Tellus (2000), 52A, 265–279
Printed in UK. All rights reserved

Copyright © Munksgaard, 2000
TELLUS
ISSN 0280–6495

An adaptive variational method for data assimilation with imperfect models

By JIANG ZHU[1,2,3]* and MASAFUMI KAMACHI[1] [1]*Meteorological Research Institute, 1–1 Nagamine, Tsukuba, Ibaraki 305–0052, Japan,* [2]*Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, Kawaguchi Center Building 12fl., 4-1-8 Honmachi, Kawaguchi, Saitama 332-0012, Japan,* [3]*Institute of Atmospheric Physics, Chinese Academy of Science, Beijing 100029, China*

(Manuscript received 8 February 1999; in final form 18 October 1999)

ABSTRACT

The solutions of the weak constraint data assimilation problems depend on a priori error covariance. If a priori error covariance have poor quality, a posteriori evaluation may have negative impact and solutions are not optimal. A novel variational data assimilation method is proposed, which does not assume the model is perfect, and can adaptively adjust model state without knowing explicitly the model error covariance matrix. Not by adjusting the initial condition in 4D-VAR, but by adjusting a steady gain matrix in a class of filters in this approach to yield a filter solution that minimize the norm of analysis innovation vector in a given span of time interval. The method enables very flexible ways to form some reduced order problems. A proper reduced-order problem not only reduces computational burden but leads to corrections that are more consistent with the model dynamics that trends to produce better forecast. It is shown that the optimal nudging can be reinterpreted as an example of the reduced order problems. The method is demonstrated using a simple nonlinear model (Burgers equation model) and simulated data. Full and several reduced order forms of the adaptive variational method are performed and compared with a simplified strong constraint 4D-VAR and the space variable optimal nudging scheme in assimilation-forecast experiments.

1. Introduction

The strong constraint variational method (also called the adjoint method or 4D-VAR) assumes the model is perfect (LeDimet and Talagrand, 1986), and yields a model solution best fit to given observations in a period of time by adjusting the model initial condition. To take model error into account in 4D-VAR methods, there are two kinds of approaches proposed. First kind is the weak constraint method first pointed by Sasaki (1970). The weak constraint methods introduce some additional variables to model equations to represent model errors. Comparing it with the strong constraint case, the cost function in the weak constraint case has some additional penalty terms that represent high cost for violating the specified model error statistics. One shortcoming of the weak constraint approaches are: the number of the model error variables, which is the number of grid points multiplies the number of time steps in the considered time period, is too large to apply to operational forecast (Dee, 1995). Consequently, only crude approximations on the model error have to be considered. It is difficult to know the model error covariance matrix, which is required by the weak constraint method. This becomes more difficult when the dimension becomes larger. The representer method proposed by Bennett

*Corresponding author.
email: jzhu@mri-jma.go.jp

(1992), Bennett et al. (1993) has reasonable efficiency when the number of observations is small. A study by Zupanski (1997) makes some assumptions on the model error statistics by assuming the model error as a first-order Markov variable and with a coarse time resolution of the model random error and showed a potential for operational application. The second kind, initially proposed by Derber (1989), are deterministic in nature. Derber (1989) considered the model deficiency in a 4D-VAR method through defining a systematic error term which is a product of a specified time dependent variable and a spatially dependent variable or more generally a linear combination of a few basis functions of time. The spatially dependent variable is the control variable. This idea is further examined by Wergen (1992), DeMaria and Jones (1993), and Zupanski (1993). The results of these studies show a considerable positive effect on the forecast.

All the above methods are variational. They all can be regarded as strong constraint problems from mathematical point of view. Different from variational methods are sequential ones. The Kalman filtering (KF), developed by Kalman (1960), and extended Kalman filtering (EKF) applied to meteorological data assimilation by Ghil et al. (1981), Cohn (1982) and Dee (1983), do not assume a perfect model and yields a time recursive solution. The KF and EKF update state estimation and its error covariance with new arrival of observations. Both weak constraint optimization and KF can deal with imperfect models but model error covariance must be specified. Although an estimation of the globally averaged numerical weather prediction model error variance is known (Bengtsson, 1991), the specification of the model error covariance remains a difficult problem. For some research in this area one can see Daley (1992) and Dee (1992). For ocean models, it is more difficult to specify the model error covariance since ocean model error also depends on atmospheric forcing errors. Since the model error covariance is poorly known at best, either the KF or the weak constraint method is not optimal. It is shown that a bad specification of model or observational errors can even lead to a divergence of the KF and the residuals grow (Jazwinsky, 1969). Another difficulty to apply the KF to meteorology or oceanography is due to high dimensionality of models. So far either simplified models or simplified KF schemes are studied (Fukumori and Malanotte-Rizzoli, 1995).

The adaptive Kalman filters (AKF) can simultaneously estimate the state and model error statistics by using the observations to improve the representation of the model error (Gelb, 1974). But this approach requires even greater computer resources than the standard KF. So only few applications of the AKF can be found in meteorological and oceanographic data assimilation literature. Dee (1983) and Dee et al. (1985) designed an AKF in a simplified meteorological context. The basic assumption is that the model error covariance can be modeled with a few parameters. Blanchet et al. (1997) compared several AKFs in a reduced space for a tropical Pacific ocean model.

In order to overcome the high dimensionality of meteorological or oceanic models and poor knowledge of model error covariance, Hoang et al. (1994) proposed a new adaptive filter (AF). This optimal AF is sought in a family of filters of a given structure, and the unknown parameters in the filter gain are adjusted recursively to converge to an optimal stationary gain to achieve a minimum prediction error. The main advantages of AF are: (1) it does not require the explicit model error covariance matrix; (2) computation cost is small and can be easily applied to high dimension models; (3) the reduction of adjusted parameter in the gain matrix can be carried out in many different ways. Hoang et al. (1995) and Hoang et al. (1998) applied the AF method to oceanic data assimilation problem using a quasi-geostrophic model and the Miami isopycnical coordinate ocean model, respectively. The disadvantages of AF are: (1) some assumptions have to be made for the gain matrix (stationary and some parameterizations); (2) both observation operator and model transition matrix are assumed time invariant (Hoang et al., 1998).

The purpose of this paper is to report a new variational data assimilation scheme motivated by previous works of Hoang et al. (1994, 1995, 1998). This adaptive variational method (AV) does not assume the model is perfect, and can adaptively adjust state variables without knowing explicitly the model error covariance matrix. In Section 2, the formulations of AV are given and discussed. The AV method is demonstrated and compared with a simplified strong constraint 4D-VAR using a simple nonlinear model (Burgers equation model) and simulated data in Section 3. Section 4 is summary and conclusions.

2. Theoretical background

2.1. *General formulation*

The variational data assimilation by incremental approach (Courtier et al., 1994) looks for the optimal increment $\delta\boldsymbol{x}^{\mathrm{a}}$ to the background field $\boldsymbol{x}^{\mathrm{b}}$, which is a solution of a nonlinear model. The resultant analysis field is

$$\boldsymbol{x}^{\mathrm{a}}=\boldsymbol{x}^{\mathrm{b}}+\delta\boldsymbol{x}^{\mathrm{a}}. \tag{1}$$

For imperfect model, the optimal incremental can be found by solving a weak constraint problem. The constraint in the weak constraint problem is the tangent linear model (TLM) with error terms, that is

$$\delta\boldsymbol{x}(t_{i+1})=M(t_{i+1},t_i)\delta\boldsymbol{x}(t_i)+\boldsymbol{c}(t_i),\quad i=0,1,\dots,N, \tag{2}$$

where $\boldsymbol{c}(t_i)$ is the model error (unbiased) and $\mathbf{Q}_i$ the model error covariance.

Observations are written as

$$\boldsymbol{y}^{\mathrm{o}}(t_i)=\boldsymbol{H}_i\boldsymbol{x}(t_i)+\eta(t_i),\quad i=0,1,\dots,N,$$

where, $\boldsymbol{x}(t_i)\in R^n$, $\boldsymbol{y}^{\mathrm{o}}(t_i)\in R^p$, $\eta(t_i)$ is observation error with zero mean, $\boldsymbol{R}_i$ is observational error covariance matrix and $\boldsymbol{H}_i$ is the observation operator (an $n\times p$ matrix).

