Hybrid Levenberg-Marquardt and weak constraint ensemble Kalman smoother method, Nonlinear Processes in Geophysics

Jan Mandel, El Houcine Bergou, Selime Gurol, Serge Gratton

Nonlin. Processes Geophys. Discuss., 2, 865–902, 2015
www.nonlin-processes-geophys-discuss.net/2/865/2015/
doi:10.5194/npgd-2-865-2015

Nonlinear Processes in Geophysics
Discussions
Open Access

Hybrid Levenberg–Marquardt and weak constraint ensemble Kalman smoother method

J. Mandel[1], **E. Bergou**[2,3], **S. Gürol**[3], **and S. Gratton**[3]

[1]University of Colorado Denver, Denver, CO, USA
[2]Institut de Mathématiques de Toulouse, Toulouse, France
[3]INP-ENSEEIHT and CERFACS, Toulouse, France

Received: 21 April 2015 – Accepted: 23 April 2015 – Published: 26 May 2015

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

Published by Copernicus Publications on behalf of the European Geosciences Union & the American Geophysical Union.

Abstract

We propose to use the ensemble Kalman smoother (EnKS) as the linear least squares solver in the Gauss–Newton method for the large nonlinear least squares 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 and the two-level quasi-geostrophic model problems.

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 very 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 using a succession of linear least square solutions. The major cost in 4DVAR iterations is in evaluating the model, tangent and adjoint opera-

NPGD
2, 865–902, 2015

tors, and solving 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, called the incremental approach (Courtier et al., 1994), leads to the Gauss–Newton method for nonlinear least squares (Bell, 1994; Tshimanga et al., 2008). However, Gauss–Newton iterations may not converge, not even locally. 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 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 at all time points from the beginning 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), similar the sequential Kalman filter.

It is well known that weak constraint 4DVAR is equivalent to the Kalman smoother in the linear case. To apply the Kalman smoother in the nonlinear case, the problem needs to be linearized, leading to variants of the extended Kalman filter and the Gauss–Newton method. 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,

NPGD
2, 865–902, 2015

with the short recursions and a forward and a backward pass, as in the KS. 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. We use the latter approach in the computations reported here.

In this paper, we propose to use the EnKS as a linear least squares solver in 4DVAR. 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. 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. The Tikhonov regularization is implemented within EnKS as a computationally cheap additional observation (Johns and Mandel, 2008). We call the resulting method EnKS-4DVAR. Theoretical convergence of the algorithm for large ensembles is studied in Bergou et al. (2014), where a rigorous proof is provided that the iterations of the EnKS-4DVAR method converge to those of the incremental 4DVAR for large ensembles in the L^p norm, for any $p \in [1, \infty)$, in the large ensemble limit.

The EnKF has become a competitive method for data assimilation. Consequently, combinations of ensemble and variational approaches have become of considerable recent interest. Estimating the background covariance for 4DVAR from an ensemble was one of the first connections (Hamill and Snyder, 2000), and it is now standard and became operational (Wang, 2010). 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), who added regularization and use LETKF-like approach to minimize the nonlinear objective function over linear combinations of the ensemble. Bocquet and Sakov (2012) scale the ensemble to approximate the derivatives (tangent operators) as in Sakov et al. (2012). They call their approach the "bundle variant", which is the same

NPGD
2, 865–902, 2015

as using finite differences to approximate derivatives. Here, we use a similar technique to approximate the derivatives. Liu et al. (2008, 2009); Liu and Xiao (2013) also minimize the (strong constraint) 4DVAR objective function over linear combinations of the ensemble by computations in the observation space. Their method, called Ens4DVAR, does not need tangent or adjoint operators also. 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).

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 and use finite difference approximations of the tangent operators, similarly as in Sakov et al. (2012) and here. 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. 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. The implementation of the EnKS in the computations reported here also uses sample covariance, and, consequently, the analysis ensemble consists of linear combinations of the forecast ensemble. However, limiting the EnKF to linear combinations only does not allow common approaches to localization (Sakov and Bertino, 2011).

The present approach is not tied to the use of sample covariance. Rather, it allows an arbitrary implementation of the EnKS to be used, which makes localization possible, e.g., by tapering the sample covariance (Furrer and Bengtsson, 2007) or replacing the sample covariance by its diagonal in a spectral space (Kasanický et al., 2015). Implementations of the EnKS with localization exist, e.g., Butala (2012), which is based

NPGD
2, 865–902, 2015

on the Bryson–Frazier version of the classical formulation the KS, with a forward and a backward pass. 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.

The paper is organized as follows. In Sect. 2, we review the formulation of 4DVAR. The EnKS for the incremental linearized squares problem is reviewed in Sect. 3. The new method without tangent operators is introduced in Sect. 4. The modifications for the regularization and the Levenberg–Marquardt method are presented in Sect. 5. Section 6 contains the results of the computational experiments, and Sect. 7 is the conclusion.

2 Incremental 4DVAR and the Gauss–Newton method

We want to estimate $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_k$, where $\boldsymbol{x}_i$ is the state at time i, from the background state, $\boldsymbol{x}_0 \approx \boldsymbol{x}_\mathrm{b}$, the model, $\boldsymbol{x}_i \approx \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right)$, and the observations $\mathcal{H}_i\left(\boldsymbol{x}_i\right) \approx \boldsymbol{y}_i$, 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 $\left(\boldsymbol{x}_0 - \boldsymbol{x}_\mathrm{b}\right)^\mathrm{T} \mathbf{B}^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_\mathrm{b}\right) \approx 0$, etc., we get the nonlinear least squares problem

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

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

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

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

NPGD
2, 865–902, 2015

For vectors $\boldsymbol{u}_i$, $i = 1, \ldots, k$, denote the composite (column) vector

$$\boldsymbol{u}_{0:k} = \begin{bmatrix} \boldsymbol{u}_0 \\ \vdots \\ \boldsymbol{u}_k \end{bmatrix}.$$

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

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

This is the Gauss–Newton method (Bell, 1994; Tshimanga et al., 2008) for nonlinear squares, known in 4DVAR as the incremental approach (Courtier et al., 1994). Denote

$$\boldsymbol{z}_{0:k} = \boldsymbol{\delta x}_{0:k}, \quad \boldsymbol{z}_\mathrm{b} = \boldsymbol{x}_\mathrm{b} - \boldsymbol{x}_0, \; \boldsymbol{m}_i = \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right) - \boldsymbol{x}_i, \quad \boldsymbol{d}_i = \boldsymbol{y}_i - \mathcal{H}_i\left(\boldsymbol{x}_i\right), \tag{3}$$
$$\mathbf{M}_i = \mathcal{M}'_i\left(\boldsymbol{x}_{i-1}\right), \quad \mathbf{H}_i = \mathcal{H}'_i\left(\boldsymbol{x}_i\right),$$

and write the auxiliary linear least squares problem (Eq. 2) as

$$\|\boldsymbol{z}_0 - \boldsymbol{z}_\mathrm{b}\|^2_{\mathbf{B}^{-1}} + \sum_{i=1}^{k} \|\boldsymbol{z}_i - \mathbf{M}_i\boldsymbol{z}_{i-1} - \boldsymbol{m}_i\|^2_{\mathbf{Q}_i^{-1}} + \sum_{i=1}^{k} \|\boldsymbol{d}_i - \mathbf{H}_i\boldsymbol{z}_i\|^2_{\mathbf{R}_i^{-1}} \to \min_{\boldsymbol{z}_{0:k}} \tag{4}$$

The function minimized in Eq. (4) is the same as the one minimized in the Kalman smoother (Bell, 1994). The Gauss–Newton method with the Kalman smoother as the linear least squares solver is known as the iterated Kalman smoother, and considerable improvements can be obtained against running the Kalman smoother, applied to the linearized problem, only once (Bell, 1994; Fisher et al., 2005).

