# Discussion of The Ensemble Kalman Filter and Earth System Predictability

Emmanuel Dibia

A scholarly paper in partial fulfillment of the requirements for the degree of Master of Science

Submitted on: November 12, 2019

Department of Atmospheric and Oceanic Science, University of Maryland

College Park, Maryland


Advisors: Dr. Xin-Zhong Liang & Dr. Daryl Kleist

## Table of Contents

**Abstract**

In this study, an introduction to ensemble Kalman filtering techniques relevant to environmental state estimation problems is presented. The historical evolution of such methods is discussed mainly in the context of numerical weather prediction with major emphasis given to theoretical formulations and to innovations that allow for operational viability. In an effort to provide insight into the workings of different schemes, the discussion is accompanied by a limited but illustrative set of original visualizations obtained from the application of such schemes to a toy model.

Ensemble Kalman filters conventionally update the state having taken into account only the ensemble mean and covariance thereby implicitly making the analysis Gaussian. Issues may arise when such a scheme is applied to practical problems likely to be dominated by highly nonlinear and inherently non-Gaussian processes since neglect of higher order information may lead to a meaningful degradation in analysis quality. Of interest is how the significance of such degradation can be reduced when data assimilation schemes employ an ensemble Kalman filter (EnKF) variant capable of non-Gaussian support. The potential improvement in model initialization due to use of a non-Gaussian assimilation scheme is further motivated by a high level description of earth system predictability derived from land surface state characterizations, the component of focus in this text being soil moisture.

# List of Figures

## List of Symbols

$\mathbf{x}$ - model state vector

$\mathbf{M}$ or $\mathcal{M}$ - model propagator or process model

$\mathbf{P}$ - error covariance matrix

$\mathbf{K}$ - Kalman gain

$\mathbf{H}$ or $\mathcal{H}$ - observation operator

$\mathbf{y}$ - observation vector

$m$ - number of ensemble members

$\mathbf{Q}$ - model-error covariance matrix

$\mathbf{R}$ - observation-error covariance matrix

BC - climatological ensemble-covariance matrix used in original examples

$\mathbf{X}$ - ensemble of normalized state vector perturbations

$\mathbf{Y}$ - ensemble of normalized observation vector perturbations

$\mathbf{d}$ - innovation; difference between observation and model's predicted observation

$\mathbf{w}$ - vector used to parameterize deterministic variants of ensemble Kalman filter analysis in subspace of the ensemble

$\mathbf{S}$ - perturbation ensemble mapped to observation space

$\mathbf{W}$ - right transformation matrix used to update ensemble forecast perturbations in deterministic variants of ensemble Kalman filter

$J$ - variational cost function

$\mathbf{B}_o$ - static background-error covariance matrix

$\delta\mathbf{x}$ - incremental 4DVar control variable; difference between arbitrary state and reference solution

$\delta\chi_{VAR}$ - alternate control variable for incremental 4DVar obtained applying control variable transform $\mathbf{L}$

$\mathbf{L}$ - control variable transform for 4DVar used to precondition Hessian of cost function

$\mathbf{B}^h$ - hybridized background-error covariance matrix used in 4DEnVar and En4DVar

$\gamma$ - scalar weight used to compute hybridized background-error covariance matrix

$\delta\boldsymbol{\chi}_{ENS}$ - alternate control variable for incremental 4DVar obtained applying control variable transform $\mathbf{X}^f$

$\delta\boldsymbol{\chi}_h$ - hybridized control variable for incremental 4DVar obtained using $\mathbf{U}_h$

$\mathbf{U}_h$ - hybridized control variable transform

# Part I

## Overview of Ensemble Kalman Filtering

Data assimilation is the process whereby observations of a system are combined with a priori information, such as a forecast output by an approximative model of that system's dynamics, so as to determine a more accurate state, the estimation of which takes into account the relative uncertainty in the model forecast and observations. It is worthwhile to review how this state estimation problem can be approached using ensemble Kalman filtering methods.

The following literature review of ensemble Kalman filtering schemes will be bolstered by a number of original examples highlighting their application to the Lorenz-63 (L63) model, a 3-variable model that exhibits chaotic behavior. Please see the appendix for experimental details.

## 1  Developmental Provocation

As follows are the forecast and analysis steps for the Kalman Filter (Kalman, 1960)

$$\mathbf{x}^f_{k+1} = \mathbf{M}_{k+1}\mathbf{x}^a_k \tag{1.1}$$

$$\mathbf{P}^f_{k+1} = \mathbf{M}_{k+1}\mathbf{P}^a_k\mathbf{M}^T_{k+1} + \mathbf{Q}_{k+1} \tag{1.2}$$

$$\mathbf{K}_{k+1} = \mathbf{P}^f_{k+1}\mathbf{H}^T(\mathbf{H}\mathbf{P}^f_{k+1}\mathbf{H}^T + \mathbf{R}_{k+1})^{-1} \tag{1.3}$$

$$\mathbf{x}^a_{k+1} = \mathbf{x}^f_{k+1} + \mathbf{K}_{k+1}(\mathbf{y}_{k+1} - \mathbf{H}\mathbf{x}^f_{k+1}) \tag{1.4}$$

$$\mathbf{P}^a_{k+1} = (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H})\mathbf{P}^f_{k+1} \tag{1.5}$$

Following the convention of Ide et al. (1997), let $\mathbf{x}$ be the model state vector and $\mathbf{P}$ be the covariance matrix of the error in the estimate $\mathbf{x}$, which are each propagated forward in time from some previous analysis point according to the process model, $\mathbf{M}$; let $\mathbf{Q}$ be the model-error covariance matrix; let $\mathbf{y}$ be the vector containing the observations to be assimilated and $\mathbf{H}$ be the observation operator, which predicts the model-equivalent

7

of the observations via a mapping of the state vector **x** to observation space; let **R** be the observation-error covariance matrix and **K** be the Kalman Gain (i.e., the influence matrix); the $f$ and $a$ superscripts refer to forecast and analysis quantities, respectively; lastly, $k$ represents the model time step, of which the explicit denotation is hereafter mostly avoided for convenience.

The Kalman filter is optimal in a least-squares sense only in the case of linear operators and Gaussian-distributed system errors. For problems characterized by minor deviations from such assumptions, the extended Kalman filter (EKF) has been shown to be a relatively more capable analysis technique (Evensen, 1992). The forecast and analysis equations for the EKF are identical to those of the standard Kalman filter but with the added constraint that the model and observation operators, $\mathcal{M}$ and $\mathcal{H}$, if nonlinear, first be expanded (via Taylor Series) about some reference solution, usually taken to be the previous analysis state.

The linearization makes it such that the EKF accounts for the error statistics only up to second order (mean and covariances), which can result in a loss of information content when considering complex non-Gaussian distributions. Neglect of higher-order moments of the background error covariance matrix, given model-dependent contributions from nonlinear effects, may lead to unphysical error growth and ultimately filter divergence (Burgers et al., 1998) (*for divergence prevention, see 4.1.2: Inflation*). One such result is given by Evensen (1992), who having carried out experiments with the EKF used in conjunction with a multi-layer QG ocean model, noted that the presence of nonlinear small-scale instabilities in the approximated background-error covariance evolution equation resulted in poor state estimates. Miller et al. (1994) achieved moderate improvements in EKF performance when third- and fourth-order moment closure equations were used with the L63 model, but it was questioned whether such improvements merit the associated increase in computational expense.

In the context of numerical weather prediction (NWP), neither the Kalman filter nor its extension is an appropriate analysis scheme. The unsuitability stems in part from the necessity to simulate highly nonlinear processes and non-Gaussian-distributed variables

such as those of relevance to the characterization of aerosols, rainfall, sea ice, and land surface states (Bishop, 2016); also posing as a hindrance to successful operational implementation is the size of the state vector $\mathbf{x}$ ($\sim O(10^7)$ in current operational NWP models), which makes the $nxn$-dim background-error covariance matrix too large to be held in computer memory, let alone evolved in time according to Eq.(1.5).

A more computationally feasible filtering technique for handling high-dimensionality is the reduced-rank Kalman filter (RRKF; Verlaan and Heemink (1997)), which for some initial system state $\mathbf{x}_0^f$, the associated error-covariance matrix is approximated as $\mathbf{P}_0^f \simeq \mathbf{D}_0^f \mathbf{D}_0^{f,T}$, where $\mathbf{D}_0^f$ is an $nxm$ matrix with columns equal to the leading $m$ principal axes of $\mathbf{P}_0^f$. The equations for the RRKF are similar to that for the standard Kalman filter expressed above with the only variation being the necessary decomposition of the background-error covariance matrix. The effectiveness of the RRKF depends on how well the error-covariance matrix is represented by a limited (ideally $m << n$) number of modes. Still problematic, though, is the assumption of linear error growth implied by the use of the tangent-linear model in the forecast step (Eq.(1.2)). A reduced-rank filter that is better able to account for nonlinearity in models is the Ensemble Kalman filter (EnKF), a Monte Carlo approximation to the deterministic EKF. Using the members of an ensemble of model states/forecasts to approximate the second moment of the background-error probability density function (pdf), the EnKF avoids the explicit forecasting of the full-rank background-error covariance matrix, thus making the EnKF especially attractive given that it circumvents the computation of the model operator's tangent-linear and adjoint.

At the highest level, there are two styles of EnKF: (i) stochastic and (ii) deterministic.

## 2   The Stochastic (Perturbed Observation) EnKF
## (Burgers et al., 1998)

Given an ensemble of model forecasts, $x_i$, $i = 1, 2, ...m$, the update equation for the mean state at the analysis step is

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}(\bar{\mathbf{y}} - \mathbf{H}\bar{\mathbf{x}}^f) \tag{2.1}$$

and that for the $ith$ member of the ensemble is

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_i^f) \tag{2.2}$$

where $\mathbf{K}$, the Kalman Gain, is given by

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \tag{2.3}$$

The background-error covariance matrix, instead of being represented explicitly by a full-rank matrix as in the standard Kalman filter or in the EKF, is now being diagnosed as the sample covariance of the forecast ensemble according to

$$\mathbf{P}^f = \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{x}^f - \bar{\mathbf{x}}^f)(\mathbf{x}^f - \bar{\mathbf{x}}^f)^T \tag{2.4}$$

where $\bar{\mathbf{x}}^f$ represents the ensemble mean. Subtracting Eq.(2.1) from Eq.(2.2) and normalizing the result by $\frac{1}{\sqrt{m-1}}$ yields normalized deviations of the analysis that can be represented by

$$\mathbf{X}^a = \mathbf{X}^f + \mathbf{K}(\mathbf{0} - \mathbf{H}\mathbf{X}_i^f) = (\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{X}_i^f \tag{2.5}$$

where $\mathbf{I_n}$ is the $nxn$ identity matrix and $\mathbf{X}^{f,a}$ is the augmented matrix of dimension $nxm$ whereby each column is given by the normalized deviation of an ensemble member's state from the ensemble mean value (i.e. $\frac{1}{\sqrt{m-1}} \left[ \mathbf{x}_1^{f,a} - \bar{\mathbf{x}}^{f,a} , \mathbf{x}_2^{f,a} - \bar{\mathbf{x}}^{f,a} , \ldots , \mathbf{x}_m^{f,a} - \bar{\mathbf{x}}^{f,a} \right]$). The expectation of the outer product of Eq.(2.5) with itself yields the low-order ensemble

approximation to the analysis-error covariance matrix

$$\mathbf{P}^a = \mathbf{X}_a\mathbf{X}_a^T = (\mathbf{I}_n - \mathbf{KH})\mathbf{X}_f\mathbf{X}_f^T(\mathbf{I}_n - \mathbf{KH})^T \tag{2.6}$$
$$= (\mathbf{I}_n - \mathbf{KH})\mathbf{P}^f(\mathbf{I}_n - \mathbf{KH})^T$$

which assumes that background and observation errors are uncorrelated with each other. Of note for comparison is the analysis-error covariance matrix equation derived in the formulation of the standard Kalman filter

$$\mathbf{P}^a = (\mathbf{I}_n - \mathbf{KH})\mathbf{P}^f(\mathbf{I}_n - \mathbf{KH})^T + \mathbf{KRK}^T \tag{2.7}$$

again assuming background and observation errors are not correlated with each other. The analysis-error covariance matrix determined from the ensemble in Eq.(2.6) does not directly take into account the observation-error information, leading to a built-in under-estimation of the analysis-error covariance that makes the EnKF more likely to diverge (i.e. ignore observations in subsequent cycles) due to an increasingly overconfident background.