The cost function is:

$$\begin{aligned}J(\delta\boldsymbol{x}(t_0),\boldsymbol{c}_1,\dots,\boldsymbol{c}_N)=&\ \frac{1}{2}\,\delta\boldsymbol{x}(t_0)^{\mathrm{T}}\boldsymbol{B}^{-1}\delta\boldsymbol{x}(t_0)\\&+\frac{1}{2}\sum_{i=0}^{N}(\boldsymbol{H}_i\delta\boldsymbol{x}(t_i)-\boldsymbol{d}_i)^{\mathrm{T}}\boldsymbol{R}_i^{-1}(\boldsymbol{H}_i\delta\boldsymbol{x}(t_i)-\boldsymbol{d}_i)\\&+\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{c}_i^{\mathrm{T}}\mathbf{Q}_i^{-1}\boldsymbol{c}_i,\end{aligned} \tag{3}$$

where $\boldsymbol{B}$ is the background error covariance matrix for the initial condition and $\boldsymbol{d}_i$ is innovation vector $\boldsymbol{y}^{\mathrm{o}}(t_i)-\boldsymbol{H}_i\boldsymbol{x}^{\mathrm{b}}(t_i)$.

There are several methods available to solve the weak constraint problem. These methods include the optimal control approach, the representer method (Bennett, 1992), the observation space method (Egbert et al., 1994; Amodei, 1995) and the Kalman smoother (Rauch et al., 1965). These methods are reviewed and discussed in the papers of Courtier (1997) and Menard and Daley (1996).

The solution of minimization of (3) provides the optimal incremental (Courtier, 1997):

$$\delta\boldsymbol{x}^{\mathrm{a}}=\boldsymbol{D}(\boldsymbol{y}^{\mathrm{o}}-\boldsymbol{H}\boldsymbol{x}^{\mathrm{b}}). \tag{4}$$

Here, $\boldsymbol{D}$ is dependent on error covariance $\boldsymbol{B}$ and $\boldsymbol{Q}=\mathrm{diag}(\boldsymbol{Q}_1,\boldsymbol{Q}_2,\dots,\boldsymbol{Q}_N)$ and therefore written as $\boldsymbol{D}=\boldsymbol{D}(\boldsymbol{B},\boldsymbol{Q})$.

2.2. *Adaptive variational formulation*

Since the model error covariance are poorly known for weather forecasting and ocean models, the specified covariance contains uncertainties that in turn may cause inconsistency of the resultant analysis. Statistical consistence contains the conditions for the analysis: the innovation vectors of analysis:

$$(\boldsymbol{y}_i^{\mathrm{o}}-\boldsymbol{H}_i\boldsymbol{x}_i^{\mathrm{a}}),\quad i=0,1,\dots,N,$$

must be time uncorrelated (there still exists some information in innovation vectors) and have norm smaller than 1σ. It is desirable that some measure can be taken when this inconsistency occurs.

The aim of adaptive variational formulation, like the adaptive Kalman filter (Mehra, 1970), is to improve the statistical consistence of analysis during data assimilation. And this is only possible with independent additional hypotheses that can not be objectively justified (Talagrand, 1998). In fact, the result of Mehra (1970) states that the covariance matrices of the model and observation errors can be determined in a Kalman filter from an infinite sequence of observations under a condition which is essentially a condition of observability. But an additional a priori hypothesis that all errors are uncorrelated in time is used. The AF approach followed by Hoang et al. (1994, 1995, 1998) determines the gain matrix of a stationary Kalman filter, independently of any quantitative hypothesis on the model or observation errors, as the matrix which a posteriori minimizes the amplitude of the innovation vector of the predicted state.

When analysis-minus-observation has a square of norm larger than observation amount times observation error variance, the following AV procedure is proposed:

First assume that $\boldsymbol{D}$ can be parameterized by a few parameters

$$\boldsymbol{D}=\boldsymbol{D}(\theta_1,\dots,\theta_L), \tag{5}$$

then adjust these parameters by minimizing

$$J(\theta_1,\dots,\theta_L)$$

$$= \sum_{i=0}^{N} (\boldsymbol{y}_i^{\mathrm{o}} - \boldsymbol{H}_i \boldsymbol{x}_i^{\mathrm{a}})^{\mathrm{T}} \boldsymbol{S}_i (\boldsymbol{y}_i^{\mathrm{o}} - \boldsymbol{H}_i \boldsymbol{x}_i^{\mathrm{a}})$$

$$= \sum_{i=0}^{N} (\boldsymbol{d}_i - \boldsymbol{H}_i \delta \boldsymbol{x}_i^{\mathrm{a}})^{\mathrm{T}} \boldsymbol{S}_i (\boldsymbol{d}_i - \boldsymbol{H}_i \delta \boldsymbol{x}_i^{\mathrm{a}}), \quad (6)$$

where (4) and (5) give the relationship between $\boldsymbol{x}_i^{\mathrm{a}}$ and $\theta_1, \dots \theta_L$, and $\boldsymbol{S}_i$ is any positive definite matrix. If the observation error covariance matrices $\boldsymbol{R}_i$ are known, $\boldsymbol{S}_i = \boldsymbol{R}_i^{-1}$. In the rest of the paper, assume that $\boldsymbol{R}_i$ is known. In fact, observation error variance is very useful for deciding when the AV should be turn on and for providing a diagnostical tool for AV.

The most important problem is how to parameterize $\boldsymbol{D}$ such that $\boldsymbol{D}$ can well represent optimal incremental associated with underlying error uncertainties.

A more ambitious formulation is to parameterize $\boldsymbol{B}$ and $\boldsymbol{Q}$ rather than $\boldsymbol{D}$. This certainly leads to explicit improvement on a priori information. However, it requires solving the weak constraint problem at first, in order to evaluate the cost function. And the evaluation of the gradient probably requires to solve the second order adjoint equation (Ngodock and LeDimet, 1998).

2.3. *An implementable approximation*

Using the equivalence of the weak constraint problem and the Kalman smoother, some reasonable approximation of (4) via parameterization of (5) can be made.

The weak constraint problem is equivalent to (Rauch et al., 1965; Bennett, 1992):

- fixed interval Kalman smoother *during the time interval*;
- Kalman filtering *at the end time.*

Using the Rauch-Tung-Striebel form of Kalman fixed interval smoother, the analysis by the weak constraint problem can be written as (Rauch et al., 1965; Gelb, 1974):

$$\delta \boldsymbol{x}^{\mathrm{a}}(t_{j-1}) = \delta \boldsymbol{x}^{\mathrm{K}}(t_{j-1}) + \boldsymbol{A}_j [\delta \boldsymbol{x}^{\mathrm{a}}(t_j) - M(t_j, t_{j-1}) \delta \boldsymbol{x}^{\mathrm{K}}(t_{j-1})],$$

$$\delta \boldsymbol{x}^{\mathrm{a}}(t_N) = \delta \boldsymbol{x}^{\mathrm{K}}(t_N), \quad j = N, N-1, \dots, 1, \quad (7)$$

where $\delta \boldsymbol{x}^{\mathrm{K}}$ is the solution of the following Kalman filtering:

$$\delta \boldsymbol{x}^{\mathrm{K}}(t_{i+1}) = M(t_{i+1}, t_i) \delta \boldsymbol{x}^{\mathrm{K}}(t_i) + \boldsymbol{K}_{i+1} [\boldsymbol{d}_{i+1} - \boldsymbol{H}_i M(t_{i+1}, t_i) \delta \boldsymbol{x}^{\mathrm{K}}(t_i)],$$

$$i = 0, 1, \dots, N. \quad (8)$$

The following approximations are made:

- a steady state Kalman Filter is assumed during the time interval;
- the backward smoother (7) is ignored.