NPGD
2, 865–902, 2015

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 (Algorithm 1), 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 $N(\boldsymbol{m}, \mathbf{A})$ 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 $\mathbf{Z}^N_{i|j} = \left[z^1_{i|j}, \ldots, z^N_{i|j}\right] = \left[z^\ell_{i|j}\right]$, 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, $\mathcal{H}_i(\boldsymbol{u}) = \mathbf{H}_i \boldsymbol{u}$.

Algorithm 1 EnKF

Initialize $z^\ell_{0|0} \sim N(z_\mathrm{b}, \mathbf{B}), \quad \ell = 1, \ldots, N$

For $i = 1, \ldots, k$, advance in time

$$z^\ell_{i|i-1} = \mathcal{M}_i(z^\ell_{i-1|i-1}) + \boldsymbol{v}^\ell_i, \quad \boldsymbol{v}^\ell_i \sim N(0, \mathbf{Q}_i), \tag{5}$$

followed by the analysis step

$$z^\ell_{i|i} = z^\ell_{i|i-1} - \mathbf{P}^N_i \mathbf{H}^\mathrm{T}_i (\mathbf{H}_i \mathbf{P}^N_i \mathbf{H}^\mathrm{T}_i + \mathbf{R}_i)^{-1} (\mathbf{H}_i(z^\ell_{i|i-1}) - \boldsymbol{d}_i - \boldsymbol{w}^\ell_i), \quad \boldsymbol{w}^\ell_i \sim N(0, \mathbf{R}_i), \tag{6}$$

where $\mathbf{P}^N_i$ is the sample covariance computed from the the ensemble $\mathbf{Z}^N_{i|i-1}$.

We write the matrices in Eq. (6) as

$$\mathbf{P}^N_i \mathbf{H}^\mathrm{T}_i = \frac{1}{N-1} \mathbf{A}(\mathbf{H}_i \mathbf{A})^\mathrm{T}, \quad \mathbf{H}_i \mathbf{P}^N_i \mathbf{H}^\mathrm{T}_i = \frac{1}{N-1} \mathbf{H}_i \mathbf{A}(\mathbf{H}_i \mathbf{A})^\mathrm{T}, \tag{7}$$

NPGD
2, 865–902, 2015

where $\mathbf{A}$ is the matrix of anomalies of the ensemble $\mathbf{Z}^N_{i|i-1}$,

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

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

$$\mathbf{H}_i a^\ell = \mathbf{H}_i \left(z^\ell_{i|i-1} - \overline{z}_{i|i-1}\right) = \mathbf{H}_i z^\ell_{i|i-1} - \frac{1}{N}\sum_{j=1}^{N} \mathbf{H}_i z^j_{i|i-1}. \tag{9}$$

Equation (9) allows the use of a nonlinear observation operator $\mathcal{H}_i$, which only needs to be evaluated on the members of the ensemble. This technique is commonly used for nonlinear observation operators, e.g., Chen and Snyder (2007); Mandel et al. (2009). The replacement of $\mathbf{H}_i$ by $\mathcal{H}_i$ in Eq. (6) is straightforward.

From Eq. (7), it follows that the analysis ensemble $\mathbf{Z}^N_{i|i}$ consists of linear combinations of the forecast ensemble, which can be written as multiplying the forecast ensemble by a suitable transformation matrix $\mathbf{T}^N_i$,

$$\mathbf{Z}^N_{i|i} = \mathbf{Z}^N_{i|i-1}\mathbf{T}^N_i, \quad \mathbf{T}^N_i \in \mathbb{R}^{N\times N}. \tag{10}$$

The EnKS is obtained by applying the same analysis step (Eq. 6) as in the EnKF to the ensemble $\mathbf{Z}_{0:i|i-1}$ of composite states from time 0 to i, conditioned on data up to time $i-1$,

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

NPGD
2, 865–902, 2015

in the place of $\mathbf{Z}_{i|i-1}$, with the observation matrix $\widetilde{\mathbf{H}}_{0:i} = [0, \ldots, \mathbf{H}_i]$. Then, Eq. (6) becomes

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

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

$$\widetilde{\mathbf{H}}_{0:i}z^{\ell}_{0:i|i-1} = [0, \ldots, 0, \mathbf{H}_i]z^{\ell}_{0:i|i-1} = \mathbf{H}_i z^{\ell}_{i|i-1} \tag{11}$$

$$\mathbf{P}^{N}_{0:i,0:i}\widetilde{\mathbf{H}}^{\mathrm{T}}_{0:i} = \mathbf{P}^{N}_{0:i,i}\mathbf{H}^{\mathrm{T}}_i, \quad \widetilde{\mathbf{H}}_{0:i}\mathbf{P}_{0:i,0:i}\widetilde{\mathbf{H}}^{\mathrm{T}}_{0:i} = \mathbf{H}_i\mathbf{P}^{N}_{i,i}\mathbf{H}^{\mathrm{T}}_i, \tag{12}$$

which gives the analysis step (Eq. 14) in the EnKS (Algorithm 2).

Algorithm 2 EnKS

Initialize $z^{\ell}_{0|0} \sim N(z_{\mathrm{b}}, \mathbf{B}), \quad \ell = 1, \ldots, N.$
For $i = 1, \ldots, k$, advance in time,

$$z^{\ell}_{i|i-1} = \mathbf{M}_i z^{\ell}_{i-1|i-1} + \boldsymbol{m}_i + \boldsymbol{v}^{\ell}_i, \quad \boldsymbol{v}^{\ell}_i \sim N(0, \mathbf{Q}_i), \tag{13}$$

followed by the analysis step

$$z^{\ell}_{0:i|i} = x^{N}_{0:i|i-1} - \mathbf{P}^{N}_{0:i,i}\mathbf{H}^{\mathrm{T}}_i(\mathbf{H}_i\mathbf{P}^{N}_{i,i}\mathbf{H}^{\mathrm{T}}_i + \mathbf{R}_i)^{-1}(\mathbf{H}_i z^{\ell}_{0:i|i-1} - \boldsymbol{d}^{\ell}_i), \quad \boldsymbol{d}^{\ell}_i \sim N(0, \mathbf{R}_i), \tag{14}$$

where $\widetilde{\mathbf{H}}_{0:i} = [0, \ldots, \mathbf{H}_i]$, and $\mathbf{P}^{N}_{i,i}$ is the sample covariance matrix of $\mathbf{Z}^{N}_{i|i-1}$.

