Journal of Zhejiang University-SCIENCE C (Computers & Electronics)
ISSN 1869-1951 (Print); ISSN 1869-196X (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Notes and correspondence on ensemble-based three-dimensional variational filters*

Hong-ze LENG†, Jun-qiang SONG, Fu-kang YIN, Xiao-qun CAO
(*College of Computer, National University of Defense Technology, Changsha 410073, China*)
†E-mail: hzleng@nudt.edu.cn

Received Jan. 20, 2013; Revision accepted Mar. 18, 2013; Crosschecked July 12, 2013

Abstract: Several ensemble-based three-dimensional variational (3D-Var) filters are compared. These schemes replace the static background error covariance of the traditional 3D-Var with the ensemble forecast error covariance, but generate analysis ensemble anomalies (perturbations) in different ways. However, it is demonstrated in this paper that they are all theoretically equivalent to the ensemble transformation Kalman filter (ETKF). Furthermore, a new method named EnPSAS is presented. The analysis shows that EnPSAS has a small condition number and can apply covariance localization more easily than other ensemble-based 3D-Var methods.

Key words: 3D-Var, Ensemble Kalman filter (EnKF), Ensemble transformation Kalman filter (ETKF), Physical space analysis system (PSAS), Ensemble data assimilation
doi:10.1631/jzus.C1300024 **Document code:** A **CLC number:** TP302.7; P409

1 Introduction

The three-dimensional variational (3D-Var) method (Parrish and Derber, 1992; Gustafsson *et al*., 2001; Haben *et al*., 2011a) is one of the most significant schemes in data assimilation (DA) for a dynamic system, and it seeks an optimal solution by minimizing a cost function

$$J(\boldsymbol{x})=\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{b}})^{\mathrm{T}}\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{b}})+\frac{1}{2}(\boldsymbol{y}^{\mathrm{o}}-h(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{y}^{\mathrm{o}}-h(\boldsymbol{x})), \tag{1}$$

where $\boldsymbol{x}^{\mathrm{b}}$ is the background state vector, $\boldsymbol{y}^{\mathrm{o}}$ is the observation vector, $h(\cdot)$ denotes the observation operator, and $\boldsymbol{B}$ and $\boldsymbol{R}$ are background and observation error covariances, respectively. However, there are some major shortcomings of 3D-Var: (1) The background error covariance is static, i.e., not flow-dependent, without dynamic information from new observations; (2) The condition number of the Hessian matrix is generally large, which yields slow convergence and an inaccurate solution (Haben *et al*., 2011b). Therefore, preconditioning technology and ensemble-based covariance are introduced.

Hamill and Snyder (2000) proposed a hybrid method that replaces the static background error covariance with a weighted sum of the original covariance and a sampling ensemble covariance. Lorenc (2003) presented another hybrid scheme by setting two sets of control variables, one of which was preconditioned upon the square root of the ensemble covariance. A similar hybrid framework was proposed by Buehner (2005), in which the ensemble covariance was incorporated into the 3D-Var system. Wang *et al*. (2007) pointed out that these hybrid ensemble-3D-Var schemes are theoretically equivalent. By combing the full-rank static covariance and the flow-dependent ensemble covariance, the schemes can obtain a better estimate than 3D-Var. However,

* Project supported by the National Natural Science Foundation of China (No. 41105063) and the Special Fund for Meteorological Scientific Research in the Public Interest (No. GYHY20100615)

these systems employ two separate DA algorithms, variational and ensemble, which will introduce large computation.

There are also some other methods (Zupanski, 2005; Liu *et al*., 2008; Tian *et al*., 2011; Pan *et al*., 2012), which replace static background error covariance only with ensemble forecast covariance, rather than the combination of static covariance and ensemble one, and we call them ensemble-based 3D-Var. It is easy to execute these methods in DA systems, and the optimal solution can be obtained in the ensemble perturbation space or the observation space, rather than the state space. Therefore, the computational cost of minimization can be significantly reduced.

It is known that state- and observation-based variational methods are similar to Kalman filters without ensemble approximations (Jazwinski, 1970; van Leeuwen and Evensen, 1996). We here want to prove that after introducing ensemble approximation, several presented ensemble-based 3D-Var methods are also equivalent to the ensemble transformation Kalman filter (ETKF). Another interesting issue with the ensemble-based methods is that nonlinear observations can be taken into account, either in an ad-hoc manner or more natural. In this paper, first, we compare three ensemble-based 3D-Var methods and show their theoretical equivalence to ETKF upon ensemble mean analysis and updated ensemble anomalies (perturbations), in a linear framework; second, a new method called the ensemble PSAS method (EnPSAS) is proposed, which has some notable advantages; third, we present a discussion on what the nonlinear observation operator does in the different methods.

2 Proof of equivalence

2.1 ETKF

ETKF, presented by Bishop *et al*. (2001), was often used as a basic and significant ensemble-based deterministic approach (Hunt *et al*., 2007; Wang *et al*., 2008; Yang *et al*., 2009; Janjić *et al*., 2011). There are two main steps in ETKF:

Step 1: Update the state vector

$$\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{b}} + \boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}})^{-1}(\boldsymbol{y}^{\mathrm{o}} - h(\boldsymbol{x}^{\mathrm{b}})), \quad (2)$$

where $\boldsymbol{x}^{\mathrm{a}}$ and $\boldsymbol{x}^{\mathrm{b}}$ denote the analysis state and the background state (the ensemble mean of forecasts here), respectively, $\boldsymbol{H}$ is the tangent linear approximation of operator $h(\cdot)$, and $\boldsymbol{P}$ indicates the ensemble-based forecast error covariance.

