Recent Progress of Data Assimilation Methods in Meteorology

Tadashi TSUYUKI and Takemasa MIYOSHI

Japan Meteorological Agency, Tokyo, Japan

(Manuscript received 7 February 2007, in final form 28 March 2007)

Abstract

Data assimilation is a methodology for estimating accurately the state of a time-evolving complex system like the atmosphere from observational data and a numerical model of the system. It has become an indispensable tool for meteorological researches as well as for numerical weather prediction, as represented by extensive use of reanalysis datasets for research purposes. New advances of data assimilation methods emerged from the 1980s. This review paper presents the theoretical background and implementation of two advanced data assimilation methods: four-dimensional variational assimilation (4D-Var) and ensemble Kalman filtering (EnKF), which currently draw much attention in the meteorological community. Recent research results in Japan on those methods are reviewed, especially on mesoscale applications of 4D-Var and tests of the local ensemble transform Kalman filter (LETKF). Comparison of 4D-Var and EnKF is also briefly discussed. An outline of the mesoscale 4D-Var system of the Japan Meteorological Agency, which is the first operational 4D-Var for a mesoscale model, is given in Appendix.

1. Introduction

Four-dimensional data assimilation (hereafter referred to as data assimilation) is a methodology to estimate the state of a time-evolving complex system like the atmosphere as accurately as possible from observational data and known physical laws. The physical laws are primarily represented by a numerical model of the system, which provides a short-range forecast as the first guess for estimation. The first guess is regarded as information from the past observations propagated by the model to the present. The numerical model is also used for evolving the error covariance of first guess in advanced data assimilation methods. Data assimilation has been widely applied to several areas of earth science such as meteorology, oceanography, and atmospheric chemistry.

Corresponding author: Tadashi Tsuyuki, Japan Meteorological Agency, 1-3-4 Otemachi, Chiyoda-ku, Tokyo 100-8122, Japan.
E-mail: tsuyuki@met.kishou.go.jp

Although data assimilation stemmed out from the necessity of providing initial conditions for numerical weather prediction (NWP), it has become an indispensable tool for meteorological researches by providing high-quality four-dimensional datasets of the atmosphere. Datasets from operational data assimilation systems at NWP centers had been widely used until reanalysis datasets appeared in the late 1990s after a proposal of long-term reanalysis by Bengtsson and Shukla (1988). The National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) released the NCEP/NCAR R1 reanalysis (Kalnay et al. 1996), and the European Centre for Medium-Range Forecasts (ECMWF) released the ERA-15 reanalysis (Gibson et al. 1997). They were upgraded to NCEP/DOE R2 (Kanamitsu et al. 2002) and ERA-40 (Uppala et al. 2005) respectively, with much better quality. Those global reanalysis datasets have been contributing much to climatological and meteorological studies in planetary and synoptic scales. In addition, the Japa-

nese 25-year Reanalysis (JRA-25) has been completed and become available to the meteorological community (Onogi et al. 2007). On the other hand, progress of data assimilation for mesoscale and cloud-resolving models made it possible to use data assimilation methods for investigating the mechanism of mesoscale phenomena such as a mesoscale convective system (e.g., Kawabata et al. 2007).

The operational implementation of NWP was started in the 1950s. The procedure for preparing initial conditions from observational data was called objective analysis in contrast to subjective analysis, because it was conducted numerically by the aid of a computer. The numerical schemes of the first generation of objective analysis were the function fitting (Panofsky 1949; Gilchrist and Cressman 1954), the method of successive corrections (Bergthorsson and Doos 1955; Cressman 1959), and the optimal interpolation (Eliassen 1954; Gandin 1963). Available observational data at that time were primarily from radiosonde and surface observations at synoptic times, and those three methods were procedures for two-dimensional interpolation of observational data into regular grid points of numerical models. The optimal interpolation (OI) is based on the statistical estimation theory in contrast to the other two methods, and gained popularity in the 1970s with sophistication to multivariate three-dimensional interpolation. According to Table 5.1 of Gustafsson (1981), seven out of ten operational NWP centers adopted OI, which is also called statistical interpolation, for objective analysis at that time.

It is to be noted that Gilchrist and Cressman (1954) suggested the use of short-range forecasts from numerical models as the first guess for objective analysis. This procedure became almost universal in the 1960s. Thompson (1961) noted that information from observations in data-rich areas propagates to data-void areas through the first guess during a data assimilation cycle. Those were important steps to the concept of data assimilation. Another important concept is initialization. The use of primitive equations models in NWP in stead of filtered equation models started in the 1960s, and it needed initialization procedures to remove spurious gravity waves from analyzed fields. The Nonlinear normal mode initialization introduced by Machenhauer (1977) and Baer and Tribbia (1977) were widely used at operational NWP centers.

The meteorological observation network was getting more complex by containing aircraft reports, drifting buoys, geostationary satellites, polar-orbiting satellites, and so on. The Global Weather Experiment called the First GARP Global Experiment (FGGE) in 1979 made an epoch-making contribution to the evolution of the observation network. A lot of asynoptic observational data such as vertical temperature sounding by polar-orbiting satellites were a new challenge to objective analysis. To meet this challenge, a new form of objective analysis and initialization procedures was conceived. It was called continuous data assimilation. Charney et al. (1969) proposed that numerical models be used to assimilate such asynoptic data in the following way. A numerical model is integrated forward in time from some preceding time up to the analysis time, and asynoptic data are inserted into the model at their observation times during the course of time integration. Their method is known as direct insertion, and another method proposed is nudging (Hoke and Anthes 1976). The continuous data assimilation more clearly articulates the basic concept of data assimilation, but both of the direct insertion and the nudging are rather empirical methods and lack a theoretical basis of statistical estimation.

New advances of data assimilation methods emerged from the 1980s, when considerable attention began to be paid to the four-dimensional variational assimilation (4D-Var) and the Kalman filter. Those assimilation methods use a flow-dependent background error covariance for estimating the atmospheric state, and can assimilate indirect observational data such as satellite radiances without transforming them into analysis variables. Those two points are major advantages over OI, in which a statistically estimated background error covariance is used and observational data that can be assimilated are limited to observations of analysis variables. Furthermore, 4D-Var and the Kalman filter are continuous data assimilation methods based on the statistical estimation theory.

The application of calculus of variations to meteorological analysis was first studied by

Sasaki (1958), and it was extended to include the time dimension by Sasaki (1969, 1970) and Thompson (1969). This extension may be regarded as a prototype of 4D-Var, but the Euler-Lagrange equations resulted from variational problems are difficult to solve for large-scale problems. The adjoint method introduced by Lions (1971) and Marchuk (1975a, b) made it possible to efficiently solve those problems, and application of the adjoint method to meteorological data assimilation was started in the 1980s. Its theoretical aspects were presented by Le Dimet and Talagrand (1986) and Talagrand and Courtier (1987).

The first operational implementation of variational methods to NWP was realized in June 1991, when the National Meteorological Center (NMC) of the United States adopted three-dimensional variational assimilation (3D-Var) for global analysis (Parrish and Derber 1992). Although 3D-Var does not include the time dimension, it has an advantage over OI in common with 4D-Var; any observed variables that are simulated by a numerical model can be assimilated in principle. The NMC started the direct assimilation of satellite radiance data in October 1995 by making use of this advantage and improved the quality of global analysis (Derber and Wu 1998).

Results from 4D-Var experiments with large-scale numerical models were published in the early 1990s (Thepaut et al. 1991; Navon et al. 1992; M. Zupanski 1993). Thepaut et al. (1993) and Andersson et al. (1994) demonstrated the ability of 4D-Var to generate flow-dependent and baroclinic structure functions in meteorological analysis. An incremental approach proposed by Courtier et al. (1994) made it possible to significantly reduce the computational costs of 4D-Var. An operational 4D-Var system was first introduced at ECMWF for global analysis in November 1997 (Rabier et al. 2000; Mahfouf and Rabier 2000; Klinker et al. 2000). The first application of 4D-Var to operational mesoscale analysis was realized by the Japan Meteorological Agency (JMA) in March 2002 (Ishikawa and Koizumi 2002). 4D-Var is getting the most popular data assimilation method for NWP, and five operational NWP centers currently adopt 4D-Var for global analysis. JMA implemented a global 4D-Var system in February 2005 (Kadowaki 2005).

The Kalman filter was introduced by Kalman (1960) to linear systems in discrete time and extended by Kalman and Bucy (1961) to systems in continuous time. This approach has been widely applied to signal processing and aerospace problems. A concerted attempt for applying the Kalman filter to meteorological data assimilation was started in the 1980s by several investigators such as Ghil et al. (1981) and Parrish and Cohn (1985). Although its heavy computational costs make it difficult to directly apply the Kalman filter to meteorological data assimilation, it has provided a reliable theoretical basis for development of data assimilation (e.g., Daley 1992a, b). Suboptimal Kalman filters with less computational costs required have been proposed by several investigators (e.g., Evensen 1994; Cohn and Todling 1995; Fisher 1998a; Pham et al. 1998). Among them the ensemble Kalman filter (EnKF) proposed by Evensen (1994) has been attracting much attention in the meteorological community in the 2000s as an advanced data assimilation method that is much easier to implement than 4D-Var. EnKF also generates a set of initial conditions for ensemble prediction. The Canadian Meteorological Centre (CMC) first introduced EnKF to its operational ensemble prediction system in January 2005 after investigations by Houtekamer et al. (2005).

This review paper is not intended as a comprehensive review on data assimilation methods that are referred to in the above brief history, since there are several excellent reviews available on this subject (e.g., Bengtsson et al. 1981; Hollingsworth 1987; Daley 1991; Ghil and Malanotte-Rizzoli 1991; Ghil et al. 1997; Bouttier and Courtier 1999; Kalnay 2003; Lewis et al. 2006; Evensen 2006). The primary purpose is to concisely present the basic concept of 4D-Var and EnKF, which currently draw much attention in the meteorological community. Although the quality control of observational data and the initialization are important ingredients of data assimilation systems, they are out of scope of this review paper. Sections 2 and 3 are devoted to presentation of the theoretical background and implementation of 4D-Var and EnKF, respectively. Recent research results in Japan on those methods are also reviewed in the two sections, especially on mesoscale applications of 4D-Var and tests of

the local ensemble transform Kalman filter (LETKF). Section 4 gives summary and concluding remarks, including a comparison of the two methods. Since a technical outline of the mesoscale 4D-Var system of JMA has not been well documented in the literature, it is presented in Appendix. A unified notation in atmospheric and ocean data assimilation proposed by Ide et al. (1997) is adopted where appropriate.

2. Variational assimilation

2.1 Basic formulation

In the original variational method proposed by Sasaki (1958), a cost function consisted of a weighted summation of squared differences between analysis variables and observational data and additional constraint terms. The optimal estimate of the analysis variables was obtained from the minimum point of the cost function. A major advantage of the variational method was thought to be its easiness to take account of dynamical balance and smoothness by adding appropriate constraint terms. Lorenc (1986) and others noted that the variational assimilation is based on the Baysian probability theory. The use of a quadratic cost function is justified when probability density functions (PDFs) of background and observational errors are Gaussian.

a. 3D-Var

The formulation of 3D-Var represents the basic concept of the variational assimilation. Let $\mathbf{x}$, $\mathbf{x}^b$, and $\mathbf{y}^o$ denote the column vectors of the state variables of a numerical model at an analysis time, the first guess of $\mathbf{x}$, and the observational data at the analysis time, respectively. The variational method gives an optimal estimate of the state variables $\mathbf{x}^a$ as the maximum a posteriori (MAP) estimate:

$$\mathbf{x}^a = \arg\max_{\mathbf{x}} \; p(\mathbf{x}\,|\,\mathbf{x}^b, \mathbf{y}^o), \tag{1}$$

where $p(\mathbf{x}\,|\,\mathbf{x}^b, \mathbf{y}^o)$ is the conditional PDF of $\mathbf{x}$ given $\mathbf{x}^b$ and $\mathbf{y}^o$. Application of the Bayes' theorem and an assumption of statistical independence between the errors of the first guess and observational data leads to the following estimate:

$$\mathbf{x}^a = \arg\max_{\mathbf{x}} [p(\mathbf{x}^b\,|\,\mathbf{x}) p(\mathbf{y}^o\,|\,\mathbf{x}) p(\mathbf{x})], \tag{2}$$

where $p(\mathbf{x})$ is the climatological PDF of $\mathbf{x}$. For the sake of convenience a cost function $J(\mathbf{x})$ defined as the minus logarithm of the term in the bracket of Eq. (2) is introduced.

$$J(\mathbf{x}) = -\log p(\mathbf{x}^b\,|\,\mathbf{x}) - \log p(\mathbf{y}^o\,|\,\mathbf{x}) - \log p(\mathbf{x}). \tag{3}$$

The optimal estimate is given by the minimum point of the cost function.

If the first two conditional PDFs on the right hand side of Eq. (3) are Gaussian and the errors in $\mathbf{x}^b$ and $\mathbf{y}^o$ do not have bias, the following expression of the cost function is obtained with irrelevant constant terms omitted.

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}[H(\mathbf{x}) - \mathbf{y}^o]^T \mathbf{R}^{-1} [H(\mathbf{x}) - \mathbf{y}^o] + J_c(\mathbf{x}), \tag{4}$$

where $\mathbf{B}$ and $\mathbf{R}$ are the error covariance matrices of the first guess (background) and observational data, respectively, and H the observation operator to convert state variables into observed variables and to interpolate from model grid points to observation points. The observation error consists of instrument error, representativeness error, and transformation error if observed variables are different from state variables. The first and second terms on the right hand side of Eq. (4) are referred to as a background term and an observation term, respectively. The third term $J_c(\mathbf{x})$ is called a penalty term, which corresponds to the last term of Eq. (3). Since the functional form of $p(\mathbf{x})$ is not well known, the main function of the penalty term is to suppress high-frequency gravity waves in analyzed fields.

The assumption of Gausian PDFs seems to be theoretically justified from the Central Limit Theorem if error statistics do not strongly depend on the atmospheric state, but its validity should be examined empirically for each state variable and observational data. For instance, it is known that the error distribution of specific humidity is far from Gaussian, and that relative humidity is a better choice for moisture analysis from a statistical point of view (Dee and Da Silva 2003). In the mesoscale 4D-Var system of JMA a non-Gaussian PDF is adopted

for observation error of precipitation data (Koizumi et al. 2005; see Appendix). Another point to be noted in the above derivation is that although the penalty term is usually regarded as an ad hoc term introduced for a practical reason it has a statistical basis. However, the error of the climatological estimate is usually much larger than the errors of the first guess and observational data, and the linear dynamical balance among state variables is represented in the background error covariance (e.g., Parrish and Derber 1992; Derber and Bouttier 1999). Therefore, the penalty term may be safely removed from the cost function of 3D-Var, and then the MAP estimate is reduced to the maximum likelihood estimate.

As mentioned in Section 1, one of the advantages of the variational assimilation over OI is that the observation operator is not limited to spatial interpolation. The variational method makes it possible to directly assimilate various kinds of observational data such as radiances from satellite observations (Derber and Wu 1998), radio refractivity from GPS occultation observations (Zou et al. 1995), precipitable water from ground-based GPS measurements (Kuo et al. 1996), and radial wind from Doppler radar (Sun and Crook 1994). Another advantage is that analysis fields are dynamically well balanced so that the analysis is not much modified by initialization to eliminate gravity wave noise. Figure 1 compares the differences between initialized fields and analysis fields for vertical p-velocity at 850 hPa from the JMA global analysis with 3D-Var and OI. Shaded areas show modifications by the initialization by more than 2 hPa h^{-1}. Though the OI analysis is modified almost everywhere, the 3D-Var analysis is not mush modified except over mountainous regions and around two typhoons near Taiwan and south of Japan. The OI method uses only a small set of observational data near a grid point to be analyzed, and it is equivalent to solve a minimization problem similar to (4) very locally, so that dynamical balance is not well secured.

