A review of operational methods of variational and ensemble-variational data assimilation

Bannister, R. N. (2017) A review of operational methods of variational and ensemble-variational data assimilation. Quarterly Journal of the Royal Meteorological Society, 143 (703). pp. 607-633. ISSN 1477-870X doi: https://doi.org/10.1002/qj.2982 Available at https://centaur.reading.ac.uk/68685/

It is advisable to refer to the publisher’s version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1002/qj.2982

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR
Variational and ensemble methods have been developed separately by various research and development groups and each brings its own benefits to data assimilation. In the last decade or so, various ways have been developed to combine these methods, especially with the aims of improving the background-error covariance matrices and of improving efficiency. The field has become confusing, even to many specialists, and so there is now a need to summarize the methods in order to show how they work, how they are related, what benefits they bring, why they have been developed, how they perform, and what improvements are pending. This article starts with a reminder of basic variational and ensemble techniques and shows how they can be combined to give the emerging ensemble-variational (EnVar) and hybrid methods. A key part of the article includes details of how localization is commonly represented.

There has been a particular push to develop four-dimensional methods that are free of linearized forecast models. This article attempts to provide derivations of the formulations of most popular schemes. These are otherwise scattered throughout the literature or absent. It builds on the nomenclature used to distinguish between methods, and discusses further possible developments to the methods, including the representation of model error.

Key Words: variational data assimilation; ensemble data assimilation; hybrid data assimilation; flow-dependent background-error covariances; localization; model error; nomenclature

1. Data assimilation and uncertainty

Dealing with uncertainty is at the heart of data assimilation (DA). Forecast models (here those used in numerical weather prediction, NWP) use initial conditions that are imperfect, and are based upon imperfect representations of physical processes. It is well known that a free-running model will accumulate errors until its forecast is no longer useful (Tribbia and Baumhefner, 2004). The only way to restore usefulness is to allow the model to be influenced by observations (Leith, 1993). DA (Daley, 1991; Kalnay, 2002; Rabier, 2005) is the procedure of maintaining the link between evolving models and reality by updating the model with fresh observations. Researchers are working towards formal and robust mathematical methods that work in ways that are consistent with the model, the data, and their degree of uncertainty. DA is related to approaches used in other fields and which go by different names, e.g. state estimation (Wunsch, 2012); optimization (Biegler, 1997); history matching (Emerick, 2012); retrieval production (Rodgers, 2000); inverse modelling (Tarantola, 2005), and there are additionally many ways of solving a DA problem. Most known DA methods are based on probabilistic theories (most, if not all, exploit Bayes' Theorem (Lorenc, 1986)) and each is made practical by making approximations.

Traditionally there are three Bayesian-based strategies that allow the DA problem to be solved in approximate (hence sub-optimal) ways. These may be categorized as the following.

(i) Variational DA (Var, or its specific implementation 3D-Var or 4D-Var) implements an algorithm to minimize a cost function. In its basic form, Var gives a single, (quasi-) optimal analysis state based on an a priori state (the background or forecast), some observations, and the prescribed Gaussian uncertainty statistics for the background and observations. Var requires linearized versions of the observation operators and 4D-Var requires a linearized version of the forecast model.

(ii) Ensemble DA, based on the ensemble Kalman filter (EnKF), offers a priori error statistics from an ensemble instead of from a prescribed source, which changes as the system evolves, and does not require use of linearized models.

(iii) Monte-Carlo methods additionally allow the assimilation of information from sources that have non-Gaussian errors.
Approaches (i) and (ii) have been successful in NWP, but have limitations (sections 1.2 and 1.3). In recent years this has led to an explosion of techniques that combine them in the hope of maximizing the benefits and eliminating the inadequacies of the separate methods. This includes so-called ‘pure EnVar’ schemes (section 1.5) and ‘hybrid EnVar’ schemes (section 1.6). The modern DA toolbox now includes a spectrum of methods that are often subtly different in their name, but may be profoundly different in their treatment of uncertainty (especially in nonlinear systems), in their efficiency, and in the way that the equations are solved (e.g. Buehner et al., 2015b; Liu and Xue, 2016). The boxes in Figure 1 represent the main methods, including Var, ensemble, pure EnVar and hybrid EnVar methods and how they are linked (the Figure caption gives details). The suggested nomenclature to distinguish the methods (Lorenc, 2013) is shown and the Figure will be explained throughout the article.