Step 2: Update the ensemble anomalies (perturbations)

$$(\boldsymbol{P}^{\mathrm{a}})^{1/2} = \boldsymbol{P}^{1/2}(\boldsymbol{I} + \boldsymbol{C})^{-\mathrm{T}/2}, \quad (3)$$

where $\boldsymbol{P}^{\mathrm{a}}$ is the analysis error covariance matrix, $(\cdot)^{1/2}$ indicates a square root matrix, $\boldsymbol{I}$ denotes the identity matrix, and $\boldsymbol{C}$ equals $(\boldsymbol{H}\boldsymbol{P}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{P}^{1/2}$.

In the following, we want to prove that ensemble-based 3D-Var (En3DVAR), maximum likelihood ensemble filter (MLEF), and ensemble-based 3D-Var processing in the observation space (En3DPOS) are all theoretically equivalent to ETKF, under the linearity assumption.

2.2 En3DVAR

Generally speaking, in the traditional 3D-Var, the condition number of the Hessian matrix is large, which results in slow convergence. To solve this problem, Lorenc (1997) presented a method named the preconditioning control variable method. Since $\boldsymbol{B}$ is a symmetric positive definite matrix, it can be decomposed as $\boldsymbol{B}=\boldsymbol{B}^{1/2}(\boldsymbol{B}^{1/2})^{\mathrm{T}}$. A new control variable $\boldsymbol{\zeta}$ is then introduced:

$$\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^{\mathrm{b}} = \boldsymbol{B}^{1/2}\boldsymbol{\zeta}. \quad (4)$$

Substituting Eq. (4) into Eq. (1), one can obtain a new cost function

$$J(\boldsymbol{\zeta}) = \frac{1}{2}\boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{\zeta} + \frac{1}{2}[\boldsymbol{y}^{\mathrm{o}} - h(\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{B}^{1/2}\boldsymbol{\zeta})]^{\mathrm{T}} \cdot \boldsymbol{R}^{-1}[\boldsymbol{y}^{\mathrm{o}} - h(\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{B}^{1/2}\boldsymbol{\zeta})]. \quad (5)$$

The Hessian matrix of the function is $\boldsymbol{I}+(\boldsymbol{B}^{1/2})^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{B}^{1/2}$, which has a relatively small condition number (Haben *et al*., 2011b). For the definition of the condition number, refer to Appendix A.

However, there is still a shortcoming of not using a flow-dependent background error covariance as in 3D-Var. Liu *et al*. (2008) extended Lorenc's method and replaced the background error covariance $\boldsymbol{B}$ with an ensemble-based covariance $\boldsymbol{P}$ in the so-called En4DVAR. The analysis solution is calculated in the ensemble perturbation space, rather than the state space, which greatly reduces the computation cost. En3DVAR is a reduced version from the original En4DVAR. We now give proof of theoretical equivalence between En3DVAR and ETKF.

In En3DVAR, the forecast error covariance $\boldsymbol{P}$ is defined as

$$\boldsymbol{P}=\boldsymbol{P}^{1/2}(\boldsymbol{P}^{1/2})^{\mathrm{T}}, \tag{6}$$

where

$$\boldsymbol{P}^{1/2}=\frac{1}{\sqrt{I-1}}(\boldsymbol{x}_1-\boldsymbol{x}^{\mathrm{b}},\boldsymbol{x}_2-\boldsymbol{x}^{\mathrm{b}},\cdots,\boldsymbol{x}_I-\boldsymbol{x}^{\mathrm{b}}), \tag{7}$$

I indicates the number of ensemble members. Here, we call $\boldsymbol{P}^{1/2}$ the forecast ensemble anomalies (perturbations). Like in Eq. (4), control variable $\boldsymbol{\zeta}$ is introduced:

$$\delta\boldsymbol{x}=\boldsymbol{x}-\boldsymbol{x}^{\mathrm{b}}=\boldsymbol{P}^{1/2}\boldsymbol{\zeta}. \tag{8}$$

Substituting Eq. (8) into Eq. (1), and using a linear approximation

$$\begin{aligned}\boldsymbol{y}^{\mathrm{o}}-h(\boldsymbol{x}^{\mathrm{b}}+\boldsymbol{P}^{1/2}\boldsymbol{\zeta})&\approx\boldsymbol{y}^{\mathrm{o}}-h(\boldsymbol{x}^{\mathrm{b}})-\boldsymbol{HP}^{1/2}\boldsymbol{\zeta}\\&=\delta\boldsymbol{y}^{\mathrm{o}}-\boldsymbol{HP}^{1/2}\boldsymbol{\zeta},\end{aligned} \tag{9}$$

where $\delta\boldsymbol{y}^{\mathrm{o}}=\boldsymbol{y}^{\mathrm{o}}-h(\boldsymbol{x}^{\mathrm{b}})$, one can obtain a new cost function

$$\begin{aligned}J(\boldsymbol{\zeta})&=\frac{1}{2}(\boldsymbol{P}^{1/2}\boldsymbol{\zeta})^{\mathrm{T}}\boldsymbol{P}^{-1}(\boldsymbol{P}^{1/2}\boldsymbol{\zeta})\\&\quad+\frac{1}{2}(\delta\boldsymbol{y}^{\mathrm{o}}-\boldsymbol{HP}^{1/2}\boldsymbol{\zeta})^{\mathrm{T}}\boldsymbol{R}^{-1}(\delta\boldsymbol{y}^{\mathrm{o}}-\boldsymbol{HP}^{1/2}\boldsymbol{\zeta})\\&=\frac{1}{2}\boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{\zeta}+\frac{1}{2}(\delta\boldsymbol{y}^{\mathrm{o}}-\boldsymbol{HP}^{1/2}\boldsymbol{\zeta})^{\mathrm{T}}\boldsymbol{R}^{-1}(\delta\boldsymbol{y}^{\mathrm{o}}-\boldsymbol{HP}^{1/2}\boldsymbol{\zeta}),\end{aligned} \tag{10}$$

and its gradient

$$\begin{aligned}\nabla J&=\boldsymbol{\zeta}-(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}(\delta\boldsymbol{y}^{\mathrm{o}}-\boldsymbol{HP}^{1/2}\boldsymbol{\zeta})\\&=[\boldsymbol{I}+(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{HP}^{1/2}]\boldsymbol{\zeta}-(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}}\\&=(\boldsymbol{I}+\boldsymbol{C})\boldsymbol{\zeta}-(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}},\end{aligned} \tag{11}$$

where

$$\boldsymbol{C}=(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{HP}^{1/2}, \tag{12}$$

$$\begin{aligned}\boldsymbol{HP}^{1/2}\approx[&h(\boldsymbol{x}_1)-h(\boldsymbol{x}^{\mathrm{b}}),h(\boldsymbol{x}_2)-h(\boldsymbol{x}^{\mathrm{b}}),\ldots,\\&h(\boldsymbol{x}_I)-h(\boldsymbol{x}^{\mathrm{b}})].\end{aligned} \tag{13}$$