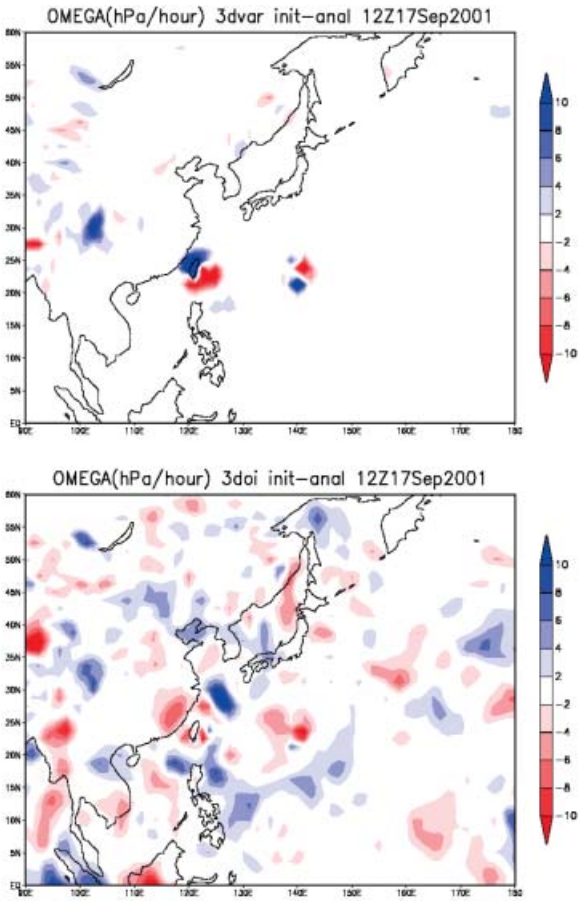

Fig. 1. Changes in vertical p-velocity (unit: hPa h^{-1}) at 850 hPa when nonlinear normal mode initialization is applied to global analyses by 3D-Var (top) and OI (bottom). Analysis time is 12 UTC 17 September 2001. (Adapted from Takeuchi and Tsuyuki 2002)

b. 4D-Var

4D-Var seeks the MAP estimate for the state variables of a numerical model during a certain period that is called an assimilation window. Although the formulation of 4D-Var is obtained from a simple extension of that of 3D-Var, there is a distinguishing advantage of 4D-Var that the full model dynamics is used for analysis. The OI and 3D-Var conduct multivariate analysis by using linear dynamical balance conditions such as geostrophic balance. However, the linear balance is a crude approximation of the model dynamics, and the use of model dynamics itself is more appropriate.

The numerical model is written as

$$\mathbf{x}_i = M_{i0}(\mathbf{x}_0) \quad (i = 1, \ldots, n), \tag{5}$$

where subscripts denote time levels, $\{M_{i0}\}$ the prediction operator of the model, and $\mathbf{x}_0$ and

$\mathbf{x}_n$ the state variables at the beginning and end of the assimilation window, respectively. A standard cost function of 4D-Var is given under a couple of assumptions; observational data are available at all the time levels, the errors of observational data at different time levels are uncorrelated with each other, and observed variables are simulated from state variables at the same time level as observations. If the governing equation of the numerical model (5) is used as a strong constraint, the cost function for Gaussian PDFs is given by

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \sum_{i=0}^{n}\frac{1}{2}[H_i(\mathbf{x}_i) - \mathbf{y}_i^o]^T\mathbf{R}_i^{-1}[H_i(\mathbf{x}_i) - \mathbf{y}_i^o] + J_c(\mathbf{x}_0, \ldots, \mathbf{x}_n). \quad (6)$$

The cost function is a function of the state variable at the beginning of the assimilation window only. In other words, the control variable of minimization is $\mathbf{x}_0$. All of the assumptions made above are easily relaxed to apply 4D-Var to more general cases such as a non-Gaussian PDF, serially-correlated observation error and assimilation of accumulated precipitation data.

Since dynamical balance represented in the background error covariance is imposed on $\mathbf{x}_0$ only, the penalty term is necessary for 4D-Var to suppress gravity wave noise that may be caused by observation error. Several types of the penalty terms are proposed by Courtier and Talagrand (1990), Zou et al. (1993c) and Polavarapu et al. (2000). Another point to be noted is that lateral boundary conditions for limited-area models could be modified so as to be more consistent with observational data in the minimization process (Zou and Kuo 1996; see Appendix). More generally if a numerical model has ambiguous parameters, optimum values of them could be estimated. For those purposes the cost function (6) is regarded as a function of the parameters as well as $\mathbf{x}_0$, and the minimum point of the cost function gives an optimal estimate of those parameters. If the first guess of parameters is available, a background term for them may be added to the cost function.

One of the advantages of 4D-Var over 3D-Var is that flow-dependent background error covariance is implicitly used in analysis. This fact is easily shown when the numerical model and the observation operator are linear and the penalty term is omitted. For this purpose observational data are assumed to be available only at the end of assimilation window. The analysis from 3D-Var at that time level is given by minimizing the cost function (4) without the penalty term:

$$\begin{aligned}\mathbf{x}_n^a &= \mathbf{x}_n^b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} \times [\mathbf{y}_n^o - H_n(\mathbf{x}_n^b)] \\ &= \mathbf{x}_n^b + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \times [\mathbf{y}_n^o - H_n(\mathbf{x}_n^b)], \end{aligned} \quad (7)$$

where $\mathbf{H}$ is the tangent linear operators of H_n. The corresponding analysis from 4D-Var is given by minimizing the cost function (6) without the penalty term and by conducting time integration:

$$\begin{aligned}\mathbf{x}_n^a &= M_{n0}(\mathbf{x}_0^a) \\ &= \mathbf{x}_n^b + \mathbf{M}\mathbf{B}\mathbf{M}^T\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{M}\mathbf{B}\mathbf{M}^T\mathbf{H}^T)^{-1} \times [\mathbf{y}_n^o - H_n(\mathbf{x}_n^b)], \end{aligned} \quad (8)$$

where $\mathbf{M}$ is the tangent linear operator of M_{n0}. A comparison of Eqs. (7) and (8) shows that the background error covariance matrix $\mathbf{B}$ in 3D-Var is replaced by $\mathbf{M}\mathbf{B}\mathbf{M}^T$ in 4D-Var. The latter matrix is the background error covariance evolved by the tangent linear model with an initial value $\mathbf{B}$.

Figure 2 compares the analysis increments (differences between analysis and first guess) of wind from a single pseudo observation in mesoscale analyses with 3D-Var and 4D-Var. The observation is put at the end of a 3-hour assimilation window of 4D-Var, and the analysis increment at that time is compared. The observation site is located upstream of a wind shear line seen in the first guess, and the observed wind is more westerly than the first guess. The analysis increment of 3D-Var shows a pattern that may be inferred from geostrophic balance, while the analysis increment of 4D-Var suggests that the analyzed shear line is shifted eastward. The 4D-Var analysis seems more plausible than 3D-Var.

Since a numerical model generally has a bias and random errors, it is more appropriate to use model equations as a weak constraint in 4D-Var. If a model with a two-time-level

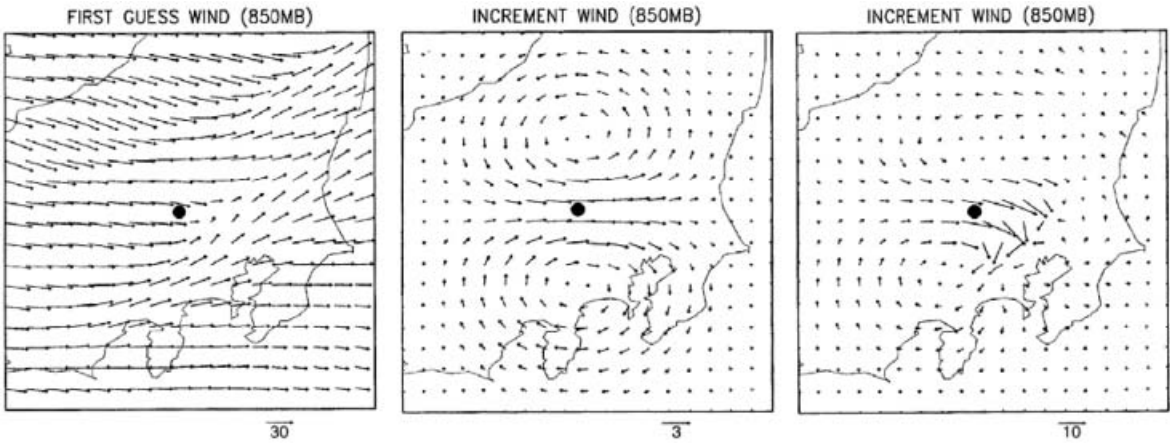

Fig. 2. First guess (left) and analysis increments of wind at 850 hPa (unit: m s^{-1}) by 3D-Var (middle) and 4D-Var (right) from a single pseudo observation of wind at 850 hPa located at a closed dot. (Adapted from Tsuyuki et al. 2002)

scheme has a bias, the model equation is written as

$$\mathbf{x}_i = M_{i-1}(\mathbf{x}_{i-1}) + \mathbf{c} \quad (i = 1, \ldots, n). \tag{9}$$

The optimum value of $\mathbf{c}$ is estimated in the process of minimization as mentioned above (Derber 1989). A numerical model including random errors $\{\varepsilon_i\}$ may be written as

$$\mathbf{x}_i = M_{i-1}(\mathbf{x}_{i-1}) + \varepsilon_{i-1} \quad (i = 1, \ldots, n). \tag{10}$$

If the model error is represented by a white Gaussian sequence with no bias, a cost function is given by

$$\begin{aligned} J(\mathbf{x}_0; \varepsilon_0, \ldots, \varepsilon_{n-1}) &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) \\ &\quad + \sum_{i=0}^{n} \frac{1}{2}[H_i(\mathbf{x}_i) - \mathbf{y}_i^o]^T \mathbf{R}_i^{-1}[H_i(\mathbf{x}_i) - \mathbf{y}_i^o] \\ &\quad + \sum_{i=0}^{n-1} \frac{1}{2}\varepsilon_i^T \mathbf{Q}_i^{-1}\varepsilon_i + J_c(\mathbf{x}_0, \ldots, \mathbf{x}_n), \end{aligned} \tag{11}$$

where $\{\mathbf{Q}_i\}$ is the error covariance matrices of $\{\varepsilon_i\}$. More general random error such as a first-order auto-regressive Gaussian sequence may be adopted (Zupanski 1997). A problem with 4D-Var with a weak constraint is that the number of variables to be estimated is much increased in comparison with 4D-Var with a strong constraint. This problem may be alleviated by reducing the spatial resolution or update frequency of errors. However, the operational 4D-Var systems at NWP centers have not yet adopted 4D-Var with a weak constraint.

2.2 *Adjoint method*

The minimum point of the cost function is numerically searched with large-scale minimization algorithms such as the quasi-Newton method and the conjugate gradient method (e.g., Navon and Legler 1987; Zou et al. 1993b; Fisher 1998b). Those methods are iterative algorithms and need the gradient of the cost function as well as its value at each iteration. The adjoint method provides an efficient algorithm to calculate the gradient.

The essence of the adjoint method may be concisely presented for an ordinary differential equation system:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(t, \mathbf{x}(t); \mathbf{c}), \tag{12}$$

where t denotes time, and $\mathbf{c}$ is a vector of parameters that is included for generalization. The goal here is to calculate the linear sensitivity of the following response function to the initial condition and the parameters.

$$R(\mathbf{x}(0), \mathbf{c}) = \int_0^T r(t, \mathbf{x}(t); \mathbf{c})\, dt. \tag{13}$$

The adjoint method gives the following result.

$$\nabla_{\mathbf{x}(0)} R = \mathbf{p}(0), \tag{14}$$

$$\nabla_{\mathbf{c}} R = \int_0^T \left(\frac{\partial \mathbf{F}}{\partial \mathbf{c}}\right)^T \mathbf{p}(t)\, dt + \int_0^T \frac{\partial r}{\partial \mathbf{c}} dt, \tag{15}$$

where $\mathbf{p}(t)$ is obtained by integrating backward in time the adjoint equation of Eq. (12) with a forcing term:

$$-\frac{d\mathbf{p}}{dt} = \left(\frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right)^T \mathbf{p}(t) + \frac{\partial r}{\partial \mathbf{x}}, \tag{16}$$

from the "initial" condition

$$\mathbf{p}(T) = \mathbf{0}. \tag{17}$$

The above result is derived by making use of the tangent linear equation of Eq. (12):

$$\frac{d\delta\mathbf{x}}{dt} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \delta\mathbf{x}(t) + \frac{\partial \mathbf{F}}{\partial \mathbf{c}} \delta\mathbf{c}. \tag{18}$$

A modification of the first variation of Eq. (13) by using Eqs. (16)–(18) and integration by parts leads to Eqs. (14)–(15) as follows:

$$\begin{aligned}
\delta R(\mathbf{x}(0), \mathbf{c}) &= \int_0^T \left(\delta\mathbf{x}(t)^T \frac{\partial r}{\partial \mathbf{x}} + \delta\mathbf{c}^T \frac{\partial r}{\partial \mathbf{c}}\right) dt \\
&= \int_0^T \delta\mathbf{x}(t)^T \left[-\frac{d\mathbf{p}}{dt} - \left(\frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right)^T \mathbf{p}(t)\right] dt \\
&\quad + \delta\mathbf{c}^T \int_0^T \frac{\partial r}{\partial \mathbf{c}} dt \\
&= -[\delta\mathbf{x}(t)^T \mathbf{p}(t)]_0^T \\
&\quad + \int_0^T \left(\frac{d\delta\mathbf{x}}{dt} - \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \delta\mathbf{x}(t)\right)^T \mathbf{p}(t)\, dt \\
&\quad + \delta\mathbf{c}^T \int_0^T \frac{\partial r}{\partial \mathbf{c}} dt \\
&= \delta\mathbf{x}(0)^T \mathbf{p}(0) \\
&\quad + \delta\mathbf{c}^T \left[\int_0^T \left(\frac{\partial \mathbf{F}}{\partial \mathbf{c}}\right)^T \mathbf{p}(t)\, dt + \int_0^T \frac{\partial r}{\partial \mathbf{c}} dt\right].
\end{aligned} \tag{19}$$

Equations (14)–(15) provide the linear sensitivity of the response function to a lot of parameters as well as the initial condition (Caccusi 1981). It is obtained from one forward integration of the nonlinear model and one backward integration of the adjoint model. On the other hand, a pair of forward integrations of the nonlinear model with different values of a certain parameter provides the nonlinear sensitivity of various output variables to the parameter. Therefore, those two sensitivity analysis methods may be regarded as complementary to each other. The adjoint sensitivity has been applied to meteorological researches by several investigators (e.g., Hall et al. 1982; Errico and Vukicevic 1992; Zou et al. 1993a). It is also used to estimate sensitive areas for targeted observations (Langland and Rohaly 1996).

Application of the adjoint method to 4D-Var in discrete time is conducted in the same way as in continuous time. For a discrete dynamical system with a two-time-level scheme:

$$\mathbf{x}_i = M_{i-1}(\mathbf{x}_{i-1}) \quad (i = 1, \ldots, n), \tag{20}$$

the gradient of the cost function (6) with respect to the initial condition is given by

$$\nabla_{\mathbf{x}_0} J = \mathbf{p}_0, \tag{21}$$

where $\mathbf{p}_0$ is calculated by the following adjoint equation for the discrete system (20):

$$\mathbf{p}_i = \mathbf{M}_i^T \mathbf{p}_{i+1} + \frac{\partial J}{\partial \mathbf{x}_i} \quad (i = 0, \ldots, n), \tag{22}$$

with the "initial" condition:

$$\mathbf{p}_{n+1} = \mathbf{0}. \tag{23}$$

The matrix $\mathbf{M}_i$ is the tangent linear operator of M_i. If a model or a cost function explicitly contains parameters as in Eq. (9) or Eqs. (10)–(11), the gradient of the cost function with respect to those parameters is given by an equation similar to Eq. (15). For the latter case, the following gradient is obtained:

$$\nabla_{\varepsilon_i} J = \mathbf{p}_{i+1} + \frac{\partial J}{\partial \varepsilon_i} \quad (i = 0, \ldots, n-1). \tag{24}$$

The adjoint method needs the adjoint of a numerical model, which requires linearization of the numerical model around a time-evolving basic state calculated by the model. A computer code of the adjoint model is obtained by transformation of a computer code of the tangent linear model. The rules for generating adjoint codes are available in several documents such as Yang and Navon (1995) and Giering and Kaminski (1998). However, generation of adjoint codes by hand is time-consuming and error-

prone. Recently several automatic differentiation tools such as Odyssee (Rostaing et al. 1993), ADIFOR (Bischof et al. 1996) and TAMC (Giering 1999) are developed and applied to atmospheric models.

The adjoint method assumes differentiability of the model equations with respect to the state variables. Parameterization schemes of physical processes used in atmospheric models are far more nonlinear than dynamical processes, and sometimes contain on/off switches that make a cost function non-differentiable or discontinuous. A number of investigators attempted to deal with the non-differentiability or discontinuity associated with on/off switches. Proposed technical modifications include neglecting perturbations to on/off switches (Zou et al. 1993d), vertical smoothing to the layers of transitions (D. Zupanski 1993), spline fitting over discontinuities (Verlinde and Cotton 1993), and removing several types of discontinuities from parameterization schemes (Tsuyuki 1996a, b). It is to be noted that for practical applications to 4D-Var the accuracy of the tangent linear approximation should be checked for finite perturbations comparable in magnitude to the typical size of uncertainties. Errico and Raeder (1999) demonstrated that straightforward linearization of some parameterization schemes leads to noise problems. The current convention in operational 4D-Var systems is to develop linear physical parameterizations that behave more regularly than straightforward linearization through simplifying original parameterization schemes (e.g., Janiskova et al. 1999; Mahfouf 1999).