The aim of this article is to firstly remind ourselves of the basic forms of (i) and (ii), to then review the methods that combine them (pure EnVar and hybrid EnVar), and to summarize the nomenclature. The similarities, differences, advantages, disadvantages and successes of each new method are reviewed. It is beyond the scope of this article to review in depth the numerous ways of implementing each traditional method (such as parallelizable implementations of 4D-Var, and the many implementations of the EnKF such as the stochastic and square-root formulations (Ehrendorfer, 2007; Houtekamer and Zhang, 2016). A review of (iii) is given by van Leeuwen (2009).

The structure of this article is as follows. In the remainder of this section we outline the basic principles of each traditional and combined approach, and mention the advantages, disadvantages and limitations of each. In section 2 we derive common forms of pure variational methods. In section 3 we review the basic principles of pure EnKF-related methods. In section 4 we review how DA systems can be formulated by replacing the fixed background-error statistics of Var with ensemble-derived statistics, while maintaining the variational framework. Broadly, there are two approaches – one that requires the linear models of 4D-Var and another that does not. In section 5 we review hybrid methods, which are DA systems that merge the prescribed and ensemble-derived background-error covariances, again in ways that may or may not require linearized models. For reasons of simplicity, up to this point in this article the inevitable
ensemble sampling noise is not considered. In section 6 we review state-space localization methods which reduce sampling noise. In section 7 we give localized forms of the methods discussed in sections 3–5, and we present a summary of accepted nomenclature. In section 8 we discuss how they could handle model error. In section 9 we compare how they perform in practice and in section 10 we summarize which schemes are used at leading centres and groups. In section 11 we provide a brief summary and some closing comments. In the appendices we give some important derivations which may help understanding of the theory and efficient implementation of the new assimilation schemes.

1.1. Data assimilation challenges

The objective of DA is to produce information about the posterior probability density function (PDF). This is produced as a first moment (the analysis), or as information about the second moment (e.g. from an ensemble). These can be used as initial conditions for weather forecasts. Observations influence the posterior by matching synthetic versions of the observations (found from the model state via observation operators) to the real observations. This process depends on the nature of each observation (whether it is in situ or remotely sensed) and the time of each observation relative to the analysis time. Most observations nowadays are asynoptic (or asynchronous), meaning that respecting the detailed temporal nature of the problem is essential. This problem poses many challenges, including the following.

- Analyses must be produced even in unobserved regions and times, hence the need for prior information (from a background forecast or an ensemble).
- All data are imperfect and uncertainties between different pieces of data can be correlated. The statistics of this information are needed for DA to work well, so uncertain data, for example, should be given lower weight than more precise data. In Var the background- and observational-error covariance matrices are specified and in the EnKF the background-error covariance matrix is estimated from the ensemble. The background-error covariances can significantly affect the analysis (e.g. Navon et al., 2005) and if they are incorrectly specified then the DA is sub-optimal. An incorrect background-error covariance matrix can actually lead to analyses being worse than the prior (Morr and Emanuel, 2002).
- All NWP models are imperfect. This is accounted for in weak-constraint 4D-Var (section 2.1) and in EnKF methods that include stochastic processes in the NWP model (section 3.2).
- Many modes can be excited by DA; these are associated with processes of very different space/time-scales. Modifying the right modes is an issue that goes right back to Richardson’s forecast in the 1920s. The use of 4D methods allows an appropriate background-error covariance matrix, and initialization (Temperton, 1988) all help with this issue.
- All DA problems for NWP are large and so methods must be able to deal with huge numbers of degrees of freedom (typically \( \geq 10^{14} \)) pieces of prior information and similar numbers of observations. Building DA systems that are practicable is often the task that takes the most effort.

The methods dealt with in this review are designed to deal with these points in different ways.

1.2. Variational data assimilation

Variational DA (Talagrand and Courtier, 1987; Schlatter, 2000) is a tool used to estimate a single initial state and a single trajectory of an NWP model.* These are found by minimizing a cost function to optimize the fit of (i) the initial conditions to the background and (ii) the model’s version of the observations to the actual observations over a time window.

Variational methods have been used in NWP for many years (Courtier et al., 1994; Park and Zupanski, 2003; Rawlins et al., 2007) and they allow assimilation of a wide range of observations, including those remotely sensed (e.g. from satellite and radar). Variational methods are restricted in their ability to quantify adequately the flow-dependent background and model error statistics, which are prescribed in Var by a parametrized scheme (Trémolé, 2007; Bannister, 2008). Such schemes will almost inevitably be incapable of representing properly the true space and time structure of error statistics. The Var equations are derived in section 2. In 4D-Var the background-error covariance statistics do evolve with the flow within each window, and weak constraint formulations can account for model errors. Normally only the first moment (the mode) of the analysis is found and 4D-Var relies on linearized and adjoint versions of the numerical model.