Although linear approximation is made in Eq. (9), there is no need to use explicit $\boldsymbol{H}$, due to the use of Eq. (13). The Hessian matrix is $\boldsymbol{I}+\boldsymbol{C}$, according to Eq. (11).

Given the solution $\boldsymbol{\zeta}^{\mathrm{a}}$ of function (10), the analysis increment can be calculated as follows:

$$\begin{aligned}\delta\boldsymbol{x}^{\mathrm{a}}&=\boldsymbol{x}^{\mathrm{a}}-\boldsymbol{x}^{\mathrm{b}}=\boldsymbol{P}^{1/2}\boldsymbol{\zeta}^{\mathrm{a}}\\&=\boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-1}(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}}\\&=\boldsymbol{P}^{1/2}[\boldsymbol{P}^{\mathrm{T}/2}(\boldsymbol{P}^{-1}+\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})\boldsymbol{P}^{1/2}]^{-1}\boldsymbol{P}^{\mathrm{T}/2}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}}\\&=(\boldsymbol{P}^{-1}+\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}}\\&=\boldsymbol{PH}^{\mathrm{T}}(\boldsymbol{R}+\boldsymbol{HPH}^{\mathrm{T}})^{-1}\delta\boldsymbol{y}^{\mathrm{o}}.\end{aligned} \tag{14}$$

The last identity can be derived using Sherman-Morrison-Woodbury formulation (Golub and van Loan, 1996).

Denote the true state by $\boldsymbol{x}^{\mathrm{t}}$, and calculate the analysis ensemble anomalies as follows:

$$\boldsymbol{x}^{\mathrm{t}}-\boldsymbol{x}^{\mathrm{a}}=(\boldsymbol{x}^{\mathrm{t}}-\boldsymbol{x}^{\mathrm{b}})-(\boldsymbol{x}^{\mathrm{a}}-\boldsymbol{x}^{\mathrm{b}})=\boldsymbol{P}^{1/2}(\boldsymbol{\zeta}^{\mathrm{t}}-\boldsymbol{\zeta}^{\mathrm{a}}), \tag{15}$$

$$\begin{aligned}\boldsymbol{P}^{\mathrm{a}}&=\left\langle(\boldsymbol{x}^{\mathrm{t}}-\boldsymbol{x}^{\mathrm{a}})(\boldsymbol{x}^{\mathrm{t}}-\boldsymbol{x}^{\mathrm{a}})^{\mathrm{T}}\right\rangle\\&=\left\langle\boldsymbol{P}^{1/2}(\boldsymbol{\zeta}^{\mathrm{t}}-\boldsymbol{\zeta}^{\mathrm{a}})(\boldsymbol{\zeta}^{\mathrm{t}}-\boldsymbol{\zeta}^{\mathrm{a}})^{\mathrm{T}}(\boldsymbol{P}^{1/2})^{\mathrm{T}}\right\rangle\\&=\boldsymbol{P}^{1/2}\boldsymbol{P}^{\zeta,\mathrm{a}}(\boldsymbol{P}^{1/2})^{\mathrm{T}},\end{aligned} \tag{16}$$

$$(\boldsymbol{P}^{\mathrm{a}})^{1/2}=\boldsymbol{P}^{1/2}(\boldsymbol{P}^{\zeta,\mathrm{a}})^{1/2}=\boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}, \tag{17}$$

where $\boldsymbol{P}^{\zeta,\mathrm{a}}$ denotes the analysis error covariance with respect to $\boldsymbol{\zeta}$, and it equals the inversion of the Hessian matrix of cost function (10).

One can see that both the analysis state and the analysis ensemble anomalies have the same forms as those in ETKF.

2.3 MLEF

By maximizing the likelihood of the posterior probability distribution, MLEF obtains its analysis solution from minimization of a cost function (Zupanski, 2005; Zupanski *et al*., 2008). MLEF is also one of the deterministic ensemble filters, since it makes use of non-perturbed observations.

In MLEF, a new control variable $\boldsymbol{\zeta}$ is defined as follows:

$$\delta \boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^{\mathrm{b}} = \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta}, \quad (18)$$

where $\boldsymbol{C}$ is the same as that in Eq. (12).

By substituting for $\boldsymbol{x}$ in Eq. (1) using $\boldsymbol{x}^{\mathrm{b}}+\boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta}$, one can obtain a different cost function:

$$\begin{aligned} J(\boldsymbol{\zeta}) &= \frac{1}{2}[\boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta}]^{\mathrm{T}}\boldsymbol{P}^{-1}[\boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta}] \\ &\quad + \frac{1}{2}[\boldsymbol{y}^{\mathrm{o}} - h(\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta})]^{\mathrm{T}}\boldsymbol{R}^{-1} \\ &\quad \cdot[\boldsymbol{y}^{\mathrm{o}} - h(\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta})] \\ &= \frac{1}{2}\boldsymbol{\zeta}^{\mathrm{T}}(\boldsymbol{I}+\boldsymbol{C})^{-1}\boldsymbol{\zeta} \\ &\quad + \frac{1}{2}[\boldsymbol{y}^{\mathrm{o}} - h(\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta})]^{\mathrm{T}}\boldsymbol{R}^{-1} \\ &\quad \cdot[\boldsymbol{y}^{\mathrm{o}} - h(\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta})], \end{aligned} \quad (19)$$

and its gradient

$$\begin{aligned} \nabla J &= (\boldsymbol{I}+\boldsymbol{C})^{-1}\boldsymbol{\zeta} - (\boldsymbol{I}+\boldsymbol{C})^{-1/2}(\boldsymbol{H}\boldsymbol{P}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1} \\ &\quad \cdot[\boldsymbol{y}^{\mathrm{o}} - h(\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta})]. \end{aligned} \quad (20)$$