Then a computational simple formulation for adaptive variational data assimilation is to minimize

$$J(\theta_1, \dots, \theta_L) = \sum_{i=0}^{N} (\boldsymbol{H}_i \hat{\boldsymbol{x}}(t_i) - \boldsymbol{d}_i)^{\mathrm{T}} \boldsymbol{R}_i^{-1} (\boldsymbol{H}_i \hat{\boldsymbol{x}}(t_i) - \boldsymbol{d}_i), \quad (9)$$

subject to

$$\hat{\boldsymbol{x}}(t_{i+1}) = M(t_{i+1}, t_i) \hat{\boldsymbol{x}}(t_i) + \boldsymbol{K} [\boldsymbol{d}_{i+1} - \boldsymbol{H}_i M(t_{i+1}, t_i) \hat{\boldsymbol{x}}(t_i)], \quad i = 0, 1, \dots, N, \quad (10)$$

where $\hat{\boldsymbol{x}}$ is an approximation of $\boldsymbol{x}^{\mathrm{a}}$, $\boldsymbol{K}$ is time-invariant, and

$$\boldsymbol{K} = \boldsymbol{K}(\theta_1, \dots, \theta_L). \quad (11)$$

If $M(t_{i+1}, t_i)$, observation operator and data error statistics change slightly with time during the time interval of data assimilation, the time invariant gain $\boldsymbol{K}$ is a good approximation. Noticed that this steadiness-during-a-fixed-interval approximation is weaker than the steady state Kalman filter assumption, which requires stronger conditions (Fukumori and Malanotte-Rizzoli, 1995).

It is not very clear how the ignorance of backward smoother (7) affects the final result. However there are two main reasons for ignoring smoother (7). First, if smoother equations (7) remain as constraints, along with the Kalman filter equations (8), both $\boldsymbol{K}_i$ and $\boldsymbol{A}_j$ will be parameterized by some few parameters. These parameterizations must be consistent with each other. However the consistence seems to be difficult, since $\boldsymbol{K}_i$ and $\boldsymbol{A}_j$ are related by estimation error covariance and TLM in a quite complicated way. Another reason to ignore the smoother equations (7) is due to the fact the Kalman filter and the Kalman smoother give the same state estimation at the end time of the interval. For numerical forecast, the state at end time of the assimilation interval is the initial condition for forecast. Hence ignoring the smoother equations does not have any effects on forecast.

2.4. Reduced order problems: parameterizations of K

Now we consider the problem of how to parameterize $\boldsymbol{K}$ in (11). There are some basic principles for parameterization.

(1) The parameters $(\theta_1, \ldots, \theta_L)$ can derive the state of system (10) to small innovation norms. This is related to controllability.

(2) We assumed that the Kalman gains in (8) are approximated by $\boldsymbol{K}(\theta_1, \ldots, \theta_L)$. And the Kalman gains in (8) contain information from TLM and model error $\boldsymbol{Q}$. Parameterization of $\boldsymbol{K}(\theta_1, \ldots, \theta_L)$ should be consistent with information of TLM contained in the Kalman gains.

(3) Since a posteriori validation of an estimation system is impossible without a priori, unverifiable, hypotheses, the parameters $\theta_1, \ldots, \theta_L$ should be consistent with some legitimate hypotheses on model errors.

These principles are very abstract and may not be used as practical guidelines, but the following remarks show some usefulness of them.

Remark 1. Generally speaking, the principle (1) requires more degree of freedom for parameterization of $\boldsymbol{K}$, and (3) requires less degree of freedom. The more degree of freedom, the smaller the minimal value of the cost function (9). It is known that, on statistical average

$$E[J_{\min}] = p,$$

which provides a diagnostic tool for the parameterization of $\boldsymbol{K}$.

Remark 2. The simplest way to parameterize $\boldsymbol{K}$ is probably the so-called optimal nudging method that was studied by Zou et al. (1992) and Stauffer and Bao (1993) as parameter estimation problem. Here it can be reinterpreted as a reduced order adaptive variational problem.

In the optimal nudging,

$$\boldsymbol{K} = \boldsymbol{E}\boldsymbol{N}, \tag{12}$$

where $\boldsymbol{E}$ is the normal embedding of observation space R^p into model space R^n (i.e., $\|\boldsymbol{E}\| = 1$ and $\boldsymbol{E}^{\mathrm{T}}\boldsymbol{E} = \boldsymbol{I}$), an $n \times p$ matrix and $\boldsymbol{N}$ is a $p \times p$ diagonal matrix. The parameters to be estimated are diagonal elements of $\boldsymbol{N}$. If all diagonal elements are further assumed to be the same, it is called optimal nudging with space-invariant nudging coefficient. If all diagonal elements are independent, it is called optimal nudging with space variable nudging coefficients (Zou et al., 1992). In optimal nudging, the approximated Kalman gain $\boldsymbol{K}$ is parameterized as a strictly localized matrix which reflects the underlying assumption of that $M(t_{i+1}, t_i)$ is diagonal and background and model covariance are both diagonal. This is not a good parameterization by the above principle (2) in AV formulation.

The concept of reduced order AF (Hoang et al., 1995, 1998) can be directly used in the AV to form some reduced order AV (ROAV) problems (and not reduced AV problems are called full order AV or FOAV). Here are some examples of parameterization of $\boldsymbol{K}$.

A commonly used class of parameterizations of $\boldsymbol{K}$ in reduced order state estimation is decomposition of the form (Bernstein and Hyland, 1985; Hoang et al., 1995, 1998):

$$\boldsymbol{K} = \boldsymbol{P}_r \boldsymbol{K}_e, \tag{14}$$

where $\boldsymbol{P}_r$ is a known matrix $(n \times r)$ $(r \geqslant p)$, which spread information from observation space onto model space: $\|\boldsymbol{P}_r\| > 1$. And $\boldsymbol{K}_e$ is an $r \times p$ matrix (sometimes $r = p$ and diagonal) to be estimated. For reduced problems some elements of $\boldsymbol{K}_e$, say, diagonal elements, are control variables.

If $r = p$, $\boldsymbol{K}_e$ is a diagonal matrix of $p \times p$ order. The only parameters to be estimated are diagonal elements of $\boldsymbol{K}_e$. In this case, $\boldsymbol{P}_r$ can be interpreted as an interpolation operator that makes correction at every model grid point from those only at observation locations. In this case, the number of control variables is p, the number of the observations which is much less than the number of state variables in meteorology and oceanography. As a consequence the optimization of the AV may require less iterations than the 4D-VAR. The operator $\boldsymbol{P}_r$ plays the rôle as an information propagator. To guess an appropriate projection matrix $\boldsymbol{P}_r$ may not be more difficult than to guess the model error covariance matrices in some problems. In oceanography, $\boldsymbol{P}_r$ propagates the surface information downward to infer subsurface temperature, salinity or even velocity information. $\boldsymbol{P}_r$ is usually determined, more or less empirically, by statistical or dynamical method. The ways of projection of altimeter data into subsurface model variables are suggested by several researchers and show improved results (Mellor and Ezer, 1991; Haines, 1991; Hoang et al., 1998).

2.5. Extension to nonlinear models

For non-incremental approach, the TLM and the innovation of background in problem (9)–(11) are replaced by the nonlinear model and observation, respectively. The new problem is:

$$\text{minimize} \quad J_{\text{adp}}(\theta_1, \ldots, \theta_L) = \sum_{i=0}^{N} (\boldsymbol{H}_i\hat{\boldsymbol{x}}(t_i) - \boldsymbol{y}_i^{\text{o}})^{\text{T}} \boldsymbol{R}_i^{-1} (\boldsymbol{H}_i\hat{\boldsymbol{x}}(t_i) - \boldsymbol{y}_i^{\text{o}}) \tag{9'}$$

subject to

$$\hat{\boldsymbol{x}}(t_{i+1}) = M_i[\hat{\boldsymbol{x}}(t_i)] + \boldsymbol{K}[\boldsymbol{y}_{i+1}^{\text{o}} - \boldsymbol{H}_i M_i[\hat{\boldsymbol{x}}(t_i)]], \quad i = 0, 1, \ldots, N, \tag{10'}$$

$$\boldsymbol{K} = \boldsymbol{K}(\theta_1, \cdot \ldots, \theta_L). \tag{11'}$$

The algorithms to solve the AV problems are the classical adjoint method in parameter estimation or optimal control (regard $\boldsymbol{K}$ or $\boldsymbol{K}_e$ as a parameter vector in (5)). The adjoint model integration gives the gradient of J with respect to $\boldsymbol{K}$ or $\boldsymbol{K}_e$. Some unconstraint optimization algorithms can be used to find an optimal $\boldsymbol{K}$ or $\boldsymbol{K}_e$. This issue is addressed by many authors in meteorology and oceanography. For more details, see Lions (1971), Wunsch (1996, chapter 6). Here only the adjoint equation and the equation for calculation of gradient are given below:

$$\frac{\partial J}{\partial K_{k,j}} = \sum_{t=0}^{N} \hat{\boldsymbol{x}}^*(k, t_i) d(j, t_i), \tag{15}$$

where $K_{k,j}$ is the (k, j)th element of matrix $\boldsymbol{K}$, $\hat{x}^*(k, t_i)$ and $y(j, t_i)$ are kth and jth element of $\hat{\boldsymbol{x}}^*(t_i)$ and $\boldsymbol{d}(t_i)$ respectively and $\hat{\boldsymbol{x}}^*$ is the solution of the adjoint equation:

$$\hat{\boldsymbol{x}}^*(t_i) = \boldsymbol{L}_i^{\text{T}} \hat{\boldsymbol{x}}^*(t_{i+1}) - \boldsymbol{H}_i^{\text{T}} \boldsymbol{R}_i^{-1} [\boldsymbol{H}_i \hat{\boldsymbol{x}}(t_i) - \boldsymbol{y}^{\text{o}}(t_i)], \tag{16}$$

$$\boldsymbol{L}_i = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H}_i) M(t_{i+1}, t_i), \tag{17}$$

with final time condition: $\hat{\boldsymbol{x}}^*(t_N) = 0$. The associated gradient and the adjoint equation for ROAV are similar.

