMSE than the EnKF in the case when the time between observation is larger and thus the behavior of the model is nonlinear. Only in the ~~situation, when the time between~~ case when the cycle length between the observation is 0.05 time unit, i.e., the model behavior is nearly linear, EnKF gives a comparable result as the proposed method.

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

The EnKS-4DVAR algorithm has been implemented into the Object Oriented Prediction System (OOPS) (Trémolet, 2013), which is a data assimilation framework developed by the European Centre for Medium-Range Weather Forecasts (ECMWF). Numerical experiments are performed by using the simple two-layer quasi-geostrophic model ~~of~~ in the 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 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 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

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

where ~~ψ is the stream function~~ψ_1, ψ_2 are the stream functions, ∇^2 is the two-dimensional Laplacian, R_s represents orography or heating, β is the (non-dimensionalised) northward variation of the Cori-

olis parameter at the fixed latitude y, f_0 is the Coriolis parameter at the southern boundary of the domain. L is the typical length scale of the motion we wish to describe, H_1 and H_2 are the depths of the two layers, $g' = g\Delta\theta/\overline{\theta}$ is the reduced gravity where $\overline{\theta}$ 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:

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

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

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

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

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

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

and

$$u_i = -\frac{\partial \psi_i}{\partial y}, \qquad v_i = \frac{\partial \psi_i}{\partial x}, \tag{38}$$

are the horizontal velocity components in each layer. Therefore, the potential vorticity at each time 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 the north–south direction. For the experiments in this paper, we choose $L = 10^6$ m, $\bar{U} = 10\,\mathrm{m\,s^{-1}}$, $H_1 = 6000$ m, $H_2 = 4000$ m, $f_0 = 10^{-4}\,\mathrm{s^{-1}}$, $\beta_0 = 1.5 \times 10^{-11}\,\mathrm{s^{-1}\,m^{-1}}$. For more details on the model and its solution, we refer to Fisher et al. (2011).

The domain for the experiments is $12000\,\mathrm{km}$ by $6300\,\mathrm{km}$ for both layers. The horizontal discretization consists of 40×20 points, so that the east–west and the north–south resolution is approximately $300\,\mathrm{km}$. 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).

The truth is generated from a model with layer depths of $D_1 = 6000\,\mathrm{m}$ and $D_2 = 4000\,\mathrm{m}$, and the time step is set to $300\,\mathrm{s}$, whereas the assimilating model has layer depths of $D_1 = 5500\,\mathrm{m}$ and $D2 = 4500\,\mathrm{m}$, and the time step is set to $3600\,\mathrm{s}$. 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 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 independent 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. The observation operator is the bi-linear interpolation of the model fields to horizontal observation locations.

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 $10^6\,\mathrm{m}$ 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 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 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 tangent and adjoint models in each iteration, with no ensembles involved. The numerical results are presented as follows.

Figure 6 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 model for a long window (10 days). In such a case, as explained in Sect. 4, a convergence to a sta-

tionary 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 ?? and ?? show~~ Figure 7 shows the objective function values along iterations for eight different choices of γ. RMSE values along the iterations for the same experiments performed with 4DVAR and EnKS-4DVAR are presented in Table 3.

It can be seen from ~~Figures ?? and ??~~ Figure 7 that when $\gamma = 0$, the iterations ~~diverging as expected(~~diverge as expected, since we do not use regularization and we only approximate the linearized subproblem using ensembles~~)~~. For small values of γ (~~for instance~~ e.g., $\gamma \leq 10^{-1}$), the objective function is not monotonically decreasing, hence the iterations are still diverging even if we use the regularization. Therefore, small values of γ can not guarantee the convergence. For large values of γ (~~for instance~~ e.g., $\gamma \geq 10$), 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~~ beyond an optimal value results in higher RMSE values~~. For large values of γ, for example $\gamma \geq 10$~~, and the reduction in RMSE values ~~is very small. For smaller values of γ the reduction is faster however RMSE values oscillates~~ becomes very slow. In any case, the RMSE values oscillate along the iterations. We ~~want to note also~~ note 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

We have proposed a stochastic solver for the incremental 4DVAR ~~weak constraint~~ 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~~ 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 sup-

ply 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 futher 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.

For the second part of the experiments, we have shown the performance of the EnKS-4DVAR method with regularization on the two-level quasi-geostropic 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~~ assimilation window length, and that the 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.

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. However, there is *a priori* 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 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-ENSEEIHT and CERFACS, and when Elhoucine Bergou, Serge Gratton, and Ivan Kasanický 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.

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.

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.

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.

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.

~~Chen, Y. and Snyder, C.: Assimilating Vortex Position with an Ensemble Kalman Filter, Monthly Weather Review, 135, 1828–1845, , 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.

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

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.

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

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

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

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.