If $h(\cdot)$ is linear, the gradient can be rewritten as

$$\begin{aligned} \nabla J &= (\boldsymbol{I}+\boldsymbol{C})^{-1}\boldsymbol{\zeta} - (\boldsymbol{I}+\boldsymbol{C})^{-1/2}(\boldsymbol{H}\boldsymbol{P}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1} \\ &\quad \cdot[\delta\boldsymbol{y}^{\mathrm{o}} - \boldsymbol{H}\boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta}] \\ &= [(\boldsymbol{I}+\boldsymbol{C})^{-1} + (\boldsymbol{I}+\boldsymbol{C})^{-1/2}(\boldsymbol{H}\boldsymbol{P}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1} \\ &\quad \cdot \boldsymbol{H}\boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}]\boldsymbol{\zeta} \\ &\quad - (\boldsymbol{I}+\boldsymbol{C})^{-1/2}(\boldsymbol{H}\boldsymbol{P}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{d} \\ &= (\boldsymbol{I}+\boldsymbol{C})^{-1}[\boldsymbol{I} + (\boldsymbol{H}\boldsymbol{P}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{P}^{1/2}]\boldsymbol{\zeta} \\ &\quad - (\boldsymbol{I}+\boldsymbol{C})^{-1/2}(\boldsymbol{H}\boldsymbol{P}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}} \\ &= \boldsymbol{\zeta} - (\boldsymbol{I}+\boldsymbol{C})^{-1/2}(\boldsymbol{H}\boldsymbol{P}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}}. \end{aligned} \quad (21)$$

The Hessian matrix is simply set to be $\boldsymbol{I}$, whose condition number is 1 and much smaller than that of the traditional 3D-Var (Haben *et al*., 2011b).

With the solution $\boldsymbol{\zeta}$, one can obtain an analysis increment

$$\begin{aligned} \delta\boldsymbol{x}^{\mathrm{a}} &= \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta}^{\mathrm{a}} \\ &= \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}(\boldsymbol{I}+\boldsymbol{C})^{-1/2}(\boldsymbol{H}\boldsymbol{P}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}} \\ &= \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-1}\boldsymbol{P}^{\mathrm{T}/2}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}} \\ &= \boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{R}+\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}})^{-1}\delta\boldsymbol{y}^{\mathrm{o}}. \end{aligned} \quad (22)$$

The last identity is the same as that in Eq. (14).

Calculate the analysis error covariance and ensemble anomalies:

$$\begin{aligned} \boldsymbol{P}^{\mathrm{a}} &= \left\langle (\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{a}})(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{a}})^{\mathrm{T}} \right\rangle \\ &= \left\langle \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}(\boldsymbol{\zeta}^{\mathrm{t}} - \boldsymbol{\zeta}^{\mathrm{a}})(\boldsymbol{\zeta}^{\mathrm{t}} - \boldsymbol{\zeta}^{\mathrm{a}})^{\mathrm{T}} \right. \\ &\quad \left. \cdot(\boldsymbol{I}+\boldsymbol{C})^{-1/2}(\boldsymbol{P}^{1/2})^{\mathrm{T}} \right\rangle \\ &= \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{P}^{\boldsymbol{\zeta},\mathrm{a}}(\boldsymbol{I}+\boldsymbol{C})^{-1/2}(\boldsymbol{P}^{1/2})^{\mathrm{T}}, \end{aligned} \quad (23)$$

where $\boldsymbol{P}^{\zeta,\mathrm{a}}=\boldsymbol{I}$. Then

$$(\boldsymbol{P}^{\mathrm{a}})^{1/2} = \boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}.$$

It is shown that MLEF has the same form as the transformation used in ETKF.

Note that if we define

$$\boldsymbol{\zeta}_{\mathrm{En3DVar}} = (\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}\boldsymbol{\zeta}_{\mathrm{MLEF}}, \quad (24)$$

and substitute Eq. (24) into Eq. (10), we can obtain a transformation from En3DVAR to MLEF.

2.4 En3DPOS

If we define the state increment as shown in PSAS (Cohn *et al*., 1998),

$$\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^{\mathrm{b}} = \boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{w}, \quad (25)$$

and substitute it into Eq. (1), a new cost function and its gradient can be obtained:

$$\begin{aligned} J(\boldsymbol{w}) &= \frac{1}{2}(\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{w})^{\mathrm{T}}\boldsymbol{P}^{-1}(\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{w}) + \frac{1}{2}(\delta\boldsymbol{y}^{\mathrm{o}} - \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{w})^{\mathrm{T}} \\ &\quad \cdot \boldsymbol{R}^{-1}(\delta\boldsymbol{y}^{\mathrm{o}} - \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{w}) \\ &= \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{w} + \frac{1}{2}(\delta\boldsymbol{y}^{\mathrm{o}} - \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{w})^{\mathrm{T}} \\ &\quad \cdot \boldsymbol{R}^{-1}(\delta\boldsymbol{y}^{\mathrm{o}} - \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{w}), \end{aligned} \quad (26)$$

$$\begin{aligned}\nabla J &= \boldsymbol{HPH}^{\mathrm{T}}\boldsymbol{w} - \boldsymbol{HPH}^{\mathrm{T}}\boldsymbol{R}^{-1}(\delta\boldsymbol{y}^{\mathrm{o}} - \boldsymbol{HPH}^{\mathrm{T}}\boldsymbol{w}) \\ &= (\boldsymbol{HPH}^{\mathrm{T}} + \boldsymbol{HPH}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{HPH}^{\mathrm{T}})\boldsymbol{w} - \boldsymbol{HPH}^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}} \\ &= \boldsymbol{HPH}^{\mathrm{T}}\boldsymbol{R}^{-1}[(\boldsymbol{R} + \boldsymbol{HPH}^{\mathrm{T}})\boldsymbol{w} - \delta\boldsymbol{y}^{\mathrm{o}}].\end{aligned} \tag{27}$$