The algorithm to solve the AV consists of the following steps:

(1) give a first guess for gain matrix $\boldsymbol{K}$ (or $\boldsymbol{K}_e$) for FOAV (or ROAV);

(2) integrate the forward filter equation (10) with estimation of gain matrix $\boldsymbol{K}$ (or $\boldsymbol{K}_e$) over time interval $[t_0, t_N]$; the estimation of $\boldsymbol{K}$ is either from the first guess or from previous iteration;

(3) calculate the cost function defined by (9);

(4) integrate the adjoint equation (16) backward with the final time condition $\hat{\boldsymbol{x}}^*(t_N) = 0$ over the time period;

(5) calculate the gradient of cost function by (15);

(6) update the previous estimation of gain matrix $\boldsymbol{K}$ (or $\boldsymbol{K}_e$) for FOAV (or ROAV) using some gradient-based iterative optimization algorithm (e.g., quasi-Newton method or conjugated gradient method);

(7) check the stop criteria (the magnitude of reduction of the norm of gradient); if meet, stop; otherwise, go to (2).

The computation cost of AV depends on conditioning and L, the number of parameters $(\theta_1, \ldots, \theta_L)$. If only consider the size of the problem, reduced problems using (14) can be less costly in computation than 4D-VAR if $L \leqslant n$.

There are flexible ways to apply AV. It can be applied to these data assimilation problems that other methods are difficult to apply for the reasons we already discussed in introduction. It also can be used as an additional part of existing strong or weak constraint 4D-VAR system that will be turn on when strong or weak constraint 4D-VAR produces a large analysis-minus-observation norm. Adding AV to existing 4D-VAR schemes as a backup will improve ability of assimilation systems in dealing with uncertainties in background and model error statistics.

Some generalizations of AV can be made if there are more information on error statistics available. For example, when the background error covariance is known, the IC can be taken as control variables and the background term is included in the cost function.

2.6. Some potential problems

There are some potential problems related to definition of $\boldsymbol{K}$ in realistic applications. (1) In meteorology and oceanography, for satellite observations, the observation locations are different at every data insertion time. When the observation operator $\boldsymbol{H}(t)$ is time-dependent, the definition of a steady $\boldsymbol{K}$ is questionable; (2) When observation amounts change with time, a problem could become to be underdetermined if the number of observations is small and the number of adjustable parameters in K is relatively large; (3) The definition of K may be more complex due to the difference between the observed and modeled quantities (e.g., radiance instead of temperature, wind).

For problem (1), generally the assumption of the time-invariant $\boldsymbol{H}$ is not required if we can design a reduced order formulation in which the gain matrix can be expressed as

$$\boldsymbol{K}_i = \boldsymbol{K}_i(\theta_1, \ldots, \theta_L) \text{ for all } i = 1, \ldots, N-1,$$

where $\theta_1, \ldots, \theta_L$ are time-invariant parameters to be estimated. This can be illustrated more by considering an example of assimilating satellite altimeter data. The TOPEX/POSEIDON altimeter covers global oceans every 10 days. The observation operator changes from time to time and so does the associated gain matrix. If the parameters $\theta_1, \ldots, \theta_L$ describe the vertical projection coefficients, which make changes of the subsurface temperature and salinity fields, and are horizontally invariant, they are time-invariant parameters.

To overcome problem (2), the ideal way is to develop some adaptive parameterization schemes that will change the way of parameterization of K according to some characteristics of observations (e.g., observation amount) and keep the problem well-defined. Problem (3) may create difficulty for applications of a quite wide range of assimilation methods including Kalman filtering and the proposed AV to assimilation of some type of observations and should be studied in future.

3. Experiments using Burgers' equation model

3.1. The Burgers' equation model

The following Burgers' equation (Burgers 1974)

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}, \tag{18}$$

describes the one-dimensional advection-diffusion process over spatial domain $(-\infty, \infty)$. The numerical model is defined as a finite-difference leapfrog scheme, with a forward step every 15 time steps. This simple scheme gives accurate results by comparing with one analytical solution (Uboldi and Kamachi, 2000). The computation spatial domain is $[-1000, 1000]$ km, but only solution within $[-100, 150]$ km is considered. A larger computation domain is to reduce the boundary effects on interior solution. The model parameters are: $\Delta t = 60$ s; $\Delta x = 5$ km. The simulation time is $T = 16$ h. The "true" solution is generated by running the model with the initial condition (IC) (see Fig. 1, the solid line without marks):

$$u(x, 0) = \begin{cases} 0 & x < -L;\ x > L \\ u_0\left(1 + \dfrac{2x}{L}\right)\left(1 - \dfrac{x}{L}\right)^2 & 0 \leqslant x < L \\ u_0\left(1 - \dfrac{2x}{L}\right)\left(1 + \dfrac{x}{L}\right)^2 & -L < x < 0 \end{cases} \tag{19}$$

where $u_0 = 5$ m/s, and $L = 50$ km and with diffusion coefficient. $\nu = 1.0 \cdot 10^4$ m^2 s^{-1}. Denoted the "true" solution by $u^t(x, t)$, -100 km $\leqslant x \leqslant 150$ km, $0 \leqslant t \leqslant 16$ h. In data assimilation experiments with imperfect model, the diffusion coefficient is set to be $\nu = 1.4 \cdot 10^4$ m^2 s^{-1} to simulate the mode error. The errors caused by the wrong diffusion coefficient are called pure model errors which are estimated by running model twice with the same above initial condition. The spatially averaged root mean square of model errors are shown in Fig. 2 with about overall averaged error of

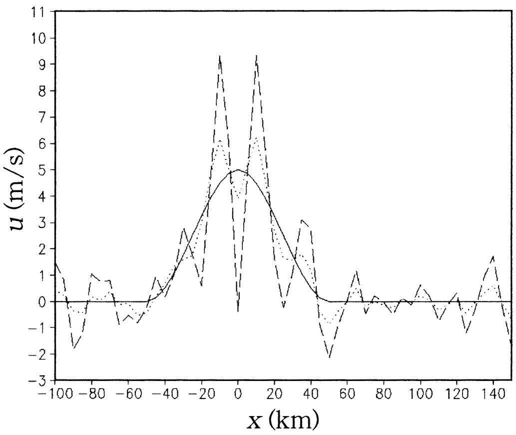


Fig. 1. Initial conditions. The solid line without mark: "true" initial condition (IC); the broken line: estimated IC from Data Set I; the dotted line: estimated IC from Data Set II.

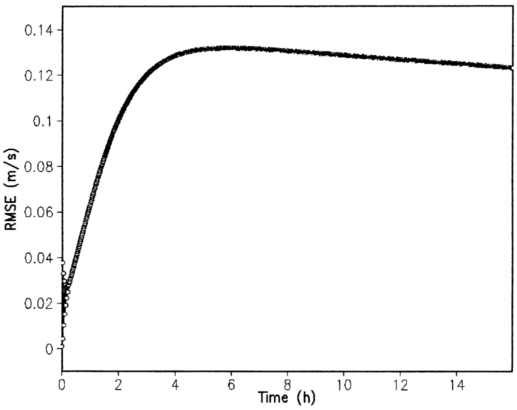


Fig. 2. RMSE caused only by model diffusion coefficient error.