2.3 *Incremental approach*

The 4D-Var method requires much more computational cost to seek the minimum point of the cost function than OI or 3D-Var, because it needs tens of iterations containing the forward integration of a nonlinear model and the backward integration of an adjoint model in an assimilation window. This problem was practically solved by adopting an incremental approach proposed by Courtier et al. (1994). The current operational 4D-Var systems at NWP centers use this approach.

The cost function (6) is equivalent to its incremental formulation to the first order of the tangent linear approximation:

$$J(\delta\mathbf{x}_0) = \frac{1}{2}(\delta\mathbf{x}_0 - \delta\mathbf{x}_0^b)^T\mathbf{B}^{-1}(\delta\mathbf{x}_0 - \delta\mathbf{x}_0^b) + \sum_{i=0}^{n}\frac{1}{2}(\mathbf{H}_i\delta\mathbf{x}_i - \mathbf{d}_i)^T\mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}_i - \mathbf{d}_i) + J_c(\delta\mathbf{x}_0, \ldots, \delta\mathbf{x}_n), \quad (25)$$

where

$$\delta\mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_i^g, \quad \delta\mathbf{x}_0^b = \mathbf{x}_0^b - \mathbf{x}_0^g, \quad \mathbf{d}_i = \mathbf{y}_i^o - H_i(\mathbf{x}_i^g). \quad (26)$$

$\{\mathbf{x}_i^g\}$ is the first guess in iteration, and the linearization is conducted around $\{\mathbf{x}_i^g\}$. To reduce the computational cost of minimization, a linear simplification operator $\mathbf{S}$ is introduced to act on $\delta\mathbf{x}_0$:

$$\delta\mathbf{w}_0 = \mathbf{S}\delta\mathbf{x}_0. \quad (27)$$

The operator $\mathbf{S}$ may be spectral truncation to large-scale or projection to selected modes so that the minimization problem is solved in a low dimensional space. Since the inverse of $\mathbf{S}$ does not exist, a generalized inverse $\mathbf{S}^{-I}$ is needed for inversion. Then the approximated minimization problem (25) may be further approximated by

$$J(\delta\mathbf{w}_0) = \frac{1}{2}(\delta\mathbf{w}_0 - \delta\mathbf{w}_0^b)^T\mathbf{B}_w^{-1}(\delta\mathbf{w}_0 - \delta\mathbf{w}_0^b) + \sum_{i=0}^{n}\frac{1}{2}(\mathbf{G}_i\delta\mathbf{w}_i - \mathbf{d}_i)^T\mathbf{R}_i^{-1}(\mathbf{G}_i\delta\mathbf{w}_i - \mathbf{d}_i) + J_c(\delta\mathbf{w}_0, \ldots, \delta\mathbf{w}_n), \quad (28)$$

where

$$\mathbf{B}_w = \mathbf{S}\mathbf{B}\mathbf{S}^T, \quad \mathbf{G}_i = \mathbf{H}_i\mathbf{S}^{-I}, \quad (29)$$

$$\delta\mathbf{w}_i = \mathbf{S}\mathbf{M}_{i0}\mathbf{S}^{-I}\delta\mathbf{w}_0 \approx \mathbf{L}_{i0}\delta\mathbf{w}_0. \quad (30)$$

$\mathbf{M}_{i0}$ is the tangent linear operator of M_{i0}, and $\mathbf{L}_{i0}$ is a low-dimensional linear operator with regularized linear physics mentioned in the previous subsection. It may be possible to replace $\{\delta\mathbf{w}_i\}$ in Eq. (30) by the difference of two nonlinear forward integrations in the low-dimensional space. This procedure is adopted in the mesoscale 4D-Var system of JMA (see Appendix).

The analysis is obtained through

$$\mathbf{x}_0^a = \mathbf{x}_0^g + \mathbf{S}^{-I}\delta\mathbf{w}_0, \quad (31)$$

$$\mathbf{x}_i^a = M_{i0}(\mathbf{x}_0^a) \quad (i = 1, \ldots, n). \quad (32)$$

The resulting $\{\mathbf{x}_i^a\}$ given by Eqs. (31)–(32) may be regarded as an update of $\{\mathbf{x}_i^g\}$, and the solution to a new minimization problem (28) with the updated $\{\mathbf{x}_i^g\}$ will give more accurate analysis.

The incremental 4D-Var may be derived as a practical implementation of the extended Kalman filter for a finite time window, in which the full nonlinear model is used to evolve the mean of a posterior PDF while a linear model is used to evolve the covariances (Lorenc 2003a). The point here is that the linear model should be designed for proper evolution of the covariances, and that it is not necessarily tangent to the full nonlinear model. This derivation justifies the use of regularized linear physics in the incremental 4D-Var.

2.4 Ambiguity removal

Since the variational assimilation is based on the Baysian probability theory, there is no restriction to the functional form of PDFs. This feature makes it possible to remove ambiguities from observational data in the minimization process. For instance, in the global 4D-Var system of ECMWF, two ambiguous wind data from satellite scatterometer measurements are simultaneously assimilated by adding the following non-standard observation term to a cost function:

$$J_o^{(SCAT)}(\mathbf{x}) = \left[\frac{J_1(\mathbf{x})^4 J_2(\mathbf{x})^4}{J_1(\mathbf{x})^4 + J_2(\mathbf{x})^4}\right]^{1/4}, \tag{33}$$

where $J_1(\mathbf{x})$ and $J_2(\mathbf{x})$ are standard observation terms based on a Gaussian PDF for the two ambiguous winds (Stoffelen and Anderson 1997).

The variational quality control (Lorenc and Hammon 1988; Ingleby and Lorenc 1993; Andersson and Jarvinen 1999) is another good example of ambiguity removal. The quality control of observational data is conducted to exclude or modify data containing gross error (rough error) based on climatological knowledge, internal consistency, and comparison with first guess and other observational data. The gross error is caused by improperly calibrated instruments, incorrect registration, incorrect coding, or telecommunication errors. In a conventional method (e.g., Gandin 1988), the procedure of quality control is conducted before the assimilation. The variational method offers a new possibility of quality control; the step of comparison with the first guess and other observational data could be done in the process of minimization. Observational data that passed the climatological check and consistency check are assimilated by using the following conditional PDF in the cost function (3):

$$p(\mathbf{y}^o \,|\, \mathbf{x}) = a p_A(\mathbf{y}^o \,|\, \mathbf{x}) + (1-a) p_N(\mathbf{y}^o \,|\, \mathbf{x}), \tag{34}$$

where a is the probability of occurrence of gross error, and $p_A(\mathbf{y}^o \,|\, \mathbf{x})$ and $p_N(\mathbf{y}^o \,|\, \mathbf{x})$ are PDFs of observational data with and without gross error, respectively. The former PDF has much larger standard deviation than the other, and is often assumed to be a uniform distribution for simplicity. The latter PDF is usually assumed to be Gaussian. The variational quality control is operational in the global 4D-Var systems of ECMWF and CMC.

Furumoto et al. (2007) adopts a similar approach to retrieve a vertical humidity profile from wind profiler measurements with a one-dimensional variational method (1D-Var). The absolute value of the refractive index gradient is inferred from the echo power, and its sign have to be determined in the retrieval. The following cost function is introduced to deal with this ambiguity in sign.

$$\begin{aligned} J(\mathbf{x}) = {} & \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) \\ & - \sum_{k=1}^{K} \log\{p_k \exp[-J_k^+(\mathbf{x})] \\ & + (1-p_k)\exp[-J_k^-(\mathbf{x})]\} \\ & + J_o^{(GPS)}(\mathbf{x}), \end{aligned} \tag{35}$$

$$J_k^{\pm}(\mathbf{x}) = \frac{1}{2\sigma_k^2}(H_k(\mathbf{x}) \mp |m_k|^o)^2, \tag{36}$$

where p_k is the climatological probability of a positive refractive index gradient, $|m_k|^o$ the observed absolute value of the gradient at the k-th level, σ_k the standard deviation of its observation error, H_k the observation operator to simulate the refractive index gradient, K the number of vertical level, and $J_o^{(GPS)}$ the observation term for precipitable water data from ground-based GPS measurements. The state vector $\mathbf{x}$ consists of the vertical profiles of tem-

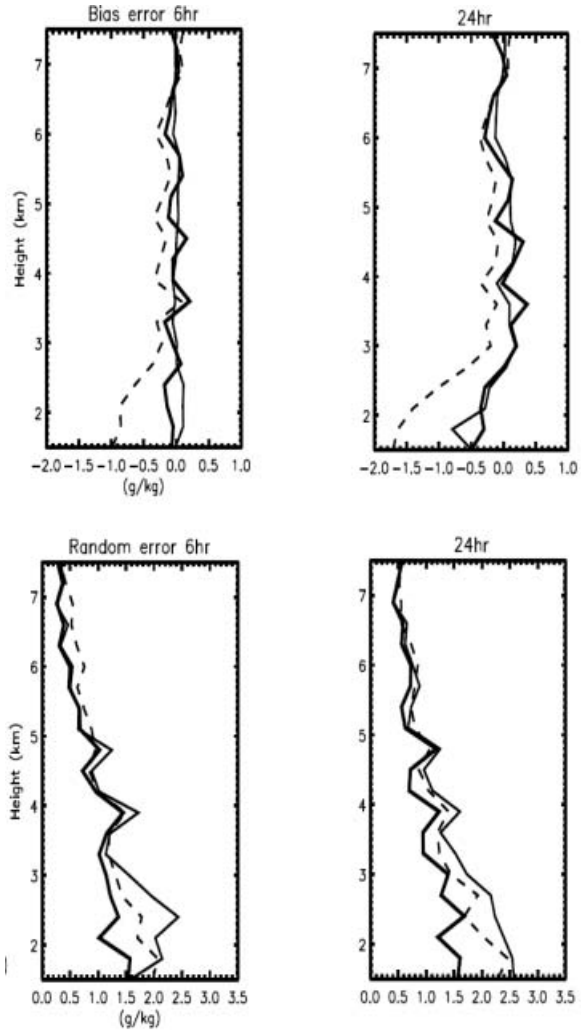

Fig. 3. Vertical profiles of mean bias errors (top) and random errors (bottom) in specific humidity analyzed by using the first guess derived from time-interpolation of 6-hourly (left) and 24-hourly (right) radiosonde observations. Thin-solid, dashed and thick-solid lines indicate the first guess, analysis by the conventional method and analysis by the variational method, respectively. (Adapted from Furumoto et al. 2007)

perature and relative humidity and the pressure at 1.5 km. Figure 3 compares the bias and random errors in specific humidity retrieved from the 1D-Var and a conventional method by Furumoto et al. (2003). The first guess used is radiosonde observations interpolated in time. It is found that the improvement over the conventional method is especially evident in the case of a large error in the first guess.

2.5 Mesoscale applications of 4D-Var

Experimental regional 4D-Var systems were developed for the NMC Eta model by M. Zupanski (1993) and for the Penn-State-NCAR MM5 model by Zou et al. (1995). However, the first operational 4D-Var system for a regional model was implemented in March 2002 at JMA. That system is operated 8 times a day to provide initial conditions for the operational mesoscale model (MSM) of JMA. A technical outline of this system is presented in Appendix. An extensive review on 4D-Var for mesoscale models is presented by Park and Zupanski (2003).

The initial condition of MSM with a horizontal resolution of 10 km had been prepared by a one-hour cycle assimilation system with OI and physical initialization (PI) before the introduction of 4D-Var. The PI was proposed by Krishnamurti et al. (1984) to improve NWP by assimilating precipitation data, and a PI method developed by Matsumura et al. (1997) was used for the initialization of MSM. The one-hour cycle assimilation system was executed for the 3-hour period just before the initial time. On the other hand, the mesoscale 4D-Var system adopted a 3-hour assimilation window, and directly assimilated precipitation data. It is found from Fig. 4 that the 4D-Var system has much benefit for precipitation forecasts over the previous system. Positive impacts were also seen in prediction of wind, temperature, and typhoon intensity (Tsuyuki et al. 2002).

The direct assimilation of precipitation data with 4D-Var was first attempted for regional models by Zupanski and Mesinger (1995) and Zou and Kuo (1996) and for a global model by Tsuyuki (1997), and they all demonstrated positive impacts. The JMA mesoscale 4D-Var system is the first operational 4D-Var that directly assimilates precipitation data. Since it is found that the statistics of departure of observation from the first guess is far from Gaussian, a non-standard observation term is used in the cost function (Koizumi et al. 2005; see Appendix). Figure 5 shows that assimilation of precipitation amount data improves precipitation forecasts throughout the whole forecast period of 18 hours for both of weak and moderate pre-

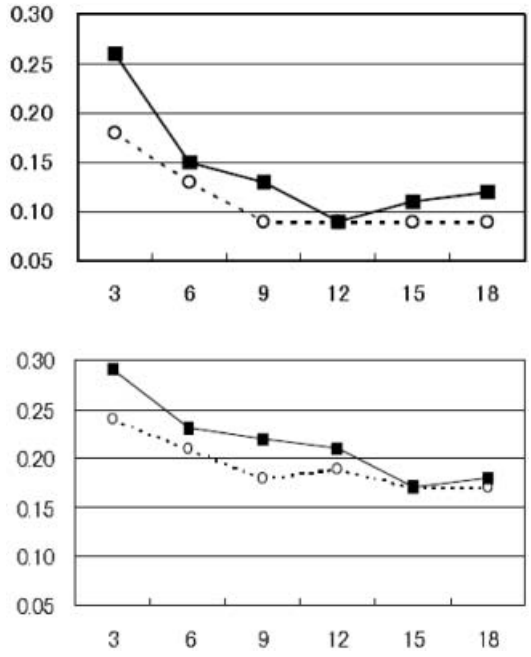

Fig. 4. Threat scores of 3-hour precipitation forecasts over Japan from MSM with 4D-Var (solid line) and OI and PI (dashed line) for June 2001 (top) and September 2001 (bottom). The threshold values are 10 mm on a 40 km grid, and abscissa is the forecast time in hour. (Adapted from Ishikawa and Koizumi 2002)

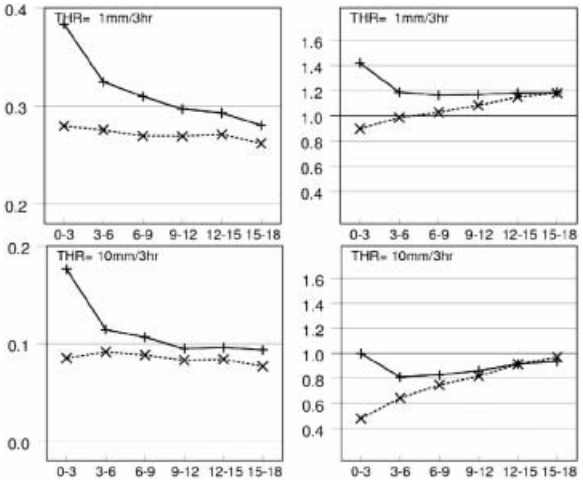

Fig. 5. Threat scores (left) and bias scores (right) of 3-hour precipitation forecasts over Japan from MSM with (solid line) and without (dashed line) precipitation assimilation for June 2001. The threshold values are 1 mm (top) and 10 mm (bottom) on a 10 km grid, and abscissa is the forecast time in hour. (Adapted from Koizumi et al. 2005)

cipitations. Impacts of precipitable water data from GPS ground receivers or satellite microwave imagers and radial wind data from Doppler radar were also investigated with the mesoscale 4D-Var system by Nakamura et al. (2004), Koizumi and Sato (2004), and Seko et al. (2004).

One of the problems with the current mesoscale 4D-Var system of JMA is that the assimilation model is different from the forecast model; the former model is a hydrostatic spectral model with a pressure-based vertical coordinate while the latter is a non-hydrostatic grid model (NHM) with microphysics and a geometrical height-based coordinate by Saito et al. (2006). A 4D-Var system based on NHM is being developed at JMA (Honda et al. 2005). The state variables of microphysics are not included in the control variables of the system. Figure 6 demonstrates a better performance of the experimental NHM-based 4D-Var than the operational 4D-Var in predicting a heavy rainfall event in Japan. This experimental system has a smaller analysis area than the operational one, and does not take an incremental approach nor assimilates precipitation data. Assimilation of precipitation data in the NHM-based 4D-Var is being tested with newly-developed adjoint codes of microphysics.

