



Introduction to Ensemble Data Assimilation

Steven J. Fletcher

Cooperative Institute for Research in the Atmosphere,

Colorado State University

Plan of talk

- Kalman Filter
- EnKF
- Inflation and Localization
- ETKF
- LETKF
- MLEF
- En4DVAR
- 4DEnVAR

KALMAN FILTER AND SMOOTHER

19

CHAPTER OUTLINE

19.1 Derivation of the Kalman Filter	766
19.2 Kalman Filter Derivation from a Statistical Approach	771
19.3 Extended Kalman Filter	774
19.4 Square Root Kalman Filter	776
19.5 Smoother	777
19.5.1 Forward Step: Kalman Filter	777
19.5.2 Backward Step: Reverse-Time Information Filter.....	778
19.5.3 Smoothing.....	779
19.6 Properties and Equivalencies of the Kalman Filter and Smoother	780
19.7 Summary	781

The discrete version of the Kalman filter first appeared in Kalman's 1960 paper: *A new approach to linear filtering and prediction problems* [244]. The basis of his work was related to problems in communications and control that are of a statistical nature [244]. The class of problems that Kalman is referring to are: (1) prediction of random signals; (2) separation of random signals from random noise; and (3) detection of signals of known forms in the presence of random noise.

Kalman makes reference to the work of Wiener [245], and says that Wiener's method to approach problems (1) and (2) above leads to what is known as the **Wiener-Hopf** integral equation. We shall not go into details about this problem here, but we note that it is part of the motivation for Kalman.

Kalman lists a series of papers that have proposed methods to solve the Wiener-Hopf integral equations as well as many different generalizations, where the objective of these series of papers was to obtain specifications of a linear dynamic system, which Kalman refers to as the **Wiener filter**, that accomplishes the prediction, separation, or detection of a random signal.

Kalman then states four problems for this method, which he shows do not afflict his new filter:

1. The optimal filter is specified by its impulse response.
2. Numerical determination of the optimal impulse response is often quite involved and poorly suited for machine computation.
3. Important generalizations require new derivations, frequently of considerable difficulty.
4. The mathematics of the derivation are not transparent.

However, Kalman says that his new approach produces the following:

1. **Optimal estimates and orthogonal projections.** His approach is to consider the Wiener filter in terms of conditional distributions and expectations. He shows that all of the statistical calculations and results are based upon first- and second-order averages; **no other statistical data are required.**
2. **Models for random processes.** In Kalman's approach, arbitrary random signals are represented as the output of a linear dynamic system that is excited by independent random signals.
3. **The solution of the Wiener problem.** Kalman uses the state-transition method, where, as a result of this approach, the single derivation then covers a large variate of problems. **Guessing** the *state* of the estimation correctly leads to a nonlinear difference/differential equation for the covariance matrix of the optimal estimation error. The solution of the equation for the covariance matrix starts at time t_0 , which is where the first observation is taken; at each later time t the solution of the equation represents the covariance of the optimal prediction error, given observations in the interval (t_0, t) . Kalman then states that

From the covariance matrix at time t we obtain at once, without further calculations, the coefficients, which could be time-varying, characterizing the optimal linear filter.

Given Kalman's motivation above we now move on to the derivation of the Kalman filter as set out in [244] and then show a direct statistical-based derivation.

19.1 DERIVATION OF THE KALMAN FILTER

The initial motivation for the derivation of the Kalman filter equations in [244] start with an example: suppose that we are given a signal $x_1(t)$ and noise $x_2(t)$; however, only the sum of the two $y(t) = x_1(t) + x_2(t)$ can be observed. Next, suppose that we have observed and know exactly the values $y(t_0), \dots, y(t)$. Now there are three possible situations with regard to the knowledge of the unobservable value of the signal at $t = t_1$:

1. If $t_1 < t$, this is a **data smoothing (interpolation)** problem.
2. If $t_1 = t$, this is a **filtering** problem.
3. If $t_1 > t$, this is a **prediction** problem.

So we now see where the terms “smoothing” and “filtering” come from [244]. It is stated in [244] that Kalman is considering all three cases and refers to them collectively as **estimation**.

Now to the mathematics and probability. We assume that the signal, noise, and their sum are random processes, where we can determine the probability with which a particular sample of the signal and noise will occur. Therefore, for any given set of measured values $y(t_0), \dots, y(t)$ of the random variable $y(t)$, it is possible to determine the probability of the simultaneous occurrence of various values, $x_1(t)$, of the random variable $x_1(t_1)$. This leads to the conditional probability of

$$P(x_1(t_1) \leq \mathbf{x}_1 | y(t_0), \dots, y(t) = \mathbf{y}(t)) = F(\mathbf{x}_1). \quad (19.1)$$

The function $F(\mathbf{x}_1)$ represents all the information that the measurement of the random variables, $y(t_0), \dots, y(t)$, have provided about the random variable $x_1(t_1)$ and is a conditional PDF.

We now denote a statistical estimate of the random variable, $x_1(t_1)$, as $X_1(t_1|t) \equiv X_1(t_1) \equiv X_1$. To be able to arrive at a way of determining $x_1(t_1)$, Kalman introduces a **penalty or loss function**, L , that should be positive and non-decreasing function of the **estimation error**, $\epsilon = x_1(t_1) - X_1(t_1)$, and the function should have the following three properties:

$$L(0) = 0, \quad L(\epsilon_1) \geq L(\epsilon_2) \geq 0 \quad (\epsilon_2 \geq \epsilon_1 \geq 0), \quad L(\epsilon) = L(-\epsilon). \quad (19.2)$$

The choice for the loss function that Kalman selects is the one that *minimized the average loss or risk*, which is defined as

$$\mathbb{E}[L(x_1(t_1) - X_1(t_1))] = \mathbb{E}[\mathbb{E}[L(x_1(t_1) - X_1(t_1)) | y(t_0), \dots, y(t)]]. \quad (19.3)$$

However, we should note that we can remove the first expectation operation on the right-hand side of (19.3) as it is not operating on X_1 , and such we have

$$\mathbb{E}[L(x_1(t_1) - X_1(t_1)) | y(t_0), \dots, y(t)]. \quad (19.4)$$

Given these assumptions, Kalman introduces the following very important theorem.

Theorem 19.1. *Given a loss function that satisfies the conditions in (19.2) and that the conditional distribution function $F(\mathbf{x})$, defined by (19.1), is symmetric about its mean, $F(\mathbf{x} - \bar{\mathbf{x}}) = 1 - F(\bar{\mathbf{x}} - \mathbf{x})$, and is also convex*

$$F(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) \leq \lambda F(\mathbf{x}_1) + (1 - \lambda) F(\mathbf{x}_2),$$

for all $\mathbf{x}_1, \mathbf{x}_2 \leq \bar{\mathbf{x}}$ and for $0 \leq \lambda \leq 1$ then the random variable $\mathbf{x}_1^(t_1|t)$ that minimizes the average loss given by (19.3) is the conditional expectation*

$$\mathbf{x}_1^*(t_1|t) = \mathbb{E}[x_1(t_1) | y(t_0), \dots, y(t)]. \quad (19.5)$$

Theorem 19.2. *Let $\{x(t)\}$ $\{y(t)\}$ be random processes with zero mean. We observe $y(t_0), \dots, y(t)$ if either the random processes are Gaussian, or the optimal estimate is restricted to be a linear function of the observed random variables and the loss function is $L(\boldsymbol{\varepsilon}) = \boldsymbol{\varepsilon}^2$; the $x^*(t_1|t)$ is the optimal estimate of $x(t_1)$ given $y(t_0), \dots, y(t)$. It is also the orthogonal projection $\bar{\mathbf{x}}(t_1|t)$ of $\mathbf{x}(t_1)$ on $\mathcal{Y}(t)$.*

The interpretation of this theorem is that the optimal estimate, given the two conditions in **Theorem 19.2**, is a linear combination of all previous observations, and can be regarded as the output of a linear filter, with the input being the actually occurring values of the observable random variables.

The next step in the derivation of the Kalman filter is to determine the models for the random processes, and this is where the control theory aspect comes in. Kalman decides that

A random function of time may be thought of as the output of a dynamic system excited by an independent Gaussian random process.

Kalman now makes the statement that a Gaussian random signal remains Gaussian after passing through a linear system. Therefore if we assume independent Gaussian primary random sources, and if the observed random signal is also Gaussian, then we have to assume that the dynamic system between the observer and the primary source is **linear**.