1.3. Ensemble data assimilation

Instead of specifying a covariance model, the EnKF (Evensen, 1994; Hamill, 2006; Ehrendorfer, 2007; Hougetekmer and Zhang, 2016) uses an ensemble of possible forecasts (section 3) that contains valuable flow-dependent information about the background-error statistics. The EnKF is designed to produce an ensemble of analyses, which will similarly contain information about the analysis-error statistics, and can be cycled to future analysis times. The EnKF does not require linearized or adjoint versions of the model, nor of the observation operators. One of the first centres to use an EnKF operationally was Environment Canada (Buehner et al., 2010a, 2010b).

Due to cost, the number of ensemble members is too small in practice. Error statistics found from small ensembles will suffer sampling error, and so the statistics will not necessarily be representative of the state. This shows itself in the form of ensemble-derived correlations that are noisy, and variances that are systematically too small (van Leeuwen, 1999; Hougetekmer and Mitchell, 2001). The issue with the correlations is problematic when errors between two variables are only weakly correlated in reality (e.g. variables that are widely separated in space).† The issue with the variances will cause the DA to believe that the forecast ensemble is more accurate that it really is. This will place too much trust in the ensemble in comparison with the observations. Left to its own devices, the ensemble will collapse effectively to a single trajectory and the observations will be ignored. This problem in DA is called ‘filter divergence’ (Hougetekmer and Mitchell, 1998). These problems can be lessened with a procedure called localization (which dampens the sample covariances at long range; Gaspari and Cohn, 1999) and ensemble inflation (which artificially increases the spread of an ensemble; Anderson and Anderson, 1999; Hamill et al., 2001) by either multiplicative or additive processes (Kalnay et al., 2007; Whitaker et al., 2008). The ad hoc nature of these fixes is largely unsatisfactory.

*Systems that deal with just one realization are often colloquially referred to as ‘deterministic’ as there is no information about alternative realities (as opposed to systems that deal with ensembles which are referred to as ‘probabilistic’). This is a different use of the term ‘deterministic’ to that used in the theory of dynamical systems. Note that, confusingly, ‘deterministic’ is also used to describe flavours of the EnKF which do not perturb observations (as opposed to ‘stochastic’ which is used to describe flavours of the EnKF which do – section 3.2.

†The expected root-mean-square error in the variance derived from an \( N \)-member ensemble is \( \sigma^2/\sqrt{N} \), where \( \sigma^2 \) is the true variance. The expected root-mean-square error in the correlation between two variables is \( (1 - \rho^2)/\sqrt{N} \) (Hougetekmer and Mitchell, 1998), where \( \rho \) is the true value of the correlation. Both errors can be large when \( N \) is small and the latter error can be large also when \( \rho \) is small.
although methods have been investigated that use adaptive (rather than imposed) moderation functions (Bishop and Hodysy, 2007, 2009a, 2011) and inflation factors that are based on innovation statistics (Bowler et al., 2008).

The EnKF is applied sequentially. The Ensemble Kalman Smoother (EnKS) on the other hand can assimilate observations at times later than the nominal analysis time (Evensen and van Leeuwen, 2000). Elements of the EnKF and the EnKS are covered in section 3.

1.4. Non-Gaussian Monte-Carlo assimilation

Monte-Carlo techniques form a range of methods that include the particle filter (PF). The PF represents a probability distribution by an ensemble (the particles) but, unlike in the EnKF, non-Gaussianity of the distribution is represented. PFs do suffer from degeneracy problems (where the filter ignores all but one particle). For this reason, PFs have to be implemented in special ways which are not used currently for operational purposes (van Leeuwen, 2009). PFs are not considered further in this article.

1.5. Combined 4D-Var/EnKF assimilation schemes (‘EnVar’)

Most of this article is about combining Var with the EnKF. Such schemes usually run a Var and an EnKF scheme in parallel with information being exchanged between them, which can be one-way or two-way.

One-way coupling can be achieved by shifting each EnKF analysis member for the ensemble mean to equal the Var analysis (‘re-centring’). In this configuration the EnKF does not transmit information to Var. This kind of coupling has been used, for example, at the Met Office (Bowler et al., 2008).

Another one-way coupling may be achieved by using the ensemble as a source of background-error covariance information (instead of or in addition to climatological covariances) in a Var system to give some flow dependency. This is the basis of En3/4DVar and 3/4DEnVar methods (section 4). Two-way coupling may be achieved, e.g. by combining En3/4DVar or 3/4DEnVar with re-centring.