The optimal solution is

$$\boldsymbol{w}^{\mathrm{a}} = (\boldsymbol{R} + \boldsymbol{HPH}^{\mathrm{T}})^{-1}\delta\boldsymbol{y}^{\mathrm{o}}, \tag{28}$$

and the analysis increment is

$$\delta\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{PH}^{\mathrm{T}}(\boldsymbol{R} + \boldsymbol{HPH}^{\mathrm{T}})^{-1}\delta\boldsymbol{y}^{\mathrm{o}}. \tag{29}$$

The analysis anomalies can be calculated as follows:

$$\begin{aligned}\boldsymbol{P}^{\boldsymbol{w},\mathrm{a}} &= (\boldsymbol{HPH}^{\mathrm{T}} + \boldsymbol{HPH}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{HPH}^{\mathrm{T}})^{-1} \\ &= [\boldsymbol{HP}^{1/2}(\boldsymbol{I} + \boldsymbol{P}^{\mathrm{T}/2}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{HP}^{1/2})\boldsymbol{P}^{\mathrm{T}/2}\boldsymbol{H}^{\mathrm{T}}]^{-1} \\ &= [\boldsymbol{HP}^{1/2}(\boldsymbol{I} + \boldsymbol{C})(\boldsymbol{HP}^{1/2})^{\mathrm{T}}]^{-1},\end{aligned} \tag{30}$$

$$\begin{aligned}(\boldsymbol{P}^{\mathrm{a}})^{1/2} &= \boldsymbol{PH}^{\mathrm{T}}(\boldsymbol{P}^{\boldsymbol{w},\mathrm{a}})^{\mathrm{T}/2} \\ &= \boldsymbol{PH}^{\mathrm{T}}(\boldsymbol{HP}^{1/2})^{-\mathrm{T}}(\boldsymbol{I} + \boldsymbol{C})^{-\mathrm{T}/2} \\ &= \boldsymbol{P}^{1/2}(\boldsymbol{I} + \boldsymbol{C})^{-\mathrm{T}/2}.\end{aligned} \tag{31}$$

One can see that this method is also equivalent to ETKF.

Note that this method solves the variational problem in the ensemble-based observation space, not in the ensemble perturbation space.

3 Ensemble PSAS

PSAS is another precondition method of 3D-Var, which defines the same $\delta\boldsymbol{x}$ as that in Eq. (25). However, PSAS has a different cost function and a different gradient from Eqs. (26) and (27):

$$J(\boldsymbol{w}) = \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}(\boldsymbol{R} + \boldsymbol{HPH}^{\mathrm{T}})\boldsymbol{w} - \boldsymbol{w}^{\mathrm{T}}\delta\boldsymbol{y}^{\mathrm{o}}, \tag{32}$$

$$\nabla J = (\boldsymbol{R} + \boldsymbol{HPH}^{\mathrm{T}})\boldsymbol{w} - \delta\boldsymbol{y}^{\mathrm{o}}. \tag{33}$$

In the following, we will introduce a new ensemble-based 3D-Var filter deriving from PSAS.

Define

$$\boldsymbol{w} = \boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-\mathrm{T}/2}\boldsymbol{\theta}, \tag{34}$$

where $\boldsymbol{D}=\boldsymbol{R}^{-1/2}\boldsymbol{HPH}^{\mathrm{T}}\boldsymbol{R}^{-\mathrm{T}/2}$,

$$\delta\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{a}} - \boldsymbol{x}^{\mathrm{b}} = \boldsymbol{PH}^{\mathrm{T}}\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-\mathrm{T}/2}\boldsymbol{\theta}. \tag{35}$$

Using Eq. (34) to substitute for $\boldsymbol{w}$ in Eq. (32), we can obtain a new cost function

$$\begin{aligned}J(\boldsymbol{\theta}) &= \frac{1}{2}[\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-\mathrm{T}/2}\boldsymbol{\theta}]^{\mathrm{T}}(\boldsymbol{R} + \boldsymbol{HPH}^{\mathrm{T}}) \\ &\quad\cdot[\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-\mathrm{T}/2}\boldsymbol{\theta}] - [\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-\mathrm{T}/2}\boldsymbol{\theta}]^{\mathrm{T}}\delta\boldsymbol{y}^{\mathrm{o}} \\ &= \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta} - \boldsymbol{\theta}^{\mathrm{T}}(\boldsymbol{I} + \boldsymbol{D})^{-1/2}\boldsymbol{R}^{-1/2}\delta\boldsymbol{y}^{\mathrm{o}},\end{aligned} \tag{36}$$

and its gradient

$$\nabla J = \boldsymbol{\theta} - (\boldsymbol{I} + \boldsymbol{D})^{-1/2}\boldsymbol{R}^{-1/2}\delta\boldsymbol{y}^{\mathrm{o}}. \tag{37}$$