Table 1. *Relations between observation error and model error*

Data set	Errors
I	obs > model
II	obs < model

0.13 m/s. The overall averaged signal is about 0.9 m/s in the time-space domain. The average model error is about 14% of the average signal. The observations are assumed to be available at every other model grid point from $x = -100$ km to 150 km. There are total 26 observation stations (total 51 model grid points). The observations are assumed to be available only at hour 1, 4, 7 and 10. The total number of the observations is 104. There are two data sets are generated. The set one (set I) contains uncorrelated noise with zero mean and standard deviation (SD) of 0.15 m/s. The set two (set II) contains noise with smaller SD of 0.06 m/s than in set I. The observational errors in set I is smaller than the model errors. See Table 1.

3.2. *The experiment description*

There are ten groups of experiments performed. Table 2 outlines these group characters. Each group of experiments includes experiments with various first guesses of IC. These first guesses of IC are generated by adding Gaussian white noise to the "true" IC. These first guesses of IC have different SDs which range from 0.20 m/s to 0.65 m/s. These first guesses are used by all experiments in the FOAV and ROAV as the ICs. The optimization algorithm used in all experiments is a quasi-Newton method with code name M1QN3. Some descriptions can be found in Gilbert and Lemarechal (1989) and Liu and Nocedal (1989).

Table 2. *Experiment list in case of imperfect model*

Experiment group	Assimilation method	Observations
EXP-1	4D-VAR	set I (0.15m/s error)
EXP-2	4D-VAR	set II (0.06m/s error)
EXP-3	FOAV	set I
EXP-4	FOAV	set II
EXP-5	ROAV-1	set I
EXP-6	ROAV-1	set II
EXP-7	ROAV-2	set I
EXP-8	ROAV-2	set II
EXP-9	SVON	Set I
EXP-10	SVON	Set II

In experiments group 1 and 2 (EXP-1 and EXP-2), a simplified strong constraint 4D-VAR is used, which we refer to "simplified 4D-VAR". The cost function in the simplified 4D-VAR differs from standard 4D-VAR in the way that does contain the background term. In most realistic applications of 4D-VAR, the information in the background field and its error covariance are considered. Since in AV or ROAVs there is no need to consider the error covariance of the background field, it will make the comparison between AV and the standard 4D-VAR awkward by posing various covariance matrices of the background field with different accuracy.

In FOAV experiments (EXP-3 and EXP-4), the first guesses for $\boldsymbol{K}_e$ are $\boldsymbol{K}_e = 0$. The ROAV experiments (EXP-5 and EXP-6) use the parameterization (7), where $\boldsymbol{P}_r$ is a known projection matrix ($n \times p$), $\boldsymbol{K}_e$ is a $p \times p$ diagonal matrix to be estimated. The $\boldsymbol{P}_r$ simply takes the form:

$$\boldsymbol{P}_r = \begin{Bmatrix} 1. & 0 & 0 & \cdots \\ .5 & .5 & 0 & \cdots \\ 0 & 1. & 0 & \cdots \\ 0 & .5 & .5 & \cdots \\ 0 & 0 & 1. & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{Bmatrix} \tag{20}$$

which represents a linear interpolation: at the grid point without observation, the value of u is interpolated from values at 2 neighboring grid points where observations are available, and the weighting (interpolation coefficient) for each values at the two neighboring grid points is 0.5. In EXP-5 and 6, only $p(=26)$ variables are estimated. The first guesses of these estimated variables are zero. In experiments EXP-7 and 8, K_e is as the same as in EXP-5 and 6, but P_r takes the form

$$\boldsymbol{P}_r = \begin{Bmatrix} 1. & 0 & 0 & \cdots \\ q_1 & .5 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & q_2 & .5 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{Bmatrix}$$

which represents a linear interpolation: at the grid point without data and parameters $q_1, q_2, \ldots$ are also estimated. The total number of parameters to be estimated is 51 in EXP-7 and 8, as the same as in 4D-VAR experiments. The first guesses for the 2 experiments are 0.5 for $q_1, q_2, \ldots$.

The EXP-9 and 10 are experiments of the space variable optimal nudging (SVON) using data sets I and II, respectively. The EXP-9 and 10 will be the same as the EXP-5 and 6 if one replaces 0.5 in (20) by 0.

Notice that there are several differences among the optimal nudging in Zou et al. (1992), Stauffer and Bao (1993) and the method used here. In Zou et al. (1992), the cost function included the background term (IC as control variables), penalty term for a priori information of nudging coefficients and the misfit-to-data term. In Stauffer and Bao (1993), there is no background term in their cost function. In SVON in this paper, the cost function only contains the misfit-to-data term.

3.3. *The results*

There is a group of first guesses with different SDs. Each first guess is used by 10 experiments listed in Table 2. We first look detailed results of the eight experiments using the first guess of IC with a SD of 0.25 m/s.

First, look at the summary of optimizations in the ten groups. Table 3 gives the details for each optimization process with IC first guess with a SD of 0.25 m/s. Each optimization process terminated normally: the test on the gradient is satisfied. The number of calls of the simulation subroutine (the simulation subroutine calculates value of the cost function and its gradient) determines the computational cost of each method. The computational costs of the FOAV and ROAVs are less than the simplified 4D-VAR.

From the values of cost function at final iterations, one can check the misfit to the observations. All experiments except simplified 4D-VARs (EXP-1 and EXP-2) achieve solutions within the range of the observation error variances (i.e., cost function <104.0). The failure of the simplified 4D-VAR is due to the ignorance of model error.

The solutions of different data assimilation methods have different characters. The initial condition estimated by the simplified 4D-VAR has a higher peak than the "true" initial condition to compensate the model error of a larger diffusion coefficient (see Fig. 1). The observation error leads to an amplification of the initial condition estimation error (Fig. 1), which is caused by the diffusion in the model.

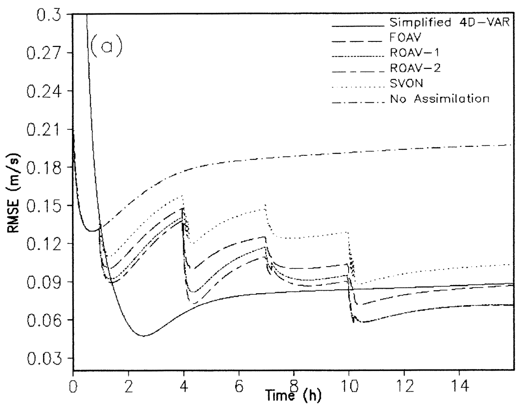
The RMSEs as a function of time are shown in Fig. 3. Fig. 3a is for Data Set I and Fig. 3b is for Data Set II. Both the FOAV and ROAV methods for both data sets outperformed over the simplified 4D-VAR after hour 10. It is clearly shown that all AV solutions reduce their RMSE at arrival times of observations, then some relaxation occurs until next data arrival time. The corrections by the FOAV or ROAVs decrease when assimilation time increase. After hour 10, the forecasts are made. ROAVs produce best forecast. The SVON for both data sets is the worst.

Fig. 4 shows time sequences of the profiles of u produced by different methods. The larger diffusion coefficient in model results in the simplified 4D-VAR solution of assimilation, which has higher peaks at early time stages and gradually becomes flat. The 4D-VAR solution at the final time t_f (i.e. hour 10) is more flat than the "true" solution (see Fig. 4a). There is also a phase delay in the simplified 4D-VAR solution. The phase delay is a consequence of weakened nonlinear interaction due to larger diffusion coefficient. The larger diffusion coefficient and a spatially more flat solution at t_f lead to an even more spatially

Table 3. *Summary of optimizations in case of imperfect model case*

	Exper. #	# of iter.	# of simul.	Cost function at 1st guess	Cost function at final	Grad. norm at 1st guess	Grad. norm at final
4D-VAR	EXP-1	63	100	2.291E2	1.072E2	2.271E2	8.799E-1
	EXP-2	44	70	9.374E2	2.488E2	1.477E3	3.522
FOAV	EXP-3	16	17	2.291E2	7.159E-7	3.562E2	8.430E-3
	EXP-4	12	14	9.374E2	3.369E-5	1.587E4	1.03E-1
ROAV-1	EXP-5	15	16	1.822E2	4.739E-6	1.242E2	7.054E-3
	EXP-6	19	19	6.048E2	3.576E-5	5.119E2	3.440E-2
ROAV-2	EXP-7	15	16	1.822E2	7.403E-6	1.257E2	9.261E-3
	EXP-8	19	20	6.048E2	3.270E-5	5.228E2	4.005E-2
SVON	EXP-9	16	18	2.291E2	2.323E-5	1.729E2	1.587E-2
	EXP-10	19	20	9.374E2	5.273E-5	8.496E2	3.957E-2

# of iter. = number of iterations; # of simul. = number of simulation calls (one simulation call requires one forward model run and one the adjoint model run).