1.6. Hybrid data assimilation

Var is suboptimal partly due to the unrealistic imposed error covariance matrix, and the EnKF is suboptimal partly due to under-sampling. Hybrid DA has been developed to accentuate the best features of each source of background-error information, namely the statistically robust background-error covariances of Var and the flow-dependency of the EnKF. Mathematical approaches have been developed to do this, and most are solved with a (deterministic) Var-style method but with added ensemble information. The simplest solves a Var-like problem, but replaces the static background-error covariance matrix with a weighted average of itself with the error covariance matrix from the ensemble. Efficient methods have been developed to deal with the very large matrices used in real problems, while at the same time attending to the ensemble sampling error problem (section 5).

2. Pure variational data assimilation

Let the $n$-element vectors $x(t)$ and $x^b$ represent a state of the model at time $t$ and the background state at time $0$ respectively (quantities with no time argument are implicitly at $t = 0$, and underlined quantities will represent augmented quantities throughout the time window). State $x(t)$ is found by propagating $x(t-1)$ forward by one time step $\delta t$ by the model, $M_{t-1,t}$:

$$x(t) = M_{t-1,t}[x(t-1) + \eta(t)],$$

(1)

where the $n$-element perturbation $\eta(t)$ is to offset model error introduced over $t-1 \rightarrow t$. Let the $p_i$-element vectors $y^t_i$ and $y^{x}_i$ be the set of observations made at time $t$, and their modelled counterparts respectively. $y^t_i$ is related to $x(t)$ via the observation operator $H_i$:

$$y^t_i = H_i(x(t)).$$

(2)

2.1. 4D-Var cost function

2.1.1. A full-form (non-incremental) cost function

Var has been the workhorse of DA for many years (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987; Thepaut and Courtier, 1991; Zupanski, 1997; Rabier et al., 2000; Rawlins et al., 2007). The 4D-Var problem may be posed as a cost function, which is a functional of the $T-1$ states in $ig = \{x(t), \ldots, x(T)\}$ (over the time window $t = 0$ to $T$) and then varying $x$ to minimize $J$ (Zupanski, 1997):

$$J(y) = \frac{1}{2} \left\| x - x^b \right\|_{R_0}^2 + \frac{1}{2} \sum_{t=0}^{T} \left\| x(t) - H_i(x(t)) \right\|_{Q_i}^2$$

(3)

$$+ \frac{1}{2} \sum_{t=1}^{T} \left\| x(t) - M_{t-1,t}[x(t-1)] \right\|_{Q_1}^2,$$

where $\left\| a \right\|_{A^{-1}}^2 = a^T A^{-1} a$.

The interpretation of this cost function is as follows.

- The first term (the background term, $h_0$) measures the deviation between $x$ and $x^b$. This is calculated in the $L^2$ norm described by the $n \times n$ background-error covariance matrix $B_0$.
- The second term (the observation term, $J_o$) measures the deviation between $y^t_i$ and $y^{x}_i$. This is calculated in the $L^2$ norm described by the $p_i \times p_i$ observation-error covariance matrix $R_i$ and is summed over the window. Equation (3) does not allow for observations that depend on the model state at more than one time (e.g. precipitation accumulation which only depends upon earlier states), although it may be adapted accordingly.
- The third term (the model error term, $J_0$) measures the deviation between $x(t)$ and $M_{t-1,t}[x(t-1)]$. This is calculated in the $L^2$ norm described by the $n \times n$ model-error covariance matrix $Q_0$. In this case we have assumed that model errors are uncorrelated in time (white noise approximation) and have a priori values of zero.

This minimization may be regarded as an inverse problem, which can be solved given the ability to solve the forward problems ($M_{t-1,t}$ and $H_i$) and then minimizing $J(y)$. Cost function (3) is weak-constraint, meaning that $x$ does not have to be exactly a model trajectory (Sasaki, 1970; Trémolet, 2006). This means that we can formally account for imperfect models in the DA problem.

A common approximation follows by setting $Q_0 = 0$ which forces the states to follow a model trajectory exactly. This is called strong constraint 4D-Var, and such systems have been used operationally for some years (e.g. Rabier et al., 2000; Rabier, 2005; Gauthier et al., 2007; Rawlins et al., 2007). In this approximation only the state at $t=0$ needs to be determined as all subsequent states follow from the model equations ($(1)$ with $\eta = 0$), and the $J_0$ term no longer appears.

© 2016 The Authors. Quarterly Journal of the Royal Meteorological Society published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.
In general, an underlined symbol represents that quantity at all relevant times over the assimilation window, and a δ prefix in brackets indicates that the quantity may be represented as a perturbation.