The stochastic variant of the EnKF handles this issue by giving each member in the ensemble a unique observation vector, $\mathbf{y}_i = \mathbf{y} + \mathbf{w}_i$ with which to perform it's analysis, where $\mathbf{w}_i$ is a vector belonging to the pdf distributed as $\mathcal{N}(\mathbf{0}, \mathbf{R})$. The observation deviation is now

$$\mathbf{Y}_i^f = \frac{\mathbf{Hx}_i^f - \mathbf{y}_i - \mathbf{H\bar{x}}^f + \bar{\mathbf{y}}}{\sqrt{m-1}} \tag{2.8}$$

and Eq.(2.5) now becomes

$$\mathbf{X}_i^a = \mathbf{X}_i^f - \mathbf{KY}_i^f = (\mathbf{I}_n - \mathbf{KH})\mathbf{X}_i^f + \mathbf{K}\frac{\mathbf{y}_i - \bar{\mathbf{y}}}{\sqrt{m-1}} \tag{2.9}$$

With the expression for the analysis deviations as given in Eq.(2.9), computation of the analysis-error covariance matrix yields a result identical to that given by Eq.(2.7)

11

thereby ensuring that for increasingly larger ensembles sizes, the analysis-error covariance matrix determined by the filter will tend to the correct value.

## 3 The Deterministic EnKF

The method of perturbing the observation vector, as in the stochastic variant of the EnKF, acts to introduce an extra source of sampling error in the analysis related to the estimation of the correct observation-error statistics. The deterministic EnKF also ensures that Eq.(2.7) holds asymptotically for larger ensemble sizes, but does so without the generation of an ensemble of observations. Firstly, the following expression becomes convenient

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{X}^f \mathbf{w} \tag{3.1}$$

whereby the analysis increment is defined as a linear combination of the forecast perturbations, as is effectively the case for the stochastic EnKF, but here is parameterized in the ensemble subspace by a vector $\mathbf{w}$. By setting Eq.(3.1) equal to Eq.(2.1), it can be found that

$$\mathbf{w} = \mathbf{S}^{f,T}(\mathbf{S}^f \mathbf{S}^{f,T} + \mathbf{R})^{-1}\overline{\delta} \tag{3.2}$$

where $\mathbf{S}^f = \mathbf{H}\mathbf{X}^f$ and $\overline{\delta} = \overline{\mathbf{y}} - \mathbf{H}\overline{\mathbf{x}}^f$. Once the mean analysis state is calculated, it becomes necessary to determine the posterior ensemble. If $\mathbf{K}$ is as given in Eq.(2.3), then $\mathbf{P}^a$ in Eq.(2.7) can be rewritten as

$$\mathbf{P}^a = (\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{P}^f \tag{3.3}$$

whereby the square root formulation for the error-covariance matrices allows for

$$\mathbf{X}^a \mathbf{X}^{a,T} = \mathbf{X}^f(\mathbf{I}_m - \mathbf{S}^{f,T}(\mathbf{S}^f \mathbf{S}^{f,T} + \mathbf{R})^{-1}\mathbf{S}^f)\mathbf{X}^{f,T} \tag{3.4}$$

and, finally

$$\mathbf{X}^a = \mathbf{X}^f (\mathbf{I}_m - \mathbf{S}^{f,T}(\mathbf{S}^f\mathbf{S}^{f,T} + \mathbf{R})^{-1}\mathbf{S}^f)^{\frac{1}{2}}\mathbf{U} \tag{3.5}$$

$$= \mathbf{X}^f \mathbf{T}^{\frac{1}{2}}\mathbf{U}$$

$$= \mathbf{X}^f \mathbf{W}$$

where $\mathbf{T} = (\mathbf{I}_m - \mathbf{S}^{f,T}(\mathbf{S}^f\mathbf{S}^{f,T} + \mathbf{R})^{-1}\mathbf{S}^f)^{\frac{1}{2}}$, $\mathbf{U}$ is some *mxm* orthogonal matrix (i.e. $\mathbf{U}\mathbf{U}^T = \mathbf{I}_m$), and $\mathbf{W}$ is the right transformation matrix of the forecast perturbation ensemble $\mathbf{X}^f$. The analysis state for each ensemble member can then be determined by adding $\bar{\mathbf{x}}^a$ as given in Eq.(3.1) to each column in the appropriately scaled analysis perturbation matrix $\mathbf{X}^a$.

## 3.1 Popular Deterministic EnKF Variants

The only constraint on the matrix $\mathbf{U}$ in Eq.(3.5) is that it be orthogonal, meaning that it is non-unique. This makes it possible for the coexistence of many different forms of the deterministic EnKF, all of which are equivalent given that Eq.(3.5) is ultimately satisfied.

### 3.1.1 Ensemble transform Kalman filter (ETKF)
### (Bishop et al., 2001)

One such variant is the ETKF. The analysis-error covariance equation as given in Eq.(3.3) can be expanded to read as

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} \tag{3.6}$$

which after having made the substitution $\tilde{\mathbf{H}} = \mathbf{R}^{-\frac{1}{2}}\mathbf{H}$ now becomes

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f\tilde{\mathbf{H}}^T(\tilde{\mathbf{H}}\mathbf{P}^f\tilde{\mathbf{H}}^T + \mathbf{I}_p)^{-1}\tilde{\mathbf{H}}\mathbf{P}^f \tag{3.7}$$

Among the most computationally expensive components of an EnKF algorithm

are those involving matrix inversions. From Eq.(3.6) and Eq.(3.7), it is clear that such inversions are performed mainly in the observation space. The ETKF is formulated such that any inversions are instead handled in a subspace of significantly lower dimension, allowing for relatively less computational demand associated with the algorithm. The ETKF accomplishes this by taking advantage of the rank-deficiency of the ensemble (in high-dimensional problems), which has a subspace spanned by at most $(m - 1)$ linearly independent vectors (since $\sum_{i=1}^{m} \mathbf{X}^{f,a} = \mathbf{0}$). In accordance with Eq.(3.5), the analysis must be a linear combination of the forecast perturbations. If an eigenvalue decomposition of $(\tilde{\mathbf{H}}\mathbf{P}^f\tilde{\mathbf{H}}^T + \mathbf{I}_p)^{-1}$ is calculated, it can be seen that at most only $(m-1)$ of the eigenvectors of $(\tilde{\mathbf{H}}\mathbf{P}^f\tilde{\mathbf{H}}^T + \mathbf{I}_p)^{-1}$ (or equivalently $\tilde{\mathbf{H}}\mathbf{P}^f\tilde{\mathbf{H}}^T$) can even be represented in the ensemble subspace. This necessary subset can be determined by taking the following eigenvalue decomposition

$$\mathbf{X}^{f,T}\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}\mathbf{X}^f = \mathbf{C}\mathbf{\Gamma}\mathbf{C}^T \tag{3.8}$$

by which the transformation matrix from Eq.(3.5) can now be written as

$$\mathbf{W} = \mathbf{C}(\mathbf{\Gamma} + \mathbf{I}_m)^{-\frac{1}{2}} \tag{3.9}$$

where the orthonormal eigenvectors of $\mathbf{X}^{f,T}\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}\mathbf{X}^f$ make up the columns of $\mathbf{C}$, and $\mathbf{\Gamma}$, a diagonal matrix, has as its entries the corresponding eigenvalues.

### 3.1.2 Local ensemble transform Kalman filter (LETKF) (Hunt et al., 2007)

The form of $\mathbf{W}$ was not made explicit by Bishop et al. (2001), whereas the LETKF, a popular derivative of the ETKF, specifies the right transformation matrix as a symmetric square root via the following eigenvalue decomposition

$$(\mathbf{W}\mathbf{W}^T)^{-1} = \mathbf{U}\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{U}^T \tag{3.10}$$

which guarantees that the analysis perturbations are centered about the mean analysis state (i.e. $\mathbf{X}^a \mathbf{1} = \mathbf{0}$ where $\mathbf{1} = (1, 1, ..., 1)^T$) with the added benefit being generally superior filter performance compared to when $\mathbf{W}$ is formulated differently. An attempt at explaining the latter point can be made by considering that $\| \mathbf{W} - \mathbf{I}_m \|_{\mathcal{F}}$ is at a minimum for when $\mathbf{W}$ is a symmetric square root; $\| \cdot \|_{\mathcal{F}}$ is the Frobenius norm defined for some $axb$-dimensional matrix $\mathbf{G}$ as $\| \mathbf{G} \|_{\mathcal{F}} = \left( \sum_{i=1}^{a} \sum_{j=1}^{b} | q_{i,j} |^2 \right)^{\frac{1}{2}}$. Consequently, in the ensemble subspace, a forecast ensemble member is minimally adjusted by the analysis, mitigating the severity with which physical balance properties in the state may be violated as a result of the update. Hunt et al. (2007) also suggest the mitigation of sampling error via localization and inflation, both of which are elaborated on more generally in later sections of this text.

### 3.1.3 Other deterministic EnKF variants

Other deterministic EnKFs exist - the singular evolutive interpolated Kalman filter (SEIK) of Pham (2001); the ensemble adjustment Kalman filter (EAKF) of Anderson (2001), which involves applying a left transformation matrix to the forecast perturbation ensemble instead of a right transformation matrix as specified in Eq.(3.5); the ensemble square root filter (EnSRF) of Whitaker and Hamill (2002), the introduction of which is accompanied by a framework that outlines the computationally efficient serial processing of observations. Again, each variant is mathematically equivalent to the other given that the analysis-error covariance of the ensemble tends asymptotically to the correct value for larger ensemble sizes. The difference between any two of the methods is restricted primarily to algorithmic implementation (Tippett et al., 2003).

## 4 Addressing Various Sources of System Error

### 4.1 Localization & Inflation

In an atmospheric or oceanic context, the number of members used in the EnKF is usually much smaller than the dimension of the model state. The rank-deficiency

associated with the ensemble approximations of the background- and analysis-error covariance matrices results in spurious long-range correlations. Two methods used to mitigate the negative effects of such correlations on filter performance are localization and inflation.

### 4.1.1 Localization

The premise behind localization is that, for a given timescale, the states of two spatially distinct points separated by some distance $r$ within a system are better approximated as being independent from one another with increasing $r$. The exploitation of this tendency allows for the analysis to be spatially localized.

One variant of localization is **R** localization (domain localization) where the **R** matrix or its inverse is multiplied by a distance-dependent function such that, given some analysis point in the model space, the errors associated with increasingly distant observations tend toward infinity. One such cutoff function is the Gaspari-Cohn function, a fifth-order piecewise rational function that resembles a Gaussian distribution but has compact support (observations outside some prescribed localization length scale are ignored during the analysis) (Gaspari and Cohn, 1999).

