MAXIMUM LIKELIHOOD ENSEMBLE FILTER: THEORETICAL ASPECTS

Milija Zupanski

Cooperative Institute for Research in the Atmosphere
Colorado State University
Foothills Campus
Fort Collins, CO 80523-1375
ZupanskiM@cira.colostate.edu

Submitted May 2003
Revised January 2004
Revised July 2004
Revised October 2004
Revised November 2004

(11 Figures)

A manuscript submitted for publication to the
Monthly Weather Review
Abstract

A new ensemble based data assimilation method, named the Maximum Likelihood Ensemble Filter (MLEF), is presented. The analysis solution maximizes the likelihood of the posterior probability distribution, obtained by minimization of a cost function that depends on a general nonlinear observation operator. The MLEF belongs to the class of deterministic ensemble filters, since no perturbed observations are employed. As in variational and ensemble data assimilation methods, the cost function is derived using a Gaussian probability density function framework. Like other ensemble data assimilation algorithms, the MLEF produces an estimate of the analysis uncertainty (e.g., analysis error covariance). In addition to the common use of ensembles in calculation of the forecast error covariance, the ensembles in MLEF are exploited to efficiently calculate the Hessian preconditioning and the gradient of the cost function. Sufficient number of iterative minimization steps is 2-3, due to superior Hessian preconditioning. The MLEF method is well suited for use with highly nonlinear observation operators, for small additional computational cost of minimization. The consistent treatment of nonlinear observation operators through optimization is an advantage of the MLEF over other ensemble data assimilation algorithms. The cost of MLEF is comparable to the cost of existing ensemble Kalman filter algorithms. The method is directly applicable to most complex forecast models and observation operators.
In this paper, the MLEF method is applied to data assimilation with one-dimensional Korteweg-de Vries-Burgers equation. The tested observation operator is quadratic, in order to make the assimilation problem more challenging. The results illustrate the stability of the MLEF performance, as well as the benefit of the cost function minimization. The improvement is noted in terms of the RMS error, as well as the analysis error covariance. The statistics of innovation vectors (observation-minus-forecast) also indicates a stable performance of the MLEF algorithm. Additional experiments suggest the amplified benefit of targeted observations in ensemble data assimilation.
1. Introduction

Since early 1960s, data assimilation in atmospheric and oceanographic applications is based on the Kalman Filtering theory (Kalman and Bucy 1961; Jazwinski 1970). Beginning with Optimal Interpolation (Gandin 1963), and continuing with three-dimensional (Parrish and Derber 1992; Rabier et al. 1998; Cohn et al. 1998; Daley and Barker 2001) and four-dimensional variational data assimilation (Navon et al. 1992; Zupanski 1993; Zou et al. 1995; Courtier et al. 1994; Rabier et al. 2000; Zou et al. 2001; Zupanski et al. 2002), data assimilation methodologies operationally used in atmospheric and oceanic applications can be viewed as an effort to approximate the Kalman filter/smooother theoretical framework (Cohn 1997). The approximations are necessary because of the lack of knowledge of statistical properties of models and observations, as well as because of a tremendous computational burden associated with high dimensionality of realistic atmospheric and oceanic data assimilation problems. So far, common approaches to realistic data assimilation were to approximate (e.g., model) error covariances, as well as to avoid the calculation of posterior (e.g., analysis) error covariance. These approximations have a common problem of not being able to use fully cycled error covariance information, as the theory suggests. The consequence is not only that the produced analysis is of reduced quality, but also that no reliable estimates of the uncertainties of the produced analysis are available.
For the first time, a novel approach to data assimilation in oceanography and meteorology pursued in recent years (Evensen 1994; Houtekamer and Mitchell 1998; Pham et al. 1998; Lermusiaux and Robinson 1999; Brasseur et al. 2000; Hamill and Snyder 2000; Evensen and van Leeuwen 2000; Keppenne 2000; Bishop et al. 2001; Anderson 2001; Van Leeuwen 2001; Haugen and Evensen 2002; Reichle et al. 2002b; Whitaker and Hamill 2002; Anderson 2003; Ott et al. 2004), based on the use of ensemble forecasting in nonlinear Kalman Filtering, offers the means to consistently estimate the analysis uncertainties. The price to pay is the reduced dimension of the analysis subspace (defined by ensemble forecasts), thus there is a concern of not being sufficient to adequately represent all important dynamical features and instabilities. Preliminary results show, however, that this may not always be a problem (e.g., Houtekamer and Mitchell 2001; Keppene and Rienecker 2002). On the other hand, it is anticipated that the ensemble size will need to be increased as more realistic and higher-resolution models and observations are used. This, however, may be feasible even on currently available computers. With the advancement in computer technology, and multiple processing in particular, which is ideally suited for ensemble framework, the future looks promising for continuing development and realistic applications of ensemble data assimilation methodology.
In achieving that goal, however, there are still few unresolved methodological and practical issues that will be pursued in this paper. Current ensemble data assimilation methodologies are broadly grouped in stochastic and deterministic approaches (Tippett et al. 2003). A common starting point to these algorithms is the use of the solution form of the Extended Kalman Filter (EKF) (e.g., Evensen 2003), obtained assuming linearized dynamics and observation operators, with Gaussian assumption regarding the measurements and control variables (e.g., initial conditions). We refer to this as a linearized solution form. Since realistic observation operators are generally nonlinear, a common approach to nonlinearity in ensemble data assimilation is to use a first order Taylor series assumption, i.e. to use a difference between two nonlinear operators in the place of a linearized observation operator. The use of linearized solution form with nonlinear observation operators, however, creates a mathematical inconsistency in treatment of nonlinear observation operators. An alternate way to deal with the nonlinearity of observation operators is to first pose a fully nonlinear problem, and then find the solution in the ensemble-spanned subspace. This is the approach adopted in this paper.

The proposed ensemble data assimilation method is based on a combination of the maximum likelihood and ensemble data assimilation, named the Maximum Likelihood Ensemble Filter (MLEF). The analysis solution is obtained as a model state that
maximizes the posterior conditional probability distribution. In practice, the calculation of the maximum likelihood state estimate is performed using an iterative minimization algorithm, thus making the MLEF approach closely related to the Iterated Kalman Filter (Jazwinski 1970; Cohn 1997). Since the cost function used to define the analysis problem is arbitrarily nonlinear, the treatment of nonlinear observation operators is considered an advantage of the MLEF algorithm. The use of optimization in MLEF forms a bond between ensemble data assimilation and control theory. Like other ensemble data assimilation algorithms, MLEF produces an estimate of the analysis uncertainty (e.g., analysis error covariance). The idea behind this development is to produce a method capable of optimally exploiting the experience gathered in operational data assimilation and the advancements in ensemble data assimilation, eventually producing a qualitatively new system. The practical goal is to develop a single data assimilation system easily applicable to the simplest, as well as to the most complex nonlinear models and observation operators.

While the maximum likelihood estimate has a unique solution for unimodal probability density functions (PDF), there is a possibility for a non-unique solution in the case of multimodal PDFs. This issue will be given more attention in future.
The method will be explained in Section 2, algorithmic details will be given in Section 3, experimental design will be presented in Section 4, results in Section 5, and conclusions will be drawn in Section 6.

2. MLEF methodology

From variational methods it is known that a maximum likelihood estimate, adopted in MLEF, is a suitable approach in applications to realistic data assimilation in meteorology and oceanography. From operational applications of data assimilation methods, it is also known that a Gaussian PDF assumption, used in derivation of the cost function (e.g., Lorenc 1986), is generally accepted and widely used. Although the model and observation operators are generally nonlinear, and observation and forecast errors are not necessarily Gaussian, the Gaussian PDF framework is still a state-of-the-art approach in meteorological and oceanographic data assimilation (e.g., Cohn 1997). This is the main reason why a Gaussian PDF framework is used in this paper.