### 2.1.2. An incremental cost function

Cost function (3) is a full-fields approach to DA. A more practical approach is incremental DA, which deals with perturbations made to known reference states (Courtier et al., 1994) and has some advantages over full-fields DA.

The cost function has a quadratic form if \( \mathcal{M}_{t_0,t_2} \) and \( \mathcal{H}_t \) are linear (which they are not in general). In practical terms, a non-quadric cost function is difficult to minimize, especially if it possesses multiple local minima. A quadratic problem emerges when realizing \( \mathcal{M}_{t_0,t_2} \) and \( \mathcal{H}_t \) about a guess (or reference) state, \( \bar{x}^b \), and formulating the problem in terms of perturbations (increments) to \( \bar{x}^b \). In this article we will take \( \bar{x}^b = \bar{x}^o \) for simplicity. The result is a modified cost function which is an approximation to the original but is easier to minimize.

The 4D state space \( \bar{x}^o \) is found by integrating the model over the assimilation window using (1) with the assumption that \( \eta(t) = 0 \):

\[
\bar{x}^b(t) = \mathcal{M}_{t_0,t_2} \left( \bar{x}^b(t-1) \right) .
\]  

This can be perturbed to give a general state \( x(t) \) as follows:

\[
x(t) = \bar{x}^b(t) + \delta x(t),
\]

\[
\eta(t) = \delta \eta(t).
\]

### Table 1. Summary of key notation used in this article, where \( n \) is the size of the state space, \( p \) is the total number of observations over the assimilation window (\( p_t \) at time \( t \)), \( T \) is the number of time steps, \( N \) is the number of ensemble members, and \( W_t \) here means ‘all times’.

| Symbol | Description | Size |
| --- | --- | --- |
| \( a \) | \( A^{-1} \) | \( a \times \) |
| \( \tilde{b} \) | \( x(t), \dot{x}, \xi \) | \( n, n, (T+1)n \) |
| \( \eta(t) \) | Model error: \( t \rightarrow \) \( \nu \) | \( n, N \) |
| \( \mathcal{M}_{t_0,t_2} \) | Nonlinear model: \( t_1 \rightarrow t_2 \) | \( n \times n \) |
| \( \mathcal{M} \) | Linear model: \( 0 \rightarrow \nu \) | \( n, n, (T+1)n \) |
| \( \mathcal{H}_t \) | Nonlinear observation operator: \( t \) | \( p \times (T+1)n \) |
| \( \mathcal{H} \) | Linear observation operator \( \nu \) | \( p 	imes n \) |
| \( \mathcal{J}_{var} \) | Perturbation to the state in control space: \( t \) | \( n \times n \) |
| \( \mathcal{J}_{ens} \) | Background-error covariance matrix \( t \) | \( p_t \times p_t \) |
| \( \mathcal{J}_{P} \) | Observation-error covariance matrix: \( t \) | \( n \times n \) |
| \( \mathcal{J}_{P} \) | Model-error covariance matrix: \( t \) | \( n \times n \) |
| \( \mathcal{J}_{P} \) | Cost function, or term therein | \( n \times n \) |
| \( \mathcal{J}_{P} \) | Innovation: \( t \), \( \nu \) | \( n \times n \) |
| \( \mathcal{J}_{P} \) | Control variable transform for the state | \( n \times n \) |
| \( \mathcal{J}_{P} \) | Control variable transform for model error: \( t \) | \( n \times n \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to variational, ensemble, hybrid components | \( n, N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to ensemble, hybrid components | \( N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to variational, ensemble, hybrid components | \( n, N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to ensemble, hybrid components | \( N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to variational, ensemble, hybrid components | \( n, N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to ensemble, hybrid components | \( N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to variational, ensemble, hybrid components | \( n, N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to ensemble, hybrid components | \( N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to variational, ensemble, hybrid components | \( n, N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to ensemble, hybrid components | \( N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to variational, ensemble, hybrid components | \( n, N, n+N \) |
| \( \mathcal{J}_{xvi} \) | Perturbation to the state in control space: \( t \) to ensemble, hybrid components | \( N, n+N \) |

Substituting these into (1) gives:

\[
x^b(t) + \delta x(t) \approx \mathcal{M}_{t_{0,t-1}} \left[ x^b(t-1) \right] + M_{t_{0,t-1}} \delta x(t-1) + \delta \eta(t),
\]

\[
\therefore \delta x(t) \approx M_{t_{0,t-1}} \delta x(t-1) + \delta \eta(t),
\]