A second approach is **B** localization (Schur localization), which directly modifies the ensemble approximation to the background-error covariance matrix such that **B** now becomes $\rho \circ \mathbf{B}$, where $\circ$ denotes the Schur (pointwise) product, and $\rho$ is some short-range correlation matrix.

In their experiments with the Simplified Parameterizations, Primitive Equation Dynamics (SPEEDY) global atmospheric model, Greybush et al. (2011) found that **B** localization has an optimal length scale greater than that for **R** localization; but **R** localization produced more balanced analyses at all length scales. The latter finding is of note because an unbalanced analysis used to initialize an atmospheric model, for example, will create spurious inertial-gravity waves that act to speed up dissipation of the memory that the model has of any assimilated observations, negatively impacting

the accuracy of subsequent forecasts.

There are techniques such as low-pass filters and nonlinear normal mode initialization that can be used to impose balance, but it is best that data assimilation be tasked with the production of physically appropriate levels of balance in the analysis, given that the use of such extra-analysis applications do not guarantee an optimal state estimate (Greybush et al., 2011).

**Adaptive Localization**   The aforementioned forms of localization require the a priori specification of a localization function. For high-dimensional and multivariate systems, a physically appropriate general set of functions and parameters may not be readily available. A form of adaptive localization proposed by Anderson (2007) employs a hierarchical EnKF in which the regression coefficient (scalar coefficient of the innovation vector in the update equation) between an observation of some variable at a fixed location and all state variables of the model is given via a distribution sampled by a long assimilation of multiple ensembles. Aided by the spread of the resulting distributions, empirically-determined regression coefficients can be provided to conventional filter configurations. Bishop and Hodyss (2009) take the ensemble correlations and raise them to a power (ECO-RAP) so as to dampen the magnitude of small correlations while making those of larger correlations more pronounced. Lei and Anderson (2013) perform a short series of observing system simulation experiments (OSSEs) that employ the desired observation network configuration to derive a spatially varying set of empirical localization functions that minimize the mean-square error of observations from the true solution; assuming the OSSE is configured well, the empirical localization functions should be appropriate for assimilating real observations.

The use of adaptive methods, a small subset of which have been mentioned, indirectly takes into account the flow-dependent dynamical information that the ensemble provides, explaining the associated improvement in filter performance when compared to the **B**- and **R**-localization (nonadaptive) variants discussed earlier.

### 4.1.2 Inflation

The localization of the analysis takes advantage of the lower dimensionality of local spaces (Pham, 2001), but still present are sampling errors due to finite ensemble size that encourage the divergence of the filter, whereby the observations are ignored leading to a model state no longer deterministically representative of the real system. The spread of the ensemble distribution can be increased, or inflated, to account for errors poorly sampled, if at all, by the ensemble.

Multiplicative inflation involves increasing the amplitude of the ensemble spread about its mean value, and can be represented by the following transformation

$$\mathbf{x}_i^a \longrightarrow \overline{\mathbf{x}}_i^a + \lambda \left( \mathbf{x}_i^a - \overline{\mathbf{x}}^a \right) \tag{4.11}$$

applied to each member of the ensemble, where $\lambda$ is some tunable parameter. Additive inflation involves adding noise to each ensemble member in a form similar to

$$\mathbf{x}_i^f \longrightarrow \mathbf{x}_i^f + \epsilon_i \tag{4.12}$$

where $\epsilon$ is of covariance $\mathbf{Q}$ (not necessarily the model error). Multiplicative inflation tends to be too aggressive for regions in the control space for which there is a sparse set of observations while additive inflation has been shown to decrease the ability of $\mathbf{B}$ to capture the flow-dependent errors-of-the-day (Kalnay, 2003). Whitaker and Hamill (2012) proposed the Relax-To-Prior-Spread (RTPS) method, a form of multiplicative inflation that inflates the posterior distribution by an amount proportional to that by which the prior is reduced due to the assimilation of observations

$$\sigma^a \longrightarrow (1 - \alpha)\, \sigma^a + \alpha \sigma^b \tag{4.13}$$

where $\alpha$ is some tunable parameter and $\sigma^b \equiv \sqrt{[1/(m-1)] \sum_{i=1}^{n} \mathbf{x}_i'^{b^2}}$ and $\sigma^a \equiv \sqrt{[1/(m-1)] \sum_{i=1}^{n} \mathbf{x}_i'^{a^2}}$ are the prior and posterior ensemble standard deviations, respectively. This is in contrast to the Relax-To-Prior-Perturbation (RTPP) method outlined in

Eq.(4.11). Whitaker and Hamill (2012) incorporated the RTPS and RTPP methods in experiments with a 2-level primitive equations spectral model and the serial ensemble square root filter (EnSRF) of Whitaker and Hamill (2002); assimilated were simulated radiosonde measurements in conjunction with the gaussian-like polynomial function of Gaspari and Cohn (1999) for $\mathbf{B}$ localization; for optimal inflation parameter values, it was concluded that the RTPS method yielded more accurate analyses than those obtained using the RTPP method when the dominant form of error present in the assimilation system was sampling error (due to use of a finite ensemble); for the scenario in which model error dominated sampling error, the RTPP method performed better than the RTPS method; if the error contributions by the model and ensemble were comparable to one another, then the analysis error was minimized by employing a combination of both methods; the results imply that it is worth considering if errors can be further compartmentalized whereby each component is treated uniquely so as to improve on current ensemble filtering schemes. In the case of model error, for example, one could manage model error due to convection with one method and model error due to boundary layer physics with another, and so on (*see 4.2: A Word on Model Error*).

**Adaptive inflation** Just as for localization, there are adaptive techniques for inflation as well. Wang and Bishop (2003) estimate a time-dependent inflation factor using the innovation statistics $\langle \mathbf{d}^T \mathbf{d} \rangle = trace(\pi \mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})$ where $\langle \cdot \rangle$ is the expectation operator, $\mathbf{d} = \mathbf{y} - \mathbf{H} \mathbf{x}^f$ is the innovation, and $\pi$ is the inflation factor; note that it is assumed that the observation-error covariance matrix is known which is a limiting factor on how much filter performance may be improved when the inflation scheme is applied in a real-observation context. Anderson (2007) uses a hierarchical Bayesian approach to develop an adaptive covariance inflation algorithm that directly incorporates the observations to be assimilated. Li et al. (2009), using the various diagnostics of system error presented by Desroziers et al. (2005), observe that the relationship between the observation-error covariance matrix and an ensemble inflation parameter is nonlinear, and suggest a method that estimates the two simultaneously; this approach assumes

that the observation-error covariance matrix is diagonal. Liang et al. (2012) propose a maximum likelihood estimation (MLE) method that is capable of inflating both the background- and observation-error covariance matrices while also allowing for the more general case whereby observation errors are spatially correlated.

## 4.2  A Word on Model Error

The ensemble analysis-error covariance may be underestimated not only because of sampling error due to finite ensemble size, but also because of the presence of extrinsic model error not accounted for elsewhere by the assimilation system. To an extent, model error can be addressed via the use of localization and inflation. This treatment is made more explicit by methods that attempt to account for model error by the application of a unique stochastic perturbation to each member of the forecast ensemble. A set of such perturbations may take a number of different forms and can target the parameterization schemes used for sub-grid scale processes (as in stochastically perturbed physics tendencies (SPPT schemes; Buizza et al. (1999))), how kinetic energy transfer between resolved and unresolved scales is approximated (as in stochastic kinetic energy backscatter (SKEB schemes; Shutts (2005))), the tendency terms in the forward model (as in stochastic total tendency perturbations (STTP schemes; Hou et al. (2006))), and/or effects associated with variability in boundary layer humidity (as in stochastic perturbed humidity (SHUM schemes; Tompkins and Berner (2008))).

## 5  Ensemble-Variational Methods

The discussion up to this point has been concerned with filters belonging exclusively to the more general class of sequential data assimilation methods. It is worthwhile to explore ways the EnKF can be implemented in tandem with other types of schemes whereby more sophisticated assimilation systems are created. Firstly, to better understand such systems, a run-through of four-dimensional variational data assimilation (4DVar) is provided.

## 5.1  4DVar Run-through

In NWP, historically, data assimilation has primarily utilized variational techniques, which determine an analysis via the minimization of a cost function. Assuming Gaussian-distributed system errors, the cost function for 4DVar, which provides for the consideration of observations (presumably acquired asynchronously) at the appropriate time, can be written as follows

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}^f)^T \mathbf{B}_o^{-1} (\mathbf{x} - \mathbf{x}^f) \, ... \tag{5.14}$$

$$+ \frac{1}{2} \sum_{k=0}^{P} (\mathcal{H}_k \mathcal{M}_k(\mathbf{x}) - \mathbf{y}_k)^T \mathbf{R}_k^{-1} (\mathcal{H}_k \mathcal{M}_k(\mathbf{x}) - \mathbf{y}_k) \, ...$$

$$+ \frac{1}{2} \sum_{k=1}^{P} (\mathbf{x}_k - \mathcal{M}_k(\mathbf{x}_{k-1}))^T \mathbf{Q}_k^{-1} (\mathbf{x}_k - \mathcal{M}_k(\mathbf{x}_{k-1}))$$

where the subscript $k$ indicates the appropriate observation time in the assimilation window over which there are a total of $P + 1$ observations. The formulation given in Eq.(5.14) is a weak-constraint 4DVar cost function whereby the analysis need not be a solution of the forward model. Omission of the terms normalized by $\mathbf{Q}_k$ in Eq.(5.14) yields a cost function for strong-constraint 4DVar, which requires that the analysis trajectory indeed be an exact solution of the forward model.

**Incremental 4DVar**  If either $\mathcal{H}_k$ or $\mathcal{M}_k$ acting on the control variable is nonlinear, the cost function may have multiple minima. The cost function can be approximated as a quadratic by using versions of the operators that have been linearized about a reference state $\mathbf{x}^g$ as is done in the incremental formulation of 4DVar. Ignoring model error and leveraging the somewhat more compact $\mathcal{L}^2$-norm notation, the incremental form of Eq.(5.14) is as given

$$J(\delta\mathbf{x}) = \frac{1}{2} \, \| \, \delta\mathbf{x} \, \|_{\mathbf{B}_o^{-1}}^2 + \frac{1}{2} \sum_{k=0}^{P} \| \, \delta\mathbf{d} \, \|_{\mathbf{R}_k^{-1}}^2 \tag{5.15}$$

where $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^g$, $\delta\mathbf{d} = \mathbf{y}_k - \mathcal{H}_k\mathbf{x}^f - \mathbf{H}_k\mathbf{M}_k\delta\mathbf{x} - \mathbf{y}_k$, and $\mathbf{x}^g$ is taken to be equal to $\mathbf{x}^f$. In practice, Eq.(5.15) uses simplified operator linearizations (e.g. reduced spatial resolution) and is minimized iteratively whereby the solution $\mathbf{x}$ is used as the reference state $\mathbf{x}^g$ in the subsequent loop (Trémolet, 2007).