Given these assumptions we now introduce the control system

$$\left. \begin{aligned} \dot{\mathbf{x}} &= \mathbf{M}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{H}(t)\mathbf{x}(t), \end{aligned} \right\} \quad (19.10)$$

19.2 KALMAN FILTER DERIVATION FROM A STATISTICAL APPROACH

There is an easier way to derive the Kalman filter equations, which involves following Kalman's framework, but without the manifolds and state transition matrices. We start with the usual concept of the background state, which is the forecast from the previous analysis state as

$$\mathbf{x}_{t|t-1}^b \equiv \mathbf{M}_{t,t-1} \mathbf{x}_{t-1|t-1}^a. \quad (19.32)$$

Let \mathbf{x}_t^t be the true state at time t , then the background/analysis error is given by

$$\boldsymbol{\varepsilon}_t^a = \mathbf{x}_{t|t-1}^b - \mathbf{x}_t^t = \mathbf{M}_{t,t-1} \mathbf{x}_{t-1|t-1}^a. \quad (19.33)$$

We know that the analysis at the previous filtering time has an associated analysis error such that

$$\mathbf{x}_{t-1|t-1}^a = \mathbf{x}_{t-1|t-1}^t + \boldsymbol{\varepsilon}_{t-1|t-1}^a. \quad (19.34)$$

This means we can write the forecast/background error as

$$\begin{aligned} \boldsymbol{\varepsilon}_t^b &= \mathbf{M} \mathbf{x}_{t-1|t-1}^t + \mathbf{M} \boldsymbol{\varepsilon}_{t-1}^a - \mathbf{x}_t^t \\ &= \mathbf{M} \boldsymbol{\varepsilon}_{t-1}^a + \boldsymbol{\varepsilon}_t^m, \end{aligned} \quad (19.35)$$

where $\boldsymbol{\varepsilon}^m$ is the model error given by $\boldsymbol{\varepsilon}_t^m \equiv \mathbf{M}\mathbf{x}_{t-1|t-1}^t - \mathbf{x}_{t|t}^t$. If we now form the background/forecast error covariance matrix, we have

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\varepsilon}_t^b, \left(\boldsymbol{\varepsilon}_t^b\right)^T\right] &= \mathbf{P}_t^b = \mathbb{E}\left[\left(\mathbf{M}\boldsymbol{\varepsilon}_{t-1|t-1}^a + \boldsymbol{\varepsilon}_t^m\right)\left(\mathbf{M}\boldsymbol{\varepsilon}_{t-1|t-1}^a + \boldsymbol{\varepsilon}_t^m\right)^T\right], \\ &= \mathbf{M}\mathbb{E}\left[\boldsymbol{\varepsilon}_{t-1|t-1}^a \left(\boldsymbol{\varepsilon}_{t-1|t-1}^a\right)^T\right]\mathbf{M}^T + \mathbb{E}\left[\boldsymbol{\varepsilon}_t^m, \boldsymbol{\varepsilon}_t^m\right], \\ &= \mathbf{M}\mathbf{P}_{t-1}^a\mathbf{M}^T + \mathbf{Q}_t.\end{aligned}\tag{19.36}$$

The expression in (19.36) is the same as for the derivation in the last section.

Now, given a predicted/forecasted state, $\mathbf{x}_{k+1|k}$, which is associated with observations up to time t_k , and assuming that we have received an observation of the state at $t = t_{k+1}$, we wish to obtain an estimate of the state at $t = t_{k+1}$ given the observation at $t = t_{k+1}$, i.e., $\mathbf{x}_{k+1|k+1}$. We assume that the estimate is a weighted sum of the prediction and the new observation given by the equation

$$\mathbf{x}_{k+1|k+1} = \mathbf{K}'_{k+1}\mathbf{x}_{k+1|k} + \mathbf{K}_{k+1}\mathbf{y}_{k+1}.\tag{19.37}$$

We next seek the gain matrices \mathbf{K}' and \mathbf{K} , such that the loss function, as Kaman called it, or given a form that we choose is minimized. Here we are minimizing the conditional mean square analysis error, which is given by

$$\boldsymbol{\varepsilon}_{k+1|k+1}^a = \mathbf{x}_{k+1|k+1} - \mathbf{x}_{k+1}^t.\tag{19.38}$$

A property that we wish for the filter to have is that it is unbiased, which then implies that we require $\mathbb{E}[\mathbf{x}_{k+1|k+1}] = \mathbb{E}[\mathbf{x}_{k-1}]$. This is verified through

$$\begin{aligned}\mathbb{E}\left[\mathbf{x}_{k+1|k+1}\right] &= \mathbb{E}\left[\mathbf{K}'_{k+1}\mathbf{x}_{k+1|k} + \mathbf{K}_{k+1}\mathbf{H}\mathbf{x}_{k+1} + \mathbf{K}_{k+1}\boldsymbol{\varepsilon}_{k+1}^o\right], \\ &= \mathbf{K}'_{k+1}\mathbb{E}\left[\mathbf{x}_{k+1|k}\right] + \mathbf{K}_{k+1}\mathbf{H}_{k+1}\mathbb{E}\left[\mathbf{x}_{k+1}\right],\end{aligned}\tag{19.39}$$

where the last term in the first line in (19.39) disappears due to the mean of the observational error being assumed to be zero, i.e., unbiased observational errors.

This then implies that our analysis state is given by

$$\begin{aligned}\mathbf{x}_{k+1|k+1} &= (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{k+1}) \mathbf{x}_{k+1|k} + \mathbf{K}_{k+1}\mathbf{y}_{k+1}, \\ &= \mathbf{x}_{k+1|k} + \mathbf{K}_{k+1} (\mathbf{y}_{k+1} - \mathbf{H}_{k+1}\mathbf{x}_{k+1|k}),\end{aligned}\tag{19.44}$$

where \mathbf{K} is the **Kalman gain matrix**.

We still have two more step to go. The first step is to derive the analysis error covariance matrix, given the updated analysis state from (19.44), which is denoted as $\mathbf{P}_{k+1|k+1}^a$ and is defined as

$$\begin{aligned}\mathbf{P}_{k+1|k+1}^a &= \mathbb{E} \left[\boldsymbol{\varepsilon}_{k+1}^a \left(\boldsymbol{\varepsilon}_{k+1}^a \right)^T \mid \mathcal{Z} \right], \\ &= \mathbb{E} \left[(\mathbf{x}_{k+1} - \mathbf{x}_{k+1|k+1}) (\mathbf{x}_{k+1} - \mathbf{x}_{k+1|k+1})^T \right], \\ &= (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{k+1}) \mathbb{E} \left[\boldsymbol{\varepsilon}_k^f \left(\boldsymbol{\varepsilon}_k^f \right)^T \right] (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{k+1}) + \mathbf{K}_{k+1} \mathbb{E} \left[\boldsymbol{\varepsilon}_{k+1}^0 \left(\boldsymbol{\varepsilon}_{k+1}^0 \right)^T \right] \mathbf{K}_{k+1}^T, \\ &= (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{k+1}) \mathbf{P}_{k+1|k}^f (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{k+1})^T + \mathbf{K}_{k+1} \mathbf{R}_{k+1} \mathbf{K}_{k+1}.\end{aligned}\tag{19.45}$$

It is possible to simplify (19.45) as follows:

$$\begin{aligned}\mathbf{P}^a &= \mathbb{E} \left[(\boldsymbol{\psi}^a - \boldsymbol{\psi}^t) (\boldsymbol{\psi}^a - \boldsymbol{\psi}^t)^T \right], \\ &= \mathbb{E} \left[\left(\boldsymbol{\psi}^f - \boldsymbol{\psi}^t + \mathbf{K} (\mathbf{y} - \mathbf{y}^t - \mathbf{H}\boldsymbol{\psi}^f + \mathbf{H}\boldsymbol{\psi}^t) \right) \left(\boldsymbol{\psi}^f - \boldsymbol{\psi}^t + \mathbf{K} (\mathbf{y} - \mathbf{y}^t - \mathbf{H}\boldsymbol{\psi}^f + \mathbf{H}\boldsymbol{\psi}^t) \right)^T \right], \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbb{E} \left[(\boldsymbol{\psi}^f - \boldsymbol{\psi}^t) (\boldsymbol{\psi}^f - \boldsymbol{\psi}^t)^T \right] (\mathbf{I} - \mathbf{K}\mathbf{H})^T \mathbf{K}\mathbf{O}\mathbf{K}^T, \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T \mathbf{K} + \mathbf{K} (\mathbf{H}\mathbf{P}^f \mathbf{H} + \mathbf{O}) \mathbf{K}^T.\end{aligned}\tag{19.46}$$

We now have to substitute the expression for the Kalman gain matrix into the left multiplication in the fourth term in (19.46), which then cancels the bracketed component of the fourth term, such that only the $\mathbf{P}^f \mathbf{H}^T$ term remains, which, when combined with \mathbf{K}^T , cancels with the third term to leave

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f. \quad (19.47)$$

The last step that we have to do here is to find the expression for the Kalman gain matrix. We know that we have to minimize the conditional mean square analysis error with respect to the Kalman gain matrix, which is equivalent to

$$\begin{aligned} L(\mathbf{e}_{k+1}^a) &= \min_{\mathbf{K}_{k+1}} \mathbb{E} \left[\mathbf{e}_{k+1}^a \left(\mathbf{e}_{k+1}^a \right) \middle| \mathbf{z}_{k+1} \right], \\ &= \min_{\mathbf{K}_{k+1}} \text{Trace} \left(\mathbb{E} \left[\mathbf{e}_{k+1}^a \left(\mathbf{e}_{k+1}^a \right) \middle| \mathbf{z}_{k+1} \right] \right), \\ &= \min_{\mathbf{K}_{k+1}} \text{Trace} \left(\mathbf{P}_{k+1|k+1}^a \right). \end{aligned} \quad (19.48)$$

Therefore, differentiating (19.45) with respect to \mathbf{K}_{k+1} and setting to zero yields

$$\frac{\partial L(\boldsymbol{\varepsilon}_{k+1}^a)}{\partial \mathbf{K}_{k+1}} = -2(\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{k+1})\mathbf{P}_{k+1|k}^f\mathbf{H}_{k+1}^T + 2\mathbf{K}_{k+1}\mathbf{R}_{k+1}^{-1} = \mathbf{0}.$$

Now, rearranging to isolate \mathbf{K}_{k+1} , we obtain the following expression for the Kalman gain matrix:

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1|k}^f\mathbf{H}_{k+1}^T \left[\mathbf{H}_{k+1}\mathbf{P}_{k+1|k}^f\mathbf{H}_{k+1}^T + \mathbf{R}_{k+1} \right]^{-1}. \quad (19.49)$$

Therefore, when dealing with the Kalman filter in the literature, it is quite often broken down into two parts: the first part is known as the **propagation, forecast, or prediction** stage, and the second part is referred to as the **update/analysis** stage. As a way to keep track of the different stages of the Kalman filter, we have provided a summary of these different stages of the Kalman filter in [Table 19.1](#).