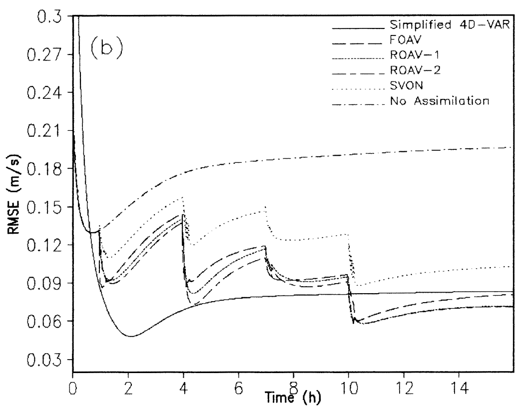


Fig. 3. Comparison of RMSEs. (a) using data set I; (b) using data set II.

flat forecast. This shortcoming of simplified 4D-VAR is overcome, to some extent, by the FOAV and ROAVs. The solutions by the FOAV and ROAVs at t_f are closer to the "true" solution than that by the simplified 4D-VAR. For AVs (FOAV and ROAV), there are less phase delays for solutions at t_f and nonlinear interactions are maintained better during the assimilation period than the simplified 4D-VAR. These lead to better forecasts.

The fact that the FOAV allows all elements of $\boldsymbol{K}$ to be estimated certainly weaken the model constraint and lead to a solution which fits well to observations at locations where data is available but may be far away from "true" state at locations where observations are not available. The data errors cause a less smooth solution of the FOAV and SVON (Figs. 4b, e). It is shown that the ROAVs produced better solution at almost grid points (Figs. 4c, d).

The gain matrix $\boldsymbol{K}$ obtained by FOAV is different from that by ROAVs. Fig. 5 shows the gain matrices obtained by the FOAV, ROAVs and SVON. Figs. 5a, b are for the FOAV using Data Sets I and II, respectively. Fig. 5c gives the gain matrix by ROAV-1 using Data Set I. Fig. 5d gives the gain matrix by SVON using data set I. Other gain matrices obtained from ROAV-1 using Data Set II and ROAV-2 using both data sets are not shown here since they are almost visually identical to (c). The gain matrices obtained by the ROAVs are less sensitive with respect to observation errors than those by the FOAV. Since the matrices by the FOAV and ROAV are quite different, the

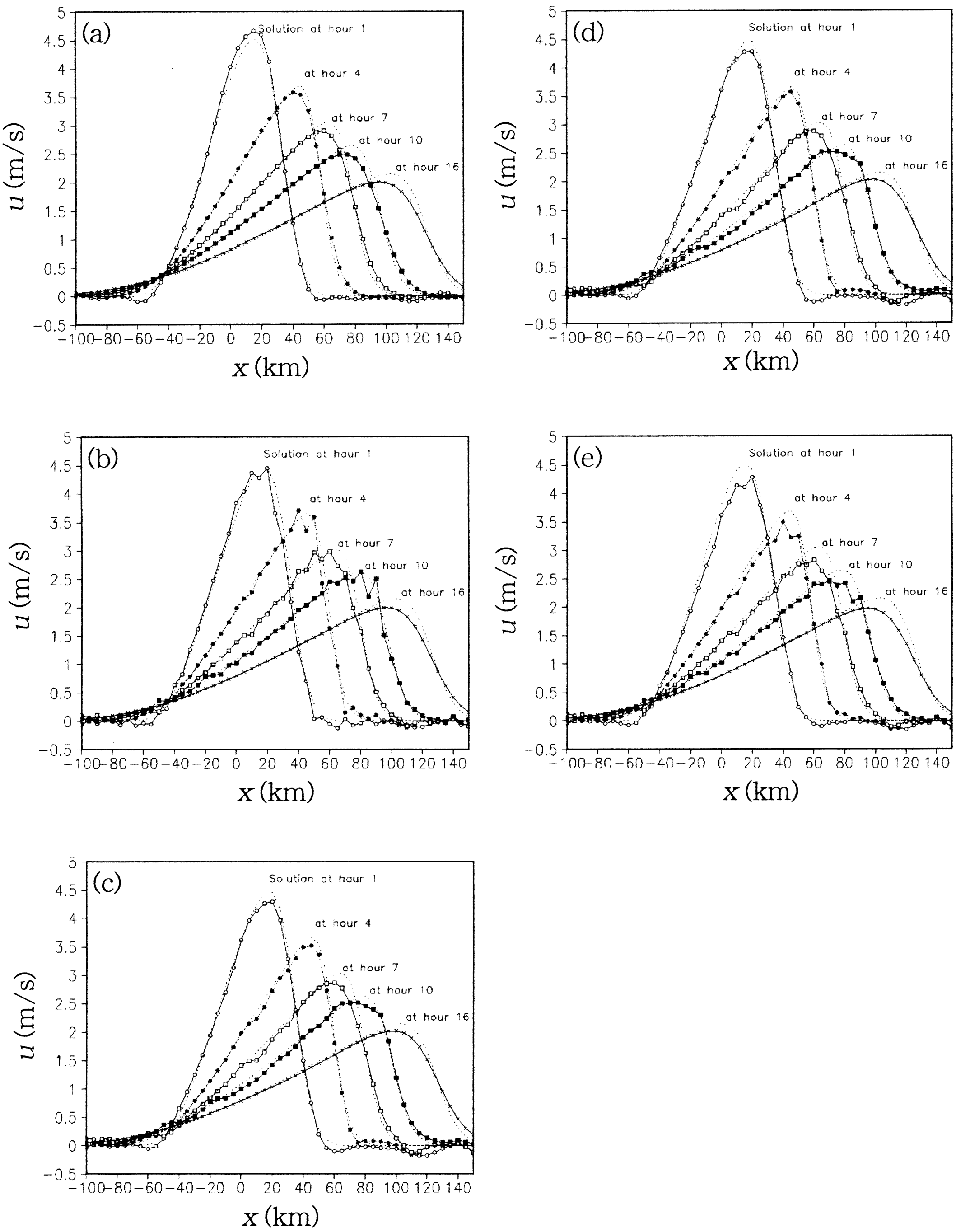


Fig. 4. The solutions with the 5 methods using data set II. (a) Solutions with 4D-VAR. From left to right, the 5 solid line curves represent the solution at time hour 1, 4, 7, 10 and 16, respectively. The dotted lines represent the "true" solution at the correspondent time. (b) As the same as (a), but for FOAV. (c) As the same as (a), but for ROAV-1; (d) As the same as (a), but for ROAV-2; (e) As the same as (a), but for SVON.