The EnKS can be implemented in a straightforward manner by applying the same transformation as in the EnKF to the composite state from times 0 to i, $\mathbf{Z}^{N}_{0:i|i} = \mathbf{Z}^{N}_{0:i|i-1}\mathbf{T}^{N}_i$, where $\mathbf{T}^{N}_i$ is the transformation matrix in Eq. (10) (Brusdal et al., 2003, Eq. 20).

NPGD
2, 865–902, 2015

4 Nonlinear EnKS-4DVAR method

The formulation of the auxiliary least squares problem (Eq. 2) relies on the linearized (i.e., tangent) model operators $\mathbf{M}_i$ and $\mathbf{H}_i$ and their adjoints. The linearized model $\mathbf{M}_i = \mathcal{M}_i'\left(\boldsymbol{x}_{i-1}\right)$ occurs only in advancing the time as action on the ensemble members $\boldsymbol{\delta x}^\ell = \boldsymbol{z}^\ell$,

$$\mathbf{M}_i \boldsymbol{z}_{i-1}^\ell + \boldsymbol{m}_i = \mathcal{M}_i'\left(\boldsymbol{x}_{i-1}\right)\boldsymbol{z}_{i-1}^\ell + \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right) - \boldsymbol{x}_i$$

Approximating by finite differences based at $\boldsymbol{x}_{i-1}$ with step $\tau > 0$, we get

$$\mathbf{M}_i \boldsymbol{z}_{i-1}^\ell + \boldsymbol{m}_i \approx \frac{\mathcal{M}_i\left(\boldsymbol{x}_{i-1} + \tau \boldsymbol{z}_{i-1}^\ell\right) - \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right)}{\tau} + \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right) - \boldsymbol{x}_i. \tag{15}$$

Thus, advancing the linearized model in time requires $N+1$ evaluations of $\mathcal{M}_i$, at $\boldsymbol{x}_{i-1}$ and $\boldsymbol{x}_{i-1} + \tau \boldsymbol{\delta x}_{i-1}^n$. The observation matrix $\mathbf{H}_i$ occurs only in the action on the deviations of the ensemble of increments, $\mathbf{H}_i\left(\boldsymbol{z}_i^\ell - \overline{\boldsymbol{z}}_i\right)$, $\overline{\boldsymbol{z}}_i = \frac{1}{N}\sum_{j=1}^N \boldsymbol{z}_i^j$. Approximating by finite differences based at $\boldsymbol{x}_i$, with step $\tau > 0$, we have in the multiplication (Eq. 7) of the ensemble covariance,

$$\mathbf{H}_i(\boldsymbol{z}_i^\ell - \overline{\boldsymbol{z}}) \approx \frac{\mathcal{H}_i\left(\boldsymbol{x}_i + \tau \boldsymbol{z}_i^\ell\right) - \mathcal{H}_i\left(\boldsymbol{x}_i\right)}{\tau} - \frac{1}{N}\sum_{j=1}^N \frac{\mathcal{H}_i(\boldsymbol{x}_i + \tau \boldsymbol{z}_i^j) - \mathcal{H}_i\left(\boldsymbol{x}_i\right)}{\tau} \tag{16}$$

Thus, evaluating the action of the linearized observation operator on the ensemble requires $N+1$ evaluations of $\mathcal{H}_i$, at $\boldsymbol{x}_i$ and $\boldsymbol{x}_i + \tau \boldsymbol{z}_i^\ell$. We call the resulting method EnKS-4DVAR (Algorithm 3).

NPGD
2, 865–902, 2015

Algorithm 3 EnKS-4DVAR

First, initialize

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

if not given already. One iteration (Eq. 2) of the incremental 4DVAR is then implemented as follows.
Given $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_k$, initialize $z^\ell_{0|0} \sim N(z_\mathrm{b}, \mathbf{B})$ with $z_\mathrm{b} = \boldsymbol{x}_\mathrm{b} - \boldsymbol{x}_0$.
For $i = 1, \ldots, k$, advance z^ℓ in time following Eq. (13), with the linearized operator approximated from Eq. (15),

$$z^\ell_{i|i-1} = \frac{\mathcal{M}_i\left(\boldsymbol{x}_{i-1} + \tau z^\ell_{i-1|i-1}\right) - \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right)}{\tau} + \mathcal{M}_i\left(\boldsymbol{x}_{i-1}\right) - \boldsymbol{x}_i + \boldsymbol{v}^\ell_i, \tag{17}$$

$\boldsymbol{v}^\ell_i \sim N(0, \mathbf{Q}_i)$, followed by the analysis step (Eq. 14), with the multiplication by the matrix $\mathbf{H}_i$ approximated by linearizing the observation operator from Eq. (16).
For $i = 1, \ldots, k$, update $\boldsymbol{x}_i \longleftarrow \boldsymbol{x}_i + \frac{1}{N}\sum_{\ell=1}^{N} z^\ell_{i|k}$.

NPGD
2, 865–902, 2015

Algorithm 4 Nonlinear EnKS

Initialize $\boldsymbol{x}^k_{0|0} \sim N(\boldsymbol{x}_\mathrm{b}, \mathbf{B})$.
For $i = 1, \ldots, k$, advance in time

$$\boldsymbol{x}^\ell_{i|i-1} = \mathcal{M}_i\left(\boldsymbol{x}^\ell_{i-1|i-1}\right) + \boldsymbol{v}^\ell_i, \quad \boldsymbol{v}^\ell_i \sim N(0, \mathbf{Q}_i) \tag{18}$$

followed by the analysis step

$$\mathbf{X}^N_{0:i|i} = \mathbf{X}^N_{0:i|i-1} - \mathbf{A}\mathbf{G}^\mathrm{T}(\mathbf{G}\mathbf{G}^\mathrm{T} + \mathbf{R}_i)^{-1}\left[\mathcal{H}_i\left(\boldsymbol{x}^\ell_{i|i-1}\right) - \boldsymbol{y}_i - \boldsymbol{w}^\ell_i\right]_{\ell=1,N}, \tag{19}$$

$$w^\ell_i \sim N(0, \mathbf{R}_i)$$

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

$$\mathbf{G} = \left[\boldsymbol{g}^1_i, \ldots, \boldsymbol{g}^N_i\right], \quad \boldsymbol{g}^\ell_i = \mathcal{H}_i\left(\boldsymbol{x}^\ell_{i|i-1}\right) - \frac{1}{N}\sum_{j=1}^N \mathcal{H}_i\left(\boldsymbol{x}^j_{i|i-1}\right) \tag{20}$$

Note that for small τ, the resulting method is asymptotically equivalent to the method with the derivatives (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 problems, that is, with the linearized advance in time (Eq. 5) replaced by application of the original, nonlinear operator $\mathcal{M}_i$ (Algorithm 4, obtained by outting together Eq. (12) with Eqs. (7)–(9). We write the nonlinear EnKS as operating on the original ensemble of the states $\mathbf{X}^N = \left[\boldsymbol{x}^\ell\right]^N_{\ell=1}$ rather than on the increments $\boldsymbol{z}^\ell = \boldsymbol{\delta x}^\ell$.) In particular, the incremental 4DVAR does not converge unless it is already at a stationary point, because each iteration delivers the same result, up to the randomness of the EnKS.