Repeating use of (7) leads to:

\[
\delta x(t) \approx M_{t_{0,t}} \delta x(0) + M_{t_{1,t}} \delta \eta(1) + M_{t_{2,t}} \delta \eta(2) + \ldots + M_{t_{t-1,t}} \delta \eta(t-1) + \delta \eta(t),
\]

\[
= M_{t_{0,t}} \delta x(0) + \sum_{\tau=1}^{t} M_{t_{\tau,t}} \delta \eta(\tau),
\]

noting that \( M_{t,t} = I \). The linearized form of (2) is found in a similar way:

\[
\mathcal{H}_t(\mathbf{x}(t)) \approx \mathcal{H}_t \left[ x^b(t) \right] + H \left[ M_{t_{0,t}} \delta x(0) + \sum_{\tau=1}^{t} M_{t_{\tau,t}} \delta \eta(\tau) \right],
\]
where $H_t = \partial y^T_t / \partial x(t)$. Substituting (9) into (3) gives the incremental cost function:

$$J_{\text{inc}}(\delta x, \delta y) = \frac{1}{2} \| \delta x \|^2_{R_t^{-1}} + \frac{1}{2} \sum_{t=0}^{T} \| \delta d_t \|^2_{Q_t^{-1}}$$

$$+ \frac{1}{2} \sum_{t=1}^{T} \| \delta y(t) \|^2_{Q_t^{-1}},$$

where $\delta d_t = d_t - H_t (M_{t-1,0} \delta x + \sum_{j=1}^{t} M_{t,j} \delta y(\tau))$.

and $d_t = y^*_t - H_t \chi^b(t)$.

These equations represent incremental weak-constraint 4D-Var (1) in Figure 1). Minimizing $J_{\text{inc}}$ (over an inner loop) gives particular values of $(\delta x, \delta y) = (\delta x, \delta y(1), \ldots, \delta y(T))$. The strong-constraint 4D-Var version is obtained by setting all $\delta y(t) = 0$ in (10) and (11), which is valid when model error is negligible over the window (2) in Figure 1).

For practical reasons Var uses a $B_0$-matrix (and weak-constraint 4D-Var uses $Q_t$ matrices) that is (are) essentially the same for each cycle, although flow-dependence can be added artificially to the way that $B_0$ is modelled by, for example, cycling variance information (Dee, 2002) and the use of (linearized) nonlinear balance relationships (Fisher, 2003; Barker et al., 2004).

2.1.3. Further simplifications

Strong-constraint 4D-Var may be further simplified to exclude part of the time dimension (Courtier et al., 1998; Gauthier et al., 1999; Lorenc et al., 2000; Gustafsson et al., 2001). The most popular is called 3D-FGAT (3D-Var with the First Guess at the Appropriate Time, but also known as just 3D-Var), which emerges from the strong-constraint version of (10) with the assumption that $M_{t-1,0} \approx I \forall t, t_2$ (2) in Figure 1):

$$J_{\text{inc}}^{3D-FGAT}(\delta x) = \frac{1}{2} \| \delta x \|^2_{R_t^{-1}} + \frac{1}{2} \sum_{t=0}^{T} \| \delta d_t \|^2_{Q_t^{-1}},$$

$$\delta d_t = d_t - H_t \delta x.$$  (13)

(Lee et al., 2004; Lawless, 2010), where $\delta x$ is here valid at the centre of the observation window. The 3D aspect is due to the loss of the time dimension through the neglect of all $M_{t-1,0}$ and the FGAT aspect is due to the preservation of time in $\delta d_t$ (12) still holds for $d_t$). Notably, 3D-FGAT was used to produce the ERA-40 reanalysis (Uppala et al., 2005).

Note that even operational 4D-Var systems in practice use approximations, e.g. the Jacobian $M_{t-1,0}$ does not normally contain linearizations of the complete model, as many physical processes are problematic to linearize (Xu, 1996).