Table 19.1 Summary of the Processes Associated With Each Stage of the Kalman Filter and the Definition of the Equivalent Matrix Operations

Propagation

Create a forecast, a priori, background, state, \mathbf{x}_k^f

Propagate the forecast error covariance matrix, \mathbf{P}_k^f

Update

Create observation innovation, \mathbf{d}_k

Create the Kalman gain matrix, \mathbf{K}_k

Update the analysis error covariance matrix, \mathbf{P}_k^a

$$\mathbf{x}_k^f = \mathbf{M}_k \mathbf{x}_{k-1}^a + \mathbf{B}_k \mathbf{u}_k$$

$$\mathbf{P}_k^f = \mathbf{M}_k \mathbf{P}_{k-1}^a \mathbf{M}_k^T + \mathbf{Q}_k$$

$$\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f$$

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T \left[\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k \right]^{-1}$$

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T.$$

ENSEMBLE-BASED DATA ASSIMILATION

20

CHAPTER OUTLINE

20.1	Stochastic Dynamical Modeling	784
20.2	Ensemble Kalman Filter	785
	20.2.1 Perturbed Observations-Based EnKF	791
20.3	Ensemble Square Root Filters	793
	20.3.1 Localization and Inflation	794
20.4	Ensemble and Local Ensemble Transform Kalman Filter	797
	20.4.1 ETKF	797
	20.4.2 LETKF	799
20.5	Maximum Likelihood Ensemble Filter	803
	20.5.1 Forecast Step	804
	20.5.2 Analysis Step	805
	20.5.3 Lyapunov and Bred Vectors	807
	20.5.4 Hybrid Lyapunov-Bred Vectors	808
20.6	Hybrid Ensemble and Variational Data Assimilation Methods	809
	20.6.1 α Control Variables	810
	20.6.2 Hybrid Ensemble Transform PSAS	814
	20.6.3 Ensembles of 4D VARs (EDA)	815
20.7	NDEnVAR	815
20.8	Ensemble Kalman Smoother	818
20.9	Ensemble Sensitivity	819
20.10	Summary	821

20.1 STOCHASTIC DYNAMICAL MODELING

Before we present the derivation of the EnKF, we shall briefly present some of the statistical/stochastic theory that Evensen uses to justify the EnKF. Following the flow of the 1994 paper, the first stochastic process we present is **Liouville's theorem**, which leads to Liouville's equations:

Liouville's theorem and equation

Theorem 20.1. *Consider a dynamical system with coordinates q_i and conjugate momenta p_i , where $i = 1, 2, \dots, N$. Then the phase space distribution $\rho(p, q)$ determines the probability $\rho(p, q) d^N q d^N p$ that a particle will be found in the infinitesimal phase space volume $d^N q d^N p$. The associated Liouville equation governs the evolution of the density in time, where now $\rho(p, q; t)$ is a function of time and the equation of the evolution is*

$$\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \sum_{i=1}^N \left(\frac{\partial\rho}{\partial q_i} \dot{q}_i + \frac{\partial\rho}{\partial p_i} \dot{p}_i \right) = 0. \quad (20.1)$$

We now introduce the notion of **Brownian motion**, which is defined as follows.

Definition 20.2. A linear Brownian motion, $b(t)$, is a real-valued stochastic process $\{b(t) : t \geq 0\}$ with the following properties:

1. $b(0) = x$.
2. The process has **independent increments**, which is to say that for all times $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$, then the increments $b(t_n) - b(t_{n-1}), b(t_{n-1}) - b(t_{n-2}), \dots, b(t_2) - b(t_1)$ are independent variables.
3. For all $t \geq 0$ and δt , the increment $b(t + \delta t) - b(t)$ are Gaussian distributed with expectation zero and variance δt .
4. The function $t \mapsto b(t)$ is continuous.

The multivariate version of Brownian motion is referred to as the n -dimensional Brownian motion and is defined as follows.

Definition 20.3. If b_1, b_2, \dots, b_n are independent linear Brownian motions started in x_1, x_2, \dots, x_n , then the stochastic process $\{b(t) : t \geq 0\}$, where

$$\{b\} = \{b_1(t), b_2(t), \dots, b_n(t)\}, \quad (20.2)$$

is called an n -dimensional Brownian motion.

Definition 20.4. A stochastic process is said to have the **Markov property** if the conditional probability distribution of the future states of the process, which could be dependent on both past and present states, depends only upon the present state.

Definition 20.5. A **Markov process** is a stochastic model that has the Markov property. It can be used to model a random system that changes states according to a transition rule that only depends on the current state. For a discrete time situation, this process is referred to as a **Markov chain**.

20.2 ENSEMBLE KALMAN FILTER

As we mentioned in the introduction to this chapter, the EnKF was first introduced in the groundbreaking paper by Geir Evensen in 1994 [8]. An important statement that is made at the start of the theory of stochastic dynamic prediction section of [8] is still relevant today:

The choice of another interpolation scheme of just different statistical parameters in the interpolation scheme will produce another initial state resulting in a different forecast even if the same deterministic model is applied. It is not possible to say that the forecast based on any interpolated initial conditions is right or wrong or better or worse, since each initial state estimate represents an individual member of an infinite ensemble of possible states that are consistent with the data.

Evensen then quotes a very telling statement from a paper by Epstein in 1969 [256]:

The different analyses will yield different forecasts, even if each were submitted to the same forecast procedure. If there is no way of determining which, if any, analysis is right, and since none is known to be wrong, there is no way of knowing in any instance, which to believe.

Here we shall leave [8] as the equations for the EnKF are not presented here, but a written explanation of the procedure is given as mentioned above. The equations for the EnKF first appear in Evensen and Van Leeuwen's paper, *Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model*, in 1996 [259]. We start by introducing the matrix \mathbf{A} that is of dimensions $n \times N$, where n is the number of state variables and N is the number of ensemble members, where the geophysical model's state from each ensemble member is stored. If we now denote the ensemble forecasts as \mathbf{A}_k^f for time step k , then we can calculate the ensemble forecast error covariance matrix at time k , denoted as \mathbf{P}_k^f , by

$$\mathbf{P}_k^f \equiv \frac{1}{N-1} \left(\mathbf{A}_k^f - \overline{\mathbf{M}}_k^f \right) \left(\mathbf{A}_k^f - \overline{\mathbf{M}}_k^f \right)^T, \quad (20.11)$$

where $\overline{\mathbf{M}}_k^f$ is an array that contains the predicted ensemble mean in each column. We should note that the rank of the error covariance matrix \mathbf{P}_k^f will be less than or equal to the number of members in the

$$\boldsymbol{\psi}_k^a = \boldsymbol{\psi}_k^f + \mathbf{K} \left(\mathbf{y}_k - \mathbf{H}_k \boldsymbol{\psi}_k^f \right), \quad (20.12)$$

$$\mathbf{y}_k = \mathbf{H}_k \boldsymbol{\psi}_k^f + \boldsymbol{\varepsilon}_k^o, \quad (20.13)$$

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k \right)^{-1}. \quad (20.14)$$

We now rewrite (20.12), the analysis equation, as

$$\boldsymbol{\psi}_k^a = \boldsymbol{\psi}_k^f + \mathbf{B}_k^T \mathbf{b}_k, \quad (20.15)$$

where $\mathbf{B}_k \equiv \mathbf{H}_k \mathbf{P}_k^f$, where the rows in \mathbf{B}_k are influence functions, but we have seen them earlier as part of the PSAS system as the representers for each observation, then the vector \mathbf{b}_k contains the amplitudes for each influence functions and is found through solving the matrix-vector system

$$\left(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k \right) \mathbf{b}_k = \mathbf{y} - \mathbf{H}_k \boldsymbol{\psi}_k^f. \quad (20.16)$$

We should note that it is not necessary to calculate the full error covariance matrices in (20.15) and (20.16), but we can apply the following procedure to evaluate them:

1. Calculate $\mathbf{S}_k = \frac{1}{N} \mathbf{H}_k \left(\mathbf{A}_k^f - \overline{\mathbf{M}} \right)$, which is a $N_o \times N$ matrix, where as always N_o represents the number of observations.
2. Form the influence functions matrix $\mathbf{B}_k = \mathbf{S}_k \left(\mathbf{A}_k^f - \overline{\mathbf{M}}_k^f \right)^T (N - 1)^{-1}$.
3. Form the representer matrix $\widehat{\mathbf{R}}_k = \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T \equiv \mathbf{H}_k \mathbf{B}_k^T \equiv \mathbf{S}_k \mathbf{S}_k^T (N - 1)^{-1}$.

Therefore to find the coefficient \mathbf{b}_k , we simply solve (20.16), but only have to calculate \mathbf{S}_k , as long as $\widehat{\mathbf{R}}_k + \mathbf{R}_k$ is non-singular.

$$\mathbf{P}_k^f = \mathbb{E} \left[\left(\psi_k^f - \psi_k^t \right) \left(\psi_k^f - \psi_k^t \right)^T \right], \quad (20.17a)$$

$$\mathbf{P}_k^a = \mathbb{E} \left[\left(\psi_k^a - \bar{\psi}_k \right) \left(\psi_k^a - \bar{\psi}_k \right)^T \right]. \quad (20.17b)$$

In the EnKF we do not know the true covariances; this is also true for all forms of data assimilation, but in the EnKF the true state is replaced by the **ensemble covariance matrices** that are calculated around the ensemble mean $\bar{\psi}$. Therefore, it is possible to approximate (20.17a) and (20.17b) by

$$\mathbf{P}_{e,k}^f = \mathbb{E} \left[\left(\psi_k^f - \psi_k^t \right) \left(\psi_k^f - \psi_k^t \right)^T \right], \quad (20.18a)$$

$$\mathbf{P}_{e,k}^a = \mathbb{E} \left[\left(\psi_k^a - \psi_k^t \right) \left(\psi_k^a - \psi_k^t \right)^T \right]. \quad (20.18b)$$

Therefore, in summary the EnKF is an approach that is equivalent to solving the Fokker-Planck (Kolmogorov's) equation for the evolution of the probability density function for the error statistics [261].