Kawabata et al. (2007) developed another NHM-based 4D-Var system for a cloud-resolving model with a horizontal resolution of 2 km. They applied their cloud-resolving 4D-Var to investigate the generation mechanism of a mesoscale convective system that caused heavy rainfall in a local area of Tokyo. Assimilated observational data are GPS-derived precipitable water, Doppler radar radial wind, and surface wind and temperature from the Automated Meteorological Data Acquisition System (AMeDAS). The assimilation window is set to a one-hour period covering the generating stage of the convective system. Figure 7 shows the analyzed vertical cross section of the convective system during the latter half period of the assimilation window. It is seen that air with high

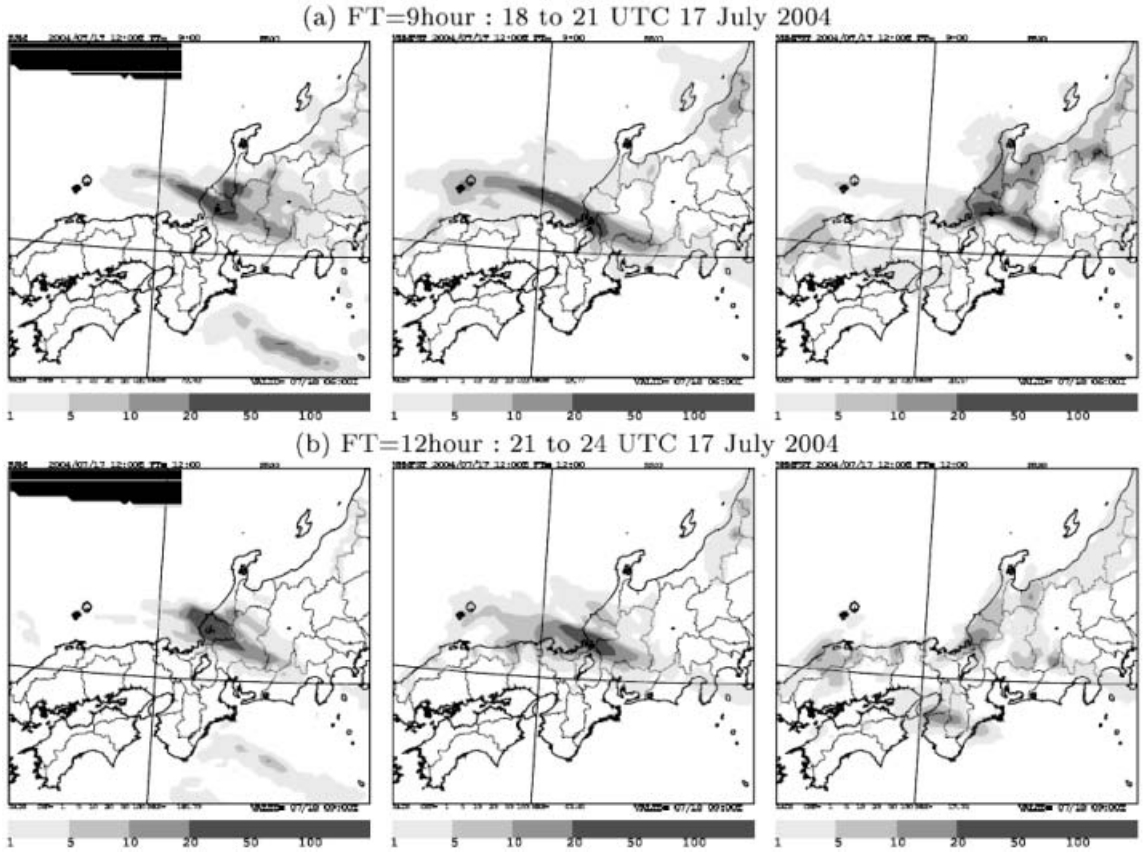

Fig. 6. Three-hour accumulated precipitation amounts (unit: mm) for (a) 18–21 UTC and (b) 21–24 UTC of 17 July 2004. The left panels are observations with data-void areas in the upper-left corners, the middle panels are prediction from NHM-based mesoscale 4D-Var, and the right panels are prediction from the operational mesoscale 4D-Var of JMA. NHM with a horizontal resolution of 10 km is used as a forecast model, and the initial time is 12 UTC 17 July 2004. (Courtesy of Y. Honda)

equivalent potential temperature was lifted by southerly sea breeze and cumulonimbus that caused heavy rainfall was generated. This study demonstrates the effectiveness of data assimilation to investigate mesoscale phenomena for which available observational data are quite limited.

3. Ensemble Kalman filtering

After first proposed by Evensen (1994), ensemble Kalman filtering (EnKF), a method of the data assimilation and the generation of ensemble perturbations, is now becoming a viable choice in the operational NWP. EnKF is advantageous in terms that EnKF analyzes the analysis errors, while other data assimilation schemes including variational methods cannot. Moreover, a main advantage of EnKF is its simple implementation and model independence, so that a wide range of applicative researches would be relatively easy. This feature would be most impressive and favorable to research communities if we consider the complexity of variational methods. In this section, a brief but thorough review of the theoretical background of EnKF is provided, followed by a review of practical applications by Miyoshi (2005) and Miyoshi and Yamane (2007).

3.1 Kalman filtering

EnKF is an approximation of Kalman filtering (KF, Kalman 1960). KF provides an algorithm to optimally control the system state and its error covariance, assuming the linear model

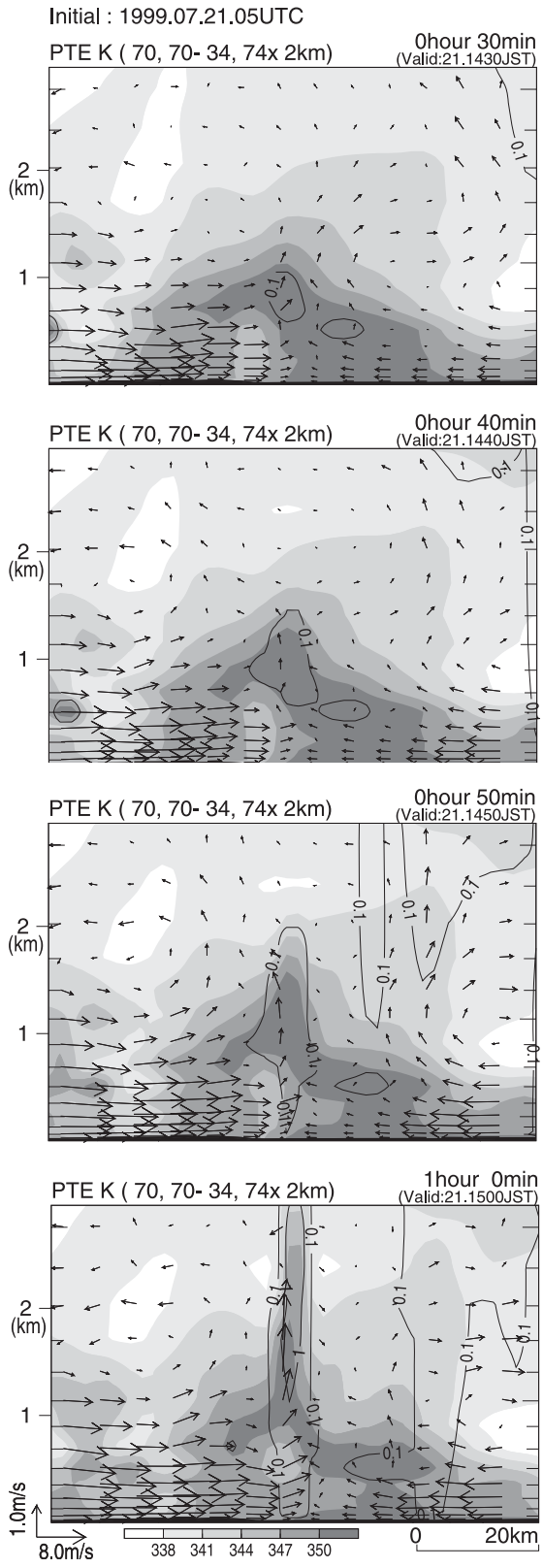

Fig. 7. Vertical cross sections in a north-south direction (left: south) of analyzed equivalent potential temperature (shade, unit: K), mixing ratio of total water (contours, unit: g/kg) and wind vectors projected on the cross section (arrows) at 1430, 1440, 1450 and 1500 JST of 21 July 1999 (from top to bottom). (Adapted from Kawabata et al. 2007)

and Gaussian PDF. Since the Gaussian PDF is parameterized by the mean and covariance, KF is regarded as the optimal data assimilation method for PDF itself. In this section, KF equations are derived in a simple manner. More mathematically precise derivations are found in Jazwinski (1970).

KF is composed of two parts: forecast and analysis. The forecast part consists of two equations:

$$\mathbf{x}_i^f = M(\mathbf{x}_{i-1}^a), \tag{37}$$

$$\mathbf{P}_i^f = \mathbf{M}\mathbf{P}_{i-1}^a\mathbf{M}^T + \mathbf{Q}, \tag{38}$$

where $\mathbf{x}$ and $\mathbf{P}$ denote the state vector and the error covariance matrix, respectively. The superscript indicates forecast (f) and analysis (a), and the subscript indicates time levels. Equation (37) means that a generally nonlinear model M evolves the analysis state $\mathbf{x}_{i-1}^a$ at time $i-1$ to the forecast state $\mathbf{x}_i^f$ at time i. Similarly, Eq. (38) evolves the error covariance matrix $\mathbf{P}$ in time. $\mathbf{Q}$ indicates the model error covariance term, which explains the unbiased system noise. If M is nonlinear, the model needs to be linearized around $\mathbf{x}_{i-1}^a$ to be applied to $\mathbf{P}$; $\mathbf{M}$ denotes the linearized model, a.k.a. tangent linear model (TLM). This nonlinear extension of KF is known as the extended KF (EKF).

The other part of KF is analysis. Suppose that the analysis state is given by a weighted mean:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}^o - H(\mathbf{x}^f)), \tag{39}$$

where the time is fixed, so the subscripts are omitted. Here, $\mathbf{y}^o$ denotes a vector composed of all observations, or a vector in the observational space. H denotes the generally nonlinear observational operator that maps a state vector onto the observational space. $\mathbf{K}$ gives the

weight between the forecast state and observations. Suppose the true state $\mathbf{x}^t$ is subtracted from both sides of Eq. (39):

$$\mathbf{x}^a - \mathbf{x}^t = \mathbf{x}^f - \mathbf{x}^t + \mathbf{K}[\mathbf{y}^o - H(\mathbf{x}^t) - (H(\mathbf{x}^f) - H(\mathbf{x}^t))]. \quad (40)$$

Let $\delta\mathbf{x}$ be the difference from the true state. Then, Eq. (40) is rewritten by:

$$\delta\mathbf{x}^a \approx \delta\mathbf{x}^f + \mathbf{K}(\delta\mathbf{y}^o - \mathbf{H}\delta\mathbf{x}^f), \quad (41)$$

where $\mathbf{H}$ denotes the linearized observational operator around $\mathbf{x}^f$. The approximation of Eq. (41) is due to the linearization of H. The error covariance is computed by

$$\begin{aligned}\mathbf{P}^a &= E(\delta\mathbf{x}^a(\delta\mathbf{x}^a)^T)\\ &\approx E([\delta\mathbf{x}^f + \mathbf{K}(\delta\mathbf{y}^o - \mathbf{H}\delta\mathbf{x}^f)]\\ &\quad \times [\delta\mathbf{x}^f + \mathbf{K}(\delta\mathbf{y}^o - \mathbf{H}\delta\mathbf{x}^f)]^T)\\ &= (\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T \quad (42)\end{aligned}$$

where $E(\bullet)$ denotes the expected value. Here, $\mathbf{R} = E[\delta\mathbf{y}^o(\delta\mathbf{y}^o)^T]$ denotes the observational error covariance matrix, and the cross terms such as $E[\delta\mathbf{y}^o(\mathbf{H}\delta\mathbf{x}^f)^T]$ are assumed to be zero, i.e., no correlation between the forecast state and observations. Since we assimilate observations to minimize analysis errors, we impose the trace of $\mathbf{P}^a$ (denoted by $tr(\mathbf{P}^a)$) to be minimum with the optimal $\mathbf{K}$. At the minimum, the differential by $\mathbf{K}$ is zero:

$$\frac{\partial}{\partial\mathbf{K}}\, tr(\mathbf{P}^a) = 0. \quad (43)$$

Using the differential formulas (cf., Eqs. (2.1–72) and (2.1–73) of Gelb et al. 1974):

$$\frac{\partial}{\partial\mathbf{X}}\, tr(\mathbf{XYX}^T) = \mathbf{X}(\mathbf{Y} + \mathbf{Y}^T), \quad (44)$$

$$\frac{\partial}{\partial\mathbf{X}}\, tr(\mathbf{XY}) = \mathbf{Y}^T, \quad (45)$$

Equation (43), $\mathbf{P}^a$ substituted by Eq. (42), is solved to obtain the optimal $\mathbf{K}$ (a.k.a. the Kalman gain):

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})^{-1}. \quad (46)$$

Here, we used the fact that the covariance matrices $\mathbf{P}^f$ and $\mathbf{R}$ are symmetric. Substituting Eq. (46) into Eq. (42), we obtain the other analysis equation for the error covariance:

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f. \quad (47)$$

Two forecast equations (37) and (38), two analysis equations (39) and (47), and the Kalman gain (46) constitute the five KF equations.

3.2 Ensemble Kalman filtering

Applying KF to atmospheric systems is problematic since the state vector $\mathbf{x}$ has $N \sim O(10^7)$ components in the current NWP models. The number of components of the error covariance matrix $\mathbf{P}$ quadruples, so that the direct computations of KF equations are prohibitive. Therefore, in EnKF, a limited number, say m, of ensemble perturbations approximates the error covariance $\mathbf{P}$ (an $N \times N$ matrix):

$$\mathbf{P} \approx \frac{\delta\mathbf{X}(\delta\mathbf{X})^T}{m-1}, \quad (48)$$

where $\delta\mathbf{X}$ denotes an $N \times m$ matrix whose columns consist of ensemble perturbations. Then, the forecast equations (37) and (38) are realized by ensemble forecasting:

$$\overline{\mathbf{X}_i^f} \approx \overline{M(\mathbf{X}_{i-1}^a)}, \quad (49)$$

$$\begin{aligned}\delta\mathbf{X}_i^f &= \mathbf{M}\delta\mathbf{X}_{i-1}^a + \sqrt{m-1}\mathbf{Q}^{1/2}\\ &\approx M(\mathbf{X}_{i-1}^a) - \overline{M(\mathbf{X}_{i-1}^a)}\\ &\quad + \sqrt{m-1}\mathbf{Q}^{1/2}, \quad (50)\end{aligned}$$

where the overbar averages each row of an $N \times m$ matrix to return an N vector, by which $\overline{\mathbf{X}}$ is the ensemble mean of an $N \times m$ matrix $\mathbf{X}$ whose columns consist of ensemble members. $M(\mathbf{X})$ denotes ensemble forecasting, which returns an $N \times m$ matrix with each column vector forecasted by the model M. To derive Eq. (50), we assumed that $\mathbf{Q}^{1/2}$ is not correlated with forecasted ensemble perturbations $\mathbf{M}\delta\mathbf{X}_{i-1}^a$.

Substituting Eq. (48) into Eq. (46), we rewrite the Kalman gain:

$$\mathbf{K} \approx \delta\mathbf{X}^f(\mathbf{H}\delta\mathbf{X}^f)^T \times [\mathbf{H}\delta\mathbf{X}^f(\mathbf{H}\delta\mathbf{X}^f)^T + (m-1)\mathbf{R}]^{-1}. \quad (51)$$

Solving the error covariance analysis equation (47) is not straightforward. There are two methods to realize Eq. (47): the perturbed observation (PO) method and square root filter (SRF) method. The PO method is employed in the early EnKF methods including Evensen (1994) and Houtekamer and Mitchell (1998, 2001). The EnKF system at CMC, the operational ensemble prediction system since Janu-

ary 2005 after the preoperational investigations by Houtekamer et al. (2005), is in this category. In the PO method, independent analysis cycles are employed. If the observations are fixed, the ensemble perturbations satisfy

$$\delta\mathbf{X}^a = (\mathbf{I} - \mathbf{KH})\delta\mathbf{X}^f. \quad (52)$$

This gives too much reduction of ensemble spreads since Eq. (52) results in the lack of the term $\mathbf{KRK}^T$ in Eq. (42). Therefore, perturbing observations to simulate the observational error covariance is necessary (Burgers et al. 1998).