2.2. The derivative of the 4D-Var cost function

The minimum of $J_{\text{inc}}$ is found iteratively (e.g. Liu and Nocedal, 1989) with algorithms that require calculation of the gradient of (10), $\nabla J_{\text{inc}}$. This is a column vector of the same structure as the argument of $J_{\text{inc}}$ in (10), i.e. $(\delta x, \delta y)$, but of derivatives as follows:

$$\nabla J_{\text{inc}} = \begin{pmatrix} \nabla J_{\text{inc}}^x \\ \nabla J_{\text{inc}}^{y(1)} \\ \vdots \\ \nabla J_{\text{inc}}^{y(T)} \end{pmatrix}$$

where the $i$th component $[\nabla J_{\text{inc}}^x]$ of the sub-vectors are:

$$[\nabla J_{\text{inc}}^x] = \partial J_{\text{inc}} / \partial [\delta x],$$

and $[\nabla J_{\text{inc}}^{y(1)}] = \partial J_{\text{inc}} / \partial [\delta y(1)]$, and

$$[\nabla J_{\text{inc}}^{y(T)}] = \partial J_{\text{inc}} / \partial [\delta y(T)].$$

There are $n$ components of each sub-vector, so $\nabla J_{\text{inc}}$ has $(T+1)n$ components in total. (For strong-constraint 4D-Var and 3D-Var, elements exist with respect to only $\delta x$ in (14) so $\nabla J_{\text{inc}}$ has just $n$ components.) The gradient is important because in phase space $-\nabla J_{\text{inc}}$ gives the direction of greatest descent of $J_{\text{inc}}$. Components of $\nabla J_{\text{inc}}$ are derived in Appendix A to be:

$$\nabla J_{\text{inc}}^x = R_t^{-1} \delta x - \sum_{t=0}^{T} M_{t,0} H_t^T H_t^{-1} \delta d_t,$$

$$\nabla J_{\text{inc}}^{y(\tau)} = - \sum_{t=1}^{T} M_{t,\tau} H_t^T H_t^{-1} \delta d_t + Q_t^{-1} \delta y(\tau).$$

4D-Var is limited in its scope for computational parallelizability. The estimation of $\delta x$ and $\delta y(t)$ as in formulation (10) – cannot be parallelized temporally (e.g. for $\delta y(t)$: this requires integration of the forward linear model from earlier times in (11), and the integration of the adjoint model from future times in (16)) – there are alternative formulations though that may allow parallelization of 4D-Var in time (M. Fisher, 2011; personal communication; Desroziers and Berre, 2012).

2.3. Control variable transforms in Var

The cost function (10), with definitions (11) and (12), requires the matrices $B_0, R_t, Q_t$ to be known explicitly, but even modern computers are incapable of dealing with such large matrices. The method of control variable transforms (CVTs) is a ‘trick’ to represent $B_0$ and $Q_t$ without needing to know them explicitly. Instead of dealing with (10) – a function of $(\delta x, \delta y)$ – an alternative cost function is formulated in terms of new variables (‘control variables’). There are number of key studies that describe the choice of control variables used for meteorological DA (e.g. Parrish and Derber, 1992; Derber and Bouttier, 1999; Berre, 2000), for ocean DA (Weaver et al., 2005), and in general (Bannister, 2008; Ménetrey and Auligné, 2015).

The following transformations may be made for $\delta x$, and $\delta y(t)$ in (10):

$$\delta x = U \delta x_{\text{var}},$$

$$\delta y(t) = V_t \delta y_{\text{var}}(t).$$

Here $\delta x_{\text{var}}$ and $\delta y_{\text{var}}(t)$ are ‘control variables’, which are associated with the model space variables $\delta x$ and $\delta y(t)$ respectively, and $U$ and $V_t$ are the CVTs. Vectors $\delta x_{\text{var}}$ and $\delta y_{\text{var}}(t)$ need not have $n$ elements each, but for the purposes of this article we shall assume that each is an $n$-element vector. Substituting (17) and (18) into (10) gives the pre-conditioned cost function:

$$J_{\text{inc}}^{\delta x_{\text{var}}, \delta y_{\text{var}}}(\delta x_{\text{var}}, \delta y_{\text{var}}) = \frac{1}{2} \| \delta x_{\text{var}} \|^2_{R_t^{-1} U^T} + \frac{1}{2} \sum_{t=0}^{T} \| \delta d_t \|^2_{Q_t^{-1}},$$

and substituting into (11) gives:

$$\delta d_t = d_t - H_t (M_{t,0} U \delta x_{\text{var}} + \sum_{t=1}^{T} M_{t,t} V_t \delta y_{\text{var}}(\tau)).$$