NPGD
2, 865–902, 2015

Theorem 1 If $\tau = 1$, then one step of EnKS-4DVAR (Algorithm 3) becomes the nonlinear EnKS (Algorithm 4). In particular, in that case, the result of the step does not depend on the previous iterate.

Proof. Indeed, Eq. (17) becomes

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

hence, $\boldsymbol{x}_i + z^{\ell}_{i|i-1} = \mathcal{M}_i\left(\boldsymbol{x}_{i-1} + z^{\ell}_{i-1|i-1}\right) + \boldsymbol{v}^{\ell}_i$, which is exactly the same as advancing the ensemble member ℓ following Eq. (18) with $\boldsymbol{x}^{\ell}_{i-1|i-1} = \boldsymbol{x}_{i-1} + z^{\ell}_{i-1|i-1}$. Similarly, Eq. (16) becomes with $\tau = 1$,

$$\boldsymbol{a}^{\ell}_i = \frac{\mathcal{H}_i\left(\boldsymbol{x}_i + z^{\ell}_{i|i-1}\right) - \mathcal{H}_i\left(\boldsymbol{x}_i\right)}{1} - \frac{1}{N}\sum_{j=1}^{N}\frac{\mathcal{H}_i\left(\boldsymbol{x}_i + z^{j}_{i|i-1}\right) - \mathcal{H}_i\left(\boldsymbol{x}_i\right)}{1} \tag{21}$$

$$= \mathcal{H}_i\left(\boldsymbol{x}_i + z^{\ell}_{i|i-1}\right) - \frac{1}{N}\sum_{j=1}^{N}\mathcal{H}_i\left(\boldsymbol{x}_i + z^{j}_{i|i-1}\right), \tag{22}$$

which is exactly the same as Eq. (20) with $\boldsymbol{x}^{\ell}_{i|i-1} = \boldsymbol{x}_i + z^{\ell}_{i|i-1}$. Finally, the innovation term in Eq. (14) becomes using Eq. (3),

$$\mathbf{H}_i z^{\ell}_{i|i-1} - \boldsymbol{d}_i = \frac{\mathcal{H}_i\left(\boldsymbol{x}_i + z^{\ell}_{i|i-1}\right) - \mathcal{H}_i\left(\boldsymbol{x}_i\right)}{1} - \left[\boldsymbol{y}_i - \mathcal{H}_i\left(\boldsymbol{x}_i\right)\right] = \mathcal{H}_i\left(\boldsymbol{x}^{\ell}_{i|i-1}\right) - \boldsymbol{y}_i,$$

which is exactly the same as in Eq. (19). □

NPGD
2, 865–902, 2015

5 Tikhonov regularization and the Levenberg–Marquardt method

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 $\|\boldsymbol{\delta x}_i\| \le \varepsilon$ leads to globally convergent trust region methods (Gratton et al., 2013). Here, we add $\boldsymbol{\delta x}_i$ in 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) $\boldsymbol{x}_{0:k} \leftarrow \boldsymbol{x}_{0:k} + \boldsymbol{\delta x}_{0:k}$, where

$$\|\boldsymbol{\delta x}_0 - \boldsymbol{z}_\mathrm{b}\|^2_{\mathbf{B}^{-1}} + \sum_{i=1}^{k} \|\boldsymbol{\delta x}_i - \mathbf{M}_i \boldsymbol{\delta x}_{i-1} - \boldsymbol{m}_i\|^2_{\mathbf{Q}_i^{-1}} + \sum_{i=1}^{k} \|\boldsymbol{d}_i - \mathbf{H}_i \boldsymbol{\delta x}_i\|^2_{\mathbf{R}_i^{-1}} + \gamma \sum_{i=0}^{k} \|\boldsymbol{\delta x}_i\|^2_{\mathbf{S}_i^{-1}} \to \min_{\boldsymbol{\delta x}_{0:k}} \tag{23}$$

Under suitable technical assumptions, the Levenberg–Marquardt method is guaranteed to converge globally if the regularization parameter $\gamma \ge 0$ 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 terms $\gamma\|\boldsymbol{\delta x}_i\|^2_{\mathbf{S}_i^{-1}}$ in Eq. (23) as arising from additional independent observations $\delta \boldsymbol{x}_i \sim N\left(0, \gamma^{-1}\mathbf{S}_i\right)$, which can be assimilated separately, resulting in a mathematically equivalent but often more efficient two-stage method – simply run the EnKF analysis (Eq. 6) 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).

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 (Eq. 23), so all distributions

NPGD
2, 865–902, 2015

are gaussian and the equivalence of solving Eq. (23) at once and assimilating the observations sequentially is statistically exact.

6 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).

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 6.1 describes the Lorenz 63 model and presents numerical results on the convergence. Using the same model, in Sect. 6.2, 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. 6.3, 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 $\mathbf{S}_i = \mathbf{I}$.

6.1 Numerical experiments using Lorenz 63 model

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

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

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

NPGD
2, 865–902, 2015

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

To evaluate the performance of EnKS-4DVAR method (Algorithm 3), 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 $\mathrm{truth}_i = \mathcal{M}(\mathrm{truth}_{i-1})$ at each time 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.

We perform numerical experiments without model error. The initial truth is set to $\mathrm{truth}_0 = [1, 1, 1]^{\mathrm{T}}$ 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 $k = 50$ time steps (5 time units). The observation operator is defined as $\mathcal{H}_i(x, y, z) = \left(x^2, y^2, z^2\right)$. At each time i, the observations are constructed as follows: $\boldsymbol{y}_i = \mathcal{H}_i(\mathrm{truth}_i) + \boldsymbol{v}_i$, where $\boldsymbol{v}_i$ is sampled from $N(0, \mathbf{R})$ with $\mathbf{R} = \mathbf{I}_3$. Observations are taken for each time step ($i = 1, \ldots 50$). The ensemble size is fixed to $N = 100$.

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

$$\mathrm{RSE}_i^{(j)} = \sqrt{\frac{1}{n}(\mathrm{truth}_i - \boldsymbol{x}_i^{(j)})^{\mathrm{T}}(\mathrm{truth}_i - \boldsymbol{x}_i^{(j)})}, \quad j = 1, \ldots, 5, \tag{24}$$

NPGD
2, 865–902, 2015

where truth_i is the true vector state at time i, $\boldsymbol{x}_i^{(j)}$ is the jth iterate of Algorithm 3 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

$$\text{RMSE}^{(j)} = \frac{1}{k}\sum_{i=0}^{k} \text{RSE}_i^{(j)}, \quad j = 1,\ldots,5, \tag{25}$$

where k is the number of time steps.