Whitaker and Hamill (2002) pointed out the additional source of sampling errors due to perturbing observations. Instead, they suggested possible advantages of the SRF method without perturbing observations; the method is a current mainstream. In the SRF method, the analysis equation (39) for the state vector is solved for the ensemble mean:

$$\overline{\mathbf{X}^a} = \overline{\mathbf{X}^f} + \mathbf{K}(\mathbf{y}^o - H(\overline{\mathbf{X}^f})). \quad (53)$$

Then the analysis ensemble perturbations $\delta\mathbf{X}^a$ are generated by a linear mapping $\mathbf{T}$ (an $m \times m$ matrix) of the forecast perturbations $\delta\mathbf{X}^f$:

$$\delta\mathbf{X}^a = \delta\mathbf{X}^f\mathbf{T}. \quad (54)$$

Substituting Eq. (54) into Eq. (47), we get

$$\delta\mathbf{X}^f\mathbf{T}\mathbf{T}^T(\delta\mathbf{X}^f)^T = (\mathbf{I} - \mathbf{KH})\delta\mathbf{X}^f(\delta\mathbf{X}^f)^T. \quad (55)$$

There are many possible implementations; at least several SRF methods are proposed so far (cf., a review by Tippett et al. 2003): e.g., an ensemble adjustment Kalman filter (EAKF, Anderson 2001), an ensemble transform Kalman filter (ETKF, Bishop et al. 2001), a serial ensemble square root filter (serial EnSRF, Whitaker and Hamill 2002), and a local ensemble Kalman filter (LEKF, Ott et al. 2002, 2004). In this paper, the implementation of a local ensemble transform Kalman filter (LETKF, Hunt 2005) is described in Subsection 3.4.

3.3 Practical techniques

Although theoretical background of EnKF has been described above, it is not enough for EnKF to be stable in practice. We have missed three problems: big sampling errors due to the limited ensemble size, underestimation of the error covariance, and model errors. There are practical techniques to treat them, described in this subsection.

The state vector has $N \sim O(10^7)$ components, and the error covariance matrix estimated by ensemble perturbations has $N^2 \sim O(10^{14})$ components. However, the ensemble size is limited to at most $m \sim O(10^2)$, so that the sampling errors are problematic. The key idea to treat the sampling error problem is that the covariance matrix contains a large proportion of redundant components such as covariance between distant points, for which intrinsically we can assume zero covariance. In practice, we apply the "covariance localization", where we multiply weights on the estimated covariance components according to the distance. This localization weighting is known as the Schur product in mathematics. There are many possible choices of the weighting function for the Schur product. A current mainstream is the fifth-order piecewise rational function (Gaspari and Cohn 1999; Hamill et al. 2001) which is very close to the Gaussian function but compactly supported. It is known that the localization length scale has major impacts on the analysis accuracy (e.g., Houtekamer and Mitchell 2001; Miyoshi and Yamane 2007). The covariance localization limits the effects of observations within a certain distance, typically about 1400 km radius for global NWP models. Therefore, some information of observations related to teleconnection patterns is ignored. However in the NWP applications, such information in large scales has little impacts on the synoptic scale weather structures, thus sampling errors dominate such large-scale error covariances.

There is another problem that KF with nonlinear models usually underestimates the error covariance even with the perfect model. A possible reason is that the nonlinearity of the model introduces a discrepancy from the linear model, which would act as system noise. As a result, KF estimates forecast errors smaller than what they should, thus observations have less weights. This is a common reason of the "filter divergence", a failure of KF data assimilation. A common way to treat the covariance underestimation is simply to inflate the error covariance manually, the method known as the "covariance inflation". This is equivalent to considering the model error covariance term $\mathbf{Q}$ in the covariance forecast equation (38). A commonly used method in EnKF is to multiply the inflation parameter ρ slightly larger than 1.0

to inflate the forecast ensemble spread, the method known as the "multiplicative spread inflation":

$$\delta\mathbf{X}_i^f = \rho\mathbf{M}\delta\mathbf{X}_{i-1}^a. \tag{56}$$

Comparing with Eq. (50), we get

$$\sqrt{m-1}\mathbf{Q}^{1/2} = (\rho-1)\mathbf{M}\delta\mathbf{X}_{i-1}^a, \tag{57}$$

indicating that the multiplicative spread inflation actually corresponds to the model error covariance term $\mathbf{Q}$. A typical value for the inflation parameter in the perfect model case is about 1.04. As described below, inflation as large as 1.2 is used in the real case with model imperfections.

When we assimilate real observations, there is a big problem: model errors. Models are assumed to be perfect in the above theoretical review. However, it is known that EnKF is generally more vulnerable to model errors than 3D-Var (Miyoshi et al., pers. comm.). Dee (1995) pointed out that quantitative information of the model errors is indispensable in KF assimilating real observations. There are several approaches proposed thus far to consider the model errors in EnKF, divided into two groups: explaining the model errors by the model error covariance $\mathbf{Q}$ (e.g., Houtekamer et al. 2005; Whitaker et al. 2006; Miyoshi and Yamane 2007), estimating and correcting the model error bias (e.g., Dee and Da Silva 1998; Danforth et al. 2007). The former approach is rather practical and well-tested, simply enlarging forecast errors so that observations have larger weights. In practice, Miyoshi and Yamane (2007) made the multiplicative spread inflation parameter ρ larger than the perfect model case. Houtekamer et al. (2005) added random samples to ensemble members, the method known as the "additive spread inflation". Here, the random samples are selected to realize the 3D-Var background error covariance with smaller amplitude. Whitaker et al. (2006) applied three approaches simultaneously: large multiplicative spread inflation, additive spread inflation, and relaxation to prior. The last one mixes forecast and analysis ensemble perturbations to enlarge the analysis ensemble spreads. Alternatively, the second approach, estimating and correcting the model error bias, is based on a different idea. The first approach aims to enlarge the weights on observations by considering the model error covariance. However, the model error covariance corresponds to random noise of the model, not to the systematic model bias. The second approach aims to estimate the systematic model bias and subtract it from the forecasts before analysis. Although several studies have been published on the second approach thus far, it is still at a researching stage and does not seem matured enough at this moment.

3.4 *Local ensemble transform Kalman filter (LETKF)*

As mentioned in Subsection 3.2, there have been at least several SRF methods proposed thus far. As an example of the implementation, the local ensemble transform Kalman filter (LETKF) method proposed by Hunt (2005) is introduced in this subsection.

The LETKF separates the entire physical domain of the forecast model into local patches. Each grid point has its own local patch, and they are assumed to be independent of each other. Analysis equations are solved explicitly for each patch; the computations could be done simultaneously. Thus, the algorithm is ideal for parallel computing environments. In fact, Miyoshi and Yamane (2007) reported that their LETKF system achieved as large as 99.99% parallelization ratio on the Earth Simulator.

In the analysis at a local patch, observations only inside the patch are assimilated, thus the localization is automatically included. The localization weights are Heaviside function, which could introduce discontinuity and cause dynamical imbalance in the analysis. Hunt (2005) proposed a way to apply a smooth localization weighting in the LETKF algorithm. The method, known as the "observation localization", weights observational errors by the inverse of the smooth weighting function. Miyoshi (2005) tested the method and found that the LEKF with observation localization performs as well as the serial EnSRF method.

The analysis equations are solved in the LETKF as follows. Equation (51) is substituted into Eq. (55) and transformed:

$$\begin{aligned}(\mathbf{I}-\mathbf{KH})\delta\mathbf{X}^f(\delta\mathbf{X}^f)^T &= \delta\mathbf{X}^f(m-1)[(m-1)\mathbf{I} \\ &\quad + (\mathbf{H}\delta\mathbf{X}^f)^T\mathbf{R}^{-1}\mathbf{H}\delta\mathbf{X}^f]^{-1}(\delta\mathbf{X}^f)^T.\end{aligned} \tag{58}$$

Comparing with the left hand side of Eq. (55), we get

$$\mathbf{T}\mathbf{T}^T = (m-1)[(m-1)\mathbf{I} + (\mathbf{H}\delta\mathbf{X}^f)^T\mathbf{R}^{-1}\mathbf{H}\delta\mathbf{X}^f]^{-1}. \quad (59)$$

Eigenvalue decomposition:

$$(m-1)\mathbf{I} + (\mathbf{H}\delta\mathbf{X}^f)^T\mathbf{R}^{-1}\mathbf{H}\delta\mathbf{X}^f = \mathbf{U}\mathbf{D}\mathbf{U}^T \quad (60)$$

makes the algorithm greatly efficient. Since the left hand side of Eq. (60) is real symmetric, eigenvalues and eigenvectors are all real, and $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ is satisfied. Therefore, Eq. (59) is solved for $\mathbf{T}$ with almost no cost:

$$\mathbf{T} = \sqrt{m-1}\mathbf{U}\mathbf{D}^{-1/2}\mathbf{U}^T. \quad (61)$$

Moreover, the Kalman gain (46) is written using $\mathbf{P}^a$:

$$\mathbf{K} = \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1}. \quad (62)$$

Using the same eigenvalue decomposition (60), the Kalman gain is written as

$$\mathbf{K} = \delta\mathbf{X}^f\mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T(\mathbf{H}\delta\mathbf{X}^f)^T\mathbf{R}^{-1}. \quad (63)$$

Finally, we obtain the ensemble analysis equation of the LETKF:

$$\begin{aligned}\mathbf{X}^a &= \overline{\mathbf{X}^f} + \mathbf{K}(\mathbf{y}^o - \overline{H(\mathbf{X}^f)}) + \delta\mathbf{X}^a \\ &= \overline{\mathbf{X}^f} + \delta\mathbf{X}^f[\mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T(\mathbf{H}\delta\mathbf{X}^f)^T\mathbf{R}^{-1} \\ &\quad \times (\mathbf{y}^o - \overline{H(\mathbf{X}^f)}) + \sqrt{m-1}\mathbf{U}\mathbf{D}^{-1/2}\mathbf{U}^T].\end{aligned} \quad (64)$$

Recently, Miyoshi et al. (2007) proposed a new implementation of LETKF without local patches. In the above implementation with local patches, we analyze all variables in each local patch. However, we keep only variables at the local patch center to construct global analysis. Thus, there are lots of redundant computations. The new implementation without local patches analyzes only required variables at the local patch center. Precisely, Eq. (64) along with Eq. (60) indicates that the matrix

$$\mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T(\mathbf{H}\delta\mathbf{X}^f)^T\mathbf{R}^{-1}(\mathbf{y}^o - \overline{H(\mathbf{X}^f)}) + \sqrt{m-1}\mathbf{U}\mathbf{D}^{-1/2}\mathbf{U}^T \quad (65)$$

is computed only with variables in the observational space, not in the model grid space. Therefore, once we obtain $\mathbf{H}\delta\mathbf{X}^f$ and $\overline{H(\mathbf{X}^f)}$, the analyzed variable $\mathbf{X}$ in Eq. (64) can be redefined to be only variables at a grid point, which greatly reduces the computations. The observational space is defined independently for each analyzed grid point, thus it is essential to have an efficient algorithm to seek observations to be assimilated at each grid point. The details of the algorithmic design are described in Miyoshi et al. (2007). Instead of using local patches, accurate physical distances between observations and the analyzed grid point are used for the observation localization. This localization with physical distance is preferable in the Polar Regions where the physical distance between two successive grid points in longitudinal direction becomes small. Moreover, the local patch size is no longer necessary, thus the number of localization tuning parameters is reduced. Miyoshi et al. (2007) reported that the new implementation reduces analysis discontinuities in the Polar Regions and that the computation is greatly accelerated.

3.5 *Practical applications of LETKF*

There have been many researches thus far applying EnKF to atmospheric models. Several researches have shown promising results with EnKF assimilating real observations (e.g., Houtekamer et al. 2005; Whitaker et al. 2006;

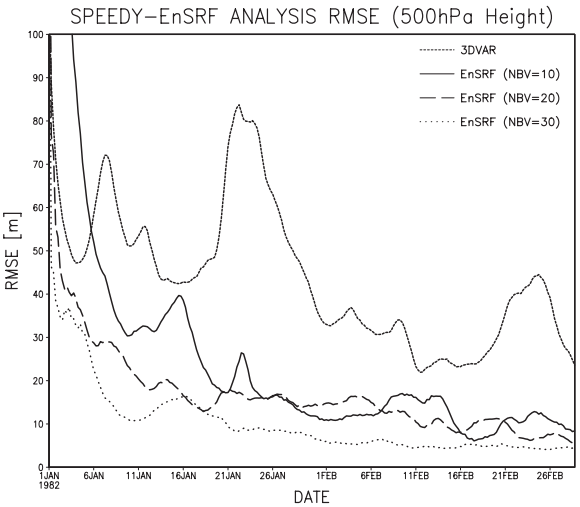

Fig. 8. Time series of the analysis RMSE of 500 hPa height (m) with the SPEEDY model for 3D-Var (short-dashed), serial EnSRF with 10 (solid), 20 (long-dashed), 30 members (dotted). (Adapted from Miyoshi 2005)

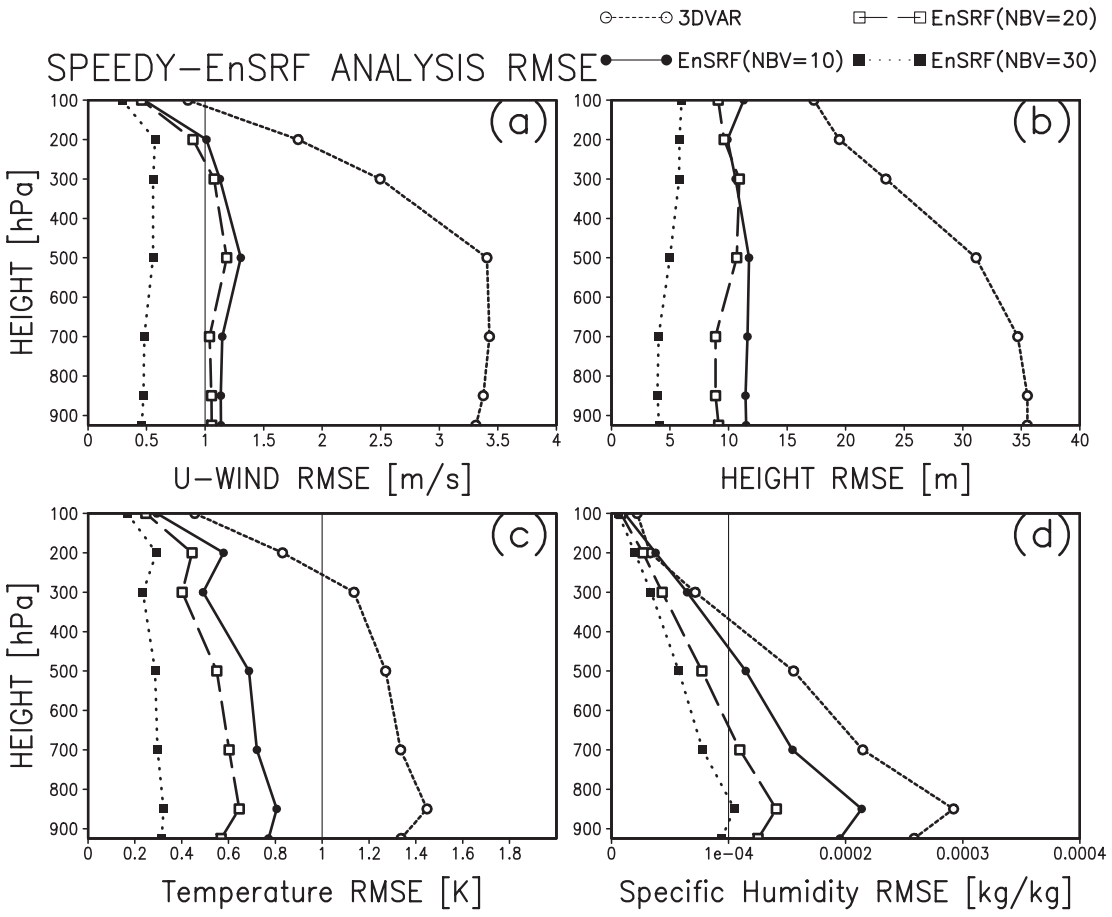

Fig. 9. Vertical profiles of analysis RMSE, temporally averaged for 1 month in February, of (a) zonal winds (m s^{-1}), (b) height (m), (c) temperature (K), and (d) specific humidity (kg/kg) with the SPEEDY model for 3D-Var (short-dashed with open circle), serial EnSRF with 10 (solid with closed circle), 20 (long-dashed with open square), and 30 members (dotted with closed square). Thin solid vertical line in panels (a), (c), (d) indicate observational errors. (Adapted from Miyoshi 2005)

Miyoshi and Yamane 2007). In this paper, the results by Miyoshi (2005) and Miyoshi and Yamane (2007) are reviewed.

a. Comparison with 3D-Var

Miyoshi (2005) developed three data assimilation systems (3D-Var, serial EnSRF, and LEKF) with an AGCM known as the SPEEDY model (Molteni 2003). The SPEEDY model has a T30L7 resolution, corresponding to a grid of $96 \times 48 \times 7$. The model variables are horizontal wind components (u, v), temperature T, specific humidity q, and surface pressure p_s. Since there is no significant difference between serial EnSRF and LEKF, results with serial EnSRF (just "EnKF", hereafter) are shown in this paper. Perfect model data assimilation experiments were performed for two months. The observing network is constant in time and chosen to be regular but sparse; the observing stations are located every 4×4 grid points. At each station, all model variables are observed at all levels. Observational errors are given to be 1.0 m s^{-1}, 1.0 K, 0.1 g/kg, and 1.0 hPa for winds, temperature, specific humidity, and surface pressure, respectively.