From Eqs. (36) and (37), one can see that the Hessian matrix of the cost function is simply set to be $\boldsymbol{I}$, as in MLEF. Accurate solution $\boldsymbol{\theta}^{\mathrm{a}}$ can be obtained by minimizing Eq. (36) with anticipated quick convergence. The analysis increment is

$$\begin{aligned}\delta\boldsymbol{x}^{\mathrm{a}} &= \boldsymbol{x}^{\mathrm{a}} - \boldsymbol{x}^{\mathrm{b}} = \boldsymbol{PH}^{\mathrm{T}}\boldsymbol{w}^{\mathrm{a}} \\ &= \boldsymbol{PH}^{\mathrm{T}}\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-\mathrm{T}/2}\boldsymbol{\theta}^{\mathrm{a}} \\ &= \boldsymbol{PH}^{\mathrm{T}}\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-1/2}\boldsymbol{R}^{-1/2}\delta\boldsymbol{y}^{\mathrm{o}} \\ &= \boldsymbol{PH}^{\mathrm{T}}[\boldsymbol{R}^{1/2}(\boldsymbol{I} + \boldsymbol{D})\boldsymbol{R}^{\mathrm{T}/2}]^{-1}\delta\boldsymbol{y}^{\mathrm{o}} \\ &= \boldsymbol{PH}^{\mathrm{T}}(\boldsymbol{R} + \boldsymbol{HPH}^{\mathrm{T}})^{-1}\delta\boldsymbol{y}^{\mathrm{o}}.\end{aligned} \tag{38}$$

Analogous to the calculations in Eqs. (15)–(17), the analysis anomalies can be obtained:

$$\begin{aligned}(\boldsymbol{P}^{\mathrm{a}})^{1/2} &= \boldsymbol{PH}^{\mathrm{T}}\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-\mathrm{T}/2}\boldsymbol{P}^{\boldsymbol{\theta},\mathrm{a}} \\ &= \boldsymbol{PH}^{\mathrm{T}}\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-\mathrm{T}/2} \\ &= \boldsymbol{P}^{1/2}(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I} + \boldsymbol{D})^{-\mathrm{T}/2},\end{aligned} \tag{39}$$

where the analysis error covariance $\boldsymbol{P}^{\boldsymbol{\theta},\mathrm{a}}$ with respect to $\boldsymbol{\theta}$ equals the identity matrix. We call this new method EnPSAS.

The optimal analysis $\boldsymbol{x}^{\mathrm{a}}$ is equivalent to that in ETKF; however, the analysis anomalies are much different from those in ETKF. In our method, the transformation matrix is an $I\times M$ matrix, whereas the size of transformation in ETKF is $I\times I$, in which M is the dimension of the observation vector.

4 Discussions

When ensemble-based variational methods are employed, nonlinear observations can be taken into account in a more actual way, rather than ad-hoc as in ETKF.

In MLEF, there is no linearization assumption of the observation operator; that is, the full nonlinear operator is used in the whole algorithm. In En3DVAR, the nonlinear observation operator is linearized near $\boldsymbol{x}^{\mathrm{b}}$, as shown in Eq. (9); however, when calculating $\boldsymbol{HP}^{1/2}$, the full nonlinear operator is used, according to Eq. (13). Similarly, in En3DPOS, the nonlinear operator employs a linear approximation when calculating $h(\boldsymbol{x})$, but one can use the full nonlinear operator to compute $\boldsymbol{HP}^{1/2}$. In EnPSAS, the cost function is derived under the assumption of linearity; however, it also uses the full nonlinear operator when computing $\boldsymbol{HP}^{1/2}$. By updating the background state and the background error covariance, one can obtain a more accurate solution from iterations of minimizing the cost function.

Note that if a nonlinear $h(\cdot)$ is used, the solution can be close to the actual minimum of the cost function, while the updated ensemble-based covariance represents only the inverse of the Hessian, not the posterior covariance. This is because the true posterior probability density function (PDF) is not Gaussian.

Spatial localization is another important issue in ensemble-based filters. In EnPSAS, $\boldsymbol{HPH}^{\mathrm{T}}$ and $\boldsymbol{PH}^{\mathrm{T}}$ can be localized easily in the following way: $\rho\circ(\boldsymbol{HPH})^{\mathrm{T}}$, $\rho\circ(\boldsymbol{PH})^{\mathrm{T}}$, where $\rho\circ$ denotes a Schur product (Hamill *et al*., 2001). The correlation coefficient could respond to the distance between observation points or between each observation point and each grid point, respectively. In contrast, the localization implementation of $(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{HP}^{1/2}$, a matrix which should be calculated in those three schemes (En3DVAR, MLEF, En3DPOS), is much different. One way to make the localization is to use a method with a truncated correlation matrix (Buehner, 2005; Liu *et al*., 2009); however, it is more expensive in computation and will introduce inaccurate correlation. Another way is to consider only the observations from a region around the location, as in Hunt *et al*. (2007) and Yang *et al*. (2009). Note that the latter scheme uses local ensemble anomalies for each grid point, and the observations are chosen within a certain distance to the grid point.

The total computation cost of EnPSAS may be higher than that of En3DVAR or MLEF (computational costs of these methods are shown in Appendix B), since its solution is obtained in the observation space, not in the ensemble perturbation space, and the transformation matrix has larger size than that in ETKF.

5 Summary

In this paper we have compared three ensemble-based 3D-Var methods. The results showed that they are theoretically equivalent to ETKF, with both the analysis state and ensemble anomalies having the same forms as those of ETKF.

Note that MLEF can be transformed from En3DVAR, but with a better Hessian matrix whose condition number is much smaller than En3DVAR's. For MLEF and En3DVAR, the minimization problem is solved in the ensemble perturbation space, whereas for En3DPOS, the analysis solution is obtained in the observation space.

In addition, we presented a new method named EnPSAS, which also minimizes the cost function in the observation space. Although EnPSAS is more expensive than MLEF, it is still worth using it in some DA systems where the dimension of the observation vector is much smaller than that of the state vector.