20.2.1 PERTURBED OBSERVATIONS-BASED EnKF

The starting point of the perturbed observations-based EnKF is the ensemble estimates for the forecast and analysis error covariance matrices from (20.17a) and (20.17b), where the argument for perturbing the observations comes from the assumption that there is no representative error involved here. If we recall that the analysis error covariance matrix from the Kalman filter is defined as

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H}^T) \mathbf{P}^f, \quad (20.19)$$

then the analysis update for the EnKF is based upon the equation above; however, if we take an ensemble of model states such that the error covariances of the forecasted ensemble mean coincides with the ensemble covariance and then we perform an analysis on each ensemble member, then the error covariance of the analyzed ensemble mean is given by (20.19) [8]. However, the **ensemble covariance** is reduced too much unless the observations are treated as random variables. The reason for this statement is because in the expression for the analysis ensemble covariance there is no analog to the term $\mathbf{K}\mathbb{E}[(\mathbf{d} - \mathbf{d}^f)(\mathbf{d} - \mathbf{d}^f)^T]\mathbf{K}^T = \mathbf{K}\mathbf{O}\mathbf{K}^T$ in (20.19), and as such this leads to spurious correlations because all of the ensemble members are updated with the same observations. This then results in the covariance of the analyzed ensemble being

$$\mathbf{P}_e^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f(\mathbf{I} - \mathbf{K}\mathbf{H})^T, \quad (20.20)$$

where we can see that we have too many $(\mathbf{I} - \mathbf{K}\mathbf{H})^T$ terms. The reason for this is that the original analysis scheme for the EnKF was based upon (20.17a) and (20.17b), while it should have been based upon (20.18a) and (20.18b) [260].

Therefore, the basis of the updated version of the EnKF from [260] is now to treat the observations as random variables by generating an ensemble of observations, where this ensemble is generated from a distribution with a mean that is equal to the first-guess observations and covariance that is equal to \mathbf{O} .

If we now define the new observations as

$$y_j = y + \epsilon^o, \quad (20.21)$$

where $j = 1, 2, \dots, n$, we now modify the analysis step of the EnKF as

$$\psi_j^a = \psi_j^f + \mathbf{K}_e \left(y_j - \mathbf{H}\psi_j^f \right), \quad (20.22)$$

where the ensemble Kalman gain matrix, \mathbf{K}_e is given by

$$\mathbf{K}_e = \mathbf{P}_e^f \mathbf{H}^T \left(\mathbf{H} \mathbf{P}_e^f \mathbf{H}^T + \mathbf{O} \right)^{-1}. \quad (20.23)$$

We can now express the mean analysis state in terms of the mean forecast state, observation, and model representative as

$$\bar{\psi}^a = \bar{\psi}^f + \mathbf{K}_e \left(\bar{y} - \mathbf{H}\bar{\psi}^f \right). \quad (20.24)$$

Given the expressions above, we can easily show that the ensemble analysis error covariance matrix is given by

$$\begin{aligned} \mathbf{P}_e^a &= \mathbb{E} \left[\left(\psi^a - \bar{\psi}^a \right) \left(\psi^a - \bar{\psi}^a \right)^T \right], \\ &= \left(\mathbf{I} - \mathbf{K}_e \mathbf{H} \right) \mathbf{P}_e^f, \end{aligned} \quad (20.25)$$

where

$$\psi^a - \bar{\psi}^a = \left(\mathbf{I} - \mathbf{K}_e \mathbf{H} \right) \left(\psi^f - \bar{\psi}^f \right) + \mathbf{K}_e \left(y - \bar{y} \right). \quad (20.26)$$

An important feature to note here is that the ensemble of the observations do not affect the update to the ensemble means, since this term is not in (20.24). Another feature to note here is that each ensemble member evolves according to a model such as

$$\psi_j^{k+1} = \mathcal{M} \left(\psi_j^k \right) + d\mathbf{q}_j^k, \quad (20.27)$$

where $d\mathbf{q}$ is a stochastic forcing that represents the model error from a probability distribution with mean zero and covariance \mathbf{Q} . The **ensemble covariance** matrix of the error in the model equations is given by

$$\begin{aligned}\mathbf{P}_e^a &= \mathbb{E} \left[\left(\psi^a - \bar{\psi}^a \right) \left(\psi^a - \bar{\psi}^a \right)^T \right], \\ &= (\mathbf{I} - \mathbf{K}_e \mathbf{H}) \mathbf{P}_e^f,\end{aligned}\tag{20.25}$$

where

$$\psi^a - \bar{\psi}^a = (\mathbf{I} - \mathbf{K}_e \mathbf{H}) \left(\psi^f - \bar{\psi}^f \right) + \mathbf{K}_e (\mathbf{y} - \bar{\mathbf{y}}).\tag{20.26}$$

An important feature to note here is that the ensemble of the observations do not affect the update to the ensemble means, since this term is not in (20.24). Another feature to note here is that each ensemble member evolves according to a model such as

$$\psi_j^{k+1} = \mathcal{M} \left(\psi_j^k \right) + d\mathbf{q}_j^k,\tag{20.27}$$

where $d\mathbf{q}$ is a stochastic forcing that represents the model error from a probability distribution with mean zero and covariance \mathbf{Q} . The **ensemble covariance** matrix of the error in the model equations is given by

$$\mathbf{Q}_e = \mathbb{E} \left[\left(d\mathbf{q}^k - \bar{d\mathbf{q}}^k \right) \left(d\mathbf{q}^k - \bar{d\mathbf{q}}^k \right)^T \right].\tag{20.28}$$

The ensemble mean then evolves according to the equation

$$\begin{aligned}\bar{\psi}^{k+1} &= \overline{\mathcal{M} \left(\psi^k \right)}, \\ &= \mathcal{M} \left(\bar{\psi}^k \right) + NLT,\end{aligned}\tag{20.29}$$

where NLT stands for nonlinear term if \mathcal{M} is nonlinear.

20.3 ENSEMBLE SQUARE ROOT FILTERS

The motivation for the derivation of the EnSRF, as we just mentioned, was firstly to avoid the sampling error caused by perturbing the observations, which was to avoid reduction of the ensemble covariance being too severe, but secondly to find a formulation of the ensemble Kalman gain matrix such that the analysis error covariance matrix for the ensemble satisfies

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^b(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T = (\mathbf{I} - \mathbf{KH})\mathbf{P}^b. \quad (20.30)$$

If we recall the work of [260], then we have that if all of the members of the ensemble are updated with the same observations using the same gain matrix, then the covariance of the analyzed ensemble is shown to be

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^b(\mathbf{I} - \mathbf{KH})^T. \quad (20.31)$$

The challenge now is to define an approximation $\tilde{\mathbf{K}}$ for \mathbf{K} from the ensemble such that when $\tilde{\mathbf{K}}$ is substituted into (20.31), we obtain a form similar to that on the right-hand side of (20.30).

It is shown in [263] that an expression for $\tilde{\mathbf{K}}$ that satisfies the required property is given by

$$\tilde{\mathbf{K}} = \mathbf{P}^b \mathbf{H}^T \left(\left(\sqrt{\mathbf{HP}^b \mathbf{H}^T + \mathbf{R}} \right)^{-1} \right)^T \left(\sqrt{\left(\mathbf{HP}^b \mathbf{H}^T + \mathbf{R} \right) + \sqrt{\mathbf{R}}} \right)^{-1}. \quad (20.32)$$

We have to note that the square roots of the matrices in (20.32) are not unique, they can be computed in different ways, say with a singular vector decomposition or through a Cholesky factorization.

If we now consider an individual observation, then $\mathbf{HP}^b \mathbf{H}^T$ and \mathbf{R} reduce to scalars and the equation

$$(\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{P}^b(\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})^T = (\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{P}^b, \quad (20.33)$$

becomes

$$\frac{\mathbf{HP}^b \mathbf{H}^T}{\mathbf{HP}^b \mathbf{H}^T + \mathbf{R}} \tilde{\mathbf{K}}\tilde{\mathbf{K}} - \mathbf{K}\tilde{\mathbf{K}}^T - \tilde{\mathbf{K}}\mathbf{K}^T + \mathbf{K}\mathbf{K}^T = \mathbf{0}. \quad (20.34)$$

If we now set $\tilde{\mathbf{K}} = \alpha\mathbf{K}$, where α is a constant, then it is possible to factorize out the $\mathbf{K}\mathbf{K}^T$ term, which results in a quadratic equation for α such that the solution that is between 0 and 1 can be shown to be

$$\alpha = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^b\mathbf{H} + \mathbf{R}}} \right)^{-1}. \quad (20.35)$$

Thus the mean and departure from the mean are updated independently according to

$$\begin{aligned} \mathbf{K} &= \mathbf{P}^b\mathbf{H}^T (\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}, \\ \bar{\mathbf{x}}^a &= \bar{\mathbf{x}}^b + \mathbf{K} (\bar{\mathbf{y}} - \mathbf{H}\bar{\mathbf{x}}^b), \\ \mathbf{x}^{a'} &= \mathbf{x}^{b'} - \tilde{\mathbf{K}} (\mathbf{H}\mathbf{x}^{b'}), \end{aligned}$$

where $\tilde{\mathbf{K}} = \alpha\mathbf{K}$. We should note that the covariance matrix here is the ensemble-based approximations.

One important feature of the EnSRF is that it processes the observations sequentially, which makes it possible to implement the covariance localization. This improves the analysis while preventing filter divergence in small ensembles. A summary of the EnSRFs up to 2003 can be found in Tippet et al. [248].