From Table 1 and Figs. 2 and 3, it can be seen that the iterates of Algorithm 3 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.

6.2 The impact of the finite difference parameter

In this section, 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 of Algorithm 3 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}.

The mean of the objective function values are shown in Table 2. The box plots of the objective function for the first four iterations are shown in Fig. 4 and for the last four iterations are shown in Fig. 5.

These figures and table show the impact of the parameter τ on the objective function minimization along the iterations. For $\tau = 1$ (when we use the classical non linear EnKS), the results are almost the same after the first iteration, in this case performing more iterations do not improve the results. However, when $\tau \leq 10^{-1}$ the objective function is decreasing along iterations. For $\tau = 10^{-1}$, more iterations are needed to reduce the objective function significantly. In the case of small values of τ, for instance

NPGD
2, 865–902, 2015

$\tau \leq 10^{-2}$, few iterations are enough to reduce the objective function significantly. When $\tau = 10^{-3}$, the results are slightly different from the results with smaller values of τ.

As a conclusion for these experiments, it is better to choose $\tau \leq 10^{-3}$, such that the results will be less sensitive to the value of τ. Note that this value is problem dependent. In practice, to avoid divergence due to the finite difference approximations, it is better to choose τ as small as possible, and since the computers use finite-precision arithmetic, we need to be careful to the effects of computer rounding.

We can also observe that for the first iteration, the best decrease in objective function is obtained when $\tau = 1$, and the worst decrease is obtained for $\tau = 10^{-6}$. Moreover, for the first four iterations the bigger τ is, the better results will be, but for the last four iterations, the smaller τ is, the better results will be. Hence, an adaptive τ can be a better choice than a fixed τ over iterations. For instance, for these experiments we can start with $\tau = 1$, then decrease its value along iterations. Exploration of the best strategy to choose τ over iterations will be studied in the future works.

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

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. 6.3.1 and 6.3.2 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. 6.3.3.

6.3.1 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

NPGD
2, 865–902, 2015

captures an important aspect of large-scale dynamics in the atmosphere: in the horizontal direction of the atmospheric flow, the Coriolis force caused by the rotation of the Earth, and the pressure gradient force are in approximate balance.

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

$$q_1 = \nabla^2\psi_1 - \frac{f_0^2 L^2}{g' H_1}(\psi_1 - \psi_2) + \beta y, \qquad q_2 = \nabla^2\psi_2 - \frac{f_0^2 L^2}{g' H_2}(\psi_2 - \psi_1) + \beta y + R_s, \tag{26}$$

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 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. Details of the derivation of these non-dimensional equations can be found in Fandry and Leslie (1984); Pedlosky (1979).

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

$$\frac{D_i q_i}{Dt} = 0, \quad i = 1,2. \tag{27}$$

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

$$\frac{D_i}{Dt} = \frac{\partial}{\partial t} + u_i\frac{\partial}{\partial x} + v_i\frac{\partial}{\partial y} \tag{28}$$

and

$$u_i = -\frac{\partial\psi_i}{\partial y}, \qquad v_i = \frac{\partial\psi_i}{\partial x}, \tag{29}$$

NPGD
2, 865–902, 2015

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. (27). 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. (26) can be solved for the stream function at each gridpoint and then the velocity fields obtained through Eq. (29). 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, $H_1 = 6000$ m, $H_2 = 4000$ m, $f_0 = 10^{-4}\,s^{-1}$, $\beta = 1.5$. For more details on the model and its solution, we refer the reader to Fisher et al. (2011).

The domain for the experiments is 12 000 km by 6300 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 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.

6.3.2 Experimental setup

The performance of EnKS-4DVAR with regularization is analyzed by using twin experiments (Sect. 6.1).

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 $D2 = 4500$ 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 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.

NPGD
2, 865–902, 2015

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

The background error covariance matrix (**B** matrix) and the model error covariances (matrices $\mathbf{Q}_k$) used in these experiments correspond to vertical and horizontal correlations. The vertical and horizontal structures are assumed to be separable and the vertical correlation function is assumed to be constant over the grid. 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 **B**, the SD and the horizontal correlation length scale in this experiments was set to 0.8 and 10^6 m respectively. The vertical correlation function value was taken as 0.2. 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 function value was taken as 0.5.

The window length is set to 10 days, with two sub-windows of 5 days ($k = 2$). The ensemble size is chosen to be $N = 3000$.

6.3.3 Numerical results

Figure 6 shows the objective function values along iterations, when using incremental 4DVAR method. The objective function is oscillating along the iterations, therefore incremental 4DVAR method without regularization is diverging.

Figures 7 and 8 show the objective function values along iterations for eight different choices of regularization parameter γ. It can be seen from these two figures that when $\gamma = 0$, Algorithm 3 is 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 Algorithm 3 is 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 $\gamma \geq 10$),

NPGD
2, 865–902, 2015

we can observe the decrease on the objective function along iterations. Moreover, the fastest decrease on the objective function is obtained for $\gamma = 10$.

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). 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.

7 Conclusions

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 (Algorithm 3) 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, bit 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 a 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

NPGD
2, 865–902, 2015

equivalent to those of obtained from the standard nonlinear EnKS. Based on computational experiments, we have concluded that in the first iteration, it is better to use the classical ensemble Kalman smoother as proposed in Evensen and van Leeuwen (2000) (i.e., $\tau = 1$, Algorithm 4 here), and then to decrease τ over the iterations.

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 standard model problems for atmospheric circulation. We have observed that the incremental 4DVAR method is not converging for a long time 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 nothing to prevent the use of more sophisticated variances of EnKS with localization and the covariance inflation, and square root filters instead of EnKS with data perturbation. 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 Mandel was visiting INP-ENSEEIHT and CERFACS, and when Bergou and Gratton were visiting the University of Colorado Denver.