The mathematical framework of the MLEF algorithm is presented in two parts, the forecast and the analysis steps, followed by a brief comparison with related data assimilation methodologies.

a. Forecast step
The forecast error covariance evolution of the discrete Kalman filter with Gaussian error assumption can be written (Jazwinski 1970) as

\[ P_f(k) = M_{k-1,k} P_a(k-1) M_{k-1,k}^T + Q(k-1) \]  

(1)

where \( P_f(k) \) is the forecast error covariance at time \( t_k \), \( M_{k-1,k} \) is the linearized forecast model (e.g., Jacobian) from time \( t_{k-1} \) to time \( t_k \), \( P_a(k-1) \) is the analysis error covariance at time \( t_{k-1} \), and \( Q(k-1) \) is the model error covariance at time \( t_{k-1} \). The model error will be neglected in the remainder of this paper. With this assumption, and after dropping the time indexing, the forecast error covariance is

\[ P_f = MP_a M^T = (MP_a^{1/2})(MP_a^{1/2})^T = (P_f^{1/2})(P_f^{1/2})^T \]  

(2)

Let us assume the square-root analysis error covariance is a column-matrix

\[ P_a^{1/2} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_S \end{pmatrix} \]

\[ p_i = \begin{pmatrix} p_{1,i} \\ p_{2,i} \\ \vdots \\ p_{N,i} \end{pmatrix} \]  

(3)

where the index \( N \) defines the dimension of the model state (e.g., initial conditions), and the index \( S \) refers to the number of ensembles. In practical ensemble applications, \( S \) is
much smaller than \( N \). Using (4) in definition (2), the square-root forecast error covariance is

\[
P_f^{1/2} = (b_1 \ b_2 \ \cdot \ \cdot \ b_S) \quad b_i = M(x_{k-1} + p_i) - M(x_{k-1}) \approx Mp_i
\]  

where \( x_{k-1} \) is the analysis from the previous analysis cycle, at time \( t_{k-1} \). Note that each of the columns \( \{b_i: i=1, \ldots, S\} \) has \( N \) elements. The ensemble square-root forecast error covariance \( P_f^{1/2} \) can be obtained from \( S \) nonlinear ensemble forecasts, \( M(x_{k-1} + p_i) \), plus one control forecast, \( M(x_{k-1}) \) (e.g., Eq.(4)). The forecast error covariance definition (4) implies the use of control (deterministic) forecast instead of an ensemble mean, commonly used in other ensemble data assimilation methods. Ideally, the control forecast represents the most likely dynamical state, thus it is intrinsically related to the use of maximum likelihood approach. In principle, however, the use of an ensemble mean instead of the most likely deterministic forecast is also possible.

Important to note is that the availability of an ensemble square-root analysis error covariance \( P_a^{1/2} \), provided by data assimilation algorithm, is critical for proper coupling between analysis and forecast. In addition to data assimilation cycles, the \( P_a^{1/2} \) columns could be used as initial perturbations for ensemble forecasting, in agreement with (4).

\( b. \) Analysis step
In the MLEF method, the analysis solution is obtained as a maximum likelihood estimate, i.e. a model state that maximizes the posterior probability distribution. With the Gaussian PDF assumption implied in definition of the cost function, the maximum likelihood problem is redefined as the minimization of an arbitrary nonlinear cost function of the form (e.g., Lorenc 1986)

\[
J(x) = \frac{1}{2} [x - x_b]^T P_f^{-1} [x - x_b] + \frac{1}{2} [y - H(x)]^T R^{-1} [y - H(x)]
\]  

(5)

where \(x\) is the model state vector, \(x_b\) denotes the prior (background) state, and \(y\) is the measurement vector. The background state \(x_b\) is an estimate of the most likely dynamical state, thus it is a deterministic forecast from the previous assimilation cycle. The nonlinear observation operator \(H\) represents a mapping from model space to observation space, \(R\) is the observation error covariance matrix.

Note that the error covariance matrix \(P_f\) is defined in the ensemble subspace (e.g. Eq(4)), thus it has a much smaller rank than the true forecast error covariance. Therefore, the cost function definition (5) only has a similar form as the three-dimensional variational cost function (e.g., Parrish and Derber 1992), however it is defined in the ensemble subspace only. Strictly speaking, the invertibility of \(P_f\) in (5) is preserved only in the range of \(P_f\), implying that the cost function (5) is effectively defined in the range of \(P_f\) as well. Same reasoning and definitions are implicit in other ensemble data
assimilation methods, with the exemption of hybrid methods (e.g., Hamill and Snyder 2000).

Hessian preconditioning is introduced by a change of variable

\[ x - x_b = P_f^{1/2} (I + C)^{-T/2} \zeta \]  

(6)

where the vector \( \zeta \) is the control variable defined in ensemble subspace, and

\[ C = P_f^{T/2} H^T R^{-1} H P_f^{1/2} = (R^{-1/2} H P_f^{1/2})^T (R^{-1/2} H P_f^{1/2}) \]  

(7)

The notation \( P_f^{T/2}=(P_f^{1/2})^T \) is used in the above formula. A closer inspection reveals that the change of variable (6) is a perfect preconditioner in quadratic minimization problems (Axelsson 1984), i.e. assuming linear observation operators. This means that, with the change of variable (6) and linear observation operators, the solution is obtained in a single step of minimization iteration. The matrix defined in (6) is the square-root of an inverse Hessian of (5). The matrix \( C \) is commonly neglected in Hessian preconditioning in variational problems, due to high dimensionality and associated computational burden.

The practical problem is now to define the matrices appearing in (6). The square-root forecast error covariance is calculated from previous ensemble forecasts (e.g., Eq. (4)). The calculation of the matrix \( (I + C)^{-T/2} \), however, requires some attention. Since
the columns of the square-root forecast error covariance are known, the $i$-th column of the matrix appearing in (7) is

$$z_i = (R^{1/2} HP_f^{1/2})_i = R^{1/2} Hb_i \approx R^{1/2} H(x + b_i) - R^{1/2} H(x)$$

(8)

Note that each of the column-vectors $z_i$ has the dimension of observation space. The matrix $C$ can be then defined as

$$C = \begin{pmatrix}
  z_1^T z_1 & z_1^T z_2 & \ldots & z_1^T z_S \\
  z_2^T z_1 & z_2^T z_2 & \ldots & z_2^T z_S \\
  \vdots & \vdots & \ddots & \vdots \\
  z_S^T z_1 & z_S^T z_2 & \ldots & z_S^T z_S 
\end{pmatrix}$$

(9)

The matrix $C$ is an $S \times S$ symmetric matrix, thus it has small dimensions defined by the number of ensembles. In order to efficiently calculate the inversion and the square-root involved in $(I + C)^{-T/2}$, an EigenValue-Decomposition (EVD) of the matrix $C$ may be used. One obtains $C = VAV^T$ where $V$ denotes the eigen-vector matrix, and $A$ is the eigenvalue matrix. Then

$$(I + C)^{-T/2} = V(I + A)^{-1/2} V^T$$

(10)
Note that the definition of matrix $C$ and subsequent EVD are equivalent to the matrix transform introduced in ETKF (Bishop et al. 2001). The change of variable (6) can be now easily accomplished. The use of ensembles is consistently introduced by (4) and (8).