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, 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.

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.~~

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.

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.

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, 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-4DEnVar and Hybrid-4DVar Data Assimilation Methods for Global NWP, Monthly Weather Review, 143, 212–229, doi:10.1175/MWR-D-14-00195.1, 2014.

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/S033427000000120X, 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.

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~~ assimilation window length is 50 ~~time steps~~cycles. 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~~ assimilation window length is 50 ~~time steps~~cycles.

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

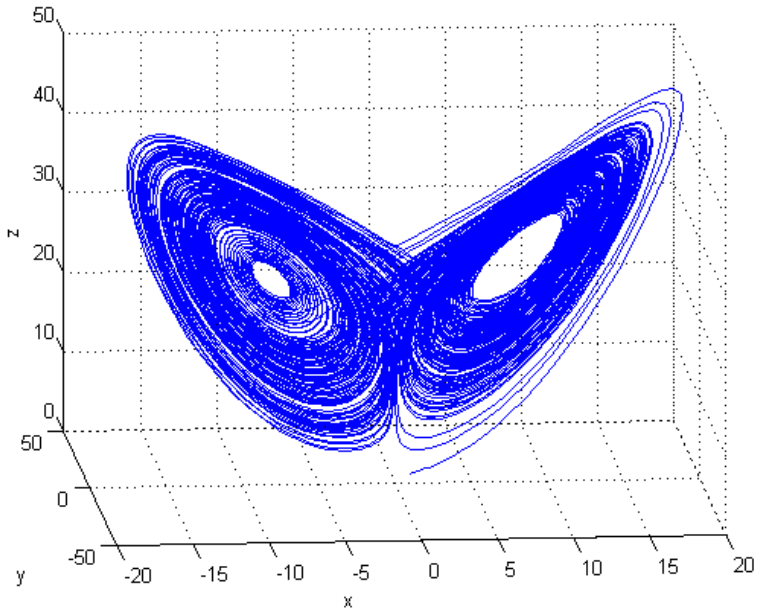

Figure 1. The Lorenz attractor, initial values $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$~~, discretization time step is dt = 0.1 time unit~~.

~~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}.~~

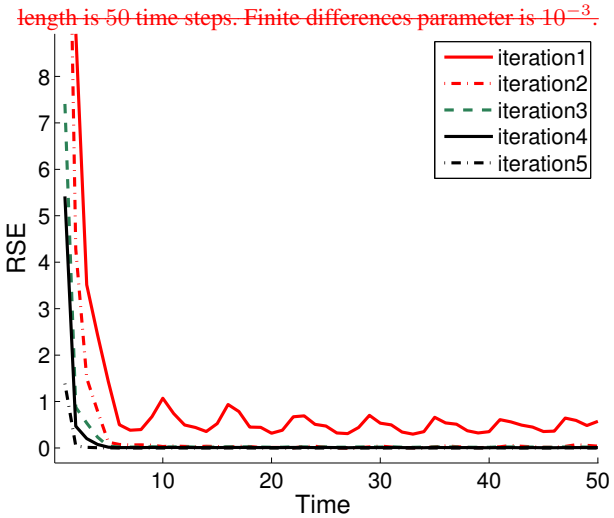

Figure 2. Root square error given by Eq. (33) for the first five Gauss–Newton iterations from Lorenz 63 problem. The ~~problem setting~~ initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$. The cycle length is dt $= 0.1$ time unit. The observations are the ~~same as in Fig~~full state at each time step. ~~??~~The ensemble size is $N = 100$. The assimilation window length is $L = 50$ cycles. Finite differences parameter is $\tau = 10^{-3}$.