NPGD
2, 865–902, 2015

References

Bell, B.: The iterated Kalman smoother as a Gauss–Newton method, SIAM J. Optimiz., 4, 626–636, doi:10.1137/0804035, 1994. 867, 871

Bergou, E., Gratton, S., and Mandel, J.: On the convergence of a non-linear ensemble Kalman smoother, arXiv:1411.4608, submitted, 2014. 868, 877

Bocquet, M. and Sakov, P.: Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems, Nonlin. Processes Geophys., 19, 383–399, doi:10.5194/npg-19-383-2012, 2012. 868, 869

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Q. J. Roy. Meteor. Soc., 140, 1521–1535, doi:10.1002/qj.2236, 2014. 869

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, J. Marine Syst., 40–41, 253–289, doi:10.1016/S0924-7963(03)00021-6, 2003. 874

Butala, M. D.: A localized ensemble Kalman smoother, in: 2012 IEEE Statistical Signal Processing Workshop (SSP), August 5-8, 2012, Ann Arbor, MI, USA, 21–24, IEEE, doi:10.1109/SSP.2012.6319665, 2012. 869

Chen, Y. and Snyder, C.: Assimilating vortex position with an ensemble Kalman filter, Mon. Weather Rev., 135, 1828–1845, doi:10.1175/MWR3351.1, 2007. 873

Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, Q. J. Roy. Meteor. Soc., 120, 1367–1387, doi:10.1002/qj.49712051912, 1994. 866, 867, 871

Developmental Testbed Center: NOAA Ensemble Kalman Filter Beta Release v1.0, available at: http://www.dtcenter.org/com-GSI/users/docs (last access: March 2015), 2015. 869

Dietrich, C. R. and Newsam, G. N.: Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix, SIAM J. Sci. Comput., 18, 1088–1107, 1997. 886

Evensen, G.: Data Assimilation: the Ensemble Kalman Filter, 2nd edn., Springer, Berlin, 307 pp., doi:10.1007/978-3-642-03711-5, 2009. 867, 868, 872

Evensen, G. and van Leeuwen, P. J.: An ensemble Kalman smoother for nonlinear dynamics, Mon. Weather Rev., 128, 1852–1867, doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2, 2000. 888

NPGD
2, 865–902, 2015

Fandry, C. and Leslie, L.: A two-layer quasi-geostrophic model of summer trough formation in the Australian subtropical easterlies, J. Atmos. Sci., 41, 807–817, 1984. 880, 884
Fisher, M., Leutbecher, M., and Kelly, G. A.: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation, Q. J. Roy. Meteor. Soc., 131, 3235–3246, doi:10.1256/qj.04.142, 2005. 867, 871
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, ECMWF Technical Memorandum 655, Reading, UK, 2011. 885
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. 869
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. 879
Gratton, S., Gürol, S., and Toint, P.: Preconditioning and globalizing conjugate gradients in dual space for quadratically penalized nonlinear-least squares problems, Comput. Optim. Appl., 54, 1–25, doi:10.1007/s10589-012-9478-7, 2013. 879
Hamill, T. M. and Snyder, C.: A hybrid ensemble Kalman filter–3D variational analysis scheme, Mon. Weather Rev., 128, 2905–2919, doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2, 2000. 868
Johns, C. J. and Mandel, J.: A two-stage ensemble Kalman filter for smooth data assimilation, Environ. Ecol. Stat., 15, 101–110, doi:10.1007/s10651-007-0033-0, 2008. 868, 879
Kasanický, I., Mandel, J., and Vejmelka, M.: Spectral diagonal ensemble Kalman filters, Nonlin. Processes Geophys. Discuss., 2, 115–143, doi:10.5194/npgd-2-115-2015, 2015. 869
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. 868
Levenberg, K.: A method for the solution of certain non-linear problems in least squares, Q. Appl. Math., 2, 164–168, 1944. 879
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, Mon. Weather Rev., 141, 2721–2739, doi:10.1175/MWR-D-12-00130.1, 2013. 869

NPGD
2, 865–902, 2015

Liu, C., Xiao, Q., and Wang, B.: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test, Mon. Weather Rev., 136, 3363–3373, doi:10.1175/2008MWR2312.1, 2008. 869

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), Mon. Weather Rev., 137, 1687–1704, doi:10.1175/2008MWR2699.1, 2009. 869

Lorenz, E. N.: Deterministic nonperiodic flow, J. Atmos. Sci., 20, 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2, 1963. 880

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 Contr. Syst. Mag., 29, 47–65, doi:10.1109/MCS.2009.932224, 2009. 873, 879

Mandel, J., Bergou, E., and Gratton, S.: 4D-VAR by ensemble Kalman smoother, arxiv:1304.5271, 2013. 869

Marquardt, D. W.: An algorithm for least-squares estimation of nonlinear parameters, J. Soc. Ind. Appl. Math., 11, 431–441, doi:10.1137/0111030, 1963. 879

Nowak, W., Tenkleve, S., and Cirpka, O.: Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities, Math. Geol., 35, 53–66, 2003. 886

Osborne, M. R.: Nonlinear least squares – the Levenberg algorithm revisited, J. Aust. Math. Soc. B, 19, 343–357, doi:10.1017/S033427000000120X, 1976. 879

Pedlosky, J.: Geophysical Fluid Dynamics, Springer, New York Heidelberg Berlin, 715 pp., 1979. 884

Rauch, H. E., Tung, F., and Striebel, C. T.: Maximum likelihood estimates of linear dynamic systems, AIAA J., 3, 1445–1450, 1965. 867

Sakov, P. and Bertino, L.: Relation between two common localisation methods for the EnKF, Computat. Geosci., 10, 225–237, doi:10.1007/s10596-010-9202-6, 2011. 869

Sakov, P., Oliver, D. S., and Bertino, L.: An iterative EnKF for strongly nonlinear systems, Mon. Weather Rev., 140, 1988–2004, doi:10.1175/MWR-D-11-00176.1, 2012. 868, 869

Strang, G. and Borre, K.: Linear Algebra, Geodesy, and GPS, Wellesley-Cambridge Press, Wellesley, MA, 490 pp., 1997. 867

NPGD
2, 865–902, 2015

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, J. Am. Stat. Assoc., 105, 978–990, doi:10.1198/jasa.2010.ap07636, 2010. 867

Trémolet, Y.: Model-error estimation in 4D-Var, Q. J. Roy. Meteor. Soc., 133, 1267–1280, doi:10.1002/qj.94, 2007. 866, 870

Trémolet, Y.: Object-Oriented Prediction System, available at: http://www.data-assimilation.net/Events/Year3/OOPS.pdf (last access: 15 May 2015), 2013. 883

Tshimanga, J., Gratton, S., Weaver, A. T., and Sartenaer, A.: Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation, Q. J. Roy. Meteor. Soc., 134, 751–769, doi:10.1002/qj.228, 2008. 867, 871

Wang, X.: Incorporating ensemble covariance in the gridpoint statistical interpolation variational minimization: a mathematical framework, Mon. Weather Rev., 138, 2990–2995, doi:10.1175/2010MWR3245.1, 2010. 868

Wright, S. J. and Holt, J. N.: An inexact Levenberg–Marquardt method for large sparse nonlinear least squares, J. Aust. Math. Soc. B, 26, 387–403, doi:10.1017/S0334270000004604, 1985. 879

Zhang, F., Zhang, M., and Hansen, J.: Coupling ensemble Kalman filter with four-dimensional variational data assimilation, Adv. Atmos. Sci., 26, 1–8, doi:10.1007/s00376-009-0001-8, 2009. 869

Zupanski, M.: Maximum likelihood ensemble filter: theoretical aspects, Mon. Weather Rev., 133, 1710–1726, doi:10.1175/MWR2946.1, 2005. 868

NPGD
2, 865–902, 2015

Table 1. The root mean square error given by Eq. (25) for the first six Gauss–Newton iterations of Algorithm 3, 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