After successfully accomplishing the Hessian preconditioning, the next step in an iterative minimization is to calculate the gradient in ensemble-spanned subspace. One can first redefine the cost function (5) using the change of variable (6) and then calculate the first derivative

$$g_{\zeta} = (I + C)^{-1}\zeta - (I + C)^{-1/2} [R^{-1/2} HP_f^{1/2}]^T R^{-1/2} [y - H(x_b + P_f^{1/2} (I + C)^{-T/2} \zeta)]$$

(11)

Note that the use of an adjoint (e.g., transpose) in (11) is avoided by employing (8) in calculation of the matrix $R^{-1/2} HP_f^{1/2}$.

As shown in appendix A, within a linear operator framework, the first minimization iteration calculated using the preconditioned steepest descent, is equivalent to the ensemble based reduced-rank Kalman filter (Verlaan and Heemink 2001; Heemink et al. 2001), or to the Monte-Carlo based ensemble Kalman filter (Evensen 1994). Although different in detail, the computational effort involved in calculation with ensemble based Kalman filters is comparable to the calculation of ensemble-based Hessian preconditioning and the gradient present in the MLEF algorithm. In both the
ensemble Kalman filters and the MLEF, the computational cost of the analysis step is dominated by a matrix inversion computation (e.g., A.5-A.7, appendix A).

In calculating the analysis error covariance, the MLEF employs a strategy somewhat different from other ensemble data assimilation methods. The MLEF calculates the analysis error covariance as the inverse Hessian matrix at the minimum (e.g., Fisher and Courtier 1995), generally available as a by-product of minimization. For quasi-Newton minimization algorithm, one could use the inverse Hessian update produced by minimization algorithm (e.g., Nocedal 1980). In applications with conjugate-gradient algorithm (e.g., Gill et al. 1981; Luenberger 1984), used here, one would update the matrix $C$ using the solution at the minimum (i.e. the optimized analysis $x_{opt}$), and then calculate $P_a^{1/2}$ according to

$$P_a^{1/2} = P_f^{1/2} \left[ I + C(x_{opt}) \right]^{-T/2}$$

The expression (12) has the same form as the analysis error covariance used in the ETKF (Bishop et al. 2001). The important difference, however, exists in applications with nonlinear observation operators. Since in MLEF the inverse Hessian is calculated at the minimum, the Taylor expansion of a nonlinear Hessian operator is well approximated by the first order (e.g., linear) term. This implies that the equivalence between the inverse Hessian and the analysis error covariance, valid only in linear framework, is preserved for arbitrary nonlinear operators. For linear observation operator, the analysis error...
covariance estimates from both algorithms would be the same. The columns of the matrix $P_a^{1/2}$ are then used as initial perturbations for the next assimilation cycle, according to (3) and (4), and cycling of analysis and forecast continues.

c. The MLEF and related data assimilation methodologies

The MLEF method encompasses few important existing methodologies and algorithms:

(1) VARIATIONAL DATA ASSIMILATION

The minimization of the cost function, used to derive the maximum likelihood estimate in MLEF, is inherently related to variational data assimilation algorithms. The difference is that, in the MLEF formulation, the minimization is performed in an ensemble-spanned subspace, while in variational method the full model space is used. The issue of the number of degrees of freedom is problem-dependent, and will require consideration in future realistic applications. At present, one should note that there are ways to introduce complementary degrees of freedom and obtain a unique mathematical solution (e.g., Hamill and Snyder 2000). Also, there is a practical possibility to increase the degrees of freedom by introducing more ensemble members. All these options require careful examination in problem-oriented applications.
(2) ITERATED KALMAN FILTER

Another methodology related to MLEF is the Iterated Kalman Filter (IKF - Jazwinski 1970; Cohn 1997), developed with the idea to solve iteratively the nonlinear problem. Bell and Cathey (1993) demonstrated that IKF is a Gauss-Newton method. As the MLEF, the IKF is calculating the mode (e.g., maximum likelihood approach), with underlying Gaussian assumption. An obvious difference is that MLEF is defined within an ensemble framework. Practical advantage of an iterative methodology, such as the IKF or MLEF, is fundamentally tied to the choice of minimization method. An integral part of the MLEF is the use of an unconstrained minimization algorithm, in the form of the nonlinear conjugate-gradient and the LBFGS quasi-Newton methods (e.g., Gill et al. 1981; Luenberger 1984; Nocedal 1980). The unconstrained minimization approach allows very efficient iterative solution to the problem with significant nonlinearities and large residuals (e.g., Gill et al. 1981).

(3) ENSEMBLE TRANSFORM KALMAN FILTER (ETKF)

The matrix transform and eigenvalue decomposition used for the Hessian preconditioning in MLEF (Eqs. (6)-(10)) is equivalent to the matrix transform introduced in the ETKF algorithm (Bishop et al. 2001). This approach allows an efficient reduction
of the dimensions of a matrix to be inverted. Therefore, the MLEF algorithm can be viewed as a maximum likelihood approach to the ETKF (Bishop, 2003, person. comm.).

The idea behind the MLEF is to retain only the components and concepts deemed advantageous from other algorithms, while weak components are changed or improved. For example, the cost function minimization, used in variational methods and IKF, is characterized as beneficial: minimization allows the equivalence between the inverse Hessian and analysis error covariance to be valid even for arbitrary nonlinear observation operators. Modeling of forecast error covariance, Hessian preconditioning and adjoint model development, are all considered weak points of variational methods, and are improved or avoided using ensemble framework. Hessian preconditioning introduced in the ETKF is considered advantageous as well. The ensembles framework makes the probabilistic forecasting and data assimilation with realistic prediction models and observations feasible, not possible with IKF.

The end products of the MLEF algorithm are: (i) deterministic analysis, corresponding to the model state that maximizes the posterior probability distribution, and (ii) (square-root) analysis error covariance, corresponding to an estimate of analysis uncertainty.

3. Algorithmic details
The MLEF algorithm is designed to exploit the data assimilation infrastructure in existing algorithms. For example, the innovation vectors (e.g., observation-minus-forecast residuals) are calculated as in existing data assimilation algorithms, and the minimization currently used in variational data assimilation can be used in MLEF. In order to optimize the MLEF performance in realistic applications, the multiple processor capability of parallel computing is made an important component of the algorithm.

As implied in previous section, the underlying principle in the MLEF development was to improve the computational stability of the algorithm by using only square-root matrices. The algorithmic steps in the MLEF are:

(i) Ensemble forecasting from previous to new analysis cycle

A square-root forecast error covariance is computed first. Normally, the initial ensemble perturbations are the columns of a square-root analysis error covariance, available from a previous analysis cycle. At the very start of data assimilation, however, there is no previous analysis error covariance, and one needs to provide some initial ensemble perturbations to be used in (4). Amongst many feasible options, the following strategy is adopted in MLEF: define random perturbations to initial conditions some time into the past, say the time interval of 1-2 assimilation cycles, in order to form a set of perturbed initial conditions. Then use this set to initiate ensemble forecasting. The
nonlinear ensemble forecast perturbations are computed as a difference between the ensemble forecasts and the control (e.g., unperturbed) forecast, valid at the time of first data assimilation cycle. According to (4), these perturbations are then used as columns of a square-root forecast error covariance, required for data assimilation.

Note that this step, common to all ensemble data assimilation algorithms, may contribute significantly to the computational cost of ensemble data assimilation in high-dimensional applications. It allows an efficient use of parallel computing, however, thus the actual cost can be significantly reduced in practice.