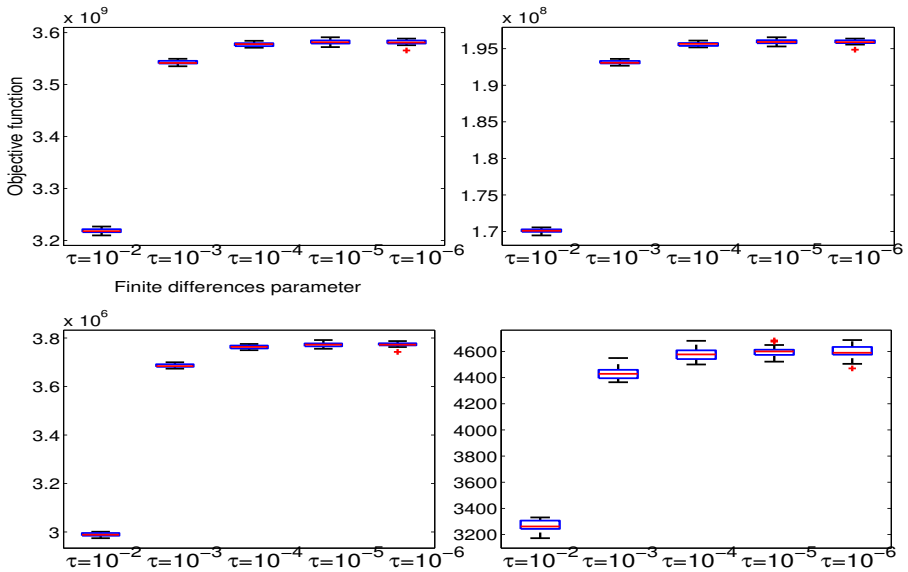

Figure 3. Box plots of objective function values for Lorenz 63 problem. From the left to the right and from the top to the bottom, the figures correspond to the results of the first, the second, the third and the fourth iteration respectively. The whole state is observed. Ensemble size is 50. The ~~time~~ assimilation window ~~length~~ is 50 ~~time steps~~cycles. 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).

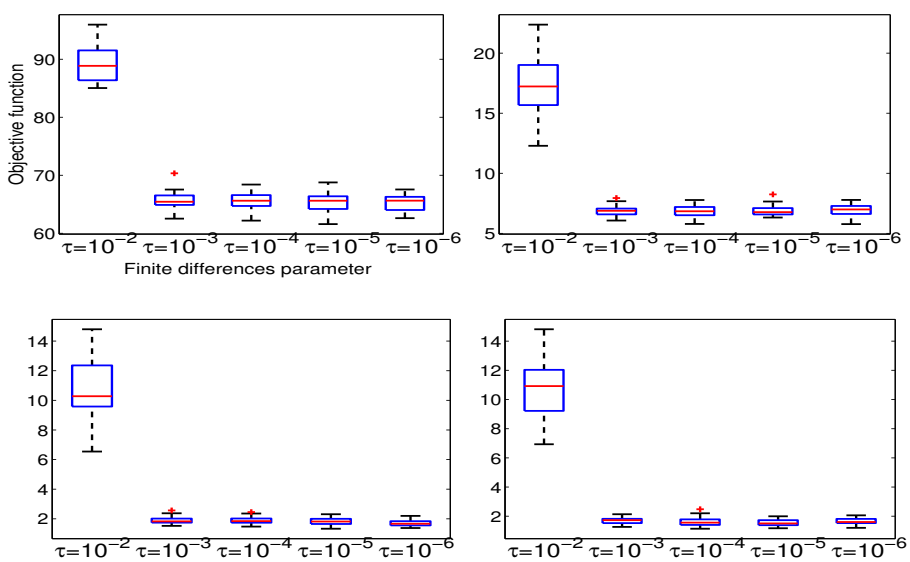

Figure 4. Same as Fig. 3, but for the fifth, the sixth, the seventh and the eighth iteration respectively.

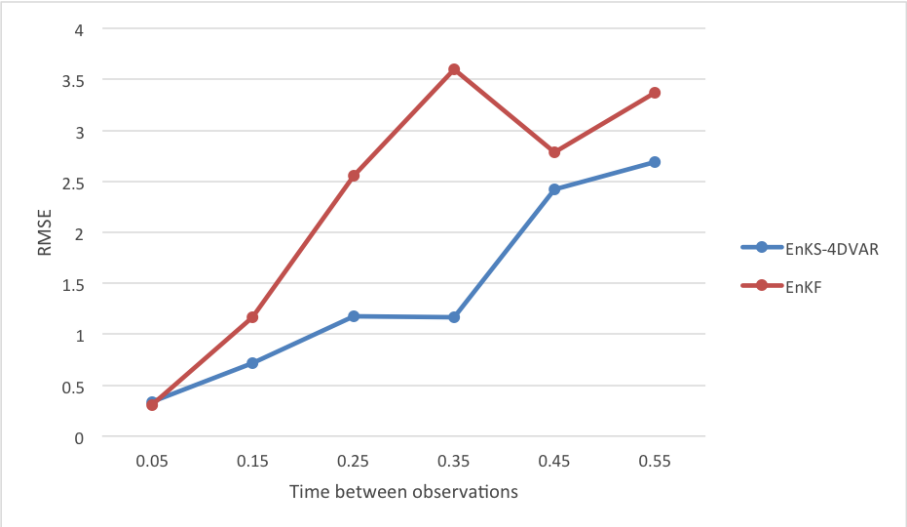

Figure 5. Comparison of RMSE between EnKF and EnKS-4DVAR from twin experiment for the Lorenz 63 model. EnKS-4DVAR has better performance for larger time interval between the observations as the model become more nonlinear. See Section 5.1.3 for further details.