Figure 8 shows time series of the analysis root mean square errors (RMSE) of 3D-Var and EnKF with various ensemble sizes. The errors are defined as the difference between the true state and analysis state (analysis ensemble mean in the case of EnKF). 3D-Var is clearly

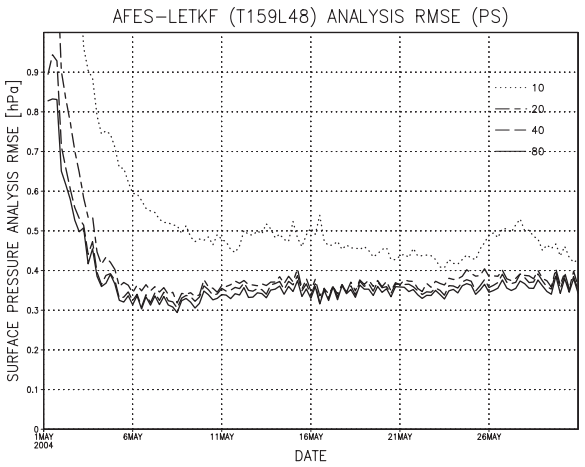

Fig. 10. Time series of the analysis RMSE of surface pressure (hPa) using LETKF with various ensemble sizes and the T159L48 AFES model. The line legend indicates ensemble sizes. (Adapted from Miyoshi and Yamane 2007)

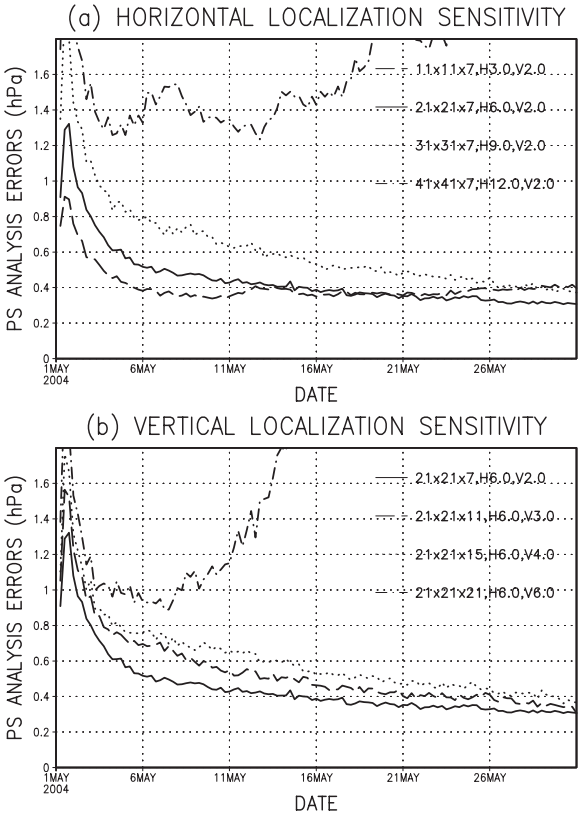

Fig. 11. Time series of the analysis RMSE of surface pressure (hPa) using LETKF with the T159L48 AFES model for various choices of localization parameters. The line legend indicates localization parameters; $21 \times 21 \times 7$, H6.0, V2.0 indicates $21 \times 21 \times 7$ local patch with 6.0-grid horizontal and 2.0-grid vertical Gaussian localization. (Adapted from Miyoshi and Yamane 2007)

worse than all EnKF cases. Increasing ensemble size improves analysis accuracy of EnKF. Vertical profiles of analysis RMSE are shown in Fig. 9. We see clear advantages of EnKF for all variables at all levels, except for specific humidity at upper levels. The improvements by increasing the ensemble size from 10 to 20 are not as large as those by increasing it from 20 to 30.

b. Experiments with a high-resolution AGCM

Miyoshi and Yamane (2007) applied the LETKF method to AFES (AGCM on the Earth Simulator, Ohfuchi et al. 2004) at a T159L48 resolution, corresponding to a grid of $480 \times 240 \times 48$. The model variables include all SPEEDY model variables and cloud water content. Even with the high resolution model, the computational time of 40-member LETKF is less than 10 minutes on the Earth Simulator with 40 computational nodes to assimilate all observational data used in JMA's operational NWP system except satellite radiances. Miyoshi and Yamane (2007) performed perfect model experiments with regular observing network, similarly to Miyoshi (2005). The three-dimensional variables are observed every $5 \times 5 \times 4$ grid points, and surface pressure is observed every 5×5 grid points. Due to the high resolution, the physical density of observing stations is higher than that of the SPEEDY model experiments.

Figure 10 shows time series of the analysis RMSE with various ensemble sizes. We see that just 10 members are many enough to stabilize the LETKF, although more accurate analysis is obtained with more members. Then, the sensitivity to the localization parameters is investigated. Figure 11 shows time series of the analysis RMSE with various choices of localization parameters. It is shown that LETKF diverges with localization parameters as large as 41×41 (equivalent to 1600 km radius) horizon-

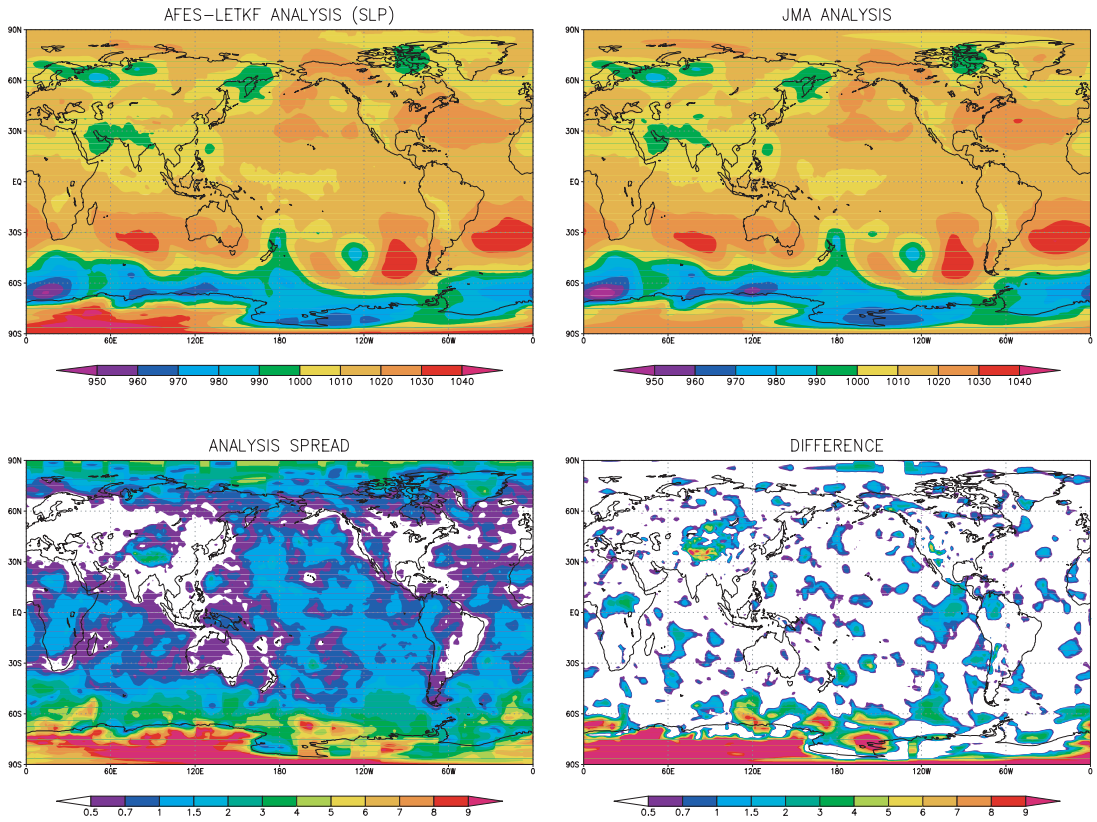

Fig. 12. Sea-level pressure analysis fields (hPa) on 00Z, August 16, 2004. Each panel shows AFES-LETKF analysis (top left), JMA operational analysis (top right), AFES-LETKF ensemble spread (bottom left), and the difference between AFES-LETKF and JMA analyses (bottom right). (Adapted from Miyoshi and Yamane 2007)

tal or 21-grid vertical local patch size due to the sampling errors in the estimated covariance between distant points. In general, there is non-negligible dependence on localization parameters, so that the tuning is required with a given ensemble size.

c. Experiments with real observations

Miyoshi and Yamane (2007) presented results with 40-member LETKF assimilating real observations except satellite radiances in August 2004. Figure 12 shows analysis fields of the LETKF and JMA's operational NWP, adapted from Fig. 15 of Miyoshi and Yamane (2007). The LETKF analysis field looks almost identical to JMA's operational analysis field. Figure 12 also shows the analysis ensemble spread, indicating how much accurate (or inaccurate) EnKF thinks about the analysis. Given more information (i.e., accurate observations), EnKF puts smaller spreads, i.e., more accuracy. This constitutes a new product of analysis errors, which has not been obtainable by other data assimilation schemes.

Since we have no true field for Mother Nature, verifying forecast against analysis is a common way to assess the system performance. Here, the LETKF analysis (only ensemble mean), JMA operational analysis, and NCEP/NCAR reanalysis (Kalnay et al. 1996) are integrated with the AFES model and verified against own analyses. Figure 13 shows the 48-hour forecast verifications, which indicates the LETKF analysis performs as well as the operational analyses. The Southern Hemisphere is greatly controlled by satellite observations, in which sense, the LETKF analysis has a disadvantage. However, the models used in the

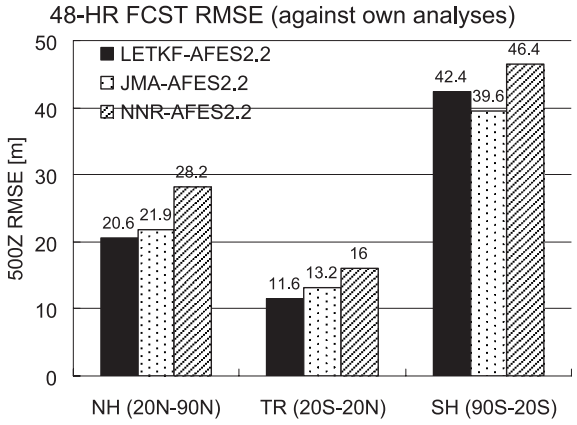

Fig. 13. 48-hour forecast verifications of the LETKF analysis, JMA operational analysis, and NCEP reanalysis, integrated by the AFES model. (Adapted from Miyoshi and Yamane 2007)

operational data assimilation cycles (T213L40 JMA and T62L28 NCEP models) are different from the AFES model, the operational analysis has disadvantages in forecasting with AFES. Therefore, the comparison is not straightforward, but at least we confirm that the LETKF analysis shows a similar level of accuracy in the forecast verification.

4. Summary and concluding remarks

This review paper describes a brief history of data assimilation methods in Introduction, and then presents the theoretical background and implementation of the two advanced data assimilation methods: 4D-Var and EnKF. Recent research results in Japan on those methods, especially on mesoscale applications of 4D-Var and tests of LETKF, are reviewed. 4D-Var is currently regarded as a well-established data assimilation method in meteorology, while EnKF is rather new and researches on this method have been actively conducted as an alternative to 4D-Var.

The variational method was proposed as early as in the 1950s, but it had not been widely used in NWP until the adjoint method was applied to efficiently calculate the gradient of a cost function. The incremental approach made it possible to significantly reduce computational costs of 4D-Var. Strong nonlinearity or non-differentiability of physical processes of numerical models had been considered as an obstacle to apply the adjoint method, but development of regularized linear physics almost solved this problem. Currently several operational NWP centers adopt 4D-Var for global or regional analyses.

If the model and observation operators are linear and PDFs are Gaussian, the analyzed values produced by 4D-Var at the end of an assimilation window is the same as those produced by the Kalman filter applied to the same time window (e.g., Lorenc 1986). The incremental 4D-Var is regarded as a practical approximation to the extended Kalman filter. However, the length of the assimilation window of 4D-Var is limited for nonlinear systems, because tangent linear approximation in the incremental 4D-Var becomes less valid with a longer time window. Even if the incremental approach is not taken, nonlinearity tends to distort a cost function and to generate multiple minima with a longer time window, and then it will be difficult to find the global minimum in the cost function (e.g., Li 1991; Miller et al. 1994). Since 4D-Var does not explicitly calculate the evolution of background error covariance, it evolves the background error covariance only within a short time window starting from a statistically derived initial value. In this respect 4D-Var is suboptimal compared to Kalman filtering.

EnKF is an approximation of Kalman filtering, and is becoming a viable choice for the operational NWP as a method of the data assimilation and the generation of ensemble perturbations. EnKF is advantageous over 4D-Var in terms that covariances are evolved indefinitely. EnKF also provides the analysis errors, while other data assimilation schemes including variational methods cannot. Moreover, a main advantage of EnKF is its simple implementation and model independence, so that a wide range of applicative researches would be relatively easy. This feature would be most impressive and favorable to research communities, if we consider the complexity of variational methods. In this paper, an example is presented showing that we have reached to the stage to assimilate real observations with good accuracy, although several problems are remaining for EnSRF to be applied to operations.

It is to be noted that EnKF assimilates ob-

servations at synoptic times only. In practice, many asynoptic observations are reported, and it is important to treat them adequately. 4D-Var provides the solution, where it uses the dynamical model for temporal propagation within the data assimilation system. 4D-Var treats asynoptic observations simultaneously to generate analysis field at any given time. Although the treatment of asynoptic observations looks a disadvantage of EnKF, Hunt et al. (2004) proposed the four-dimensional EnKF (4D-EnKF) approach. 4D-EnKF enables adequate treatment of asynoptic observations, under the same approximation as 4D-Var, but without the adjoint model. Thus, the treatment of asynoptic observations is no longer disadvantageous to EnKF.

Lorenc (2003b) and Kalnay et al. (2007) discussed theoretical comparisons between 4D-Var and EnKF. Although EnKF has several advantages over the variational methods, the major problem with EnKF is associated with a limited ensemble size. The ability to fit observations is limited by the ensemble size, although localization techniques could reduce this problem. It is not clear how well EnKF simultaneously estimates the atmospheric state in a wide range of scales compared to 4D-Var. Since the 4D-Var system is generally complicated due to the presence of tangent-linear and adjoint models, the comparative research between 4D-Var and EnKF is rather limited. Once EnKF is adopted in experimental systems at operational NWP centers, the comparison with the operational 4D-Var system would provide important research outcomes.

Theoretically, 4D-Var and EnKF for linear systems are equivalent to the Kalman filter, if 4D-Var has no limitation of assimilation window length and EnKF employs infinite ensemble size. Since those conditions cannot be achieved, they are inherent limitations. In this respect a hybrid approach may be attractive. Fisher (1998a) tried to combine 4D-Var with a reduced-rank Kalman filter to indefinitely evolve the background error covariance used in 4D-Var. An hybrid approach of 4D-Var and EnKF is proposed by Lorenc (2003b). Since larger-scale fields tend to have longer predictability, a procedure for evolving large-scale background errors indefinitely is strongly desirable.

Acknowledgments

The first author thanks Profs. T.N. Krishnamurti and I.M. Navon for their kind support when he stayed at the Florida State University for 1994–1996 to study 4D-Var. His thanks are extended to Messrs. Y. Takeuchi, Y. Ishikawa, K. Koizumi, T. Kadowaki for developing the operational variational assimilation systems of JMA. The second author thanks Prof. E. Kalnay and Drs. S. Yamane and T. Enomoto for fruitful discussions.

Appendix

Outline of the operational mesoscale 4D-Var system of JMA

a. Model

The model used in the 4D-Var system is a hydrostatic mesoscale spectral model (MSM) with a horizontal resolution of 10 km and 40 vertical levels. The state variables of the model are wind (u, v), virtual temperature T_v, surface pressure p_s and specific humidity q in spectral space. Large scale fields that satisfy the time-dependent lateral boundary conditions are subtracted from the actual fields, so that the remaining fields satisfy the rigid boundary conditions and are expanded in the double Fourier series (Tatsumi 1986). In order to reduce spurious noise near the lateral boundary, a boundary relaxation technique proposed by Davies (1976) is used.