**Preconditioning**   Minimization of Eq.(5.14) and Eq.(5.15) requires the use of a numerical optimization algorithm. The behavior of such algorithms is dependent on the condition number of the Hessian of the cost function, where given some eigenspectrum, the condition number, in this text, is defined as the ratio of the eigenvalue with the largest magnitude to that of the smallest. It is possible to precondition the Hessian so that the condition number is closer to one, which leads to better convergence rates for optimization algorithms. Consider the following

$$\delta\mathbf{x} = \mathbf{L}\delta\chi_{VAR} \tag{5.16}$$

where $\mathbf{L}$ is a control variable transform (CVT). The strong-constraint incremental cost function can now be written as

$$J(\delta\chi_{VAR}) = \frac{1}{2} \parallel \delta\chi_{VAR} \parallel^2_{\mathbf{L}^T\mathbf{B}_o^{-1}\mathbf{L}} + \frac{1}{2} \sum_{k=0}^{P} \parallel \delta\mathbf{d} \parallel^2_{\mathbf{R}_k^{-1}} \tag{5.17}$$

If $\mathbf{B}_o = \mathbf{L}\mathbf{L}^T$, then the normalization matrix $\mathbf{L}^T\mathbf{B}_o^{-1}\mathbf{L}$ in the first term of Eq.(5.17) becomes equal to the identity matrix. The usefulness of the CVT can be illustrated by first supposing that we want to assimilate a single observation that is valid at the same time as the initial background state. For such a case (three-dimensional (or stationary) variational data assimilation (3DVar)), the Hessian of Eq.(5.17) is equal to $(\mathbf{B}_o^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})$, which tends to have a very high condition number. Taking into account a CVT of the form derived earlier, the Hessian becomes $(\mathbf{I} + \mathbf{L}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L})$, the result being, generally, a better conditioned problem.

For any variational method, $\mathbf{B}_o$ is crudely estimated one of a number of ways (details on some of the more notable estimation, or calibration, methods are given in the

appendix). The key point is that $\mathbf{B}_o$ is full-rank but quasi-static. This is in contrast to the rank-deficient, flow-dependent background-error covariance description that is used in the EnKF. Combining ensemble and variational methods so as to exploit/mitigate the strengths/weaknesses of either when implemented in isolation is the goal of Ensemble-Variational (EnVar) data assimilation.

EnVar has many forms, all of which involve either (i) 1-way coupling where information is passed from one method to another but not the other way around (e.g. an analysis determined from 4DVar (with calibrated $\mathbf{B}_o$) is used to recenter an EnKF ensemble running in parallel) or (ii) 2-way coupling whereby information is exchanged with and used by both the sequential and variational subsystems (e.g. 4DVar incorporates background-error information provided by an ensemble running in parallel to determine an analysis which is then used to recenter that same aforementioned ensemble). Here, an overview of some EnVar methods with 1-way coupling will be given.

For the purposes of the following discussion, the components making up each EnVar method are assumed to be 4DVar and EnKF schemes unless stated otherwise. Such EnVar methods are subdivided into pure and hybrid variants.

## 5.2  Pure EnVar

Pure EnVar methods determine the analysis using a variational framework that incorporates the background-error information provided by an ensemble. En4DVar and 4DEnVar, the methods to be discussed here, accomplish this in slightly different ways whereby the tangent-linear of $\mathcal{M}_k$ and $\mathcal{H}_k$ are directly employed in the former but not in the latter.

**Pure En4DVAR**  Consider the following

$$\delta\mathbf{x} = \mathbf{X}^f \delta\chi_{ENS} \tag{5.18}$$

where $\mathbf{X}^f$ is the CVT. Eq.(5.15) can be rewritten to arrive at the following cost function for (pure) En4DVar

$$J(\delta\chi_{ENS}) = \frac{1}{2} \parallel \delta\chi_{ENS} \parallel_{\mathbf{I}}^2 \dots \qquad (5.19)$$
$$+ \frac{1}{2} \sum_{k=0}^{P} \parallel \mathbf{y}_k - \mathcal{H}_k\mathbf{x}^f - \mathbf{H}_k\mathbf{M}_k\mathbf{X}^f\chi_{ENS} \parallel_{\mathbf{R}_k^{-1}}^2$$

where it is implicit that $\mathbf{B}_o = \mathbf{X}^f\mathbf{X}^{f,T}$.

**Pure 4DEnVar**   For large and complex systems, it may be necessary to use simplified versions of the tangent-linear operators (e.g. due to switches in code). Additionally, the derivation and maintenance of tangent-linear code, approximated or not, can require what may be a prohibitive number of man-hours. The desire to entirely avoid the tangent-linear in (pure) En4DVar is a key motivator behind the (pure) 4DEnVar formulation. Starting with the following approximation for the $i$-th column of the forecast perturbation matrix in Eq.(5.19)

$$\mathbf{H}_k\mathbf{M}_k\mathbf{X}_i^f \approx \frac{1}{m-1}\left(\mathcal{H}_k\mathcal{M}_k\mathbf{x}_i^f - \mathcal{H}_k\mathcal{M}_k\bar{\mathbf{x}}^f\right) \qquad (5.20)$$

yielded is an expression equivalent to the $i$-th column of $\mathbf{Y}_k$, so that now

$$J(\delta\mathbf{x}) = \frac{1}{2} \parallel \delta\chi_{ENS} \parallel_{\mathbf{I}}^2 \dots \qquad (5.21)$$
$$+ \frac{1}{2} \sum_{k=0}^{P} \parallel \mathbf{y}_k - \mathcal{H}_k\mathbf{x}^f - \mathbf{Y}_k\delta\chi_{ENS} \parallel_{\mathbf{R}_k^{-1}}^2$$

for which tangent-linear operators are not needed.

## 5.3   Hybrid EnVar

The aforementioned En4DVar and 4DEnVar can be hybridized such that the $\mathbf{B}_o$ matrix is some combination of each of the hybrid's subsystem's description of the background errors ($\mathbf{B}_o$ and $\mathbf{P}^f$ for pure 4DVar and the pure EnKF, respectively). This can be

done via a simple weighted average whereby the resultant hybridized background-error covariance matrix $\mathbf{B}^h$ takes the place of $\mathbf{B}_o$ in a variational cost function and may be made explicit in similar fashion to (Hamill and Snyder, 2000) by

$$\mathbf{B}^h = (1 - \gamma)\mathbf{B}_o + \gamma \mathbf{P}^f \tag{5.22}$$

where $\gamma$ is some tunable scalar parameter with the constraint $0 \le \gamma \le 1$.

Eq.(5.22) suggests that a full-rank $\mathbf{B}_o$ needs to be provided, which may be problematic for high-dimensional systems. $\mathbf{B}^h$ can be determined implicitly by appropriately augmenting the control vector (Wang et al., 2007), which allows for reformulation of the incremental cost function

$$J([\delta\chi_{VAR} \quad \delta\chi_{ENS}]^T) = \frac{1}{2} \parallel [\delta\chi_{VAR} \quad \delta\chi_{ENS}]^T \parallel_{\mathbf{I}}^2 \dots \tag{5.23}$$
$$+ \frac{1}{2} \sum_{k=0}^{P} \parallel \delta\mathbf{d} \parallel_{\mathbf{R}_k^{-1}}^2$$

so now

$$\delta\mathbf{d} = \mathbf{y}_k - \mathcal{H}_k \mathbf{x}^f \dots \tag{5.24}$$
$$- \mathbf{H}_k \mathbf{M}_k (\sqrt{1 - \gamma} \, \mathbf{L} \, \delta\chi_{VAR} + \sqrt{\gamma} \, \mathbf{X}^f \, \delta\chi_{ENS})$$

$$\delta\mathbf{x} = \left[ \sqrt{1 - \gamma}\mathbf{L} \quad \sqrt{\gamma}\mathbf{X}^f \right] [\delta\chi_{VAR} \quad \delta\chi_{ENS}]^T \tag{5.25}$$
$$= \mathbf{U}_h \, \delta\chi_h$$

where $\mathbf{U}_h$, the hybrid CVT, is equal to $\left[ \sqrt{1 - \gamma}\mathbf{L} \quad \sqrt{\gamma}\mathbf{X}^f \right]$ and $\delta\chi_h$, the augmented control variable, is equal to $[\delta\chi_{VAR} \quad \delta\chi_{ENS}]^T$. Recalling that $\mathbf{B}^h = \langle \delta\mathbf{x}\delta\mathbf{x} \rangle$ and that the expectation $\langle \delta\chi_h \delta\chi_h^T \rangle$ was defined to yield the identity matrix (as shown in the first

term in Eq.(5.23)) allows for the following

$$\mathbf{B}^h = \langle (\mathbf{U}_h \, \delta\boldsymbol{\chi}_h) (\mathbf{U}_h \, \delta\boldsymbol{\chi}_h)^T \rangle \tag{5.26}$$

$$= \mathbf{U}_h \, \langle \delta\boldsymbol{\chi}_h \delta\boldsymbol{\chi}_h^T \rangle \mathbf{U}_h^T$$

$$= \mathbf{U}_h \mathbf{U}_h^T$$

$$= (1 - \gamma) \mathbf{L}\mathbf{L}^T + \gamma \mathbf{X}^f \mathbf{X}^{f,T}$$

$$= (1 - \gamma) \mathbf{B}_o + \gamma \mathbf{P}^f$$

an expression equivalent to that given in Eq.(5.22).

Most hybrid schemes involve averaging the background-error covariance matrices from the variational and ensemble subsystems in a manner similar to that which was shown above, but hybridization can also be achieved in different ways. Penny (2014) computes a weighted average of the Kalman gain derived from a variational method $\mathbf{K}_{VAR}$ and that of an ensemble method $\mathbf{K}_{ENS}$ to yield a hybridized gain $\mathbf{K}^h$ which is then finally used to update the state in a sequential framework (i.e. $\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}^h(\mathbf{y} - \mathcal{H}\mathbf{x}^f)$). The Ensemble Variational Integrated Localized (EVIL) method of Auligné et al. (2016) uses the analysis determined from the cost function minimization to compute an ensemble of states that adequately sample the analysis-error covariance matrix; this method relies on the assumption of linear operators, which if valid means that the inverse of the Hessian of the cost function is equal to the analysis-error covariance matrix. Pereira and Berre (2006) compute a hybridized background-error covariance matrix using the sample covariance of an ensemble of 4DVar systems instead of from an EnKF ensemble.

## 6 Accommodating Non-Gaussianity

Up to this point, all of the methods presented have assumed Gaussian-distributed system errors. Non-Gaussianity in a system can be introduced by non-Gaussian priors and/or by nonlinear operators. It is possible for conventional Gaussian systems to

mitigate the effects of non-Gaussianity and nonlinearity. For example, the EnKF uses the full nonlinear model to propagate the ensemble members in between observations (but still performs a linear analysis). The incremental formulation for 4DVar breaks up the minimization of a non-quadratic cost function into a series of smaller tasks that make use of linear approximations. Additionally, for small systems, one can use stochastic optimization algorithms (e.g. simulated annealing (Krüger 1993)) in order to ensure that the solution obtained in a variational framework is indeed the global minimum of the prescribed cost function. System nonlinearities may also be reduced by deploying observations more frequently in areas corresponding to domains in the model space dominated by chaotic behavior, whereby targeting can be guided by an analysis of the leading local Lyapunov exponents of a system (Pires 1996). Cohn (1997) shows that, via an appropriately defined anamorphosis function, one may transform a non-Gaussian random variable into a Gaussian one. This new variable can be used in an analysis of a form similar to that given in Eq.(1.4) (albeit with an extra observation bias term dependent on the specific transformation employed), whereby the result is then converted back to the non-Gaussian space using the inverse of the anamorphosis function.

There do exist mathematical formalisms that allow for the full consideration of every moment of a non-Gaussian distribution by a nonlinear data assimilation system, but their implementation requires the time evolution of the pdf of every variable in the state vector (van Leeuwen & Evensen 1996). Assuming that such functions are determinable, the associated information content introduces a complexity that does not scale well with system size, explaining in large part why the second-order moment closure (effectively Gaussian) approximation is commonly made and consequently why linear approximations are needed for a number of data assimilation methods when applied to high-dimensional problems.