NPGD
2, 865–902, 2015

Table 2. Mean of the objective function from 30 runs of the EnKS-4DVAR algorithm (Algorithm 3) 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.	$\tau = 1$	$\tau = 10^{-1}$	$\tau = 10^{-2}$	$\tau = 10^{-3}$	$\tau = 10^{-4}$	$\tau = 10^{-5}$	$\tau = 10^{-6}$
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

NPGD
2, 865–902, 2015

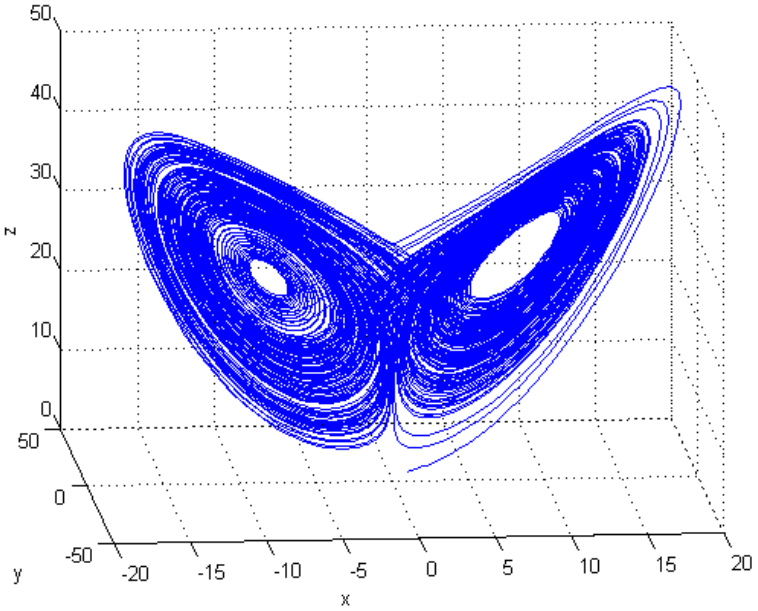


Figure 1. The Lorenz attractor, initial values $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, discretization time step is dt = 0.1 time unit.

NPGD
2, 865–902, 2015

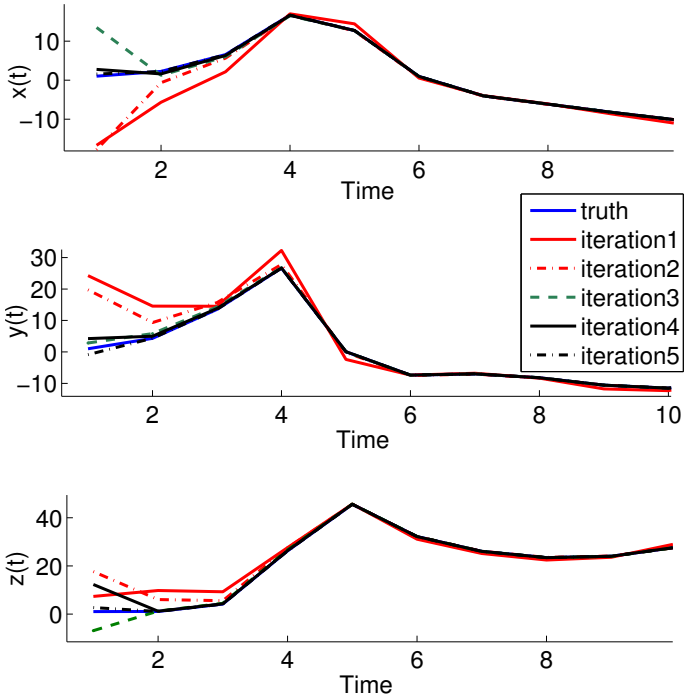


Figure 2. The three components x, y, z of the truth and the first five Gauss–Newton iterations from Lorenz 63 problem, for the first 10 time steps. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$. Time step is dt $= 0.1$ time unit. Observations are the full state at each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

NPGD
2, 865–902, 2015

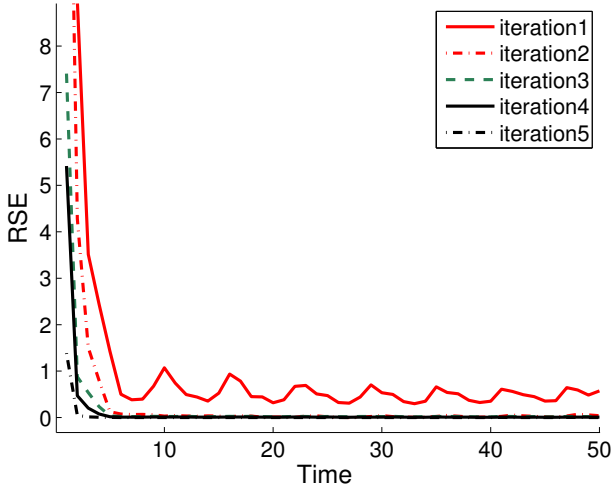


Figure 3. Root square error given by Eq. (24) for the first five Gauss–Newton iterations from Lorenz 63 problem. The problem setting is the same as in Fig. 2.

NPGD
2, 865–902, 2015

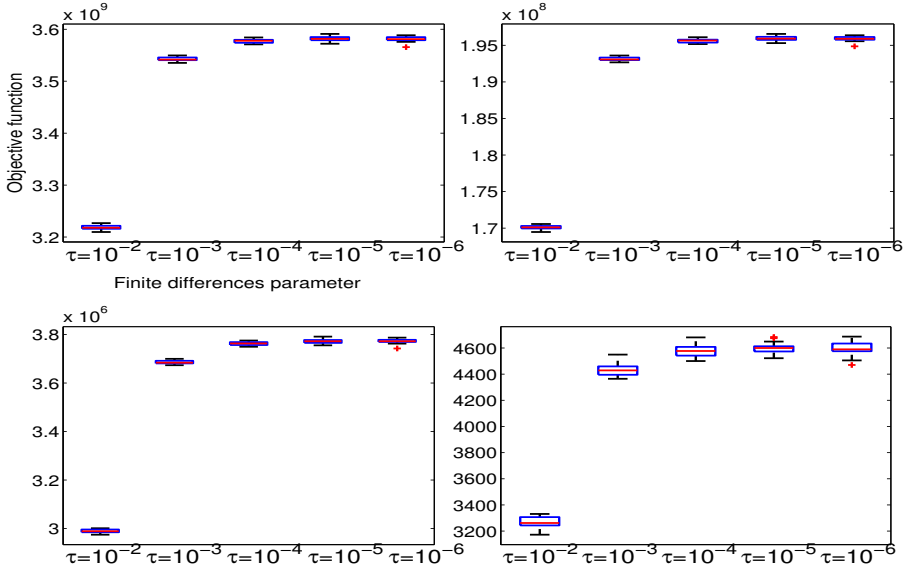


Figure 4. Box plots of objective function values for Lorenz 63 problem. From the left to the right and from the top to the bottom, the figures correspond to the results of the first, the second, the third and the fourth iteration respectively. The whole state is observed. Ensemble size is 50. The time 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).