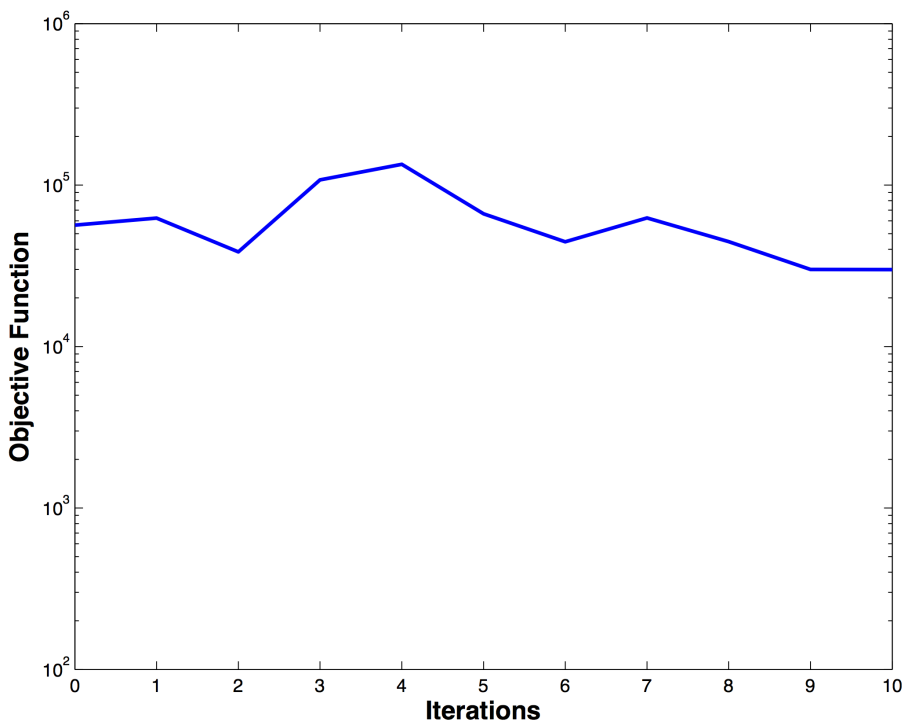

Figure 6. Objective function values along incremental 4DVAR iterations ~~,~~ for the two-level quasi-geostrophic problem from Sect. 5.2.2.

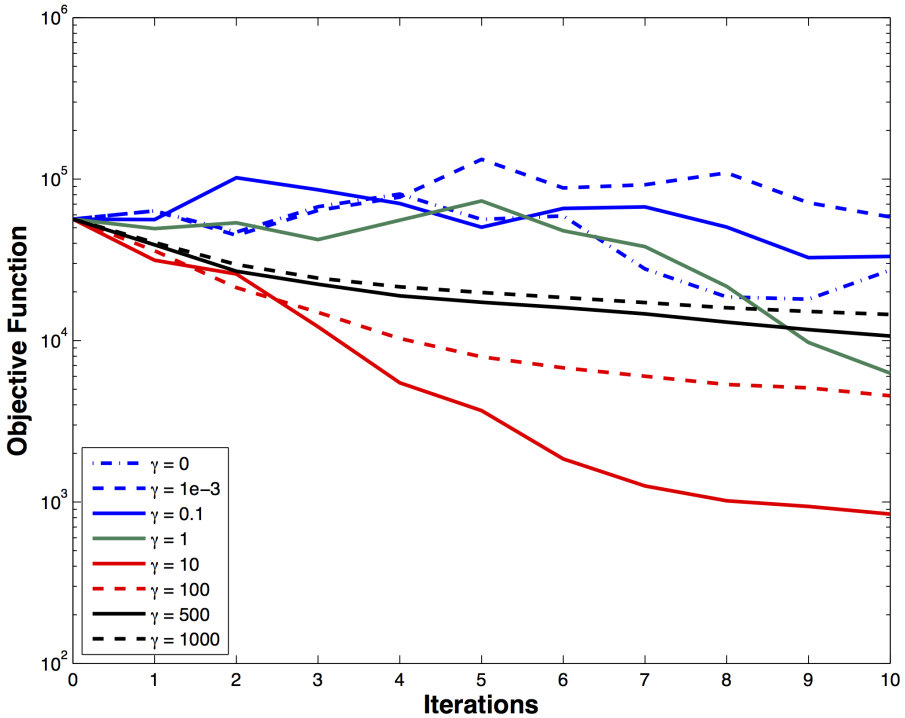

Figure 7. Objective function values along EnKS-4DVAR with regularization iterations for the two-level quasi-geostrophic problem (Sect. 5.2.2). ~~From the left to the right and from the top to the bottom: $\gamma = 0$, $\gamma = 0.001$, $\gamma = 0.1$, $\gamma = 1$.~~

~~As Fig. **??**, but for $\gamma = 10$, $\gamma = 100$, $\gamma = 500$, $\gamma = 1000$, respectively.~~