It is of interest to consider data assimilation methods that entirely avoid assumptions of Gaussianity. Such avoidance is exemplified by the particle filter. In similar spirit to how the EnKF uses ensemble members to parameterize Gaussian distributions, the particle filter enlists ensemble members, or particles, that are capable of parameterizing

any arbitrarily-shaped distribution.

## 6.1 The Gist of Particle Filtering

The particle filter can be derived from Bayesian first principles as follows

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} \tag{6.27}$$

where the posterior pdf $p(\mathbf{x}|\mathbf{y})$ is obtained once the prior pdf $p(\mathbf{x})$ has been updated with new information via the likelihood $p(\mathbf{y}|\mathbf{x})$. The marginal distribution $p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}$ acts as a normalization factor ensuring that the integral $\int p(\mathbf{x}|\mathbf{y})d\mathbf{x}$ evaluates to 1. The formula given in Eq.(6.27) is commonly known as Bayes' rule, which imposes no restrictions on any of the distributions. The particle, or discrete, representation of the posterior pdf is as follows

$$p(\mathbf{x}|\mathbf{y}) = \sum_{i=1}^{m} w_i \delta(\mathbf{x} - \mathbf{x}_i) \tag{6.28}$$

where $\delta$ is the dirac-delta function and

$$w_i = \frac{p(\mathbf{y}|\mathbf{x_i})}{\sum_{j=1}^{m} p(\mathbf{y}|\mathbf{x_j})} \tag{6.29}$$

Of note is that the sum of the weights expressed in Eq.(6.29) is equal to 1.

**Resampling** Eq.(6.28) and Eq.(6.29) represent the most generic form of the particle filter (the bootstrap variant of Gordon et al. (1993)). Given some observation $\mathbf{y}$, each particle (forecast) $\mathbf{x_i}$ is updated with its respective likelihood $p(\mathbf{y}|\mathbf{x_i})$. The particle filter beautifully avoids matrix inversions and provides for a fully nonlinear and non-Gaussian data assimilation framework. Unfortunately, the naive implementation of the particle filter is likely to be ineffective due to an inherent degeneracy problem whereby weights tend to vanish with each analysis cycle (Berliner and Wikle, 2007). This is related to the curse of dimensionality which states that a given number of particles tend to be

less representative of systems with increasingly large state spaces. One way to avoid or at least postpone weight-collapse is to produce a new sample of uniformly-weighted particles via a resampling of the degenerate distribution. Intuitively, such a resampling is not useful if the forecast is deterministic as this would only lead to multiple copies of a single particle.

## 6.2 Nonlinear Ensemble Transform Filter (NETF)
## (Tödter and Ahrens, 2015)

Several ensemble-based filters have been developed that bypass the Gaussian constraint of the EnKF variants by requiring the analysis ensemble to have the same mean and covariance as that given by the particle filter. Such moment-matching ensemble filters include the Particle Filter with Gaussian Resampling of Xiong et al. (2006), the Merging Particle Filter of Nakano et al. (2007), and the Nonlinear Ensemble Transform Filter (NETF) of Tödter and Ahrens (2015), the last of which will be overviewed as a representative example.

The NETF assumes that the posterior mean and covariance are equivalent to the respective sample estimates given by an ensemble of priors (forecasts)

$$\overline{\mathbf{x}}^a = \sum_{i=1}^m w_i \mathbf{x}_i^f \tag{6.30}$$

$$\mathbf{P}^a = \sum_{i=1}^m w_i (\mathbf{x}_i^f - \overline{\mathbf{x}}^a)(\mathbf{x}_i^f - \overline{\mathbf{x}}^a)^T \tag{6.31}$$

where the weights $w_i$ are normalized likelihoods as given in Eq.(6.29). Eq.(6.30) can be rewritten as $\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{X}^f \mathbf{w}$ (i.e. Eq.(3.1)), where here $\mathbf{w}$ is an $m$-dim column vector containing all of the individual $w_i$. This allows the following expansion of the

analysis-error covariance matrix

$$\mathbf{P}^a = \sum_{i=1}^{m} w_i((\mathbf{x}_i^f - \overline{\mathbf{x}}^f) - \mathbf{X}^f\mathbf{w})((\mathbf{x}_i^f - \overline{\mathbf{x}}^f) - \mathbf{X}^f\mathbf{w})^T \tag{6.32}$$

$$= \sum_{i=1}^{m} w_i(\mathbf{x}_i^f - \overline{\mathbf{x}}^f)(\mathbf{x}_i^f - \overline{\mathbf{x}}^f)^T - \sum_{i=1}^{m} w_i(\mathbf{x}_i^f - \overline{\mathbf{x}}^f)\mathbf{w}^T\mathbf{X}^{f,T}...$$

$$- \sum_{i=1}^{m} w_i\mathbf{X}^f\mathbf{w}(\mathbf{x}_i^f - \overline{\mathbf{x}}^f)^T + \sum_{i=1}^{m} w_i\mathbf{X}^f\mathbf{w}\mathbf{w}^T\mathbf{X}^{f,T}$$

$$= \mathbf{X}^f\mathbf{W}\mathbf{X}^{f,T} - \mathbf{X}^f\mathbf{w}\mathbf{w}^T\mathbf{X}^{f,T}$$

where $\mathbf{W} = diag(\mathbf{w})$ (not to be confused with the right-transform matrix introduced in Eq.(3.5)). Recalling the requirement that $\mathbf{P}^a$ be equal to the sample-error covariance of the updated ensemble, the following transformation follows

$$\frac{1}{m-1}\mathbf{X}^a\mathbf{X}^{a,T} = \frac{m}{m-1}\mathbf{X}^f(\mathbf{W} - \mathbf{w}\mathbf{w}^T)\mathbf{X}^{f,T} \tag{6.33}$$

$(\mathbf{W} - \mathbf{w}\mathbf{w}^T)^{\frac{1}{2}}$ is the right-transform matrix of the prior perturbation ensemble analogous to that given in Eq.(3.5) for the deterministic EnKF variants discussed earlier.

Having also implemented localization and inflation for the NETF in 80-variable Lorenz-96 model experiments, Tödter and Ahrens (2015) noted that filter stability was better maintained when the posterior ensemble was rotated (while still maintaining the proper second-order statistics). Why this is helpful can be understood by noting that the NETF, given the Bayesian update from which the necessary transform matrix is derived, is subject to the same degeneracy issues as the particle filter. The rotation mitigates the threat of degeneracy by effectively reducing the variance in the weights assigned to the particles.

# Part II

**Improving Predictability of the Earth System**

Forecasting at subseasonal-to-seasonal (S2S; 14 days - 12 months) timescales is a research area that has received substantial attention from the environmental modeling community. This interest is largely motivated by the increasingly forceful societal demand that actionable information concerning high-impact weather situations (e.g. floods, droughts, and blizzards) be provided as far in advance of their onset as possible. A potential source of S2S predictability in the earth system is the land surface, which has the ability to modulate the large-scale atmospheric flow via significant moisture and energy fluxes. Capitalization of this potential improvement in forecast skill requires that the land surface initialization be accurate. Some issues that can lead to degraded model representations of land surface states are associated with the limiting assumptions on which conventional land data assimilation schemes are based and with the assimilated observations themselves. The material that follows briefly expands on the problem presented in the specific context of the characterization of soil moisture in land surface models (LSMs).

## 7    Model Representation of Soil Moisture

Soil moisture acts as a limiting factor for evapotranspiration, which can have a notable influence on boundary layer dynamics. Koster and Suarez (2003) found that Boreal summer precipitation is affected strongly by land surface initialization in wet-dry transition regions like the US Great Plains due to a confluence of factors: (i) high soil moisture variability, (ii) a likelihood for soil moisture anomalies to induce evaporation anomalies, and (iii) a likelihood for evaporation anomalies to induce precipitation anomalies. Stated differently, improvements in model soil moisture initialization may translate to improvements in forecasts of near-surface sensible weather variables. The realization of such potential requires an accurate description of the soil moisture state, which is something that LSMs have been noted to struggle with. For example, using 7 different LSMs,

Koster et al. (2009) found soil moisture to be best described as a model-specific quantity having no directly comparable real-world analogue but that its temporal variability could still be useful for improving S2S forecasts of low-level atmospheric fields.

In experiments with the Noah 3.3 LSM, Kumar et al. (2014) noted that satellite soil moisture retrieval data, when assimilated with a one-dimensional (Gaussian) EnKF, yielded small but statistically significant improvements in the model prognostic analogs over the continental United States (CONUS) compared to when no assimilation was used. Additionally, for the case in which soil moisture retrieval data were assimilated, percentile-based drought area estimates obtained from forecasts of model-derived root-zone soil moisture were of comparable utility with more sophisticated drought monitoring resources in depicting the intensity and spatial orientation of droughts, especially for less extreme events. Such results showcase how soil moisture information can be of use as a predictability source for high-impact weather.

## 8   Current Land Data Assimilation Practices

In-situ soil moisture measurements can help describe the near-surface soil moisture profile, but taking into account the inadequate global-scale spatial coverage of current ground-based observing networks and the point-scale nature of the information, it becomes attractive to supplement those observations with estimates derived from satellite retrievals. The inclusion of remotely sensed data allows for an appreciable increase in spatiotemporal coverage of the earth's surface. Soil moisture retrievals, due to their dependence on poorly estimated parameters like surface roughness and surface temperature, tend to be error prone but can still extend invaluable assistance to the approximation of the true state when their error properties are taken into consideration by a capable data assimilation system.

LSMs, unlike atmospheric models, are not sensitive to initial conditions. Initialized with a number of unique starting points, the resulting LSM forecasts will all collapse to a trajectory largely determined by the boundary conditions (e.g. meteorological forcing

data). Also, LSMs typically do not account for horizontal flow. Such factors make it natural that certain data assimilation methods are preferred over others for operational applications. The European Center for Medium-Range Weather Forecasts (ECMWF) uses a land data assimilation system (LDAS) that implements a simplified EKF (SEKF) in conjunction with screen-level observations from ground-based networks to analyze soil moisture and separately employs a 2-dim optimal interpolation (OI) scheme (OI is exactly the Kalman Filter but with background-error covariances that are prescribed instead of propagated temporally) to analyze snow parameters (Rosnay et al., 2013). The same quantities are analyzed exclusively using the stochastic EnKF at the Canadian Meteorological Center (CMC) via the Canadian LDAS (CaLDAS) (Carrera et al., 2015). Given that LSM code tends to be littered with switches, tangent-linear operators are determined via finite differences whereby a perturbed model run is needed for every variable in the state vector.

It has been shown that for an adequately-sized ensemble, the EnKF is able to produce more accurate land surface state estimates than those obtained by the EKF (Reichle et al., 2002). Despite this, the assumption of Gaussian-distributed system errors used in the conventional formulation of the EnKF mandates that, by ignoring the higher-order information of the ensemble distribution, the yielded state estimate may be only a sub-optimal approximation for non-Gaussian systems.

## 9 Summary & Conclusions

The EnKF has been shown to be a computationally affordable filter suitable for high-dimensional state estimation problems. The threat of filter divergence is mitigated significantly by ad hoc impositions such as inflation and localization. Due to only the first two moments of the ensemble being explicitly updated in the traditional formulation of the EnKF, the scheme is implicitly Gaussian. Presumably, non-Gaussian variables and processes would be better described using an assimilation scheme with non-Gaussian support. It is of interest to observe how the predictability of the earth system is affected

when one or more components of a coupled framework use an assimilation scheme that can accommodate deviations from normality in an optimal way. The discussion given here focused on the soil moisture given its being able to modulate the near-surface atmospheric flow in certain regions on S2S timescales. Soil moisture is a bounded variable with an error distribution that can be highly skewed. These factors suggest that better estimates of soil moisture states would be obtained via the use of a data assimilation scheme not dependent on the assumption of normality. The implications of non-Gaussian assimilation on model initialization make it worth further research.