(ii) Forward ensemble run to observation location – innovation vector calculation

Once the ensemble forecasting step is completed, producing square-root forecast error covariance columns, the analysis step begins. An essential component of this step is the calculation of innovation vectors, i.e. the observation minus first guess differences for each ensemble member. In practice, the vectors $z_i$ (Eq. (8)) are computed as nonlinear ensemble perturbations of innovation vectors

$$z_i = R^{-1/2}H(x + b_i) - R^{-1/2}H(x) = R^{-1/2}[y - H(x)] - R^{-1/2}[y - H(x + b_i)]$$

(13)

where the vectors $b_i$ are obtained from previously completed ensemble forecasts (Eq. (4)). This means that each ensemble forecast is interpolated to observation location, using
the same observation operator available in an existing variational data assimilation
algorithm. The calculation of innovation vector perturbations is done without
communication between processors, thus it is efficiently scalable on parallel computers.

(iii) Hessian preconditioning and $C$-matrix calculation

This step is done only in first minimization iteration. The matrix $C$ is computed
from ensemble perturbations around the initial forecast guess, and is used for Hessian
preconditioning. The innovation vectors calculated in step (ii) are then used to calculate
the elements of the matrix $C$ (Eq.(8)). The elements of $C$ are computed through an inner-
product calculation, and this represents the second dominant computational effort in
MLEF (most dominant being the ensemble forecasting). Note that an equivalent
computational effort is involved in the ETKF algorithm. Although $C$ is an $S \times S$
symmetric matrix ($S$ being the ensemble size), there are still $S(S+1)/2$ inner products to be
calculated. If parallel computing is available, each of the inner products can be efficiently
calculated on separate processors, essentially with no communication between the
processors, thus significantly reducing the computational cost. The EVD calculation of $C$
is of negligible cost, $C$ being a small-dimensional matrix. Standard EVD subroutines for
dense matrices, commonly available in a general mathematical library, such as the Linear
Algebra Package (LAPACK – Anderson et al. 1999), or similar, may be used. As shown
by (10), the matrix inversion involved in the change of variable (6) is easily accomplished.

(iv) Gradient calculation

The gradient calculation requires a repeated calculation of innovation vector perturbations $z_i$ in each minimization iteration, however without the need to update the matrix $C$. The components of the gradient vector in ensemble space (Eq. (11)) are essentially the control forecast innovation vector components projected on each ensemble perturbation. With mentioned good parallel scalability of innovation vectors calculation, the cost of the gradient calculation is relatively small.

(v) Analysis error covariance

As stated earlier, the required square-root of analysis error covariance is obtained as a by-product of minimization algorithm. The actual computation method depends on the employed minimization algorithm. For example, if a quasi-Newton algorithm is used, one could use the inverse Hessian update formula (e.g., Nocedal 1980) to update the analysis error covariance. In this work, however, we employed a nonlinear conjugate gradient algorithm (e.g., Luenberger 1984), with the line-search algorithm as defined in Navon et al. (1992). In order to obtain a satisfactory square-root analysis error
covariance, the relation (12) is used, with $C$ computed around the optimal analysis. Otherwise, the calculation is identical to the step (iii). Since $C$ is computed close to the true minimum, the nonlinear part of the Hessian is negligible, and a good estimate of the analysis error covariance can be obtained. The columns of the square-root analysis error covariance are then used as perturbations to ensemble forecasting in step (i), and the new analysis cycle begins.

Note that error covariance localization, not employed in the current MLEF algorithm, is an important component of most ensemble based data assimilation algorithms (e.g., Houtekamer and Mitchell 1998; Hamill et al. 2001; Hamill and Withaker 2002). The idea is that, if the forecast error covariance is noisy and has unrealistic distant correlations, these correlations should be removed. The noisy error covariance is anticipated if the number of ensembles is very small. In the MLEF applications presented here, however, initially noisy error covariances were localized anyway after only few analysis cycles, without any need for additional localization procedure. For that reason, the issue of error covariance localization is left for future work.

4. Experimental design
The MLEF method will be used in a simple one-dimensional example, in order to illustrate the anticipated impact in realistic applications.

\textit{a. Model}

The forecast model used in this paper is a one-dimensional Korteweg-de Vries-Burgers (KdVB) model

\[
\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + 6 \frac{\partial^3 u}{\partial x^3} = \nu \frac{\partial^2 u}{\partial x^2}
\]  

(14)

where \( u \) is a non-dimensional model state vector, and \( \nu \) is a diffusion coefficient. The numerical solution is obtained using centered finite-differences in space, and the 4-th order Runge-Kutta scheme for time integration (Marchant and Smyth 2002). The model domain has dimension \( N=101 \), with the grid-spacing \( \Delta x=0.5 \) non-dimensional units, and the time-step is \( \Delta t=0.01 \) non-dimensional units. The periodic boundary conditions are used. In the control experiment the diffusion coefficient is \( \nu=0.07 \).

The KdVB model includes few desirable characteristics, such as the nonlinear advection, dispersion, and diffusion. It also allows the solitary waves (e.g., solitons), a nonlinear superposition of several waves, with damping due to diffusion. Various forms of this model are being used in hydrodynamics, nonlinear optics, plasma physics, and
elementary particle physics (see overview by Bullough 1988). An interesting weather-related application of a coupled KdV-based system of equations can be found in Gottwald and Grimshaw (1999a,b). Also implied by Mitsudera (1994) in applications to cyclogenesis, the KdV-based system supports baroclinic instability, and it models realistically a nonlinear interaction between the flow and topography.

In the experiments presented here, a two-soliton analytic solution of the Korteweg-de Vries equation (Vvedensky 1993) is chosen for the initial conditions

\[
\begin{align*}
\mathbf{u}(x,t) &= 2(\beta_1 - \beta_2) \left( \text{Sech} \left[ \sqrt{\beta_1} \left( \frac{x - 2t\beta_1}{\sqrt{2}} \right) \right] \right)^2 \beta_1 + \text{Csch} \left[ \sqrt{\beta_2} \left( \frac{x - 2t\beta_2}{\sqrt{2}} \right) \right] \right)^2 \beta_2 \\
&= -\sqrt{2\beta_2} \text{Coth} \left[ \sqrt{\beta_2} \left( \frac{x - 2t\beta_2}{\sqrt{2}} \right) \right] + \sqrt{2\beta_1} \text{Tanh} \left[ \sqrt{\beta_1} \left( \frac{x - 2t\beta_1}{\sqrt{2}} \right) \right] \right)^2
\end{align*}
\]  

where \( x \) refers to distance and \( t \) to time. The parameters \( \beta_1 \) and \( \beta_2 \) reflect the amplitude of the two solitons, and are chosen \( \beta_1 = 0.5 \) and \( \beta_2 = 1.0 \). The solitons progress with the speed proportional to their amplitude, and the specific choice of the parameters assures that the solitons will often interact during the time integration of the model.

Note that the model run defined as \textit{truth} is using \( \beta_1 = 0.5 \) and \( \beta_2 = 1.0 \), while the initial conditions used in assimilation experiments are defined using \( \beta_1 = 0.4 \) and \( \beta_2 = 0.9 \), as well as the time parameter \( t \) is lagging behind the truth by one time unit (e.g.,
100 model time-steps). The initial forecast error covariance is defined using ensemble forecasts (e.g., (4)), initiated from a set of random perturbations two cycles prior to the first observation time. The initial perturbations are formed by randomly perturbing parameters of the solution (15), such as the time, and $\beta_1, \beta_2$ parameters, around the values used in assimilation run, i.e. using $\beta_1=0.4$ and $\beta_2=0.9$.

b. Observations

The observations are chosen as random perturbations to the truth (i.e., the forecast run with initial conditions using $\beta_1=0.5$ and $\beta_2=1.0$ in (15)), with the error $\varepsilon_{\text{obs}}=0.05$ non-dimensional units. Note that such choice implies a perfect model assumption. The observation error covariance $R$ is chosen to be diagonal (e.g., variance), with elements $\varepsilon_{\text{obs}}^2$. There are approximately 10 irregularly spaced observations available at each analysis time. Two types of the experiments are performed: (i) in-situ observations, fixed at one location at all times, and (ii) targeted observations, with observations following the solitons’ peaks throughout the integration. Initially, however, both the in-situ and targeted observations are chosen identical.