An incremental approach (Courtier et al. 1994) is adopted by using a hydrostatic MSM with a horizontal resolution of 20 km as an inner-loop model. A full-physics nonlinear model and a reduced-physics adjoint model are used for inner-loop forward and backward integrations, respectively. The current length of assimilation windows is 6 hours, and the number of iterations for minimization, which is limited by a wall clock time, is about 40. Although the forward and backward models are not consistent with each other, a convergence problem in iterations has not been encountered.

Physics contained in the adjoint model are large-scale condensation, middle-level convection of a moist convective adjustment type, simplified vertical diffusion, simplified surface friction and simplified long wave radiation only. Although the hydrostatic MSM contains a prognostic Arakawa-Shubert scheme to parameter-

ize deep cumulus convection, it is not contained in the adjoint model. Since most of precipitation in the hydrostatic MSM is found to come from the large-scale condensation, the absence of deep cumulus convection in the adjoint model does not cause a serious problem in the 4D-Var system.

b. Cost function

The control variables of the 4D-Var system are the initial condition and the lateral boundary conditions at the beginning and end of the assimilation window. The lateral boundary conditions at the other time levels are obtained from linear interpolation. The background term of the cost function used is

$$J_b(\mathbf{x}_0^{(S)}, \mathbf{x}_0^{(L)}, \mathbf{x}_n^{(L)}) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}(\mathbf{x}_n^{(L)} - \mathbf{x}_n^{(L)b})^T \mathbf{B}^{-1}(\mathbf{x}_n^{(L)} - \mathbf{x}_n^{(L)b}), \quad \text{(A1)}$$

where $\mathbf{x}_0^{(S)}$ is the state variables of the hydrostatic MSM at the beginning of the assimilation window with large-scale fields subtracted, $\mathbf{x}_0^{(L)}$ and $\mathbf{x}_n^{(L)}$ are the state variables of large-scale fields at the beginning and end of the assimilation window and used as the lateral boundary conditions. The actual state $\mathbf{x}$ is given by a linear combination of $\mathbf{x}^{(S)}$ and $\mathbf{x}^{(L)}$. The same background error covariance matrix $\mathbf{B}$ is used for $\mathbf{x}_0$ and $\mathbf{x}_n^{(L)}$, and correlations between the background errors of $\mathbf{x}_0$ and $\mathbf{x}_n^{(L)}$ are neglected.

The observation terms of the cost function assume that PDFs for observation errors are Gaussian except for precipitation amount data. Observational data are thinned so that correlations among the errors of observational data are neglected. The following observation term is used for one-hour precipitation amount data (Koizumi et al. 2005).

$$J_o^{(PREC)}(\mathbf{x}) = \sum_{\substack{j \\ where\ y_j^o \geq 0.5\,\mathrm{mm}}} \frac{(H_j(\mathbf{x}) - y_j^o)^2}{2\sigma(y_j^o)^2}, \quad \text{(A2)}$$

$$\sigma(y_j^o) = \begin{cases} r_c & (y_j^o \geq H_j(\mathbf{x})) \\ 3r_c & (y_j^o < H_j(\mathbf{x})) \end{cases}, \quad \text{(A3)}$$

where H_j denotes the observation operator that converts the state variables $\mathbf{x}$ to one-hour accumulated precipitation at the j-th grid point, and y_j^o the observed precipitation at the grid point. If $y_j^o < 1$ mm, r_c is a constant value of standard deviation of departure (observation minus first guess) previously calculated for the observational data less than 1 mm, otherwise r_c is proportional to observed precipitation amount. Observations of one-hour precipitation less than 0.5 mm are not assimilated, since it is found that the quality of such observations is rather poor in snowfall cases.

In order to reduce undesirable gravity wave noise, the following penalty term is added to the cost function.

$$J_c(\mathbf{x}) = \frac{1}{2}\rho \sum_{i=0}^{n} \left| \frac{d}{dt}(\mathbf{D}_i - \mathbf{D}_i^b) \right|^2, \quad \text{(A4)}$$

where ρ is a penalty parameter, and $\mathbf{D}_i$ is a vector of horizontal divergence at all grid points at the i-th time level.

c. Background error covariance

The background error covariance matrix is such a huge matrix that enough statistics are not available to determine all of its elements. Therefore, its structure is modeled with much less parameters to be determined by using several assumptions.

The background error covariance is represented in grid space to easily handle the lateral boundary conditions. The following unbalanced wind (u_U, v_U) is used in stead of wind itself to neglect correlations with (T_v, p_s).

$$\begin{pmatrix} u_U \\ v_U \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} - \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \begin{pmatrix} u_g \\ v_g \end{pmatrix}, \quad \text{(A5)}$$

where $\{r_{ij}\}$ are regression coefficients that depend on vertical levels only, and (u_g, v_g) is the geostrophic wind. The background error covariance matrix is constructed in terms of $\{u_U, v_U, (T_v, p_s), q\}$, and the background errors of u_U, v_U, (T_v, p_s) and q are assumed to be uncorrelated with each other.

To simplify the calculation of the inverse of the background error covariance matrix, a reversible linear transformation is applied to the above variables by making use of the following relation:

$$(\xi - \xi^b)^T \mathbf{B}_\xi^{-1}(\xi - \xi^b) = (\tilde{\xi} - \tilde{\xi}^b)^T \mathbf{B}_{\tilde{\xi}}^{-1}(\tilde{\xi} - \tilde{\xi}^b) = \chi^T \chi, \tag{A6}$$

$$\chi = \mathbf{L}^{-1}(\tilde{\xi} - \tilde{\xi}^b), \tag{A7}$$

where ξ denotes a vector of one of the original variables $\{u_U, v_U, (T_v, p_s), q\}$, $\mathbf{B}_\xi$ its background error covariance matrix, $\tilde{\xi}$ the transformed variable, $\mathbf{B}_{\tilde{\xi}}$ its background error covariance matrix, and $\mathbf{L}$ the lower triangular matrix of the Cholesky decomposition of $\mathbf{B}_{\tilde{\xi}}$. The variable χ defined by Eq. (A7) is used as the control variable in the minimization for a preconditioning procedure to accelerate convergence.

The original variable is normalized by its background error standard deviations, which are assumed to be horizontally homogeneous. Then it is expanded in the eigenvectors of the background error vertical correlation matrix, which is assumed to be homogeneous in the horizontal direction. The resulting linear transformation is given by

$$\tilde{\xi}^{(k)} - \tilde{\xi}^{(k)b} = \sum_{l=1}^{K} (\mathbf{U}^T)_{kl} \frac{\xi^{(l)} - \xi^{(l)b}}{\sigma_l} \quad (k = 1, \ldots, K), \tag{A8}$$

where $\xi^{(l)}$ denotes the original variable at all horizontal grid points at the l-th vertical level, $\xi^{(l)b}$ the first guess of $\xi^{(l)}$, σ_l the background error standard deviation at that level, $\mathbf{U}$ the orthogonal matrix consisting of the eigenvectors, $\tilde{\xi}^{(k)}$ the k-th expansion coefficients at all horizontal grid points, $\tilde{\xi}^{(k)b}$ the first guess of $\tilde{\xi}^{(k)}$, and K the number of vertical levels. The background error covariance of the transformed variable is given by the following block diagonal matrix:

$$\mathbf{B}_{\tilde{\xi}} = \begin{pmatrix} \mathbf{C}^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{(2)} & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}^{(K)} \end{pmatrix}, \tag{A9}$$

where $\mathbf{C}^{(k)}$ is the horizontal correlation matrix among the k-th expansion coefficients at all horizontal grid points.

The horizontal correlation of the expansion coefficients is assumed to be horizontally homogeneous. Let M and N denote the numbers of grid points in the zonal and meridional directions, respectively. Then the horizontal correlation matrix for the k-th expansion coefficients in Eq. (A9) is simplified to

$$\mathbf{C}^{(k)} = \begin{pmatrix} \mathbf{C}_0^{(k)} & \mathbf{C}_1^{(k)} & \cdots & \mathbf{C}_{N-1}^{(k)} \\ \mathbf{C}_1^{(k)} & \mathbf{C}_0^{(k)} & & \mathbf{C}_{N-2}^{(k)} \\ \vdots & & \ddots & \vdots \\ \mathbf{C}_{N-1}^{(k)} & \mathbf{C}_{N-2}^{(k)} & \cdots & \mathbf{C}_0^{(k)} \end{pmatrix}, \tag{A10}$$

where $\mathbf{C}_j^{(k)}$ is the $M \times M$-correlation matrix between the k-th expansion coefficients at the grid points in two rows that lie in the zonal direction and are separated by j grids in the meridional direction. $\mathbf{C}_0^{(k)}$ is the correlation matrix among the M grid points in the same row.

Horizontal correlation functions are assumed to be Gaussian with different correlation lengths in zonal and meridional directions. Then the correlation matrices $\{\mathbf{C}_j^{(k)}\}$ in Eq. (A10) are proportional to $\mathbf{C}_0^{(k)}$:

$$\mathbf{C}_j^{(k)} = \exp\left[-\frac{1}{2}\left(\frac{j}{l_m^{(k)}}\right)^2\right] \mathbf{C}_0^{(k)}, \tag{A11}$$

where $l_m^{(k)}$ is the meridional correlation length of the k-th expansion coefficient in unit of the meridional grid interval. Let $\mathbf{L}_0^{(k)}$ be the $M \times M$-lower triangular Cholesky decomposition matrix of $\mathbf{C}_0^{(k)}$:

$$\mathbf{C}_0^{(k)} = \mathbf{L}_0^{(k)} \mathbf{L}_0^{(k)^T}. \tag{A12}$$

Then the Cholesky decomposition of $\mathbf{C}^{(k)}$ is expressed as follows:

$$\mathbf{C}^{(k)} = \mathbf{L}^{(k)} \mathbf{L}^{(k)^T}, \tag{A13}$$

$$\mathbf{L}^{(k)} = \begin{pmatrix} a_{00}^{(k)} & 0 & \cdots & 0 \\ a_{10}^{(k)} & a_{11}^{(k)} & & 0 \\ \vdots & & \ddots & \vdots \\ a_{N-1\,0}^{(k)} & a_{N-1\,1}^{(k)} & \cdots & a_{N-1\,N-1}^{(k)} \end{pmatrix} \otimes \mathbf{L}_0^{(k)}, \tag{A14}$$

where $\otimes$ denotes the tensor product, and the coefficients $\{a_{ij}^{(k)}\}$ are calculated from Eqs. (A10)–(A13). The Cholesky decomposition matrix $\mathbf{L}$ in Eq. (A7) is given by a block diagonal matrix consisting of $\{\mathbf{L}^{(k)}\}$.

The background error statistics are obtained from differences between 18- and 6-hour forecasts for the same valid time using the NMC method (Parrish and Derber 1992). The elements of the correlation matrix Eq. (A10) of

which absolute values are less than 0.0001 are neglected to save computational time.

References

Anderson, J.L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Andersson, E. and H. Jarvinen, 1999: Variational quality control, *Quart. J. Roy. Meteor. Soc.*, **125**, 697–722.

Andersson, E., J. Pailleux, J.-N. Thepaut, J.R. Eyer, A.P. McNally, G.A. Kelly, and P. Courtier, 1994: Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **120**, 627–653.

Baer, F. and J.J. Tribbia, 1977: On complete filtering of gravity modes through non-linear initialization. *Mon. Wea. Rev.*, **105**, 1536–1539.

Bengtsson, L., M. Ghil, and E. Kallen (eds), 1981: *Dynamic Meteorology: Data Assimilation Methods*, Springer-Verlag, 330 pp.

Bengtsson, L. and J. Shukla, 1988: Integration of space and in situ observations to study global climate change. *Bull. Amer. Meteor. Soc.*, **69**, 1130–1143.

Bergthorsson, P. and B. Doos, 1955: Numerical weather map analysis. *Tellus*, **7**, 329–340.

Bischof, C., A. Carle, P. Khademi, and A. Mauer, 1996: ADIFOR 2.0: Automatic differentiation of Fortran 77 programs. *IEEE Comput. Sci. Eng.*, **3**, 18–32.

Bishop, C.H., B.J. Etherton, and S.J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman Filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.

Bouttier, F. and P. Courtier, 1999: *Data assimilation concepts and methods*. Meteorological Training Course Lecture Series, ECMWF, 75 pp.

Burgers, G., P.J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Caccusi, D.G., 1981: Sensitivity theory for non-linear systems. I: Nonlinear functional analysis approach. *J. Math. Phys.*, **22**, 2794–2803.

Charney, J., M. Halem, and R. Jastrow, 1969: Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci.*, **26**, 1160–1163.

Cohn, S.E. and R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteor. Soc. Japan*, **74**, 63–75.

Courtier, P. and O. Talagrand, 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus*, **42A**, 531–549.

Courtier, P., J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387.

Cressman, G.P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.

Daley, R., 1992a: The effect of serially correlated observation and model error on atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 164–177.

Daley, R., 1992b: The lagged innovation covariance: A performance diagnostic for atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 178–196.

Danforth, C.M., E. Kalnay, and T. Miyoshi, 2007: Estimating and correcting global weather model error. *Mon. Wea. Rev.*, **135**, 281–299.

Davies, H.C., 1976: A lateral boundary formulation for multi-level prediction models. *Quart. J. Roy. Meteor. Soc.*, **102**, 405–418.

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

Dee, D.P. and A.M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **126**, 269–295.

Dee, D.P. and A.M. da Silva, 2003: The choice of variable for atmospheric moisture analysis. *Mon. Wea. Rev.*, **131**, 155–171.

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

Derber, J.C. and F. Bouttier, 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A**, 195–221.

Derber, J.C. and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, **126**, 2287–2299.

Eliassen, A., 1954: *Provisional Report on Calculation of Spatial Covariance and Autocorrelation of the Pressure Field*. *Rep. No. 5*, Videnskaps-Akademiets Institutt for Vaer-Og Klimaforskning, Oslo (reprinted in Bengtsson et al. 1981, 319–330).

Errico, R.M. and T. Vukicevic, 1992: Sensitivity analysis using an adjoint of the PSU-NCAR mesoscale model. *Mon. Wea. Rev.*, **120**, 1644–1660.

Errico, R.M. and K.D. Raeder, 1999: An examination of the accuracy of the linearization of a mesoscale model with moist physics. *Quart. J. Roy. Meteor. Soc.*, **125**, 169–195.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10143–10162.

Evensen, G., 2006: *Data Assimilation: The Ensemble Kalman Filter*, Springer, 279 pp.

Fisher, M., 1998a: Development of a Simplified Kalman Filter. *Technical Memorandum*, No. 260, ECMWF, 16 pp.

Fisher, M., 1998b: Minimization algorithms for variational data assimilation. *Proc. of ECMWF Seminar on Recent Development in Numerical Methods for Atmospheric Modelling (7–11 September 1998, Reading, UK)*, ECMWF, 364–385.

Furumoto, J., K. Kurimoto, and T. Tsuda, 2003: Continuous observations of humidity profiles with the MU radar-RASS combined with GPS and radiosonde measurements. *J. Atmos. Oceanic Technol.*, **20**, 23–41.

Furumoto, J., S. Imura, T. Tsuda, H. Seko, T. Tsuyuki, and K. Saito, 2007: The variational assimilation method for the retrieval of humidity profiles with the wind-profiling radar. *J. Atmos. Oceanic Technol.* (accepted).

Gandin, L.S., 1963: Objective Analysis of Meteorological Fields, Gidrometeorologicheskoe Izdatelstvo, Leningrad (in Russian), English translation by Israel program for scientific translations, Jerusalem, 242 pp.

Gandin, L.S., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.*, **116**, 1137–1156.

Gaspari, G. and S.E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.

Gelb, A., J.F. Kasper, R.A. Nash, C.F. Price, and A.A. Sutherland, 1974: *Applied Optimal Estimation*, The M.I.T. Press, 374 pp.

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, 1981: Application of estimation theory to numerical weather prediction. *Dynamic Meteorology: Data Assimilation Methods* (L. Bengtsson, M. Ghil, and E. Kallen, eds), 139–224.

Ghil, M., K. Ide, A. Bennett, P. Courtier, M. Kimoto, M. Nagata, M. Saiki, and N. Sato (eds), 1997: *Data Assimilation in Meteorology and Oceanography: Theory and Practice*, *Special issue of J. Meteor. Soc. Japan*, **75**, 386 pp.

Ghil, M. and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33**, 141–266.

Gibson, J.K., P. Kållberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ERA Description. *ECMWF ERA-15 Project Report Series*, **1**, 71 pp.

Giering, R. and T. Kaminski, 1998: Recipes for adjoint code construction. *Trans. Math. Software*, **4**, 437–474.