## A    Original Examples

### A.1    System Design

A description of the L63 dynamics is given by the nonlinear system of equations

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma\,(x - y) \tag{A1}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \rho x - y - xz \tag{A2}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z \tag{A3}$$

where $x = x(t)$, $y = y(t)$, $z = z(t)$. $\sigma$, $\rho$, $\beta$ are constants, the values of which are set equal to 10, 28, and 8/3, respectively, for the experiments presented, all of which were carried out exlusively in Python (v3.7.3). The L63 equations were discretized using the Modified-Euler method (a second-order Runge-Kutta scheme) with $\Delta t = 0.02$.

The EKF data assimilation system used was inspired largely by the Python FilterPy module created and maintained by Roger Labbe. Plotting functions and the accommodation of user-defined observation frequencies were added. Additionally, effort was made to create a more model agnostic framework. All of the ensemble and variational methods to be shown here were coded independently making use of various other Python modules, namely mpi4py and SymPy.

A general workflow for the code is as follows
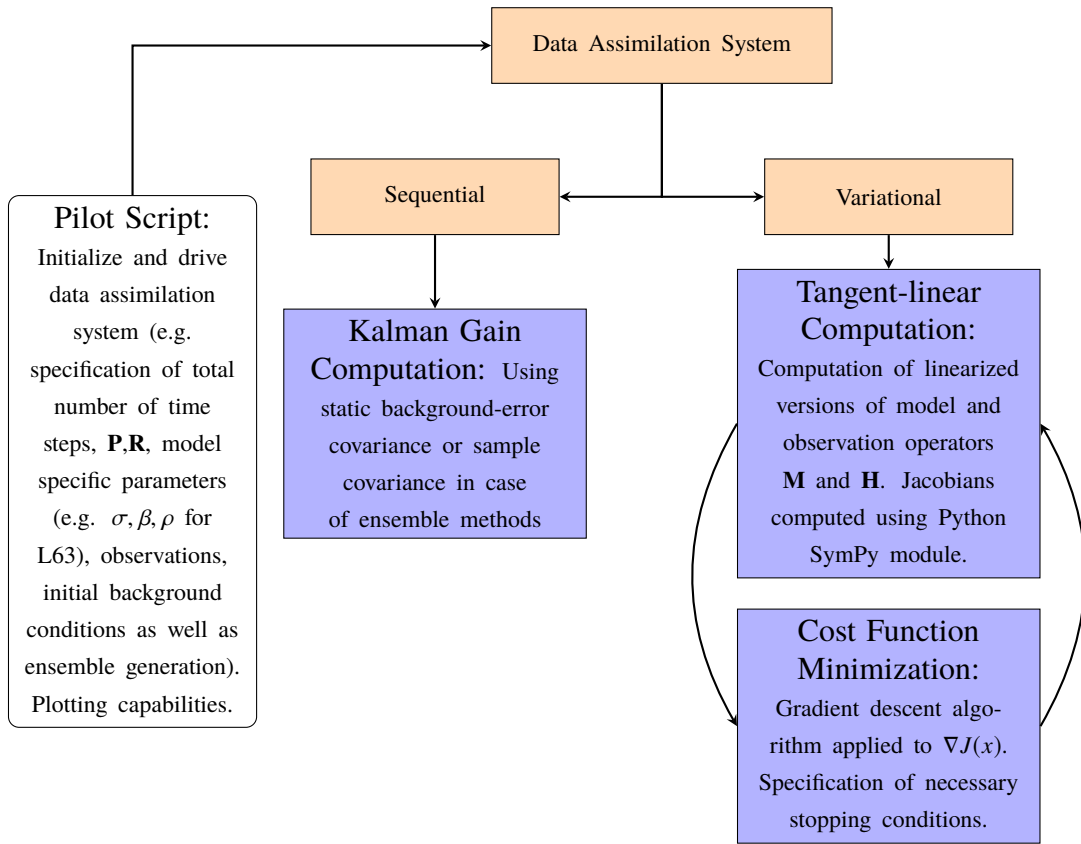


with the notable exception of the EKF, which requires both the determination of the Kalman gain and the tangent-linear model, but does not make use of any cost function minimization. The feedback loop between the Tangent-linear Computation script and the Cost Function Minimization script is only for the case of 4DVar, which needs to recompute linearizations about each new first guess, whereas this minimization is carried out only once in 3DVar. Where possible, user readability of code took priority over computational efficiency.

## A.2    Some Results & Comments

**EKF | Figure 1**    Pictured is the variable $x$. The black solid line represents the truth trajectory. The prior trajectory is given by the green dotted line. The value of the prior used to compute the analysis at the observation time is given by a green circle. An observation is denoted by a red x. Computed analyses are given by black asterisks. All variables in L63
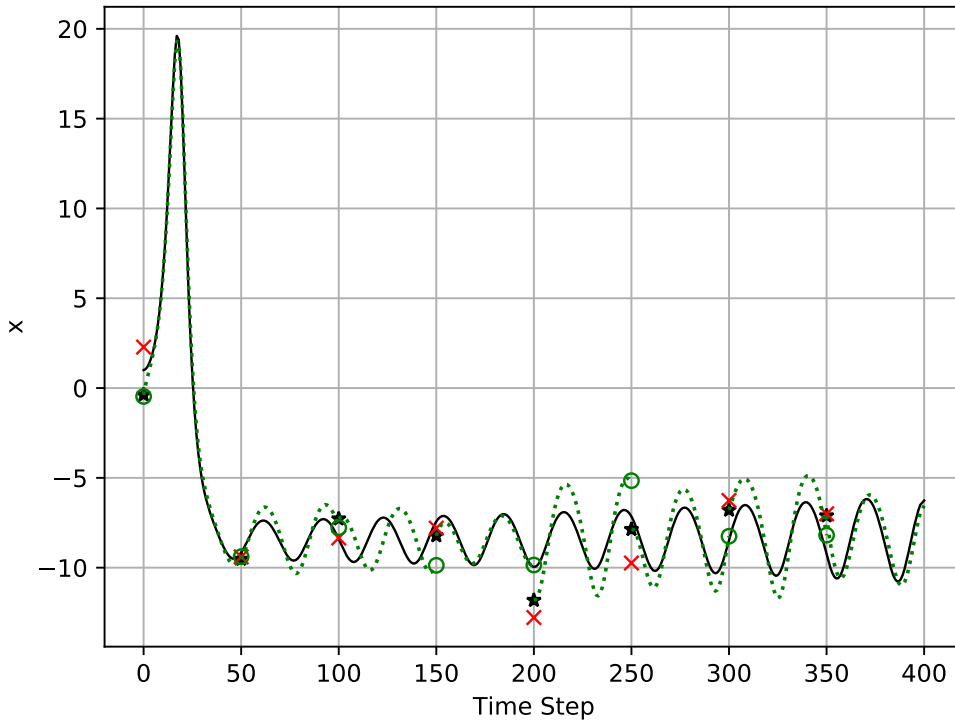
**Figure 1:** Extended Kalman filter (EKF) as applied to L63 model.

are directly observed and assimilated (i.e. $H = I_3$, where $I_g$ is the $g \times g$ identity matrix). Observations are obtained by perturbing the truth trajectory with random samples of Gaussian-distributed noise consistent with $R = 4 * I_3$, which was also the observation-error covariance matrix prescribed to the EKF algorithm. The model error **Q** was chosen arbitrarily to be equal to $0.3 * I_3$. Observations are assimilated at a frequency of one every 50 time steps. The TLM and model adjoint were computed analytically using Python's SymPy module. The difference between the TLM and the full nonlinear process model was found to be smaller for smaller perturbation magnitudes, a result that supports the correctness of the TLM. The adjoint was simply computed as the transpose of the tangent-linear operators. The same process for determining the tangent-linear and adjoint was used for the variational methods as well. The background-error covariance matrix was set equal to a climatological ensemble-covariance matrix determined using a set of optimally-tuned inflation parameters for the stochastic ensemble Kalman filter

(EnKF) (i.e. that which minimized the root-mean-square analysis-error over a long run of the assimilation system; data from the first portion of the long run was discarded to try and correct for EnKF spin-up). Using this empirically determined climatology (hereafter denoted BC) as a proxy for the true (and unknown) background-error covariance matrix, the initial conditions were determined from a random sample of an approximate multivariate Gaussian distribution centered about the truth trajectory. The EKF handles nonlinear regimes well. Intuitively, performance was better when analyses were more frequent.
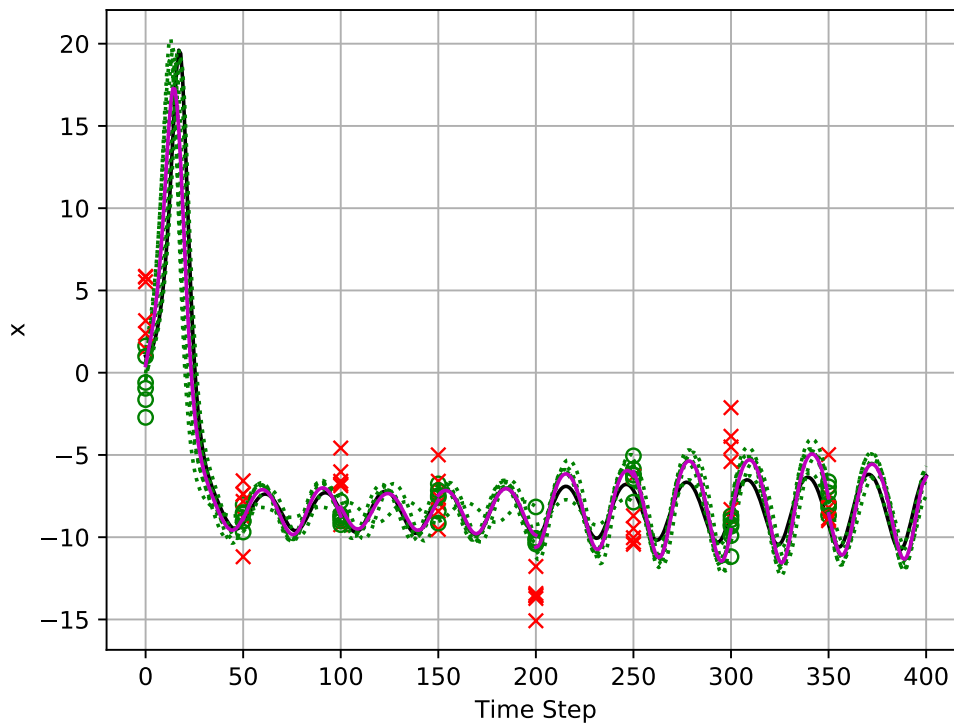