In the theory that we have just presented, we saw one of the problems associated with the EnKF, which was rank deficiencies. But there is also another problem associated with the **sampling** of the ensembles. This undersampling can lead to smaller scales not being resolved correctly; we saw this in the VAR subcomponent chapter where we could dynamically excite spurious gravity waves.

While the EnKF does not generate guaranteed balanced initial conditions unless some constraint is applied, there is the problem coming from the undersampling. In the next section we shall introduce a couple of techniques that have been introduced to compensate for the undersampling because, as with the balance constraints in VAR, it is possible for the filters to *diverge*. These techniques are referred to as **localization and inflation**.



In Houtekamer and Mitchell's 2001 paper *A sequential ensemble Kalman filter for atmospheric data assimilation* [264], they introduce an ensemble error localization so as to be able to filter the forecast error covariance matrix of the small background-error correlations that come from using too small an ensemble. They achieve this localization by applying a Schur product, which is also referred to as the Hadamard product, which is an element-by-element matrix product, of the covariances of the forecast error that has been calculated from the ensemble and a correlation function that has local support.

Specifically, Houtekamer and Mitchell redefine the ensemble Kalman gain matrix \mathbf{K} as

$$\mathbf{K} = \left((\rho \circ \mathbf{P}^f) \mathbf{H}^T \right) \left(\mathbf{H} (\rho \circ \mathbf{P}^f) \mathbf{H}^T + \mathbf{R} \right)^{-1}, \quad (20.36)$$

where it is possible to interchange the order of the forward interpolation and the Schur product such that

$$\mathbf{k} = \left(\rho \circ \left(\mathbf{P}^f \mathbf{H}^T \right) \right) \left(\rho \circ \left(\mathbf{H} \mathbf{P}^f \mathbf{H}^T \right) + \mathbf{R} \right)^{-1}, \quad (20.37)$$

where

$$\mathbf{P}^f \mathbf{H}^T \equiv \frac{1}{N-1} \sum_{i=1}^N \left(\psi_i^f - \bar{\psi}^f \right) \left(\mathbf{H} \psi_i^f - \mathbf{H} \bar{\psi}^f \right),$$
$$\mathbf{H} \mathbf{P}^f \mathbf{H}^T \equiv \frac{1}{N-1} \sum_{i=1}^N \left(\mathbf{H} \psi_i^f - \mathbf{H} \bar{\psi}^f \right) \left(\mathbf{H} \psi_i^f - \mathbf{H} \bar{\psi}^f \right).$$

The matrix ρ is a relatively broad function. For the study presented in [264] the authors use a fifth-order compactly supported piecewise rational function which comes from [226], and is also used in the work in [265]. There are a large suite of different compactly supported two- and three-dimensional correlation functions presented in [226] and the reader is referred to that paper for more details about the different piecewise functions.

Another reason for the localization in either observation space and/or ensemble covariances is to prevent what is referred to as **filter divergence**. The first instance of filter divergence, according to [265], occurred in [263]. Filter divergence is the process that the ensemble progressively ignores the observations more and more in successive cycles, which then leads to a useless ensemble. In [265] the authors state that the cause of the filter divergences is due to using the ensemble to produce reduced rank representation of background error statistics, or, as we have been calling them, forecast error statistics.

20.4 ENSEMBLE AND LOCAL ENSEMBLE TRANSFORM KALMAN FILTER

In this section we shall introduce two versions of the EnKF that are used quite extensively today. The first filter we consider is the ETKF. The ETKF first appeared in [272]; however, here we shall summarize the derivation from Bishop et al. [273], where the derivation is a bit easier to follow.

20.4.1 ETKF

As always we are going to start with one of the equations of the Kalman filter; this time we are starting with the analysis error covariance matrix as

$$\mathbf{P}_t^a = \frac{1}{K-1} \sum_{k=1}^K (\mathbf{x}_k(t_i) - \bar{\mathbf{x}}(t_i)) (\mathbf{x}_k(t_i) - \bar{\mathbf{x}}(t_i))^T = \mathbf{X}_t \mathbf{X}_t^T, \quad (20.38)$$

where the columns of \mathbf{X} are given by $\frac{\mathbf{x}_k(t_i) - \bar{\mathbf{x}}(t_i)}{\sqrt{K-1}}$. Therefore according to (20.38) the forecast error matrix at t_{i+1} is given by

$$\mathbf{P}_{t+1}^f = \mathbf{Z}_{t+1} \mathbf{Z}_{t+1}^T = \mathbf{X}_{t+1} \mathbf{T}_0 \mathbf{T}_0^T \mathbf{X}_{t+1}^T, \quad (20.39)$$

where at the initial time then \mathbf{T}_0 is equal to the identity \mathbf{I} . Thus at any later data assimilation time t_{i+m} , we have

$$\begin{aligned} \mathbf{P}_{t+m}^f &= \mathbf{Z}_{i+m} \mathbf{Z}_{i+m}^T, \\ &= \mathbf{X}_{i+m} \mathbf{T}_{t+m-1} \mathbf{T}_{t+m-1}^T \mathbf{X}_{i+m}^T, \end{aligned} \quad (20.40)$$

where \mathbf{T}_{t+m-1} is a $K \times K$ transformation matrix that is generally not equal to the identity matrix.

It is now the goal of [273] to be able to write the ensemble-based analysis error covariance matrix in the form

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} = \mathbf{Z}^f \mathbf{T} \mathbf{T}^T \mathbf{Z}^{fT}, \quad (20.41)$$

for the transformation matrix \mathbf{T} given that the forecast error covariance matrix is $\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{fT}$.

Now the next goal of the EnKFs is to avoid the size and ill-conditioning problems associated with the reduce rank approximations. With respect to the ETKF they first state that we introduce a normalized observation operator a $\tilde{\mathbf{H}} = \mathbf{R}^{-\frac{1}{2}} \mathbf{H}$. This enables us to write

$$\begin{aligned} \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{P}^f &= \mathbf{P}^f \mathbf{H}^T \left(\mathbf{R}^{-\frac{1}{2}} \left(\mathbf{R}^{-\frac{1}{2}} \mathbf{H} \mathbf{P}^f \mathbf{H}^T \mathbf{R}^{-\frac{1}{2}} + \mathbf{I} \right) \mathbf{R}^{-\frac{1}{2}} \right)^{-1} \mathbf{H} \mathbf{P}^f, \\ &= \mathbf{P}^f \tilde{\mathbf{H}}^T \left(\tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}^T + \mathbf{I} \right)^{-1} \tilde{\mathbf{H}} \mathbf{P}^f, \end{aligned} \quad (20.42)$$

where \mathbf{I} is of the dimensions of the number of observations. We now apply a eigenvector decomposition to (20.42) using the property that the eigenvectors of $\tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}^T$ are equivalent to those of $\tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}^T + \mathbf{I}$, which implies that

$$\left(\tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}^T + \mathbf{I} \right)^{-1} = \mathbf{E}^c (\mathbf{\Gamma}^c + \mathbf{I})^{-1} \mathbf{E}^c, \quad (20.43)$$

where the p columns of \mathbf{E}^c contain the complete set of orthonormal eigenvectors of $\tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}^T$ and the diagonal matrix $\mathbf{\Gamma}$ contains the corresponding eigenvalues. We only require the eigenvectors that are not in the right null space of $\mathbf{P}^f \mathbf{H}^T$. Since

$$\mathbf{P}^f \tilde{\mathbf{H}}^T = \mathbf{Z}^f \mathbf{Z}^{fT} \tilde{\mathbf{H}}^T, \quad (20.44)$$

then only the eigenvectors of $\tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}^T$ that contribute to the analysis error covariance matrix defined by (20.41) are those that can be written as a linear combination of the column vectors of $\tilde{\mathbf{H}} \mathbf{Z}^f$. However, we should note that not all of the eigenvectors are linearly independent because the sum of the K ensemble perturbations from which they are derived is equal to zero. Therefore, we seek the set of eigenvectors that are linearly independent to form the vector space.

After some manipulations that are shown in [273], we arrive at the three main equations for the ETKF that are given by

$$\mathbf{P}^f \tilde{\mathbf{H}}^T \left(\tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}^T + \mathbf{I} \right)^{-1} \tilde{\mathbf{H}} \mathbf{P}^f = \mathbf{Z} (t^a) \mathbf{C} \Gamma (\Gamma + \mathbf{I})^{-1} \mathbf{C}^T \mathbf{Z}^{fT} (t^a), \quad (20.47)$$

$$\mathbf{x}^a (t) - \mathbf{x}^f (t) = \mathbf{M} (t, t^a) \mathbf{P}^f \tilde{\mathbf{H}}^T \left(\tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}^T + \mathbf{I} \right)^{-1} \left(\mathbf{R}^{-\frac{1}{2}} \mathbf{y} - \tilde{\mathbf{H}} \mathbf{x}^f (t^a) \right), \quad (20.48)$$

$$\mathbf{x}^a (t) - \bar{\mathbf{x}}^f (t) = \mathbf{Z}^f (t) \mathbf{C} \Gamma^{\frac{1}{2}} (\Gamma + \mathbf{I})^{-1} \mathbf{E}^T \left(\mathbf{R}^{-\frac{1}{2}} \mathbf{y} - \tilde{\mathbf{H}} \mathbf{x}^f (t^a) \right), \quad (20.49)$$

where \mathbf{M} is the numerical model and \mathbf{y} is the vector of observations.