Giering, R., 1999: *Tangent Linear and Adjoint Model Compiler. Users Manual 1.4*, 64 pp (Available online: http://www.autodiff.com/tamc/).

Gilchrist, B. and G. Cressman, 1954: An experiment in objective analysis. *Tellus*, **6**, 309–318.

Gustafsson, N., 1981: A review of methods for objective analysis. *Dynamic Meteorology: Data Assimilation Methods* (L. Bengtsson, M. Ghil and E. Kallen, eds), 17–76.

Hall, M.C.G. and D.G. Cacusi, 1982: Sensitivity analysis of a radiative-convective model by the adjoint method. *J. Atmos. Sci.*, **39**, 2038–2050.

Hamill, T.M., J.S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.

Hoke, J. and R. Anthes, 1976: The initialization of numerical models by a dynamic initialization technique. *Mon. Wea. Rev*., **104**, 1551–1556.

Hollingsworth, A., 1987: Objective analysis for numerical weather prediction. *Short- and Medium-Range Numerical Weather Prediction* (T. Matsuno, ed.), *Collection of Papers Presented at the WMO/IUGG NWP Symposium, Tokyo, 4–8 August 1986*, 11–59.

Honda, Y., M. Nishijima, K. Koizumi, Y. Ohta, K. Tamiya, T. Kawabata, and T. Tsuyuki, 2005: A pre-operational variational data assimilation system for a nonhydrostatic model at Japan Meteorological Agency: Formulation and preliminary results. *Quart. J. Roy. Meteor. Soc.*, **131**, 3465–3475.

Houtekamer, P.L. and H.L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

Houtekamer, P.L. and H.L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Houtekamer, P.L., H.L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620.

Hunt, B.R., E. Kalnay, E.J. Kostelich, E. Ott, D.J. Patil, T. Sauer, I. Szunyogh, J.A. Yorke, and A.V. Zimin, 2004: Four-dimensional ensemble Kalman filtering. *Tellus*, **56A**, 273–277.

Hunt, B.R., 2005: Efficient Data assimilation for spatiotemporal chaos: a Local ensemble transform Kalman filter. arXiv:physics/0511236v1, 25 pp.

Ide, K., P. Courtier, M. Ghil, and A.C. Lorenc, 1997: Unified notation for data assimilation: operational, sequential and variational. *J. Meteor. Soc. Japan*, **75**, 181–189.

Ingleby, B. and A.C. Lorenc, 1993: Bayesian quality control using multivariate normal distributions. *Quart. J. Roy. Meteor. Soc.*, **119**, 1195–1225.

Ishikawa, Y. and K. Koizumi, 2002: One-month

cycle experiments of the JMA mesoscale 4-dimensional variational data assimilation (4D-Var) system. *Research Activities in Atmosphric and Oceanic Modelling*, WMO/TD-No. 1105, 1.26–1.27.

Janiskova, M., J.-F. Thepaut, and J.-F. Geleyn, 1999: Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation. *Mon. Wea. Rev.*, **127**, 26–45.

Jazwinski, A.H., 1970: *Stochastic Processes and Filtering Theory*, Academic Press, 376 pp.

Kadowaki, T., 2005: A 4-dimensional variational assimilation system for the JMA Global Spectrum Model. *Research Activities in Atmosphric and Oceanic Modelling*, WMO/TD-No. 1276, 1.17–1.18.

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

Kalman, R.E. and R. Bucy, 1961: New results in linear filtering and prediction theory. *Trans. ASME, Ser. D, J. Basic Eng.*, **83**, 95–108.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, 341 pp.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, B. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.

Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy, 2007: 4D-Var or ensemble Kalman filter?. *Tellus* (in press).

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J.J. Hnilo, M. Fiorino, and G.L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643.

Kawabata, T., H. Seko, K. Saito, T. Kuroda, K. Tamiya, T. Tsuyuki, Y. Honda, and Y. Wakatsuki, 2007: An assimilation experiment of the Nerima heavy rainfall with a cloud-resolving nonhydrostatic 4-dimensional variational data assimilation system. *J. Meteor. Soc. Japan*, **85**, 255–276.

Klinker, E., F. Rabier, G. Kelly, and J.-F. Mahfouf, 2000: The ECMWF operational implementation of four-dimensional variational assimilation—Part III: Experimental results and diagnostics with operational configuration. *Quart. J. Roy. Meteor. Soc.*, **126**, 1191–1215.

Koizumi, K., Y. Ishikawa, and T. Tsuyuki, 2005: Assimilation of precipitation data to JMA mesoscale model with a four-dimensional variational method and its impact on precipitation forecasts. *SOLA*, **1**, 45–48.

Koizumi, K. and Y. Sato, 2004: Impact of GPS and TMI precipitable water data on mesoscale numerical weather prediction model forecasts. *J. Meteor. Soc. Japan*, **82**, 453–457.

Krishnamurti, T.N., K. Ingels, S. Cocke, R. Pasch, and T. Kitade, 1984: Details of low latitude medium-range numerical weather prediction using a global spectral model II. Effects of orography and physical initialization. *J. Meteor. Soc. Japan*, **62**, 613–649.

Kuo, Y.-H., X. Zou, and Y.-R. Guo, 1996: Variational assimilation of precipitable water using a nonhydrostatic mesoscale adjoint model. Part I: Moisture retrieval and sensitivity experiments. *Mon. Wea. Rev.*, **124**, 122–147.

Langland, R.H. and G.D. Rohaly, 1996: Analysis error and adjoint sensitivity in prediction of a North Atlantic frontal cyclone. Preprint, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 150–152.

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

Lewis, J.M., S. Lakshmivarahan, and S.K. Dhall, 2006: *Dynamic Data Assimilation: A Least Squares Approach*, Cambridge, 654 pp.

Li, Y., 1991: A note on the uniqueness problem of variational adjustment approach to four-dimensional data assimilation. *J. Meteor. Soc. Japan*, **69**, 581–585.

Lions, J.L., 1971: *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, 396 pp.

Lorenc, A.C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.

Lorenc, A.C., 2003a: Modelling of error covariances by 4D-Var data assimilation. *Quart. J. Roy. Meteor. Soc.*, **129**, 3167–3182.

Lorenc, A.C., 2003b: The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203.

Lorenc, A.C. and O. Hammon, 1988: Objective quality control of observations using Bayesian methods: Theory, and a practical implementation. *Quart. J. Roy. Meteor. Soc.*, **114**, 515–543.

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model with applications to normal mode initialization. *Contrib. Atmos. Phys.*, **50**, 253–271.

Mahfouf, J.-F., 1999: Influence of physical processes on the tangent-linear approximation. *Tellus*, **51A**, 147–166.

Mahfouf, J.-F. and F. Rabier, 2000: The ECMWF operational implementation of four-dimensional variational assimilation—Part II: Experimental results with improved physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1171–1190.

Marchuk, G.I., 1975a: Formulation of the theory of perturbations for complicated models. *Appl. Math. Optimi.*, **2**, 1–33.

Marchuk, G.I., 1975b: Formulation of the theory of perturbations for complicated models. Part II, Weather prediction. *Geofisica International*, **15**, 169–182.

Matsumura, T., I. Takano, K. Aonashi, and Ta. Nitta, 1997: Improvement of spin-up of precipitation calculation with use of observed rainfall in the initialization scheme. *Numerical Methods in Atmospheric and Oceanic Modelling—Andre J. Robert Memorial Volume*, NRC Research Press, 353–368.

Miller, R.N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037–1056.

Miyoshi, T., 2005: Ensemble Kalman filter experiments with a primitive-equation global model. Ph.D. Dissertation, University of Maryland, 197 pp.

Miyoshi, T. and S. Yamane, 2007: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.* (in press).

Miyoshi, T., S. Yamane, and T. Enomoto, 2007: Localizing the error covariance by physical distances within a local ensemble transform Kalman filter (in preparation).

Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. *Clim. Dyn.*, **20**, 175–191.

Nakamura, H., K. Koizumi, and N. Mannoji, 2004: Data assimilation of GPS precipitable water vapor into the JMA mesoscale numerical weather prediction model and its impact on rainfall forecasts. *J. Meteor. Soc. Japan*, **82**, 441–452.

Navon, I.M. and D.M. Legler, 1987: Conjugate-gradient methods for large-scale minimization in meteorology. *Mon. Wea. Rev.*, **115**, 1479–1502.

Navon, I.M., X. Zou, J.C. Derber, and J. Sela, 1992: Variational data assimilation with an adiabatic version of the NMC spectral model. *Mon. Wea. Rev.*, **120**, 1433–1446.

Ohfuchi, W., H. Nakamura, M.K. Yoshioka, T. Enomoto, K. Takaya, X. Peng, S. Yamane, T. Nishimura, Y. Kurihara, and K. Ninomiya, 2004: 10-km mesh meso-scale resolving simulations of the global atmosphere on the Earth Simulator: Preliminary outcomes of AFES (AGCM for the Earth Simulator). *J. Earth Simulator*, **1**, 8–34.

Onogi, K., H. Koide, M. Sakamoto, S. Kobayashi, J. Tsutsui, H. Hatsushika, T. Matsumoto, N. Yamazaki, H. Kanahori, K. Takahashi, K. Kato, R. Oyama, T. Ose, S. Kadokura, and K. Wada, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–434.

Ott, E., B.R. Hunt, I. Szunyogh, M. Corazza, E. Kalnay, D. J. Patil, J.A. Yorke, A.V. Zimin, and E.J. Kostelich, 2002: Exploiting local low dimensionality of the atmospheric dynamics for efficient ensemble Kalman filtering. arXiv: physics/0203058v3, 20 pp.

Ott, E., B.R. Hunt, I. Szunyogh, A.V. Zimin, E.J. Kostelich, M. Corazza, E. Kalnay, D.J. Patil, and J.A. Yorke, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428.

Panofsky, H., 1949: Objective weather-map analysis. *J. Appl. Meteor.*, **6**, 386–392.

Park, S.K. and D. Zupanski, 2003: Four-dimensional variational data assimilation for mesoscale and storm-scale applications. *Meteor. Atmos. Phys.*, **82**, 173–208.

Parrish, D.F. and S.E. Cohn, 1985: A Kalman filter for a two dimensional shallow-water model. *Proc. of the 7th Conference on Numerical Weather Prediction (17–20 June 1985, Montreal, Canada)*, American Meteorological Society, 1–8.

Parrish, D.F. and J.C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.

Pham, D.T., J. Verron, and M.-C. Roubaud, 1998: Singular evolutive extended Kalman filter with EOF initialization for data assimilation in oceanography. *J. Mar. Syst.*, **16**, 323–340.

Polavarapu, S., M. Tanguay, and L. Fillion, 2000: Four-dimensional variational data assimilation with digital filter initialization. *Mon. Wea. Rev.*, **128**, 2491–2510.

Rabier, F., H. Jarvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation—Part I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170.

Rostaing, N., S. Dalmas, and A. Galligo, 1993: Automatic differentiation in Odysee. *Tellus*, **45A**, 558–568.

Saito, K., T. Fujita, Y. Yamada, J. Ishida, Y. Kumagai, K. Aranami, S. Ohmori, R. Nagasawa, S.

Kumagai, C. Muroi, T. Kato, H. Eito, and Y. Yamazaki, 2006: The operational JMA nonhydrostatic mesoscale model. *Mon. Wea. Rev.*, **134**, 1266–1298.

Sasaki, Y., 1958: An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, **36**, 77–88.

Sasaki, Y., 1969: Proposed inclusion of time variation terms, observational and theoretical, in numerical variational objective analysis. *J. Meteor. Soc. Japan*, **47**, 115–124.

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

Seko, H., T. Kawabata, T. Tsuyuki, H. Nakamura, K. Koizumi, and T. Iwabuchi, 2004: Impacts of GPS-derived water vapor and radial wind measured by Doppler radar on numerical prediction of precipitation. *J. Meteor. Soc. Japan*, **82**, 473–489.

Stoffelen, A. and D. Anderson, 1997: Ambiguity removal and assimilation of scatterometer data. *Quart. J. Roy. Meteor. Soc.*, **123**, 491–518.

Sun, J. and A. Crook, 1994: Wind and thermodynamic retrieval from single-Doppler measurements of a gust front observed during Phoenix II. *Mon. Wea. Rev.*, **122**, 1075–1091.

Takeuchi, Y. and T. Tsuyuki, 2002: The operational 3D-Var assimilation system of JMA for the global spectral model and the typhoon model. *Research Activities in Atmosphric and Oceanic Modelling*, WMO/TD-No. 1105, 1.59–1.60.

Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311–1328.

Tatsumi, Y., 1986: A spectral limited-area model with time dependent boundary conditions and its application to a multi-level primitive equation model. *J. Meteor. Soc. Japan*, **64**, 637–663.

Thepaut, J.-N. and P. Courtier, 1991: Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Quart. J. Roy. Meteor. Soc.*, **117**, 1225–1254.

Thepaut, J.-N., R.N. Hoffman, and P. Courtier, 1993: Interactions of dynamics and observations in a four-dimensional variational assimilation. *Mon. Wea. Rev.*, **121**, 3393–3414.

Thompson, P., 1961: A dynamical method of analyzing meteorological data. *Tellus*, **13**, 334–349.

Thompson, P., 1969: Reduction of analysis error through constraints of dynamical consistency. *J. Appl. Meteor.*, **8**, 738–742.

Tippett, M.K., J.L. Anderson, C.H. Bishop, T.M. Hamill, and J.S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.

Tsuyuki, T., 1996a: Variational data assimilation in the tropics using precipitation data. Part I: Column model. *Meteor. Atmos. Phys.*, **60**, 87–104.

Tsuyuki, T., 1996b: Variational data assimilation in the tropics using precipitation data. Part II: 3D model. *Mon. Wea. Rev.*, **124**, 2545–2561.

Tsuyuki, T., 1997: Variational data assimilation in the tropics using precipitation data. Part III: Assimilation of SSM/I precipitation rates. *Mon. Wea. Rev.*, **125**, 1447–1464.

Tsuyuki, T., K. Koizumi, and Y. Ishikawa, 2002: The JMA mesoscale 4D-Var system and assimilation of precipitation and moisture data. *Proc. of the ECMWF/GEWEX Workshop on Humidity Analysis (8–11 July 2002, Reading, UK)*, ECMWF, 59–67.

Uppala, S.M., P.W. Kållberg, A.J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J.K. Gibson, J. Haseler, A. Hernandez, G.A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P. Allan, E. Andersson, K. Arpe, M.A. Balmaseda, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B.J. Hoskins, L. Isaksen, P.A.E.M. Janssen, R. Jenne, A.P. McNally, J.-F. Mahfouf, J.-J. Morcrette, N.A. Rayner, R.W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen, 2005: The ERA-40 reanalysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.

Verlinde, J. and W.R. Cotton, 1993: An adjoint sensitivity study of blocking in a two-layer isentropic model. *Mon. Wea. Rev.*, **121**, 2776–2793.

Whitaker, J.S. and T.M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Whitaker, J.S., T.M. Hamill, X. Wei, Y. Song, and Z. Toth, 2006: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.* (submitted).

Yang, W. and I.M. Navon, 1995: *Documentation of the Tangent Linear Model and Its Adjoint of the Adiabatic Version of the NASA GEOS-1 C-Grid GCM (Version 5.2)*. Technical Report Series on Global Modeling and Data Assimilation 104606, NASA/GSFC, 60 pp.

Zou, X., A. Barcilon, I.M. Navon, J. Whitker, and D.G. Cacusi, 1993a: An adjoint sensitivity study of blocking in a two-layer isentropic model. *Mon. Wea. Rev.*, **121**, 2833–2857.

Zou, X., I.M. Navon, M. Berger, K.H. Phua, T. Schlick, and F.X. Le Dimet, 1993b: Numerical

experience with limited-memory quasi-Newton and truncated Newton methods. *SIAM J. Optim.*, **3**, 582–608.

Zou, X., I.M. Navon, and J. Sela, 1993c: Control of gravitational oscillations in variational data assimilation. *Mon. Wea. Rev.*, **121**, 272–289.

Zou, X., I.M. Navon, and J. Sela, 1993d: Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus*, **45A**, 370–387.

Zou, X., Y.-H. Kuo, and Y.-R. Guo, 1995: Assimilation of atmospheric radio refractivity using a nonhydrostatic adjoint model. *Mon. Wea. Rev.*, **123**, 2229–2249.

Zou, X. and Y.-H. Kuo, 1996: Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Mon. Wea. Rev.*, **124**, 2859–2882.

Zupanski, D., 1993: The effects of discontinuities in the Betts-Miller cumulus convection scheme on four-dimensional variational data assimilation. *Tellus*, **45A**, 511–524.

Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.*, **125**, 2274–2292.

Zupanski, D. and F. Mesinger, 1995: Four-dimensional variational assimilation of precipitation data. *Mon. Wea. Rev.*, **123**, 1112–1127.

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