References

Bishop, C.H., Etherton, B.J., Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Mon. Wea. Rev.*, **129**(3):420-436. [doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2]

Buehner, M., 2005. Ensemble-derived stationary and flow-dependent background-error covariances: evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.*, **131**(607):1013-1043. [doi:10.1256/qj.04.15]

Cohn, S.E., da Silva, A., Guo, J., Sienkiewicz, M., Lamich, D., 1998. Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev.*, **126**(11):2913-2926. [doi:10.1175/1520-0493(1998)126<2913:ATEODS>2.0.CO;2]

Golub, G., van Loan, C., 1996. Matrix Computations. The Johns Hopkins University Press, London.

Gustafsson, N., Berre, L., Hornquist, S., Huang, X.Y., Lindskog, M., Navascues, B., Mogensen, K.S., Thorsteinsson, S.,

2001. Three-dimensional variational data assimilation for a limited area model. *Tellus*, **53A**:425-446.

Haben, S.A., Lawless, A.S., Nichols, N.K., 2011a. Conditioning of incremental variational data assimilation, with application to the Met Office system. *Tellus*, **63A**:782-792.

Haben, S.A., Lawless, A.S., Nichols, N.K., 2011b. Conditioning and preconditioning of the variational data assimilation problem. *Comput. Fluids*, **46**(1):252-256. [doi:10.1016/j.compfluid.2010.11.025]

Hamill, T.M., Snyder, C., 2000. A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**(8):2905-2919. [doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2]

Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**(11):2776-2790. [doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2]

Hunt, B.R., Kostelich, E.J., Szunyogh, I., 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Phys. D*, **230**(1-2):112-126. [doi:10.1016/j.physd.2006.11.008]

Janjić, T., Nerger, L., Albertella, A., Schröter, J., Skachko, S., 2011. On domain localization in ensemble-based Kalman filter algorithms. *Mon. Wea. Rev.*, **139**(7):2046-2060. [doi:10.1175/2011MWR3552.1]

Jazwinski, A.H., 1970. Stochastic Processes and Filtering Theory. Academic Press, New York and London.

Liu, C., Xiao, Q., Wang, B., 2008. An ensemble-based four-dimensional variational data assimilation scheme. Part I: technical formulation and preliminary test. *Mon. Wea. Rev.*, **136**(9):3363-3373. [doi:10.1175/2008MWR2312.1]

Liu, C., Xiao, Q., Wang, B., 2009. An ensemble-based four-dimensional variational data assimilation scheme. Part II: observing system simulation experiments with advanced research WRF (ARW). *Mon. Wea. Rev.*, **137**(5):1687-1704. [doi:10.1175/2008MWR2699.1]

Lorenc, A.C., 1997. Development of an operational variational assimilation scheme. *J. Meteorol. Soc. Jpn.*, **75**:339-346.

Lorenc, A.C., 2003. The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Q. J. R. Meteorol. Soc.*, **129**(595):3183-3203. [doi:10.1256/qj.02.132]

Pan, X., Tian, X., Li, X., Xie, Z., Shao, A., Lu, C., 2012. Assimilating Doppler radar radial velocity and reflectivity observations in the weather research and forecasting model by a proper orthogonal-decomposition-based ensemble, three-dimensional variational assimilation method. *J. Geophys. Res.*, **117**(D17113). [doi:10.1029/2012JD017684]

Parrish, D.F., Derber, J.C., 1992. The national meteorological center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**(8):1747-1763. [doi:10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2]

Tian, X., Xie, Z., Sun, Q., 2011. A POD-based ensemble four-dimensional variational assimilation method. *Tellus*, **63A**: 805-816.

van Leeuwen, P.J., Evensen, G., 1996. Data assimilation and inverse methods in terms of probabilistic formulation. *Mon. Wea. Rev.*, **124**(12):2898-2913. [doi:10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2]

Wang, X., Snyder, C., Hamill, T.M., 2007. Notes and correspondence on the theoretical equivalence of differently proposed ensemble/3D-Var hybrid analysis schemes. *Mon. Wea. Rev.*, **135**(1):222-227. [doi:10.1175/MWR3282.1]

Wang, X., Barker, D.M., Snyder, C., Hamill, T.M., 2008. A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Mon. Wea. Rev.*, **136**(12):5116-5131. [doi:10.1175/2008MWR2444.1]

Yang, S., Kalnay, E., Hunt, B., Bowler, N.E., 2009. Weight interpolation for efficient data assimilation with the local ensemble transform Kalman filter. *Q. J. R. Meteorol. Soc.*, **135**(638):251-262. [doi:10.1002/qj.353]

Zupanski, M., 2005. Maximum likelihood ensemble filter: theoretical aspects. *Mon. Wea. Rev.*, **133**(6):1710-1726. [doi:10.1175/MWR2946.1]

Zupanski, M., Navon, I.M., Zupanski, D., 2008. The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. *Q. J. R. Meteorol. Soc.*, **134**(633): 1039-1050. [doi:10.1002/qj.251]

Appendix A: Condition number

The condition number is a property of a matrix or a problem. A high condition number may result in slow convergence to the solution, even divergence.

Assuming that $\boldsymbol{A}$ is an invertible matrix, we can define the condition number

$$\kappa(\boldsymbol{A})=\|\boldsymbol{A}\|\cdot\|\boldsymbol{A}^{-1}\|,$$

where $\|\cdot\|$ denotes the norm defined in the square-summable sequence space L^2, and the condition number satisfies

$$\kappa(\boldsymbol{A})\geq 1.$$

Another equivalent definition can be written as

$$\kappa(\boldsymbol{A})=\left|\frac{\lambda_{\max}(\boldsymbol{A})}{\lambda_{\min}(\boldsymbol{A})}\right|,$$

where $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ are the maximum and minimum eigenvalues of $\boldsymbol{A}$, respectively.

Appendix B: Computational costs

With the assumption of linear systems and Gaussian error distribution, only a single iteration is needed in variational methods, and the step length is equal to unity (Zupanski, 2005). Assume $\delta\boldsymbol{y}^{\mathrm{o}}$ and $\delta\boldsymbol{x}^{\mathrm{a}}$ are inexpensive to apply. Thus, the main computational cost results from calculating ∇J and $(\boldsymbol{P}^{\mathrm{a}})^{1/2}$.