The observation operator is a quadratic transformation operator, defined as $H(u)=u^2$. The choice of quadratic observation operator is influenced by a desire to test the algorithm with relatively weakly nonlinear observation operator, not necessarily related
to any meteorological observations. In practice, the observation operators of interest
would include highly nonlinear observations operators, such as the radiative transfer
model for cloudy atmosphere (e.g., Greenwald et al. 1999), with extensive use of
exponential functions. Also, radar reflectivity measurements of rain, snow and hail are
related to model produced specific humidity and density through logarithmic and other
nonlinear functions (Ming Xue, 2004, priv. comm.). The observations are taken at grid-
points to avoid additional impact of interpolation. The case of linear observation operator
is less interesting, since then the MLEF solution is identical to the reduced-rank ensemble
Kalman filter solution. The use of linear observation operator, however, is important for
algorithm development and initial testing. In that case, the MLEF solution is obtained in a
single minimization step due to the implied perfect Hessian preconditioning.

The observations are made available every two non-dimensional time units. Given
the model time-step of 0.01 units, each analysis cycle implies 200 model time-steps. The
time integration of the control forecast, and the observations are shown in Fig.1. Note
that no data assimilation is involved in creating these plots. The shown time evolution
corresponds to the first 10 analysis cycles, and illustrates the two-soliton character of the
solution. The shown cycles correspond to the cycles that will be shown in section 5.

c. Experiments
The control experiment includes 10 ensemble members (compared to 101 total degrees of freedom) with 10 targeted observations, and employs a quadratic observation operator. The iterative minimization employed is the Fletcher-Reeves nonlinear conjugate-gradient algorithm (e.g., Luenberger 1984). In each of the MLEF data assimilation cycles, three minimization iterations are performed to obtain the analysis. In all experiments 100 analysis cycles are performed, until the amplitude of solitary waves in the control forecast was reduced by one order of magnitude due to diffusion. Long assimilation also helps evaluating the stability of the MLEF algorithm performance.

The experiments are designed in such way to address few potentially important and challenging problems in realistic atmospheric and oceanic data assimilation: (i) impact of minimization, and (ii) impact of observation location.

d. Validation

In order to compare the results of various experiments, few validation methods are employed: (i) root-mean-square (RMS) error, (ii) analysis error covariance estimate, (iii) $\chi^2$ validation test, and (iv) innovation vector PDF statistics.

(i) In calculating the RMS error, it is assumed that the true analysis solution, denoted $u_{true}$, is given by the control forecast used to produce the observations. This is not
completely true, being dependent on the relative influence of observation and forecast errors, but it is assumed acceptable. With this assumption, the RMS error is calculated as

$$RMS = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (u_n - u_{true})^2}$$

As before, the index $N$ defines the model state dimension (i.e. the number of grid points).

(ii) The analysis error covariance is an estimate obtained from an ensemble data assimilation algorithm, and it will be shown in terms of the actual matrix elements. This is the new information produced by ensemble data assimilation, generally not available in variational data assimilation. It requires special attention, since this information is directly transferred to ensemble forecasting, and also estimates the uncertainty of the produced analysis.

(iii) The $\chi^2$ validation diagnostics (e.g., Menard et al. 2000), developed to validate the Kalman filter performance, can also be used in the context of ensemble data assimilation. This diagnostics evaluates the correctness of the innovation (observation minus forecast) covariance matrix that employs a pre-defined observation error covariance $R$, and the MLEF-computed forecast error covariance $P_f$. We adopt the
definition used in Menard et al. (2000), namely $\chi^2$ is defined in observation space, normalized by the number of observations, $N_{\text{obs}}$:

$$
\chi^2 = \frac{1}{N_{\text{obs}}} [y_k - H(x_k)]^T [HP_f H^T + R]^{-1} [y_k - H(x_k)]
$$

(17)

In the MLEF algorithm, the above formula is rewritten as

$$
\chi^2 = \frac{1}{N_{\text{obs}}} \left\{ R^{-\frac{1}{2}} [y_k - H(x_k)] \right\}^T G^{-1} \left\{ R^{-\frac{1}{2}} [y_k - H(x_k)] \right\}
$$

(18)

where the matrix $G^{-1}$ (e.g., its square-root) is defined in appendix B, $y$ denotes observations, and $x$ is the model forecast. Due to an iterative estimation of optimal analysis in MLEF, the forecast $x$ denotes the forecast from the last minimization iteration, and the matrix $C$ is calculated about the optimal state. For Gaussian distribution of innovations, and linear observation operator $H$, the conditional mean of $\chi^2$ defined by (18) should be equal to one. As in Menard et al. (2000), the conditional mean is substituted by a time mean. In this paper, a 10-cycle moving average is computed, as well as the instant values of $\chi^2$, calculated at each assimilation cycle. Due to the use of nonlinear model in calculation of $P_f$, and statistically small sample (i.e., relatively few observations per cycle), one can expect only values of $\chi^2$ close to one, not necessarily equal to one.
Another important statistical verification of an ensemble data assimilation algorithm, also related to innovation vectors, is the probability density function (PDF) of innovations (e.g., Reichle et al. 2002a). From (18), the normalized innovations are defined as

\[
Innov = \frac{1}{\sqrt{N_{\text{obs}}}} G^{-\frac{I}{2}} \left\{ R^{-\frac{1}{2}} [y_k - H(x_k)] \right\}
\] (19)

With Gaussian filtering assumptions regarding the measurements and control variables, and for linear dynamic system and observation operators, the resulting innovation PDF should have a standard normal distribution $\mathcal{N}(0,1)$. Note that, if innovations (19) are random variables with distribution $\mathcal{N}(0,1)$, then the Eqs. (17)-(18) define a $\chi^2$ distribution with $N_{\text{obs}}$ degrees of freedom.

In our applications, due to the nonlinearity of the forecast model and the observation operator $H$, and due to relatively small statistical sample, only an approximate normal distribution should be expected.

5. Results
a. Linear observation operator experiments

When linear observation operators are employed, and Gaussian error distribution assumed, in principle there is no difference between the MLEF and any related EnKF algorithm. Formally, a single minimization iteration of the MLEF is needed, with step-length equal to one. These experiments are conducted in order to develop and test the MLEF algorithm, especially the statistics of produced results, using diagnostics defined in section 4, parts (iii) and (iv). Note that perfect statistical fit cannot be expected, since the forecast model is still a nonlinear model, with diffusion, and the posterior statistics is not exact Gaussian. An obvious consequence of having few observations per cycle is that the innovation statistics may not be representative of true PDF statistics. Two experiments are performed, one with 10 (targeted) observations per cycle, and the other with all observations (e.g., 101 per cycle).

The $\chi^2$ test is shown in Fig. 2. Although in both experiments the value of $\chi^2$ is close to one, much better agreement is obtained with more observations (e.g., Fig. 2b). It also takes more analysis cycles to converge to one, which may be a sign of an increased difficulty of the KdVB model to fit numerous and noisy observations. Note that with all observations assimilated, there is a greater chance that some observations are negative, which would in turn impact the numerical stability of the model.
The innovation histogram is shown in Fig. 3, and indicates a similar impact of statistical sample. Deviations from a Gaussian PDF are more pronounced when fewer observations are used (Fig. 3a), then in the case with all observations (Fig. 3b). There is also notable right shift of the PDF, which could be the impact of diffusion (e.g., Daley and Menard (1993), or the impact of model nonlinearity.