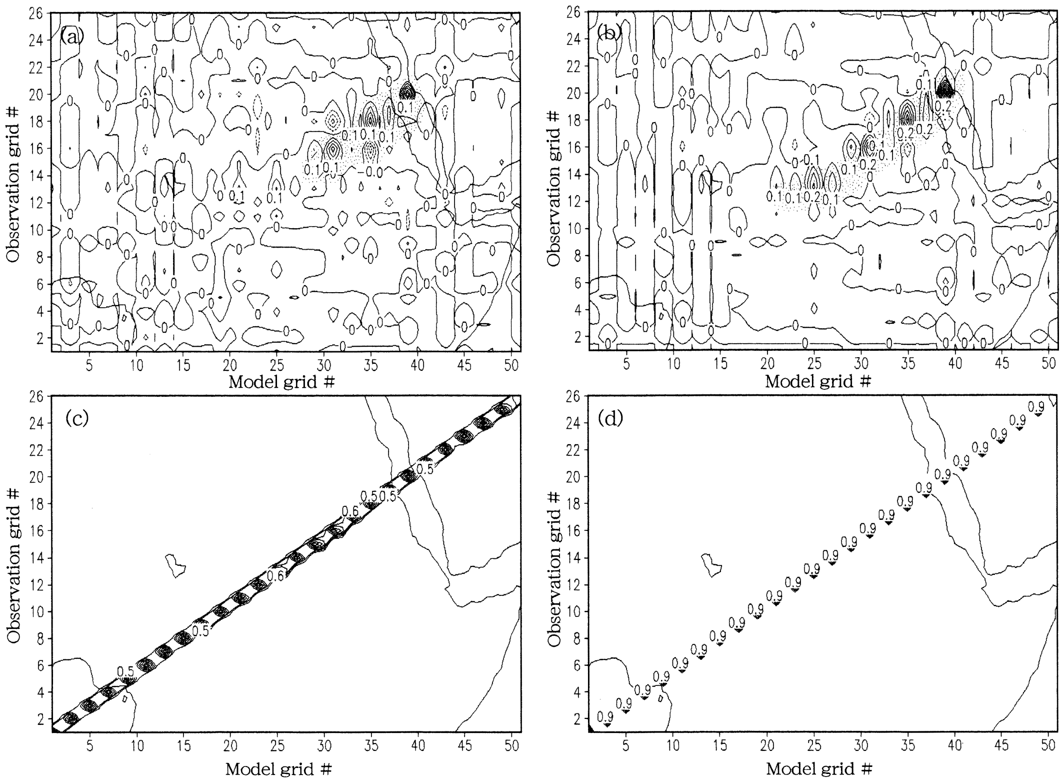
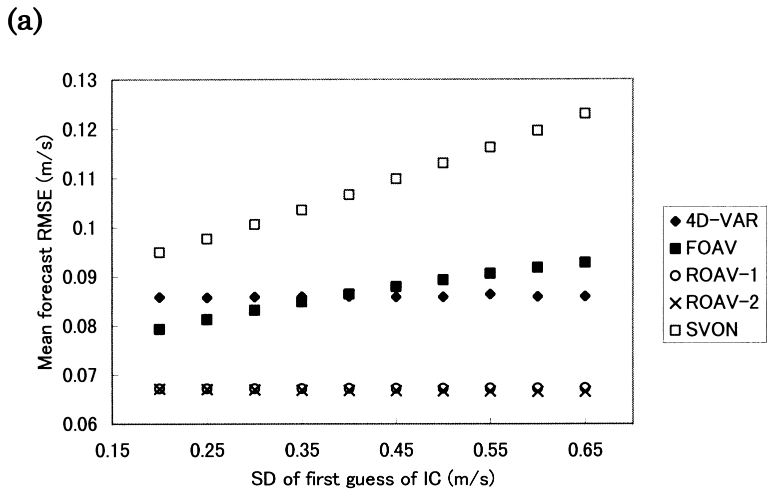


Fig. 5. The gain matrices obtained by FOAV, ROAVs and SVON. (a) FOAV, using data set I, contouring −0.3 to 0.7, interval 0.1; (b) FOAV, using data set II, contouring −0.3 to 0.7, interval 0.1; (c) ROAV-1, using data set I, contouring 0.4 to 1.0, interval 0.1; (d) SVON, using Data Set I, contouring 0.4 to 1.0, interval 0.1. The stable elements in gain matrix of FOAV are located in the dotted area in (a), (b). Along the lines across from the left-bottom corner to the right-top corner of (c) and (d), the elements are distributed in a zigzag shape, with peak values of about 1.0, but valley values for (c) are 0.5 and 0 for (d). The difference between (c) and (d) is the values of valleys.

structures of optimal gain matrices for FOAV may not be unique. The gain matrix by SVON is strictly localized (see Fig. 5d). Some elements (shown in dotted area of Figs. 5a, b) of gain matrices obtained by FOAV are not sensitive to data error; these elements are called stable elements here. Values of these stable elements are also close to those obtained by ROAVs. The model space locations of stable elements are where data has high signal to noise ratio (i.e., model solution has large amplitudes).

Above some detailed results are shown for the ten experiments using only one first guess of IC. Fig. 6 shows the hour 10–16 mean forecast RMSEs of different data assimilation methods with different first guesses of IC. In all cases, ROAVs are better than the simplified 4D-VAR, and ROAVs are better than FOAV. The FOAV is only better than the simplified 4DVAR when IC has error less than 0.40 and 0.55 m/s for data set I and II, respectively. By comparing (a) and (b) of Fig. 6, one can see that performance of FOAV close to that of ROAVs when observation error is small. It is also shown that the forecast errors by AV are more sensitive to observation errors than by other methods. The SVON remains worst among all methods.

The FOAV and SVON can be regarded as two extremes of the adaptive variational methods. The gain matrix for FOAV is widely distributed. The FOAV considered that model error covariances are rather arbitrary. The SVON considered model



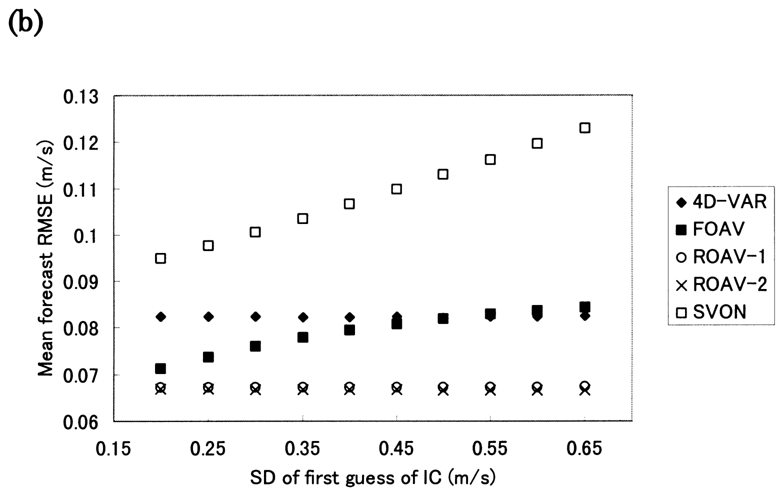


Fig. 6. The hour 10–16 mean RMSEs of forecasts with respect to various first guess of the initial condition (at hour 0). (a) using data set I; (b) using data set II.

error covariances are uncorrelated which leads to a strictly localized gain matrix. The ROAV-1 and ROAV-2 are something between and the structures of their gain matrices are more consistent with the true underlying model error covariance.

4. Summary and conclusions

A formulation of reducing non-optimal behavior of analysis using a posteriori information is presented. A priori hypotheses is that the optimal incremental can be parameterized by a few parameters in some specified forms. Then these parameters are determined by minimizing the norm of analysis innovation. An approximate, computationally simple formulation is deduced using Rauch-Tung-Striebel form of Kalman fixed interval smoother, which is equivalent to weak constraint problem. Based on this simple formulation, a novel variational data assimilation method is proposed, which does not assume the model is perfect, and can adaptively adjust model state without explicitly knowing the model error covariance matrix. Not by adjusting the initial condition in 4D-VAR, but by adjusting a steady gain matrix in a class of filters in this approach to yield a filter solution that minimizes the norm of analysis innovation vector in a given span of time interval. The method enables very flexible ways to form some reduced-order problems. A proper reduced-order problem not only reduces computational burden but leads to corrections that are more consistent with the model dynamics that trends to produce better forecast.

The method is demonstrated using a simple nonlinear model (Burgers' equation model) and simulated data. Full and several reduced order forms of the adaptive variational method are performed and compared with a simplified 4D-VAR in assimilation-forecast experiments. The results of numerical experiments show that, when the model is imperfect, the ROAVs can produce a better solution through introduction of a model error and outperform the simplified 4D-VAR method. It is shown that a proper reduced order problem not only reduces computational burden but leads to corrections that are more consistent with the model dynamics and produces better forecasts. The optimal nudging method, which does not assume a perfect model but has a crude assumption about model error covariance, performs not good as the simplified 4D-VAR. The computational cost of the ROAVs is less than the 4D-VAR from our experiments, and this is very attractive from operational point of view.

There are flexible ways to apply AV. It can be applied to these data assimilation problems that other methods are difficult to apply for the reasons we already discussed in introduction. It also can be used as an additional part of existing strong or weak constraint 4D-VAR system that will be turn on when strong or weak constraint 4D-VAR produces a large analysis-minus-observation norm.