NPGD
2, 865–902, 2015

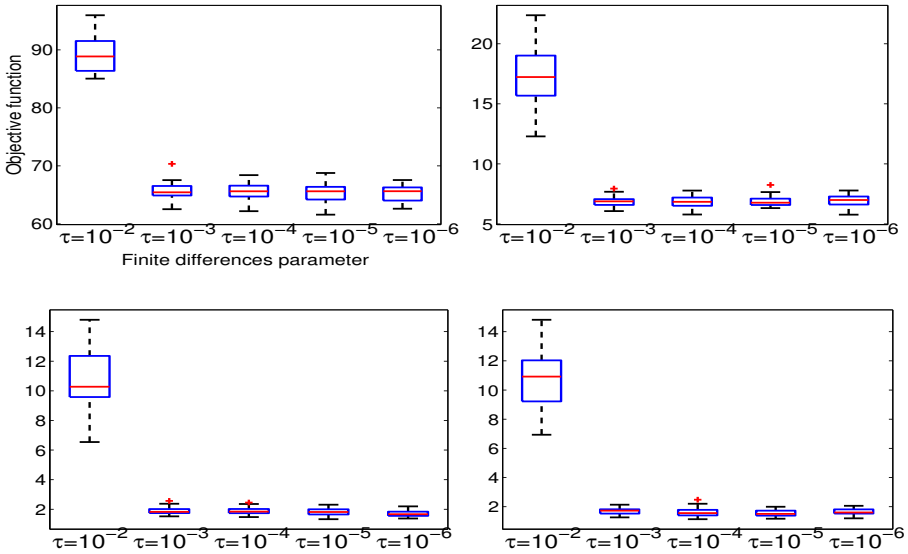
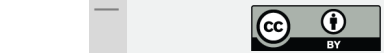


Figure 5. Same as Fig. 4, but for the fifth, the sixth, the seventh and the eighth iteration respectively.

NPGD
2, 865–902, 2015



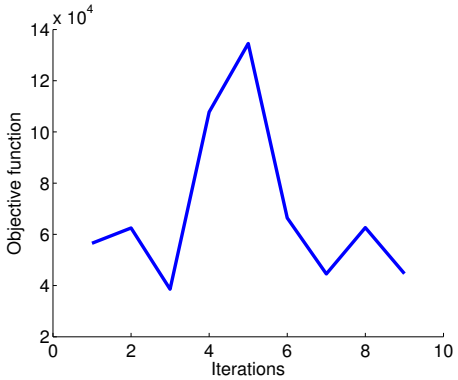


Figure 6. Objective function values along incremental 4DVAR iterations, for two-level quasi-geostrophic problem from Sect. 6.3.2.

NPGD
2, 865–902, 2015

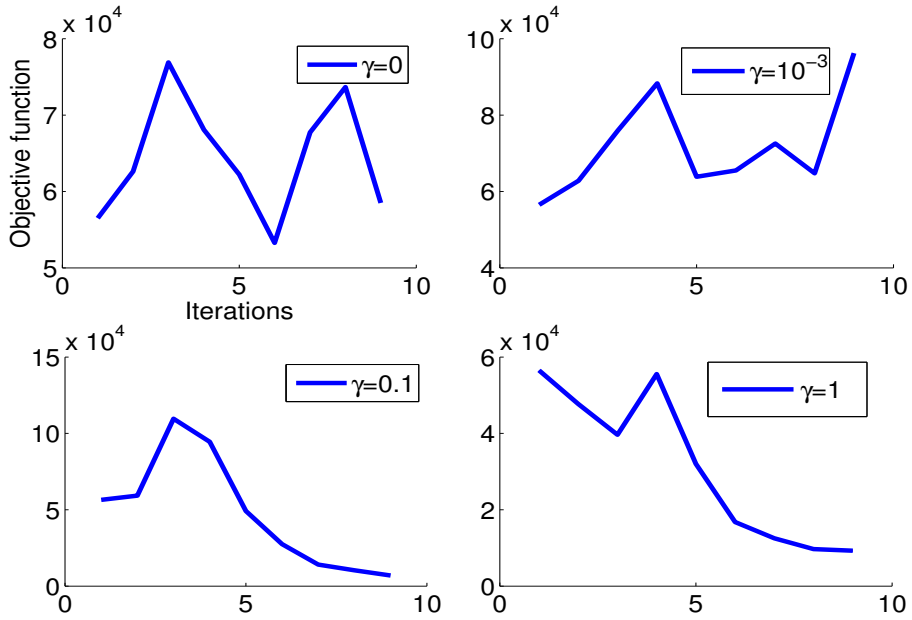


Figure 7. Objective function values along EnKS-4DVAR with regularization iterations for two-level quasi-geostrophic problem (Sect. 6.3.2). From the left to the right and from the top to the bottom: $\gamma = 0$, $\gamma = 0.001$, $\gamma = 0.1$, $\gamma = 1$.

NPGD
2, 865–902, 2015

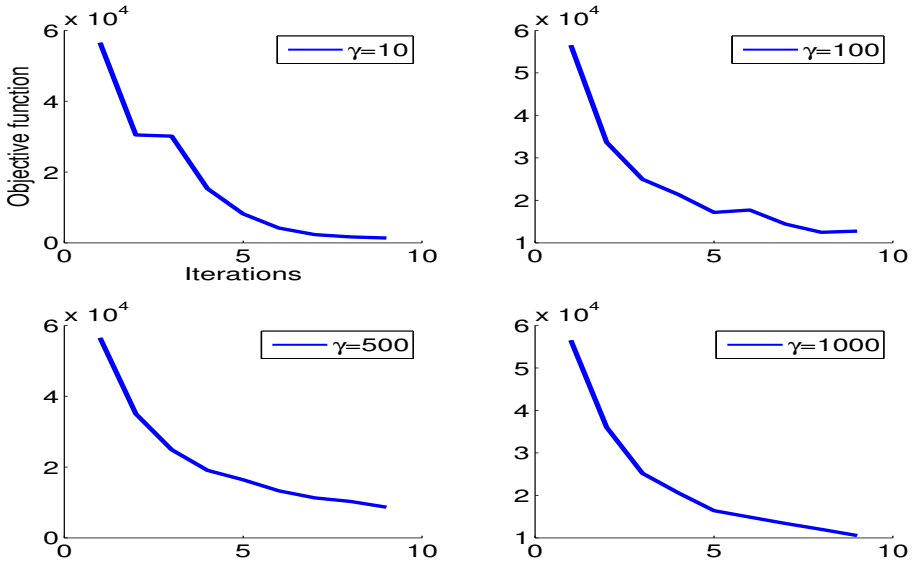


Figure 8. As Fig. 7, but for γ = 10, γ = 100, γ = 500, γ = 1000, respectively.

NPGD
2, 865–902, 2015