One can obtain $\boldsymbol{C}=\boldsymbol{V\Lambda V}^{\mathrm{T}}$, where $\boldsymbol{V}$ is the eigenvector matrix and $\boldsymbol{\Lambda}$ denotes the eigenvalue matrix. Thus, $(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}=\boldsymbol{V}(\boldsymbol{I}+\boldsymbol{\Lambda})^{-1/2}\boldsymbol{V}^{\mathrm{T}}$.

Similar decomposition can be made for $(\boldsymbol{I}+\boldsymbol{D})^{-\mathrm{T}/2}$.

1. En3DVAR

1) Calculate $\nabla J=(\boldsymbol{I}+\boldsymbol{C})\boldsymbol{\zeta}-(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}}$.

1.1) Form $\boldsymbol{HP}^{1/2}$. Cost: $O(MNI)$.

1.2) Compute $\boldsymbol{C}=(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{HP}^{1/2}$. Assume $\boldsymbol{R}^{-1}$ is inexpensive to apply. Cost: $O(M^2I+MI^2)$.

The cost of calculating ∇J: $O(MNI+M^2I+MI^2)$.

2) Calculate $(\boldsymbol{P}^{\mathrm{a}})^{1/2}=\boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}$.

2.1) Form $\boldsymbol{C}$. Cost: $O(MNI+M^2I+MI^2)$.

2.2) Calculate the eigenvalue decomposition of $\boldsymbol{C}$. Cost: $O(I^3)$.

2.3) Apply to $\boldsymbol{P}^{1/2}$. Cost: $O(NI^2)$.

The cost of calculating $(\boldsymbol{P}^{\mathrm{a}})^{1/2}$: $O(MNI+M^2I+MI^2+I^3+NI^2)$.

The total cost: $O(MNI+M^2I+MI^2+I^3+NI^2)$.

2. MLEF

1) Calculate $\nabla J=\boldsymbol{\zeta}-(\boldsymbol{I}+\boldsymbol{C})^{-1/2}(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\delta\boldsymbol{y}^{\mathrm{o}}$.

1.1) Form $\boldsymbol{HP}^{1/2}$. Cost: $O(MNI)$.

1.2) Form $\boldsymbol{C}=(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{HP}^{1/2}$. Cost: $O(M^2I+MI^2)$.

1.3) Calculate the eigenvalue decomposition of $\boldsymbol{C}$. Cost: $O(I^3)$.

The cost of calculating ∇J: $O(MNI+M^2I+MI^2+I^3)$.

2) Calculate $(\boldsymbol{P}^{\mathrm{a}})^{1/2}=\boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}$.

The cost of calculating $(\boldsymbol{P}^{\mathrm{a}})^{1/2}$: $O(MNI+M^2I+MI^2+I^3+NI^2)$.

The total cost: $O(MNI+M^2I+MI^2+I^3+NI^2)$.

3. En3DPOS

1) Calculate $\nabla J=\boldsymbol{HPH}^{\mathrm{T}}\boldsymbol{R}^{-1}[(\boldsymbol{R}+\boldsymbol{HPH}^{\mathrm{T}})\boldsymbol{w}-\delta\boldsymbol{y}^{\mathrm{o}}]$.

1.1) Form $\boldsymbol{HPH}^{\mathrm{T}}$. Cost: $O(MNI+M^2I)$.

1.2) Apply to $\boldsymbol{R}^{-1}$. Cost: $O(M^3)$.

The cost of calculating ∇J: $O(MNI+M^2I+M^3)$.

2) Calculate $(\boldsymbol{P}^{\mathrm{a}})^{1/2}=\boldsymbol{P}^{1/2}(\boldsymbol{I}+\boldsymbol{C})^{-\mathrm{T}/2}$.

The cost of calculating $(\boldsymbol{P}^{\mathrm{a}})^{1/2}$: $O(MNI+M^2I+MI^2+I^3+NI^2)$.

The total cost: $O(MNI+M^2I+MI^2+M^3+I^3+NI^2)$.

4. EnPSAS

1) Calculate $\nabla J=\boldsymbol{\theta}-(\boldsymbol{I}+\boldsymbol{D})^{-1/2}\boldsymbol{R}^{-1/2}\delta\boldsymbol{y}^{\mathrm{o}}$.

1.1) Form $\boldsymbol{D}=\boldsymbol{R}^{-1/2}\boldsymbol{HPH}^{\mathrm{T}}\boldsymbol{R}^{-\mathrm{T}/2}$. Cost: $O(MNI+M^2I)$.

1.2) Compute the eigenvalue decomposition of $\boldsymbol{D}$: $O(M^3)$.

1.3) Apply $(\boldsymbol{I}+\boldsymbol{D})^{-1/2}$ to $\boldsymbol{R}^{-1/2}$. Cost: $O(M^3)$.

The cost of calculating ∇J: $O(MNI+M^2I+M^3)$.

2) Calculate $(\boldsymbol{P}^{\mathrm{a}})^{1/2}=\boldsymbol{P}^{1/2}(\boldsymbol{HP}^{1/2})^{\mathrm{T}}\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I}+\boldsymbol{D})^{-\mathrm{T}/2}$.

2.1) Form $\boldsymbol{R}^{-\mathrm{T}/2}(\boldsymbol{I}+\boldsymbol{D})^{-\mathrm{T}/2}$. Cost: $O(MNI+M^2I+M^3)$.

2.2) Apply to $(\boldsymbol{HP}^{1/2})^{\mathrm{T}}$. Cost: $O(M^2I)$.

2.3) Apply to $\boldsymbol{P}^{1/2}$. Cost: $O(MNI)$.

The cost of forming the analysis anomalies: $O(MNI+M^2I+M^3)$.

The total cost: $O(MNI+M^2I+M^3)$.