The experiments present here are quite promising. There remains, however, to test the method in further experiments which will allow using more complicated, realistic atmospheric and oceanic models and real data. Many problems remain open related to the method. Except for those discussed in Subsection 2.6, in meteorological and oceanic applications, what a priori hypotheses can legitimately be made? And once those hypotheses are made, what can be objectively determined from the available observations? These fundamental problems are important issues for further research of data assimilation.

5. Acknowledgements

This work is supported by CREST (Core Research for Evolution Science and Technology) of JST (Japan Science and Technology Corporation). We thank two anonymous reviewers for useful comments.

REFERENCES

Albert, A. E. and Gardner, L. A., Jr. 1967. *Stochastic approximation and nonlinear regression.* The MIT Press, Cambridge, Mass.

Amodei, L. 1995. Solution approachée pour un problèm d'assimilation de donnees météorologiques avec prise en compte de l'erreur de modele. *Comptes Rendus de l'Académie des Sciénces* **321**, serie II a, 1087–1094.

Bengtsson, L. 1991. Advances and prospects in numerical weather prediction. *Q. J. R. Meteorol. Soc.* **117**, 855–902.

Bennett, A. F. 1992. *Inverse methods in physical oceanography.* Cambridge University Press, New York, 346 pp.

Bennett, A. F., Leslie, L. M., Hagelberg, C. R. and Powers, P. E. 1993. Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method. *Mon. Wea. Rev.* **121**, 1714–1729.

Bernstein, S. and Hyland, D. 1985. The optimal projection equations for reduced-order state estimation. *IEEE Trans. Autom. Contr.* **AC-30**, 583–585.

Blanchet, I., Frankignoul, C. and Cane, M. 1997. A comparison of adaptive Kalman filters for a tropical Pacific ocean model. *Mon. Wea. Rev.* **125**, 40–48.

Burgers, J. M. 1974. *The nonlinear diffusion equation.* D. Reidel Publ. Co., Dordrecht, Holland, 173 pp.

Cohn, S. E. 1982. *Methods of sequential estimation for determining initial data in numerical weather prediction.* PhD thesis. Courant Institute of Mathematical Sciences Report no. CI-6–82, New York University, New York, 183 pp.

Courtier, P. 1997. Dual formulation of 4-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.* **123**, 2449–2461.

Courtier, P. Thepaut, J.-N. and Hollingsworth, A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1388.

Daley, R. 1992. Forecast-error statistics for homogeneous and inhomogeneous observation networks. *Mon. Wea. Rev.* **120**, 627–647.

Dee, D. P. 1983. *Computational aspects of adaptive filtering and applications to numerical weather prediction.* PhD thesis. Courant Institute of Mathematical Sciences Report no. CI-6–83, New York University, New York, 150 pp.

Dee, D. P. 1992. A simple scheme for tuning forecast error covariance parameters. *Workshop on Variational assimilation, with special emphasis on three-dimensional aspects.* ECMWF, Reading (UK), 191–205.

Dee, D. P. 1995. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.* **123**, 1128–1145.

Dee, D. P., Cohn, S. E., Dalcher, A. and Ghil, M. 1985. An efficient algorithm for estimating noise covariance in distributed systems. *IEEE Trans. Automat. Control* **AC-30**, 1057–1065.

DeMaria, M. and Jones, R. W. 1993. Optimization of a hurricane track forecast model with the adjoint model equations. *Mon. Wea. Rev.* **121**, 1730–1745.

Derber, J. C. 1989. A variational continuous assimilation technique. *Mon. Wea. Rev.* **117**, 2437–2446.

Egbert, G. D. and Bennett, A. F. 1996. TOPEX/Poseidon tides estimated using a global inverse model. *J. Geophys. Res.* **99**, 24 821–24 852.

Fukumori, I. and Malanotte-Rizzoli, P. 1995. An approximate Kalman filter for data assimilation: an example with an idealized Gulf Stream model. *J. Geophys. Res.* **100**, 6777–6793.

Gelb, A. 1974. *Applied optimal estimation.* The MIT Press, 374 pp.

Gilbert, J. C. and Lemarechal, C. 1989. Some numerical experiments with variable storage quasi-Newton algorithms. *Math. Programming* **B45**, 407–435.

Ghil, M., Cohn, S. E., Tavantzis, J., Dube K. and Isaacson, E. 1981. Application of estimation theory to numerical weather prediction. In: *Dynamic meteorology: data assimilation methods* (L. Bengtsson, M. Ghil, and E. Källén, eds.). Springer Verlag, New York, 139–224.

Haines, K. 1991. A direct method for assimilating sea surface height data into ocean model with adjustments to the deep circulation. *J. Phys. Oceanogr.* **21**, 843–868.

Hoang, H. S., De Mey, P. and Talagrand, O. 1994. A simple adaptive algorithm of stochastic approximation type for system parameter and state estimation. *Proc. 33rd Conference on decision and control.* Lake Buena Vista, Florida, US. 747–752 (available from Dr. Hoang, SHOM/CMO/GRGS/CNRS, 18 Av. Edouard Belin 31055, Toulouse Cedex, France).

Hoang, H. S., De Mey, P., Talagrand, O. and Baraille, R. 1995. Assimilation of altimeter data in a multilayer quasi-geostrophic ocean model by simple nonlinear adaptive filter. *Proc. Second Int. Symp. On Assimilation of observations in meteorology and oceanography*, Tokyo, Japan, World Meteor. Org., 521–526 (available from Dr. Hoang, SHOM/CMO/GRGS/CNRS, 18 Av. Edouard Belin 31055, Toulouse Cedex, France).

Hoang, H. S., Baraille, R., Talagrand, O., Carton X. and De Mey, P. 1998. Adaptive filtering: application to assimilation of satellite data assimilation in oceanography. *Dyn. Atmo. and Oceano.* **24**, 257–281.

Jazwinsky, A. H. 1969. Adaptive filtering. *Automatika* **5**, 475–485.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82D**, 35–45.

LeDimet, F. X. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* **38**, 97–110.

Lions, J. L. 1971. *Optimal control system governed by partial differential equations.* Springer-Verlag, New York, pp. 396.

Liu, D. C. and Nocedal, J. 1989. On the limited BFGS method for large scale optimization. *Math. Programming* **B45**, 503–528.

Mehra, R. K. 1970. On the identification of variances and adaptive Kalman filtering. *IEEE Trans. Automat., Contr.* **AC-15**, 175–184.

Mellor, G. L. and Ezer, T. 1991. A Gulf stream model and an altimetry assimilation scheme, *J. Geophys. Res.* **96**, C5, 8779–8795.

Menard, R. and Daley, R. 1996. The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus* **48A**, 221–237.

Ngodock, H. E. and LeDimet, F. X. 1998. Sensitivity studies for geophysical flows. Proceedings of Int. Conf. On Compu. fluid dyn., Brenos Aires (available from Prof. LeDimet, LMC, BP 53X, 38041 Grenoble Cédex, France).

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

Sasaki, Y. 1970. Some basic formalisms on numerical variational analysis. *Mon. Wea. Rev.* **98**, 875–883.

Stauffer, D. R. and Bao, J.-W. 1993. Optimal determination of nudging coefficients using the adjoint equations. *Tellus* **45A**, 358–369.

Talagrand, O. 1998. *A Posteriori* evaluation and verification of analysis and assimilation algorithms. *Proc. of ECMWF Workshop on Diagnosis of Data Assimilation System.* Reading, UK. 17–28.

Uboldi, F. and Kamachi, M. 2000. Time-space weak-constraint data assimilation for nonlinear models. *Tellus* **52A**, in press.

Wergen, W. 1992. The effect of model errors in variational assimilation. *Tellus* **44A**, 297–313.

Wunsch, C. 1996. *The ocean circulation inverse problem*, Cambridge Univ. Press, Cambridge, UK, 442 pp.

Zou, X., Navon, I. M. and LeDimet, F. X. 1992. An optimal nudging data assimilation scheme using parameter estimation. *Q. J. R. Meteorol. Soc.* **118**, 1163–1186.

Zupanski, M. 1993. Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev.* **121**, 2396–2408.

Zupanski, D. 1997. A general weak constraint application to operational 4D-VAR data assimilation system. *Mon. Wea. Rev.* **125**, 2274–2292.