The results in Figs. 2a and 3a indicate what can be expected from the experiments with quadratic observation operator and few observations per cycle. On the other hand, in future applications of the MLEF with real models and real observations, one could expect much better innovation statistical sample, given an enormous number of satellite and radar measurements available today.

b. Quadratic observation operator

(1) CONTROL EXPERIMENT

The RMS result of the control MLEF experiment is shown in Fig. 4. Also, the RMS error from the experiment with no observations is shown. Any acceptable data assimilation experiment should have smaller RMS error than the no-assimilation experiment. During the initial 11-12 cycles, however, there is a pronounced increase of the RMS error. This suggests that the particular initial perturbation (defined as a difference from truth) is unstable during initial cycles. As the cycles continue, however,
the RMS error in the no-assimilation experiment converges to the true solution, i.e.
indicating an ultimate stability of initial perturbation. This is an artifact of diffusion,
which would eventually force all forecasts to be zero, therefore producing all RMS errors
equal to zero. One can note good RMS convergence in the MLEF experiment, within first
few cycles. The final RMS error is non-zero, since the defined truth (e.g., $u_{true}$) is just a
long-term forecast used to create observations, not necessarily equal to the actual true
analysis solution. Overall, the RMS error indicates a stable MLEF performance.

The estimate of the analysis error covariance in the control MLEF experiment is
shown in Fig.5, for the analysis cycles No. 1, 4, 7, and 10. These cycles are chosen in
order to illustrate an initial adjustment of the analysis. Each of the figures represents
actual matrix elements, with diagonal corresponding to the variance. All analysis error
covariance figures have a threshold of $\pm 1 \times 10^{-4}$ non-dimensional units, in order to ease
the qualitative comparison between the results from different experiments. Although the
true analysis error covariance is not known, it would have non-zero values, since the
observations have a non-zero error. One can immediately note how analysis error
covariance became localized by the 4th cycle, without the need for any artificial error
covariance localization. Also, the values of the covariance matrix remain relatively small
through cycles, however moving with the solitons.
Statistics of innovation vectors is important sign of the algorithm performance, especially useful when the truth is not known. The $\chi^2$ test and the innovation histogram are shown in Fig.6. As suggested earlier, due to increased nonlinearity and small statistical sample, one should expect only approximate agreement. It is clear that the $\chi^2$ value remains near the value of one throughout analysis cycles, again suggesting a stable performance of MLEF algorithm. The innovation histogram is showing close resemblance to standard normal PDF, confirming that the statistics of innovations is satisfactory.

(2) IMPACT OF ITERATIVE MINIMIZATION

The control MLEF experiment (with 3 minimization iterations) is compared to the ensemble data assimilation experiment with no explicit minimization. In both experiments 10 ensemble members and 10 observations are used, and a quadratic observation operator is employed. The only difference is that the MLEF employs an iterative minimization, while the no-minimization experiment is a single minimization iteration with the step-length equal to one (e.g., appendix A). The experiment without minimization indirectly reflects the impact of linear analysis solution, implied in ensemble Kalman filters. It should be noted, however, that there are many other details of
ensemble Kalman filters not captured in this experiment, and any direct comparison should be taken with caution.

The RMS errors are shown in Fig. 7. It is obvious that without minimization, the ensemble-based reduced-rank data assimilation algorithm is not performing well. The explanation is that the MLEF algorithm is better equipped to handle nonlinearities of observation operators, thus it creates smaller RMS errors. Most of the difficulties in the no-minimization experiment are occurring during first 11-12 cycles, coinciding with the RMS increase noted in the no-observation experiment (Fig. 4).

With fewer observations, the positive impact of iterative minimization is still notable in terms of RMS error, although the impact is somewhat smaller in magnitude (Fig. 8). Again, most of the differences occur during first cycles, both solutions reaching the same RMS in later cycles. The reduced number of observations does have a negative impact on the performance of both algorithms, as expected. The impact of minimizations is also evaluated for the in-situ observations, in terms of RMS error (Fig. 9). As before, the positive impact of minimization is notable only in first cycles, with both algorithms showing signs of difficulty. A hidden problem with no-minimization experiments with 5 observations and with in-situ observations was that the satisfactory solution was possible only for smaller initial ensemble perturbations. Therefore, the results shown in Figs. 8 and 9 imply smaller initial ensemble perturbations than in the experiments with 10
targeted observations (Fig. 7). This may be an indication of the sensitivity of the KdVB numerical solution to large perturbations, but also it may suggest a critical role of iterative minimization in situations with large innovation residuals. This issue will be further examined in future, in applications with realistic models and real observations.

Overall, the use of iterative minimization in MLEF shows a positive impact in terms of the RMS error. The impact appears to be stronger for a better observed system.

(3) IMPACT OF OBSERVATION LOCATION

An interesting problem, related to the impact of observation location on the performance of ensemble data assimilation, is now considered. The issue of targeted observations, as means to improve the regular observation network, has been thoroughly discussed and evaluated (Palmer et al. 1998; Buizza and Montani 1999; Langland et al. 1999; Szunyogh et al. 2002; Majumdar et al. 2002). Here, we indirectly address this issue by examining the impact of observation location on the performance of the MLEF algorithm.

At initial time, the in-situ and targeted observations are chosen to be identical. The two solitons may be viewed as weather disturbances with phase and amplitude important to predict, i.e. as the temperature associated with fronts, for example. Since these systems move and interact with each other, it is instructive to evaluate the impact of
targeted observations, intuitively associated with the location of the disturbances. The
Fig.10 shows the RMS errors in targeted and in-situ MLEF experiments. There is a
strong positive impact of targeted observations, being able to resolve the two
disturbances at all times. The particular location of the in-situ observations does not allow
the optimal use of observation information with regard to the two solitons. Only at cycles
when the solitons are passing through the in-situ observation network the observation
information is adequately transferred and accumulated, eventually resulting in small
RMS errors.

The analysis error covariance associated with the in-situ MLEF experiment is
shown in Fig.11, and should be compared with the control MLEF experiment (Fig.5).
The positive impact of targeted observations is now even more obvious. In first cycle, the
results are identical since the targeted and in-situ observations are identical. As the cycles
proceed, much larger uncertainties are obtained than in the control MLEF experiment,
especially near the location of solitons. Although one should not draw strong conclusions
from this simple experiment, the results appear to suggest that targeted observations
amplify the beneficial impact of ensemble data assimilation.

6. Summary and Conclusions
The Maximum Likelihood Ensemble Filter is presented, in applications to one-dimensional Korteweg-de Vries-Burgers equation with two solitons. The filter combines the maximum likelihood approach with the ensemble Kalman filter methodology, to create a qualitatively new ensemble data assimilation algorithm with desirable computational features. The analysis solution is obtained as a model state that maximizes the posterior probability distribution, via an unconstrained minimization of an arbitrary nonlinear cost function. This creates an important link between the control theory and ensemble data assimilation. Like other ensemble data assimilation algorithms, the MLEF produces an estimate of the analysis uncertainty (e.g., analysis error covariance), and employs solely nonlinear forecast model and observation operators. The use of linearized models, or adjoints, required for variational methods, is completely avoided. The impact of the MLEF method is illustrated in an example with quadratic observation operator. The innovation vector statistics (e.g., $\chi^2$ test and innovation histogram) indicates satisfactory, stable performance of the algorithm. Although in this paper the MLEF method is applied in a simple environment, all calculations and processing of observations are directly applicable to the use with state-of-the-art forecast models and arbitrary nonlinear observation operators. Since the observations assimilated in the experiments presented here are just a single realization of infinitely many possible realizations, the obtained results also depend on the particular observation realization.
The impact of targeted observations is another important issue relevant to the operational data assimilation and the use of ensembles. It appears that targeted observation network amplifies the beneficial impact of ensemble data assimilation. This is certainly an issue worth of further investigation.