Localized Ensemble Transform Kalman Filter

LETKF

1. Advance the analysis ensemble of global geophysical states to the next analysis time. This is a new background ensemble of global geophysical states.
2. Associate a local region with each grid point and, for each local region and each member of the background ensemble, form vectors of geophysical state information in that local region.
3. For each local ensemble member vector obtained from the previous step, calculate its perturbation from the background ensemble mean, and project these perturbations on to a lower dimensional subspace that best represents the ensemble in that region.
4. Perform the data assimilation in each of the local low dimensional subspaces, obtaining analysis mean and covariance in each local region.
5. From the local analysis mean and covariance, obtain a suitable local analysis ensemble of local geophysical states.
6. Use the local analyses obtained in step 5 to form a new global analysis ensemble.
7. return to 1.

The next step that Ott et al. [274] move on to is the assimilation step, which is to minimize the incremental 3D VAR cost function, but projected into ensemble space. They linearize the nonlinear observation operator about the mean ensemble background state $\bar{\mathbf{x}}_{mn}^b$, where this state is assumed to be quite close to the true state, that they denote as \mathbf{x}_{mn}^a , which is the **local analysis**. We shall not go into all of the details about the minimization but will present the important equations:

$$\mathbf{h}(\mathbf{x}_{mn}^a) \approx \mathbf{h}(\bar{\mathbf{x}}_{mn}^b) + \mathbf{H}_{mn} \Delta \mathbf{x}_{mn}^a, \quad (20.62)$$

$$\Delta \mathbf{x}_{mn}^a = \mathbf{x}_{mn}^a - \bar{\mathbf{x}}_{mn}^b, \quad (20.63)$$

$$J(\Delta \hat{\mathbf{x}}_{mn}^a) = \frac{1}{2} (\Delta \hat{\mathbf{x}}_{mn}^a)^T (\hat{\mathbf{P}}_{mn}^b)^{-1} \Delta \hat{\mathbf{x}}_{mn}^a + \frac{1}{2} (\hat{\mathbf{H}}_{mn} \Delta \hat{\mathbf{x}}_{mn}^a + \mathbf{h}_{mn}(\bar{\mathbf{x}}_{mn}^b) - \mathbf{y}_{mn})^T \mathbf{R}_{mn}^{-1} (\hat{\mathbf{H}}_{mn} \Delta \hat{\mathbf{x}}_{mn}^a + \mathbf{h}_{mn}(\bar{\mathbf{x}}_{mn}^b) - \mathbf{y}_{mn}), \quad (20.64)$$

where the superscript $\hat{\cdot}$ represents that that matrix has been projected by \mathbf{Q}_{mn} into the \mathbb{S}_{mn} subspace.

Now if minimizing the cost function in (20.64), then the state $\hat{\mathbf{x}}_{mn}^a$ is the most probable state, which is equal to

$$\hat{\mathbf{x}}_{mn}^a = \hat{\mathbf{P}}_{mn}^b \hat{\mathbf{H}}_{mn}^T \mathbf{R}_{mn}^{-1} (\mathbf{y}_{mn} - \mathbf{h}_{mn}(\bar{\mathbf{x}}_{mn}^b)), \quad (20.65)$$

Maximum Likelihood Ensemble Filter (MLEF)

20.5.1 FORECAST STEP

The MLEF comprises of two different stages. The first stage is the forecast step and is concerned with the evolution of the forecast error covariances. The starting point for this step is from the evolution equation of the discrete version of the Kalman filter [257]. Therefore, we have that

$$\mathbf{P}^f(k) = \mathcal{M}_{k-1,k} \mathbf{P}_a(k-1) \mathcal{M}_{k-1,k}^T + \mathbf{Q}(k-1), \quad (20.78)$$

where \mathbf{P}^f is the forecast error covariance matrix, k is the time index, \mathcal{M} is the nonlinear model evolution operator, and \mathbf{q} is the model error matrix which is assumed to be normally distributed. For the purpose of this work this is assumed to be zero.

A factorization of \mathbf{P}^f into a square root form can be defined as

$$\mathbf{P}^f = \mathcal{M}\mathbf{P}_a\mathcal{M}^T = \left(\mathcal{M}\mathbf{P}_a^{\frac{1}{2}}\right)\left(\mathcal{M}\mathbf{P}_a^{\frac{1}{2}}\right)^T = \mathbf{P}_f^{\frac{1}{2}}\mathbf{P}_f^{\frac{1}{2}T}. \quad (20.79)$$

The structure of the square-root analysis error covariance matrix, $\mathbf{P}_a^{\frac{1}{2}}$, is

$$\mathbf{P}_a^{\frac{1}{2}} = \left(\mathbf{p}_1 \quad \mathbf{p}_2 \quad \dots \quad \mathbf{p}_S \right) \quad \text{where } \mathbf{p}_i = \begin{pmatrix} p_{1,i} \\ p_{2,i} \\ \vdots \\ p_{N,i} \end{pmatrix}, \quad (20.80)$$

where N is the number of state variables and S is the number of ensemble members with the assumption that $S \ll N$.

Upon expanding (20.80), the square root forecast error covariance matrix, $\mathbf{P}_f^{\frac{1}{2}}$, can be expressed as

$$\begin{aligned} \mathbf{P}_f^{\frac{1}{2}} &= \left(\mathbf{b}_1 \quad \mathbf{b}_2 \quad \dots \quad \mathbf{b}_S \right), \\ \mathbf{b}_i &= \mathcal{M}(\mathbf{x}_{k-1} + \mathbf{p}_i) - \mathcal{M}(\mathbf{x}_{k-1}) \approx \mathcal{M}\mathbf{p}_i, \end{aligned} \quad (20.81)$$

where \mathbf{x}_{k-1} is the analysis state from the previous assimilation cycle, which is found from the posterior analysis PDF [5]. Therefore the MLEF evolves the square root analysis error covariance matrix through the ensemble members.

20.5.2 ANALYSIS STEP

The second step in the MLEF is the analysis step which involves solving a nonlinear cost function, similar to that of [5], which is based on a Gaussian assumption for the background and observational errors.

The associated cost function is defined in terms of \mathbf{P}_f , although this matrix is never calculated or stored in the process of the filter. This results in

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{P}_f^{-1} (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2} (\mathbf{y} - \mathbf{h}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{h}(\mathbf{x})), \quad (20.82)$$

where \mathbf{y} is the vector of observations, \mathbf{h} is the nonlinear observation operator, \mathbf{R} is the observational covariance matrix, and \mathbf{x}_b is a background state, such that $\mathbf{x}_b = \mathcal{M}(\mathbf{x}_{k-1})$.

To find the minimum of (20.82), we introduce a change of variable through a Hessian preconditioner, defined by

$$\mathbf{x} - \mathbf{x}_b = \mathbf{P}_f^{\frac{1}{2}} (\mathbf{I} + \mathbf{C})^{-\frac{T}{2}} \boldsymbol{\xi}, \quad (20.83)$$

where $\boldsymbol{\xi}$ is our vector of control variables, defined in ensemble subspace, and \mathbf{C} is the Hessian matrix of (20.82), which is

$$\mathbf{C} = \mathbf{P}_f^{\frac{T}{2}} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{P}_f^{\frac{1}{2}} = \left(\mathbf{R}^{-\frac{1}{2}} \mathbf{H} \mathbf{P}_f^{\frac{1}{2}} \right)^T \left(\mathbf{R}^{-\frac{1}{2}} \mathbf{H} \mathbf{P}_f^{\frac{1}{2}} \right), \quad (20.84)$$

where \mathbf{H} is the Jacobian matrix of \mathbf{h} evaluated at \mathbf{x}_b .

It may be the case that the observation operator is nonlinear, difficult to differentiate analytically, or even discontinuous. To overcome this problem we use information from \mathbf{P}^f to approximate the square root of \mathbf{C} , componentwise, as

$$\begin{aligned} \mathbf{z}_i &= \left(\mathbf{R}^{-\frac{1}{2}} \mathbf{H} \mathbf{P}_f^{\frac{1}{2}} \right)_i = \mathbf{R}^{-\frac{1}{2}} \mathbf{H} \mathbf{b}_i, \\ &\approx \mathbf{R}^{-\frac{1}{2}} (\mathbf{h}(\mathbf{x} + \mathbf{b}_i) - \mathbf{h}(\mathbf{x})). \end{aligned} \quad (20.85)$$

It is shown in [278] that the MLEF is a non-differentiable minimization algorithm due to these types of approximations.

A new matrix \mathbf{Z} is now defined such that

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_S \end{pmatrix}. \quad (20.86)$$

The definition above allows \mathbf{C} to be written as $\mathbf{C} = \mathbf{Z}^T \mathbf{Z}$, i.e.,

$$\mathbf{C} = \begin{pmatrix} \mathbf{z}_1^T \mathbf{z}_1 & \mathbf{z}_1^T \mathbf{z}_2 & \dots & \mathbf{z}_1^T \mathbf{z}_S \\ \mathbf{z}_2^T \mathbf{z}_1 & \mathbf{z}_2^T \mathbf{z}_2 & \dots & \mathbf{z}_2^T \mathbf{z}_S \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{z}_S^T \mathbf{z}_1 & \mathbf{z}_S^T \mathbf{z}_2 & \dots & \mathbf{z}_S^T \mathbf{z}_S \end{pmatrix}. \quad (20.87)$$

To accomplish the inversion of $(\mathbf{I} + \mathbf{C})$, required in (20.83), we apply the spectral theorem for Hermitian matrices [279]. The theorem allows for an orthogonal eigenvalue decomposition of \mathbf{C} in the form

$$\mathbf{C} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T,$$

where \mathbf{V} is a matrix whose columns are orthogonal eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of \mathbf{C} .