**Figure 2:** Same as Figure 1 but for the stochastic EnKF and without any model error.

**Stochastic EnKF | Figure 2**    There are a total of 6 ensemble members with a mean value denoted by the purple solid line. Recall that the ensemble mean is regarded as the analysis state throughout the ensemble integration. Each ensemble member receives its own observation to assimilate at the analysis time. RTPP and RTPS inflation parameters were set equal to 0.1 and 0.05, respectively. The initial ensemble was determined via

a random draw of a pdf described with a mean given by the truth trajectory and a covariance given by BC (the same climatological covariance matrix outlined in the EKF experiment).
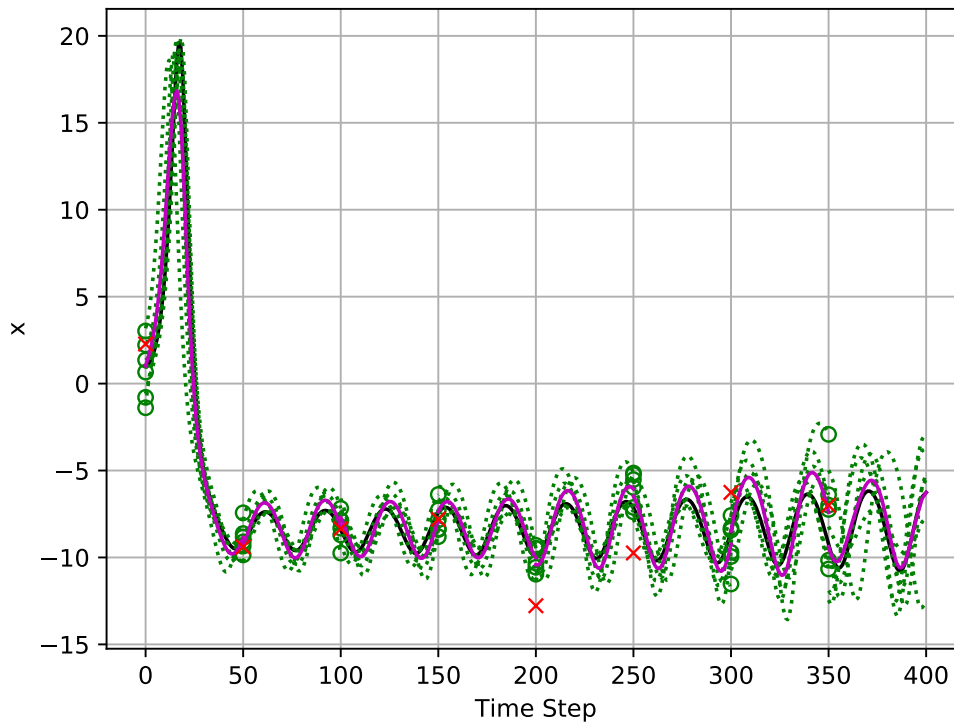


**Figure 3:** Same as Figure 2 but for the deterministic EnKF.

**Deterministic EnKF | Figure 3**    The deterministic EnKF variant shown is that of Whitaker and Hamill (2002). Here, the observations of the L63 system variables $x$, $y$, and $z$ are processed synchronously. Note, in contrast to the stochastic EnKF, that the ensemble assimilates a single observation of each variable at the analysis time. RTPP and RTPS inflation parameters were set equal to 0.8 and 0.9, respectively. Qualitatively, the filter performs well. (Of note is that the code is the most contrived of all the methods presented here.)

**Filter Divergence | Figure 4**    Such inflation (here the term "deflation" might be more appropriate) parameter values encourage a small ensemble spread and hence an over-
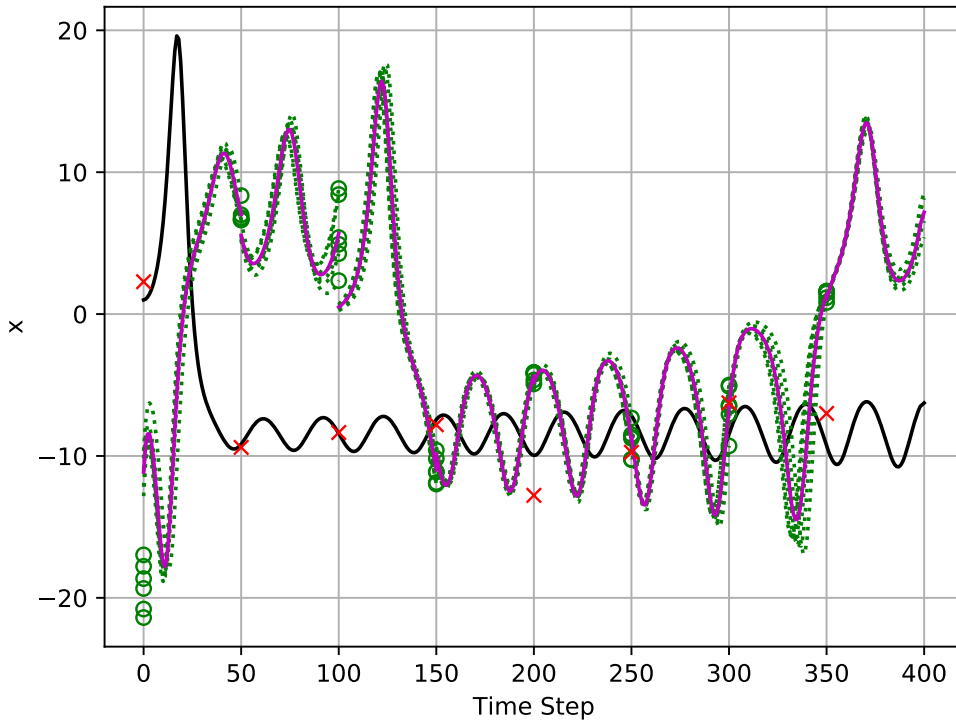
**Figure 4:** Same as Figure 3 except with RTPP and RTPS inflation parameters set equal to 0 and -0.15, respectively.

confidence in the prior that over time leads to the catastrophic filter divergence given in the figure. To help ensure that filter divergence would be observed more clearly, the initial ensemble was centered about a prior state that was significantly distant from the truth. Note that after the first few analyses, there is a complete lack of response the analysis has to the observation information.

**4DVar | Figure 5** The length of the assimilation window is 350 times steps and is demarcated by the blue vertical lines. A set of short forecasts (50 time steps in length) follows the window. The prescribed background- and observation-error covariance matrices are BC and $4 * I_3$, respectively, as in previous experiments with other assimilation schemes. The strong-constraint cost function's gradient $\nabla J(x) = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x^f}) - [\mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{d}_0 + \mathbf{M}_1^T[\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{d}_1 + \mathbf{M}_2^T[\mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{d}_2 + \ldots + \mathbf{M}_P^T\mathbf{H}_P^T\mathbf{R}_P^{-1}\mathbf{d}_P]]]$ was minimized using a bootstrapped gradient descent function with a step size of $.001 * \nabla J(x)$.
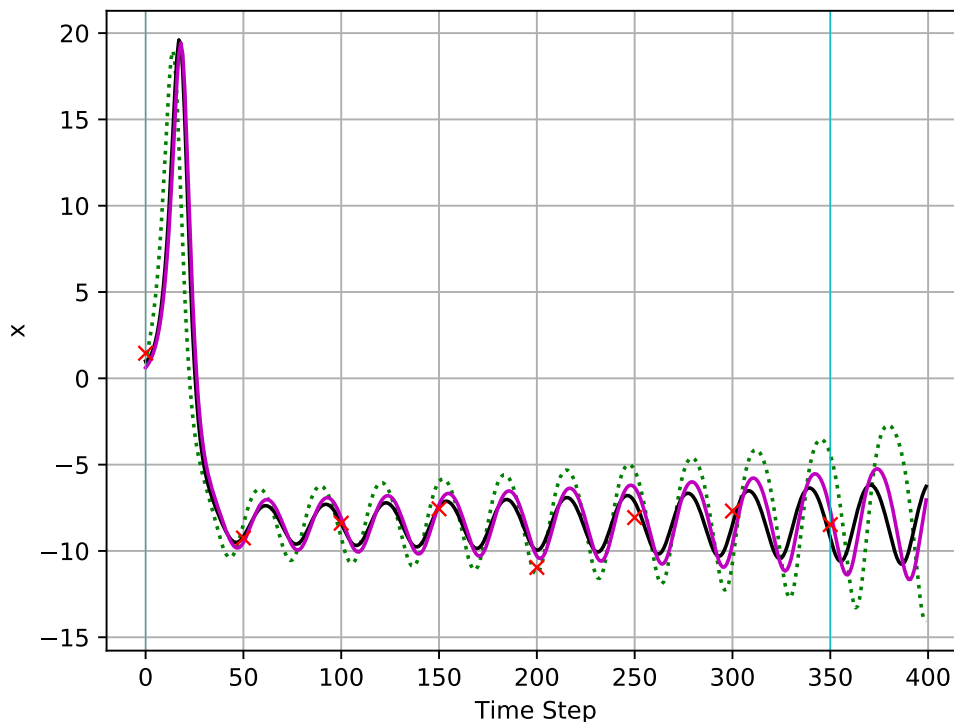
**Figure 5:** Same as Figure 2 but for strong-constraint 4DVar.

Correctness of the $\nabla J(x)$ was supported by an appropriate decrease in higher order terms for smaller perturbation magnitudes given a Taylor Series expansion of $J(x)$. Divergence of the minimization algorithm, (here specified as increasing values of $J(x)$ for a total of 3 iterations (whether successive or not)) was used as a stopping condition. The minimization would also cease if $\nabla J(x) < 0.05$ or if 500 total iterations had been reached. In experiments with the 4DVar system, it was noted that cost function minimization behavior is especially sensitive to the length of the assimilation window as well as to the number of the observations in that window. This is due to the lack of an accounting of model error in the strong-constraint formulation. Too many observations for a given assimilation window length may over-constrain the analysis trajectory and lead to divergence. The caveat is that too few observations will also likely lead to poor analysis quality as well due to a less-than-ideal amount of system information being assimilated to help guide the model solution to a more representative description of the

truth.
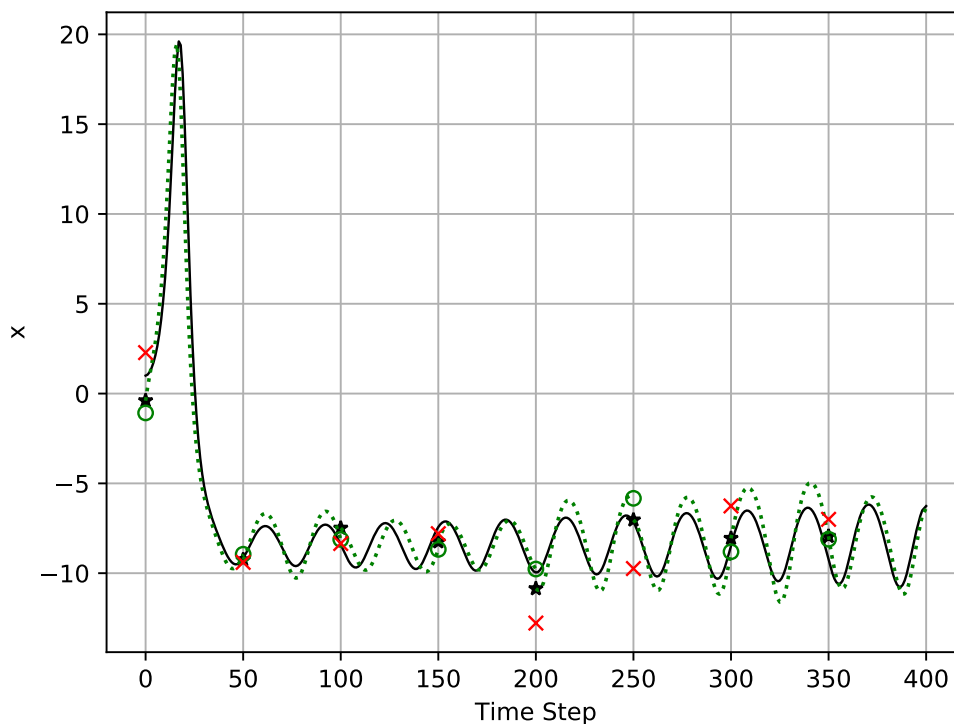


**Figure 6:** Same as Figure 5 but for 3DVar.

**3DVar | Figure 6**  The code used for 3DVar was similar to that used for 4DVar, with the primary difference being that 3DVar assimilated a single observation and minimized the cost function only one time for the determination of the analysis.