The positive impact of iterative minimization, on both the RMS error and the analysis error covariance, is obvious. The MLEF algorithm clearly benefits from the maximum likelihood component. The additional calculation involved in iterative minimization is almost negligible, compared to the cost of ensemble forecasts and the Hessian preconditioning calculations. Only 2-3 minimization iterations are anticipated in realistic applications, further relaxing possible concern of using iterative minimization.

A positive impact of minimization in the case of nonlinear observation operator suggests that an iterative minimization approach can be also used in other ensemble based data assimilation algorithms based on the use and calculation of the conditional mean (e.g., ensemble mean). Such an algorithm would be more robust with respect to nonlinear observation operators.

Due to the use of a control deterministic forecast as a first guess, the MLEF method may be more appealing in applications where a deterministic forecast is of interest. The MLEF method offers a potential advantage when computational burden forces the ensemble forecasts to be calculated in coarser resolution than desired. One can
still minimize the cost function defined in fine resolution and thus produce the control
(maximum likelihood) forecast in fine resolution. Only the ensembles, used for error
covariance calculation are defined in coarse resolution. Using the ensemble mean as a
first guess, on the other hand, may be a limiting factor in that respect, since data
assimilation problem would be defined and solved only in coarser resolution.

In forthcoming paper, the model error and model error covariance evolution will
be added to the MLEF algorithm. Applications to realistic models and observations are
also underway. For somewhat higher computational cost, the MLEF algorithm allows a
straightforward extension to smoothing, which could be relevant in applications with high
temporal frequency of observations.

In future MLEF development, the non-Gaussian PDF framework and improved
Hessian preconditioning is anticipated, to further extend the use of control theory in
challenging geophysical applications. Both the conditional mean (e.g., minimum
variance) and the conditional mode (e.g., maximum likelihood) are important PDF
estimates (e.g., Cohn 1997). Future development of the MLEF will address these issues.
Acknowledgments

I would like to thank Dusanka Zupanski for many helpful discussions and careful reading of the manuscript. My gratitude is also extended to Ken Eis for helpful comments and suggestions. I would also like to thank Thomas Vonder Haar and Tomislava Vukicevic for their continuous support throughout this work. I am greatly indebted to Rolf Reichle and an anonymous reviewer for thorough reviews that significantly improved the manuscript. This research was supported by the Department of Defense Center for Geosciences/Atmospheric Research at Colorado State University under Cooperative Agreement DAAD19-02-2-0005 with the Army Research Laboratory.
APPENDIX A

Equivalence of the Kalman gain and Hessian preconditioning-gradient calculation

The preconditioned steepest descent is often used as a first iterative step in many gradient based minimization algorithms, such as the conjugate gradient, quasi-Newton, truncated Newton algorithms (e.g., Gill et al. 1981)

\[ x = x_b - \alpha E^{-1} g \]  \hspace{1cm} (A.1)

where \( \alpha \) is the step-length, \( E \) is the Hessian and \( g \) is the gradient of the cost function (6) in first minimization iteration. Denoting the Jacobian of observation operator as \( H \)

\[ H = \left( \frac{\partial H}{\partial x} \right)_{x_b} \] \hspace{1cm} (A.2)

the gradient of the cost function in first iteration is

\[ g = \frac{\partial J}{\partial x} = -H^T R^{-1} [y - H(x_b)] \] \hspace{1cm} (A.3)

and the Hessian

\[ E = \frac{\partial^2 J}{\partial x^2} = P_f^{-1} + H^T R^{-1} H \] \hspace{1cm} (A.4)
Substituting (A.3) and (A.4) in (A.1) gives

\[ x = x_h + \alpha \left( P_f^{-1} + H^T R^{-1} H \right)^T H^T R^{-1} [y - H(x_h)] \]  \hspace{1cm} (A.5)

After employing the matrix identity (Jazwinski 1970)

\[ \left( P_f^{-1} + H^T R^{-1} H \right)^T H^T R^{-1} = P_f H^T \left( H P_f H^T + R \right)^T \] \hspace{1cm} (A.6)

The analysis update in first minimization iteration becomes

\[ x = x_h + \alpha P_f H^T \left( H P_f H^T + R \right)^T [y - H(x_h)] \] \hspace{1cm} (A.7)

For quadratic cost function, the step-length \( \alpha \) is equal to one (Gill et al. 1981). Therefore, for quadratic cost-function, the solution of the iterative minimization problem in first iteration is identical to the extended Kalman filter solution (Jazwinski 1970). In this context, the matrix identity (A.6) shows the equivalence between the Kalman gain calculation and the Hessian-gradient calculation in iterative minimization. For non-quadratic cost function, the step length is different from one, and the solution (A.7) is not identical to the extended Kalman filter solution. The Kalman gain computation, however, is still the same as the Hessian-gradient computation.
APPENDIX B

Computation of the matrix $G^{-1/2}$

The matrices $G^{-1}$ (Eq. 18) and $G^{1/2}$ (Eq. 19) are needed for computation of normalized innovations. An efficient algorithm for computing the inverse square-root matrix $G^{1/2}$ is presented here. It relies on the use of Sherman-Morrison-Woodbury (SMW) formula, as well as on the use of an iterative matrix square-root calculation procedure. This matrix is used to calculate the normalized innovations (Eq.(19)). Finally, the calculated normalized innovations are then used in calculating the $\chi^2$ sum (Eq.(18)).

From (17) and (18), one can see that

$$G^{-1} = \left[ I + R^{-1/2} H P / H^T R^{-1/2} \right]^{-1}$$  \hspace{1cm} (B.1)

Using (13) to define the columns of the matrix $Z=(z_1 \ z_2 \ \ldots \ z_S)$, one can redefine $G^{-1}$ as

$$G^{-1} = \left[ I + ZZ^T \right]^{-1}$$  \hspace{1cm} (B.2)

Note that columns of the matrix $Z$ are the same as used in the MLEF algorithm (e.g., (8) and (13)), and are available for no additional cost. The issue is how to calculate the inversion in (B.2), as well as the matrix square-root required by the Eq.(19) from text.

From the SMW formula (e.g., Golub and Van Loan 1989)
\[ G^{-1} = I - Z[I + Z^TZ]^{-1}Z^T \]  \hfill (B.3)

Note that \( C = Z^TZ \), where the matrix \( C \) is same as defined by (9) and (10) from text. This means that (B.3) can be rewritten as

\[ G^{-1} = I - ZV[I + A]^{-1}V^TZ^T \] \hfill (B.4)

Where the eigenvector matrix \( V \) and the eigenvalue matrix \( A \) are both available from the MLEF algorithm (Eq.(10)). Therefore, all matrices on the right-hand-side of (B.4) are available. In order to calculate the square-root of a positive-definite symmetric matrix \( G^{-1} \) one can exploit an iterative formula (Golub and Van Loan 1989, p.554, problem P11.2.4) which produces a unique symmetric positive-definite square-root matrix \( G^{1/2} \)

\[ X_0 = I \]

\[ X_{k+1} = \frac{1}{2}(X_k + G^{-1}X_k^{-1}) \] \hfill (B.5)

\[ X_k \to G^{1/2} \]

It is important to realize that the specific form of the matrix \( G^{-1} \) (e.g., (B.4)), and the fact that \( V \) are unitary (e.g., \( V^TV = I \)), allow a simplification of matrix inversion involved in (B.5). To see that, it is convenient to write \( G^{-1} \) in generic form.
\[ G^{-I} = I + ZV\Psi_0V^TZ^T \quad \Psi_0 = -(I + A)^{-1} \quad (B.6) \]

where \( \Psi_0 \) is a non-zero diagonal matrix, with known elements: \( \psi_i = -1/(1+\lambda_i) \). After applying (B.5) with formulation (B.6), one obtains

\[ X_1 = I + ZV\Sigma_1V^TZ^T \quad \Sigma_1 = \frac{1}{2}\Psi_0 \quad (B.7) \]