The final point about the MLEF is the updating of the square root analysis error covariance matrix by

$$\mathbf{P}_a^{\frac{1}{2}} = \mathbf{P}_f^{\frac{1}{2}} (\mathbf{I} + \mathbf{C}(\mathbf{x}_{opt}))^{-\frac{T}{2}}, \quad (20.88)$$

Hybrid 3D VAR

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + \frac{1}{2} (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}), \quad (20.95)$$

where the background error covariance matrix \mathbf{B} is decomposed in the form $\mathbf{B} = \mathbf{S}\mathbf{C}\mathbf{S}^T$, where \mathbf{S} is the transform from spectral coefficients to grid points and \mathbf{C} is the diagonal matrix of variances of the spectral coefficients and seek the analysis increments, $\mathbf{x}^a - \mathbf{x}^b$, such that

$$(\mathbf{I} + \mathbf{B}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) (\mathbf{x}^a - \mathbf{x}^b) = \mathbf{B}\mathbf{H}^T\mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}^b). \quad (20.96)$$

The hybrid EnKF-3D VAR scheme uses a weighted mean of $\mathbf{B} = \mathbf{S}\mathbf{C}\mathbf{S}^T$ and \mathbf{P}^b derived from the ensemble, which is fully time dependent and spatially inhomogeneous, such that

$$\mathbf{B} = (1 - \alpha)\mathbf{P}^b + \alpha\mathbf{S}\mathbf{C}\mathbf{S}^T. \quad (20.97)$$

By changing the value of α from 0 to 1 then the analysis changes from using only flow-dependent ensemble-based error covariances to using the original 3D VAR covariances.

The advantages of the hybrid approach as set out in [285] are:

1. The hybrid scheme allows the user to evaluate combinations of 3D VAR and ensemble-based background statistics rather than relying strictly upon one or the other.
2. Ensemble-based statistics alone will be rank deficient and subject to sampling errors, and as such blending in the 3D VAR static statistics may *fill out* the covariance matrix and ameliorate some of the sampling error problems.

Hybrid 4D Var or En4DVAR

20.6.1 α CONTROL VARIABLES

The α control variables were introduced at the United Kingdom's Met. Office in the late 1990s and the theory behind them was mentioned in Lorenc [286]. This technique has been implemented operationally at the Met. Office as well as in the National Center for Environmental Prediction (NCEP) initial hybrid system [287]. We shall consider the approach set out in Clayton et al. [288]. The motivation of the hybrid approach at the Met. Office was to capture the *errors of the day* which are the short-range errors, while the climatological or static covariance matrix of the variational component captures the large- and longer-scale errors.

For the ensemble component of the hybrid system at the Met. Office they run what is referred to as the Met. Office Global and Regional Ensemble Prediction Systems, or MOGREPS, which comprises of an ETKF as described in Wang et al. [289]. The global version of the ensemble prediction system is called MOGREPS-G.

An important feature of the hybrid system at the Met. Office is that it is completely coupled. This coupling comes from:

1. the analysis perturbations for the ensemble system are generated by centering around the deterministic analysis from the 4D VAR system; and
2. the 4D VAR system is dependent on forecast data from the ensemble system.

At the start of each 4D VAR window the necessary ensemble forecast fields are taken from MOGREPS-G and interpolated to the analysis grid of the 4D VAR system. Note that the ensemble members are run at a lower resolution to both the analysis scheme in the 4D VAR and the global forecast model as well. As we have seen with the other ensemble filters, the differences between the ensemble members fields and the ensemble mean are taken, scaled by $\frac{1}{\sqrt{K-1}}$, and stored in an array \mathbf{W} such that

$$\mathbf{P}_e^f = \mathbf{W}\mathbf{W}^T. \quad (20.99)$$

As with most ensemble-based approximations to the error covariance matrix, the approximation in (20.99) is an undersampling of the errors and as such the Met. Office applies a localization matrix \mathbf{C} such that

$$\mathbf{B}_e = \mathbf{P}_e^f \circ \mathbf{C}. \quad (20.100)$$

Therefore, given the static, climatological background error covariance matrix from the 4D VAR system, the hybrid system seeks to implement a hybrid background error covariance matrix that is a linear combination of the static and flow dependent background error covariance matrices as

$$\mathbf{B} = \beta_c^2 \mathbf{B}_c + \beta_e^2 \mathbf{B}_e, \quad (20.101)$$

where β_c^2 and β_e^2 are scalar weights. This is where the α control variables come in, or implementing an extended control variable method as it is referred to as in [288].

The process of implementing the α control variable approach starts with the decomposition that the Met. Office implements for its control variable transform. The Met. Office uses an incremental 4D VAR system that involves increments denoted as $\delta\mathbf{w}$, and that there is a control variable \mathbf{v} such that $\delta\mathbf{w} = \mathbf{U}\mathbf{v}$, where \mathbf{U} is the square root matrix of \mathbf{B}_c , which are the inverses of the transforms \mathbf{T} that make the control variables uncorrelated and have approximately unit variance. The transforms associated with the \mathbf{T} matrix are referred to as the T-Transforms and they comprise of

$$\mathbf{T} = \mathbf{T}_h\mathbf{T}_v\mathbf{T}_p, \quad (20.102)$$

where \mathbf{T}_p is the transform to combine some fields to reduce the number of fields; \mathbf{T}_v is the projection on to approximately uncorrelated vertical modes; and \mathbf{T}_h is the projection on to global spherical harmonic functions.

The \mathbf{U} transform is made up of the approximate, or exact, inverses of the \mathbf{T} transforms, such that

$$\mathbf{U} = \mathbf{U}_p\mathbf{U}_v\mathbf{U}_h. \quad (20.103)$$

Thus the increment $\delta\mathbf{w}$ is related to the static and ensemble components through

$$\delta\mathbf{w} = \beta_c \mathbf{U}_p \mathbf{U}_v \mathbf{U}_h \mathbf{v} + \beta_e \sum_{k=1}^K \mathbf{w}'_k \circ \alpha_k, \quad (20.104)$$

where \mathbf{w}'_k is the ensemble error mode.

Given the definition in (20.104), then the incremental 4D VAR cost function becomes

$$J(\mathbf{v}, \alpha_1, \alpha_2, \dots, \alpha_k) = \frac{1}{2} \mathbf{v}^T \mathbf{v} + \sum_{k=1}^K \alpha_k^T \mathbf{C} \alpha_k + J_o + J_c, \quad (20.105)$$

where the J_o term is the observational component of the cost function, while J_c is any constraint that we wish to enforce on the minimum of the cost function. It is possible to improve the conditioning of the minimization of the cost function in (20.105) by introducing an α control vector \mathbf{v}^α which is the concatenation of the K vectors \mathbf{v}_k^α such that

$$\alpha_k = \mathbf{U}^\alpha \mathbf{v}_k^\alpha, \quad (20.106)$$

where $\mathbf{U}^\alpha \equiv \mathbf{C}^{\frac{1}{2}}$.

Therefore, the cost function in (20.105) now becomes

$$J(\mathbf{v}, \mathbf{v}^\alpha) = \frac{1}{2} \mathbf{v}^T \mathbf{v} + \frac{1}{2} (\mathbf{v}^\alpha)^T (\mathbf{v}^\alpha) + J_o + J_c. \quad (20.107)$$

NDEnVAR

20.7 NDEnVAR

According to Lorenc [301], the data assimilation systems that have the designate NDEnVAR, where the ND stands for number of dimensions, i.e., 3D or 4D, are variational-based data assimilation systems that only use ensemble covariances, but in the four-dimensional case they do not use the adjoint of the tangent linear models.

A clear mathematical and algorithmic description of the NDEnVAR systems can be found in Desroziers et al. [302], but we shall present the brief description from Lorenc et al. [303].

For this section we shall consider four-dimensional trajectories, which can be seen as a sequence of three-dimensional states describing the evolution over a time window. A standard that has been introduced to represent these 4D fields in the atmospheric community is to denote the vector with an underline. Therefore let $\underline{\mathbf{x}}^b$ be the background trajectory. We then have the expected error covariance of $\underline{\mathbf{x}}^b$ as $\underline{\mathbf{P}}$. This defines a Gaussian PDF for the 4D increment $\delta\underline{\mathbf{x}}$:

$$\delta\underline{\mathbf{x}} \sim G(\mathbf{0}, \underline{\mathbf{P}}), \quad (20.122)$$

which gives the probability that $\underline{\mathbf{x}}^b + \delta\underline{\mathbf{x}}$ is the true trajectory. This means if we follow the usual assumptions for incremental 4D VAR, then we have to find the minimum of the following cost function:

$$J(\delta\underline{\mathbf{x}}) = \frac{1}{2} \delta\underline{\mathbf{x}}^T \underline{\mathbf{P}}^{-1} \delta\underline{\mathbf{x}} + \frac{1}{2} (\underline{\mathbf{y}} - \underline{\mathbf{y}}^o)^T \underline{\mathbf{R}}^{-1} (\underline{\mathbf{y}} - \underline{\mathbf{y}}^o), \quad (20.123)$$

where for the observational component in (20.123), $\underline{\mathbf{y}}^o$ is the observations in the time window, and $\underline{\mathbf{y}}$ is the nonlinear observation operator acting in time given by

$$\underline{\mathbf{y}} - \underline{\mathbf{H}}(\underline{\mathbf{x}}^b + \delta\underline{\mathbf{x}}), \quad (20.124)$$

where $\underline{\mathbf{H}}$ includes time, horizontal, and vertical interpolation, followed by potentially nonlinear calculations.