# B   A Few Background-Error Covariance Matrix Calibration Methods

Calibration of the background-error covariance matrix is important since the matrix largely determines how the observation information is distributed in state space during the analysis. As follows is a description of a few of the ways this estimation can be approached.

## B.1 Analysis of innovations (o-b)
## (Rutherford, 1972)

The observation vector equation $\mathbf{y} = \mathcal{H}\mathbf{x}_t + \epsilon_\mathbf{0}$ can be used to arrive at an expression for the innovation $\mathbf{y} - \mathbf{Hx}_b = \mathbf{H}\epsilon_b + \epsilon_o$, which uses the truncated Taylor series expansion of the observation operator about the background state $\mathcal{H}(\mathbf{x}_b + (\mathbf{x}_t - \mathbf{x}_b) = \mathcal{H}\mathbf{x}_b + \mathbf{H}(\mathbf{x}_t - \mathbf{x}_b) + ...,$ where $(\mathbf{x}_b - \mathbf{x}_t) = \epsilon_b$ is the background error. From consideration of variables $w_1$ and $w_2$ located at positions $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively, it follows that

$$\langle (\mathbf{y}(w_1, \mathbf{r}_1) - \mathbf{Hx}_b(w_1, \mathbf{r}_1))(\mathbf{y}(w_2, \mathbf{r}_2) - \mathbf{Hx}_b(w_2, \mathbf{r}_2)) \rangle \tag{A4}$$

$$\approx \langle \epsilon_o(w_1, \mathbf{r}_1)\epsilon_o(w_2, \mathbf{r}_2) \rangle + \langle \epsilon_b(w_1, \mathbf{r}_1)\epsilon_b(w_2, \mathbf{r}_2) \rangle$$

$$\approx \langle \epsilon_b(w_1, \mathbf{r}_1)\epsilon_b(w_2, \mathbf{r}_2) \rangle$$

where the correlations between the background and observations have been neglected, and it is assumed that the observation errors at two distinct points are uncorrelated. The final result is an expression that allows for the estimation of $\epsilon_b$. An obvious issue is that many samples are needed to compute the average of the innovation correlations, which is not feasible for high-dimensional systems. The analysis of innovations can still be helpful in tuning the background-error covariance matrix where appropriate.

## B.2 National Meteorological Center (NMC) Method
## (Parrish and Derber, 1992)

Suppose there are two forecasts of different lengths that verify at the same time

$$\mathbf{x}_b^{2\Delta} = \mathcal{M}_{2\Delta}\mathbf{x}_a^0 + \epsilon_\mathbf{b}^{\mathbf{2\Delta}} \tag{A5}$$

$$\mathbf{x}_b^{2\Delta} = \mathcal{M}_{\Delta}\mathbf{x}_a^{\Delta} + \epsilon_\mathbf{b}^{\mathbf{\Delta}} \tag{A6}$$

where $\mathbf{x}_a^t$ is the analysis obtained at time $t$, and $\epsilon_b^\Delta$ is the forecast error accumulated over the time interval $\Delta$. If $\delta\mathbf{x}$ = Eq.(A6) - Eq.(A5), it follows that

$$
\begin{aligned}
\langle\delta\mathbf{x}\delta\mathbf{x}\rangle &= \langle\epsilon_b^\Delta\epsilon_b^{\Delta,T}\rangle + \langle\epsilon_b^{2\Delta}\epsilon_b^{2\Delta,T}\rangle \\
&\approx 2\langle\epsilon_b\epsilon_b^T\rangle \\
&= 2\mathbf{B}_o
\end{aligned}
\tag{A7}
$$

where it is assumed that the accumulated errors of each of the forecasts are comparable to one another. This method avoids having to observe the entire state. In poorly observed regions, the background state does not change much, so error variances may be underestimated in those regions. Variances at larger scales are likely to be overestimated in the cases where forecast samples differ more than the length of the a priori state (e.g. using forecasts that differ by 24 hours to approximate background errors of 6 hour forecasts). This method is best suited for estimating climatological covariances. Also noteworthy is the lagged NMC method (Siroka et al. 2003), which seeks to incorporate into the description of $\mathbf{B}_o$ the error associated with lateral boundary conditions used by a limited area model (LAM).

### B.3 Ensemble Method

The sample covariance of an ensemble of states about its mean can be used to approximate the background-error covariance matrix

$$
\mathbf{B}_o = \langle(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T\rangle
\tag{A8}
$$

where $\mathbf{x}_i$ is a forecast and $\bar{\mathbf{x}}$ is the mean. The ensemble can be derived from an ensemble-based data assimilation system such as the EnKF or from an ensemble of variational systems. The ensemble method has been shown to produce errors with more appropriate length scales than the NMC method, especially in data-sparse regions. Such results can be explained in part by the use of forecast samples of the same duration as the prior and

by the use of flow-dependent background errors.

## B.4 Canadian Quick (CQ) Method
##    (Polavarapu et al., 2005)

A drawback to the aforementioned methods is that they all require a background-error covariance matrix to begin with. It has been suggested to use samples of the difference of single forecast and it's earlier state to estimate $\mathbf{B}_o$

$$\mathbf{B}_o \approx \frac{1}{2}\langle(\mathbf{x}_{t+\Delta t} - \mathbf{x}_t)(\mathbf{x}_{t+\Delta t} - \mathbf{x}_t)^T\rangle \tag{A9}$$

where the means of each sample have already been subtracted. Eq.(A6) assumes that the time-tendency of a forecast is a suitable placeholder for the background error, which may lead to the assignment of spuriously small/large correlations to features that evolve slowly/quickly over time.

# References

Anderson, J. L., 2001: An ensemble adjustment kalman filter for data assimilation. *Monthly Weather Review*, **129**(12), 2884–2903.

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, **230**(1), 99 – 111. Data Assimilation.

Auligné, T., B. Ménétrier, A. C. Lorenc, and M. Buehner, 2016: EnsembleâĂŞvariational integrated localized data assimilation. *Monthly Weather Review*, **144**(10), 3677–3696.

Berliner, L. M. and C. K. Wikle, 2007: Approximate importance sampling monte carlo for data assimilation. *Physica D: Nonlinear Phenomena*, **230**(1), 37 – 49. Data Assimilation.

Bishop, C. H., 2016: The gigg-enkf: ensemble kalman filtering for highly skewed nonnegative uncertainty distributions. *Quarterly Journal of the Royal Meteorological Society*, **142**(696), 1395–1412.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly Weather Review*, **129**(3), 420–436.

Bishop, C. H. and D. Hodyss, 2009: Ensemble covariances adaptively localized with eco-rap. part 1: tests on simple error models. *Tellus: Series A*, **61**(1), 84 – 96.

Buizza, R., M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ecmwf ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **125**(560), 2887–2908.

Burgers, G., P. Jan van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble kalman filter. *Monthly Weather Review*, **126**(6), 1719–1724.

Carrera, M. L., S. Bélair, and B. Bilodeau, 2015: The canadian land data assimilation system (caldas): Description and synthetic evaluation study. *Journal of Hydrometeorology*, **16**(3), 1293–1314.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, **131**(613), 3385–3396.

Evensen, G., 1992: Using the extended kalman filter with a multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research: Oceans*, **97**(C11), 17905–17924.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125**(554), 723–757.

Gordon, N., D. Salmond, and A. Smith, 1993: Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, **140**, 107–113(6).

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble kalman filter localization techniques. *Monthly Weather Review*, **139**(2), 511–522.

Hamill, T. M. and C. Snyder, 2000: A hybrid ensemble kalman filterâĂŞ3d variational analysis scheme. *Monthly Weather Review*, **128**(8), 2905–2919.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D: Nonlinear Phenomena*, **230**(1), 112–126.

Ide, K., P. Courtier, M. Ghil, and A. Lorenc, 1997: Unified notation for data assimilation : Operational, sequential and variational. *Journal of the Meteorological Society of Japan*, **75**, 181–189.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, **82**(Series D), 35–45.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation, and predictability*. Cambridge University Press, New York, 1-369.

Koster, R. D., Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, 2009: On the nature of soil moisture in land surface models. *Journal of Climate*, **22**(16), 4322–4335.

Koster, R. D. and M. J. Suarez, 2003: Impact of land surface initialization on seasonal precipitation and temperature prediction. *Journal of Hydrometeorology*, **4**(2), 408–423.

Kumar, S. V., C. D. Peters-Lidard, D. Mocko, R. Reichle, Y. Liu, K. R. Arsenault, Y. Xia, M. Ek, G. Riggs, B. Livneh, and M. Cosh, 2014: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *Journal of Hydrometeorology*, **15**(6), 2446–2469.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **135**(639), 523–533.

Liang, X., X. Zheng, S. Zhang, G. Wu, Y. Dai, and Y. Li, 2012: Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble kalman filter assimilation. *Quarterly Journal of the Royal Meteorological Society*, **138**(662), 263–273.

Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, **51**(8), 1037–1056.

Nakano, S., G. Ueno, and T. Higuchi, 2007: Merging particle filter for sequential data assimilation. *Nonlinear Processes in Geophysics*, **14**(4), 395–408.

Parrish, D. F. and J. C. Derber, 1992: The national meteorological center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, **120**(8), 1747–1763.

Penny, S. G., 2014: The hybrid local ensemble transform kalman filter. *Monthly Weather Review*, **142**(6), 2139–2149.

Pereira, M. B. and L. Berre, 2006: The use of an ensemble approach to study the background error covariances in a global nwp model. *Monthly Weather Review*, **134**(9), 2466–2489.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, **129**(5), 1194–1207.

Polavarapu, S., S. Ren, Y. Rochon, D. Sankey, N. Ek, J. Koshyk, and D. Tarasick, 2005: Data assimilation with the canadian middle atmosphere model. *Atmosphere-Ocean*, **43**(1), 77–100.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble kalman filter. *Monthly Weather Review*, **130**(1), 103–114.

Rosnay, P. d., M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel, and L. Isaksen, 2013: A simplified extended kalman filter for the global operational soil moisture analysis at ecmwf. *Quarterly Journal of the Royal Meteorological Society*, **139**(674), 1199–1213.

Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. *Journal of the Atmospheric Sciences*, **29**(5), 809–815.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, **131**(612), 3079–3102.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Monthly Weather Review*, **131**(7), 1485–1490.

Tödter, J. and B. Ahrens, 2015: A second-order exact ensemble square root filter for nonlinear data assimilation. *Monthly Weather Review*, **143**(4), 1347–1367.

Tompkins, A. M. and J. Berner, 2008: A stochastic convective approach to account for model uncertainty due to unresolved humidity variability. *Journal of Geophysical Research: Atmospheres*, **113**(D18).

Trémolet, Y., 2007: Incremental 4d-var convergence study. *Tellus A: Dynamic Meteorology and Oceanography*, **59**(5), 706–718.

Verlaan, M. and A. W. Heemink, 1997: Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrology and Hydraulics*, **11**(5), 349–368.

Wang, X. and C. H. Bishop, 2003: A comparison of breeding and ensemble transform kalman filter ensemble forecast schemes. *Journal of the Atmospheric Sciences*, **60**(9), 1140–1158.

Wang, X., C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble-3dvar hybrid analysis schemes. *Monthly Weather Review*, **135**(1), 222–227.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, **130**(7), 1913–1924.

Whitaker, J. S. and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, **140**(9), 3078–3089.

Xiong, X., I. M. Navon, and B. Uzunoglu, 2006: A note on the particle filter with posterior gaussian resampling. *Tellus: Series A*, **58**(4), 456 – 460.