With the help of SWM formula

\[ X_1^{-I} = I - ZVT\Sigma_1V^TZ^T \quad \Gamma_1 = \Sigma_1(I + A\Sigma_1)^{-1} \quad (B.8) \]

If the procedure is continued, it soon becomes clear that both \( X_k \) and \( X_k^{-1} \) keep the same form, and only diagonal matrices \( \Sigma_k \) and \( \Gamma_k \) are updated during iterations. This greatly simplifies the computational burden of a matrix square-root calculation. A recursive (iterative) algorithm for \( G^{1/2} \) can then be defined:

Define \( \Psi_0 = -(I + A)^{-1} \quad \Sigma_0 = 0 \quad \Gamma_0 = 0 \quad (B.9) \)

For \( k = 1, N \) do

\[ \Sigma_k = \frac{1}{2}[\Sigma_{k-1} + \Psi_0 - \Gamma_{k-1} - \Psi_0 A\Gamma_{k-1}] \]

\[ \Gamma_k = \Sigma_k(I + A\Sigma_k)^{-1} \quad (B.10) \]

end do
The recursive formula (B.9)-(B.11) is computationally very efficient, since the iterative procedure (B.10) employs only diagonal matrices. Once it is determined that the algorithm converged, a square-root matrix is formed (e.g., (B.11)).

In experiments conducted in this paper, a satisfactory convergence was found after only 3 iterations (e.g., N=3 in (B.10)). The above algorithm is stable, and convenient for the matrix square-root calculations in the context of MLEF.
References


Bullough, R.K., 1988: The wave par excellence, the solitary, progressive great wave of


Evensen, G., and P.J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear


Greenwald, T.J., S.A. Christopher, J. Chou, and J.C. Liljegren, 1999: Inter-comparison of


Luenberger, D.L., 1984: *Linear and Non-linear Programming*. 2nd ed. Addison-Wesley,


Vvedensky, D., 1993: *Partial Differential Equations with Mathematica*. Addison-


Figure Captions

Fig.1. Time integration of the KdVB model and observations: (a) Targeted observations, (b) In-situ observations. The triangles denote the observations. The horizontal axis represents the model domain, and the ordinate axis is the amplitude. The shown cycles are No. 1, 4, 7 and 10. Note how the targeted observations follow the solitons, while the in-situ observations remain in one location.

Fig.2. $\chi^2$ statistics in linear observation operator assimilation experiment, with: (a) 10 observations per cycle, and (b) 101 observations per cycle. The dashed line represents instant values of $\chi^2$ from each analysis cycle, while the solid line represents a 10-cycle moving average.

Fig. 3. Innovation histogram in linear observation operator assimilation experiment, with: (a) 10 observations per cycle, and (b) 101 observations per cycle. The solid line represents the normal distribution $\mathcal{N}(0,1)$.

Fig.4. The RMS error in the control MLEF experiment, with quadratic observation operator and 10 observations (thin solid line). The horizontal axis denotes the analysis
cycles, and the ordinate axis the RMS error. Also, the RMS error in the no-assimilation experiment is shown (thick solid line).

Fig.5. The analysis error covariance in the control MLEF experiment: (a) analysis cycle No. 1, (b) analysis cycle No. 4, (c) analysis cycle No. 7, and (d) analysis cycle No. 10. Each point represents the \((i,j)\)-th matrix element, \(p_{ij}\), with the horizontal axis denoting the \(i\)-index, and the ordinate axis the \(j\)-index. Dark-shaded area represents positive covariance, and the light-shaded area represents the negative covariance, using the threshold of \(\pm 1 \times 10^{-4}\) non-dimensional units. The contour interval is: (a) 20, (b) 2, (c) 2, and (d) 2 non-dimensional units.

Fig.6. Innovation statistics in the control MLEF experiment: (a) \(\chi^2\) test, and (b) PDF histogram. The notation is same as in Figs. 2 and 3.

Fig.7. Impact of minimization on the MLEF performance. The figure shows the RMS errors of ensemble data assimilation without minimization (solid line), and the control MLEF for comparison (dashed line). The horizontal axis denotes the analysis cycles, and the ordinate axis the RMS error.
Fig. 8. Same as in Fig. 7, except for the experiment with 5 observations.

Fig. 9. Same as in Fig. 7, except for in-situ observation experiment with 10 observations.

Fig. 10. Impact of observation location. The dashed line represents the RMS errors obtained with in-situ observations, while the solid line is the RMS error from the control MLEF experiment (e.g., targeted observations).

Fig. 11. Same as in Fig. 5, except for in-situ assimilation experiment. The contour interval is: (a) 20, (b) 2, (c) 2, and (d) 20 non-dimensional units.
Fig. 1. Time integration of the KdVB model and observations: (a) Targeted observations, (b) In-situ observations. The triangles denote the observations. The horizontal axis represents the model domain, and the ordinate axis is the amplitude. The shown cycles are No. 1, 4, 7 and 10. Note how the targeted observations follow the solitons, while the in-situ observations remain in one location.
Fig 2. $\chi^2$ statistics in linear observation operator assimilation experiment, with: (a) 10 observations per cycle, and (b) 101 observations per cycle. The dashed line represents instant values of $\chi^2$ from each analysis cycle, while the solid line represents a 10-cycle moving average.
Fig. 3. Innovation histogram in linear observation operator assimilation experiment, with: (a) 10 observations per cycle, and (b) 101 observations per cycle. The solid line represents the normal distribution $\mathcal{N}(0,1)$. 

63
Fig. 4. The RMS error in the control MLEF experiment, with quadratic observation operator and 10 observations (thin solid line). The horizontal axis denotes the analysis cycles, and the ordinate axis the RMS error. Also, the RMS error in the no-assimilation experiment is shown (thick solid line).
Fig. 5. The analysis error covariance in the control MLEF experiment: (a) analysis cycle No. 1, (b) analysis cycle No. 4, (c) analysis cycle No. 7, and (d) analysis cycle No. 10. Each point represents the \((i,j)\)-th matrix element, \(p_{ij}\), with the horizontal axis denoting the \(i\)-index, and the ordinate axis the \(j\)-index. Dark-shaded area represents positive covariance, and the light-shaded area represents the negative covariance, using the threshold of \( \pm 1 \times 10^{-4} \) non-dimensional units. The contour interval is: (a) 20, (b) 2, (c) 2, and (d) 2 non-dimensional units.
Fig. 6. Innovation statistics in the control MLEF experiment: (a) $\chi^2$ test, and (b) PDF histogram. The notation is same as in Figs. 2 and 3.
Fig. 7. Impact of minimization. The figure shows the RMS errors of ensemble data assimilation without minimization (dashed line), and the control MLEF for comparison (solid line). The horizontal axis denotes the analysis cycles, and the ordinate axis the RMS error.
Fig. 8. Same as in Fig. 7, except for the experiment with 5 observations.
Fig. 9. Same as in Fig. 7, except for *in-situ* observation experiment with 10 observations.
Fig. 10. Impact of observation location. The dashed line represents the RMS errors obtained with *in-situ* observations, while the solid line is the RMS error from the control MLEF experiment (e.g., *targeted* observations).
Fig. 11. Same as in Fig. 5, except for \textit{in-situ} assimilation experiment. The contour interval is: (a) 20, (b) 2, (c) 2, and (d) 20 non-dimensional units.