As we have seen, the incremental 4D VAR system replaces the true state \mathbf{x}^t by $\mathbf{x}^b + \delta\mathbf{x}$ and then makes the assumption

$$\underline{\mathcal{M}}(\underline{\mathbf{x}}^b + \delta\underline{\mathbf{x}}) = \underline{\mathcal{M}}(\underline{\mathbf{x}}^b) + \delta\underline{\mathbf{x}}, \quad (20.125a)$$

$$\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\delta\underline{\mathbf{x}}. \quad (20.125b)$$

Applying a reduction of the control variable technique means that the 4D covariance matrix, which is implicit inside of 4D VAR, is given by

$$\underline{\mathbf{P}} = \underline{\mathbf{M}}\underline{\mathbf{P}}\underline{\mathbf{M}}^T, \quad (20.126)$$

where $\underline{\mathbf{P}}$ is the 3D error covariance matrix at the beginning of the window. If we apply the $\underline{\mathbf{T}}$ and $\underline{\mathbf{U}}$ transforms as we showed earlier in the summary of the Met. Office's En4D VAR system, it is possible to rewrite the cost function in terms of the control variables \mathbf{v} such that $\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\underline{\mathbf{U}}\mathbf{v}$,

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}\mathbf{v}^T + \frac{1}{2}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o)\underline{\mathbf{R}}^{-1}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o). \quad (20.127)$$

The associated 4D error covariance matrix is given by

$$\underline{\mathbf{P}} = \underline{\mathbf{M}}\underline{\mathbf{U}}\underline{\mathbf{U}}^T\underline{\mathbf{M}}^T. \quad (20.128)$$

Ensemble 4D VAR implicitly uses localized ensemble error instead of the climatological covariance \mathbf{B} :

$$\mathbf{P} = \mathbf{C} \circ \mathbf{X}\mathbf{X}^T, \quad (20.129)$$

where $\mathbf{X} = [\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_N]$ is the array of normalized ensemble perturbations valid at the beginning of the window. We apply the α control variable technique

$$\delta \underline{\mathbf{x}} = \underline{\mathbf{M}} \sum_k \alpha \circ \mathbf{x}'_k. \quad (20.130)$$

Each α_k is smooth using the technique $\alpha_k = \mathbf{U}^\alpha \mathbf{v}_k^\alpha$, such that

$$\mathbf{C} = \mathbf{U}^\alpha (\mathbf{U}^\alpha)^T. \quad (20.131)$$

All of the \mathbf{v}_k^α are concatenated into a single vector \mathbf{v} so that the background component of the cost function is transformed into

$$\frac{1}{2} \mathbf{v}\mathbf{v}^T = \frac{1}{2} \sum_k \mathbf{v}_k^\alpha (\mathbf{v}_k^\alpha)^T. \quad (20.132)$$

The important feature to note here is that the error covariance matrix now becomes

$$\underline{\mathbf{P}} = \underline{\mathbf{M}} \left(\mathbf{C} \circ \mathbf{X}\mathbf{X}^T \right) \underline{\mathbf{M}}^T. \quad (20.133)$$

Hybrid 4D VAR is constructed through a weighted average combining the traditional 4D VAR with the ensemble 4D VAR above, such that

$$\underline{\delta \mathbf{x}} = \underline{\mathbf{M}} \left(\beta_c \mathbf{U} \mathbf{v} + \beta_e \sum_{\mathbf{k}} \mathbf{U}^\alpha \mathbf{v}_{\mathbf{k}}^\alpha \circ \mathbf{x}'_{\mathbf{k}} \right). \quad (20.134)$$

Therefore, the error covariance matrix for the hybrid 4D VAR is given by

$$\underline{\mathbf{P}} = \underline{\mathbf{M}} \left(\beta_c^2 \mathbf{B} + \beta_e^2 \mathbf{C} \circ \mathbf{X} \mathbf{X}^T \right) \underline{\mathbf{M}}^T. \quad (20.135)$$

We should note here that the square of the hybrid weights are in the definition above, and assuming that they both give independent valid estimates of \mathbf{P} , then we obtain the condition that $\beta_c^2 + \beta_e^2 = 1$.

4DEnVAR – Four Dimensional Ensemble VAR

The approach presented here is similar to that of Ensemble 4D VAR (En4DVAR) but was designed to **AVOID** the TLM and adjoints. 4DEnVAR extends to four dimensions for all the En4DVAR equations, but now applying them to the sequence of states in trajectories rather than at the single time at the beginning of the window. This implies that we can use an implicit localized ensemble covariance

$$\underline{\mathbf{P}} = \underline{\mathbf{C}} \circ \underline{\mathbf{X}}\underline{\mathbf{X}}^T, \quad (20.136)$$

where $\underline{\mathbf{X}} = [\underline{\mathbf{x}}'_1, \underline{\mathbf{x}}'_2, \dots, \underline{\mathbf{x}}'_N]$ contains the normalized ensemble perturbations valid through the window as $\underline{\mathbf{X}} = \frac{\underline{\mathbf{x}}_k - \bar{\underline{\mathbf{x}}}}{\sqrt{N-1}}$. This method uses the α control variables but now in the form of $\underline{\alpha}_k$, which define the local weight given to each perturbation trajectory, so that

$$\delta \underline{\mathbf{x}} = \sum_k \underline{\alpha}_k \circ \underline{\mathbf{x}}'_k. \quad (20.137)$$

Since each $\underline{\mathbf{x}}'_k$ is a normalized difference of nonlinear forecasts, then (20.137) is a linear combination of nonlinear forecasts, using localized weights. It is a linear function of the $\underline{\alpha}_k$ s and hence is a different kind of linear model to that in (20.125b).

Since each $\underline{\mathbf{x}}'_k$ is a normalized difference of nonlinear forecasts, then (20.137) is a linear combination of nonlinear forecasts, using localized weights. It is a linear function of the $\underline{\alpha}_k$ s and hence is a different kind of linear model to that in (20.125b).

In its initial implementation of 4DEnVAR, the Met. Office does not allow for the variation in time of the $\underline{\alpha}_k$ s, but instead uses a persistence forecast $\underline{\mathbf{I}}$ such that $\underline{\alpha}_k = \underline{\mathbf{I}}\underline{\alpha}_k$. A smoothing technique is still applied to the α control variables. Therefore the 4D incremental trajectory is given by

$$\delta \underline{\mathbf{x}} = \beta_c \underline{\mathbf{I}} \underline{\mathbf{U}} \underline{\mathbf{x}} \underline{\mathbf{v}}^c + \beta_e \sum_{\mathbf{k}} \underline{\mathbf{I}} \underline{\mathbf{U}}^{\alpha} \underline{\mathbf{v}}_{\mathbf{k}}^{\alpha} \circ \underline{\mathbf{x}}'_{\mathbf{k}}, \quad (20.138)$$

and the implicit covariance still contains neither the square of the weights and is given by

$$\underline{\mathbf{P}} = \beta_c^2 \underline{\mathbf{I}} \underline{\mathbf{B}} \underline{\mathbf{I}}^T + \beta_e^2 \underline{\mathbf{C}} \circ \underline{\mathbf{X}} \underline{\mathbf{X}}^T, \quad (20.139)$$

where we can clearly see that the expression in (20.139) neither contains the tangent linear model nor the adjoint.

While this appears to be a great alternative to 4D VAR, most results have shown that while 4DEnVAR beats the static 4D VAR and the hybrid 3D VAR, it does not appear, at the time of writing, to be able to beat En4D VAR. One possible reason, as mentioned in [303], is the fact that the 4DEnVAR does not create very well-balanced increments, but another reason could be the fact that persistence of the static component through the window is not a good choice as it is hard to believe that the large-scale motions would have the same error covariances throughout the length of the window, given the fact that incremental 4D VAR evolves the static component of the background error covariance matrix throughout the window. We know that even the static 4D VAR evolves the static covariance matrix through the Hessian of the cost function. There is a detailed study performed with the NCEP Global Forecasting System with different versions of En3D VAR, En4D VAR, 3DEnVAR, and 4DEnVAR in Kleist and Ide [304,305], and we recommend the reader to these papers for a more detailed explanation of the performances that they discovered of these different configurations of the hybrid systems.

Table 20.1 Summary of the Increments and the Analysis Covariance Matrices for Each Data Assimilation Scheme

Scheme	Increment	Analysis Covariance Matrix
4D VAR	$\delta \underline{\mathbf{x}} = \underline{\mathbf{M}} \delta \mathbf{x}$	$\underline{\mathbf{P}} = \underline{\mathbf{M}} \underline{\mathbf{P}} \underline{\mathbf{M}}^T$
EnVAR	$\delta \underline{\mathbf{x}} = \underline{\mathbf{M}} \sum_k \alpha_k \circ \mathbf{x}'_k$	$\underline{\mathbf{P}} = \underline{\mathbf{M}} (\underline{\mathbf{C}} \circ \underline{\mathbf{X}} \underline{\mathbf{X}}^T) \underline{\mathbf{M}}$
En4D VAR	$\delta \underline{\mathbf{x}} = \underline{\mathbf{M}} (\beta_c \underline{\mathbf{U}} \mathbf{v} + \beta_e \sum_k \underline{\mathbf{U}}^\alpha \mathbf{v}_k^\alpha \circ \mathbf{x}'_k)$	$\underline{\mathbf{P}} = \underline{\mathbf{M}} (\beta_c^2 \underline{\mathbf{B}} + \beta_e^2 \underline{\mathbf{C}} \circ \underline{\mathbf{X}} \underline{\mathbf{X}}^T) \underline{\mathbf{M}}^T$
4DEnVAR	$\delta \underline{\mathbf{x}} = \beta_c \underline{\mathbf{I}} \underline{\mathbf{U}} \mathbf{v}^c + \beta_e \sum_k \underline{\mathbf{I}} \underline{\mathbf{U}}^\alpha \circ \mathbf{x}'_k$	$\underline{\mathbf{P}} = \beta_c^2 \underline{\mathbf{I}} \underline{\mathbf{B}} \underline{\mathbf{I}}^T + \beta_e^2 \underline{\mathbf{C}} \circ \underline{\mathbf{X}} \underline{\mathbf{X}}^T$