\footnote{The $R_t$ matrices are often assumed diagonal. Over all time steps, there are $p = \sum_{t=0}^{T} p_t$ observations in total, which is a practical number of diagonal elements. There is evidence though that the correlations between errors (including representivity) of different observations should be accounted for (Stewart et al., 2014; Waller et al., 2014).}
Consider a population of uncertainty with an ensemble. We show briefly how the EnKF (EnKF) circumvent this problem by representing background error statistics change in time. Ensemble methods (e.g. the start of each cycle), even though in reality background-error from the previous analysis, generically implied $U_i\varphi (n)$ can be cycled to the next observation time. The key idea is that (28) does not need the explicit $n \times n$ matrix $P_{\text{bg}}^{(n)}$. Equation (28) is a stochastic formulation of the EnKF (Houtekamer and Mitchell, 1998; Burgers et al., 1998) – so-called because the observations are perturbed stochastically to give the correct spread of the analysis ensemble. Other important variations of the EnKF (e.g. the ensemble transform Kalman filter) use deterministic formulations (Anderson, 2001; Bishop et al., 1998; Whitaker and Hamill, 2003) which do not require perturbed observations. (It is outside the scope of this article to review the many flavours of ensemble filters.) The EnKF is $\mathcal{E}$ in Figure 1.

The last line of (28) is written so that it does not require the adjoint of the observation operator. The linear operation $H x^b$ is also often approximated by the nonlinear operations $\{H(x^{b,i}) - H(\bar{x})\}/\sqrt{N-1}$ (approximating the $i$th column of $H X_b$). The EnKF can therefore be implemented without the need for linearized operators (this includes the linearized forecast model in the case of the ensemble smoother – section 3.3). This represents a significant reduction of development effort compared to Var (Lorenc, 2003). Many studies show that the EnKF is a useful tool in NWP, often comparable or better than Var (e.g. Houtekamer et al., 2005; Zhang et al., 2006; Meng and Zhang, 2007; Bonavita et al., 2008; Meng and Zhang, 2008a, 2008b; Szunyogh et al., 2008, Whitaker et al., 2008; Hamill et al., 2011a, 2011b; Zhang et al., 2011, Houtekamer et al., 2014). The differences between EnKF and Var can be most evident in data-sparse regions (Whitaker et al., 2008).

The background-error covariance matrix in the ordinary KF is assumed to include predictability errors (propagation of analysis error from the previous analysis, generically $M P M^T$) and a model error contribution ($Q$). Since the model producing the $x^{b,i}$ is imperfect, model error should be allowed to affect the spread of the ensemble background states (the system simulation approach of Houtekamer et al., 1996). There are many strategies for accounting for model error, including covariance inflation (Whitaker et al., 2002) and additive error (Houtekamer et al., 2005). These approaches have been tested in an EnKF with a simplified primitive-equation model to represent the model error introduced due to unresolved scales (Hamill and Whitaker, 2005). However the difficulty is choosing the degree of inflation and the structure of the additive error. Hamill and Whitaker (2005) chose to prescribe inflation factors a little over unity (although more advanced approaches are possible, e.g. Raynaud et al., 2012), and they chose additive errors from three options: random samples of differences between forecasts of different resolutions, random samples of differences between forecasts and climatology, and random samples of forecast tendencies. It is also possible to relax the analysis ensemble spread closer to the spread of the forecast ensemble (Zhang et al., 2004).
3.3. Ensemble smoothers

In the EnKF, \( x^b_{t|0}, y^b_{t|0} \) and \( x^b_t \) are valid at \( t = 0 \), but information is propagated from one time to the next via cycling (sequential DA). An EnKS (Evensen, 2007) on the other hand introduces the time dimension into the analysis problem itself so it can deal directly with observations spanning a time window \( (0 \leq t \leq T) \). We discuss two ways of implementing an EnKS, which later influence how EnVar is developed.

3.3.1. An EnKS with a 3D state vector

The EnKF (28) may be adapted by keeping its 3D state vector (valid at the start of the window), but incorporating the forecast model into the observation operator (as in 4D-Var) for observations made at \( t > 0 \):

\[
x^b_{t|0} = x^b_{0|0} + x^b \left[ M_x X^b \right]^T \left[ R + M_x X^b \left( M_x X^b \right)^T \right]^{-1} d^b_{t|0},
\]

where \( d^b_{t|0} = y^b_{t|0} - \mathcal{H}(x^b_{t|0}) \) (see below). Objects (not underlined) are for \((t = 0) \ t \geq 0; \ y^b_{t|0} \) contains the perturbed observations (error covariance \( R \) over the window in chronological order, \( \mathcal{H}(x^b_{t|0}) \) is the time-distributed observation operator, and \( M_x \) is its Jacobian:

\[
\mathcal{H}(x^b_{t|0}) = \begin{pmatrix} H_0(x^b_{t|0}) \\ \vdots \\ H_T(x^b_{t|0}) \end{pmatrix},
\]

\[
H = HM = \begin{pmatrix} H_0 \ 0 \\ 0 \\ \vdots \\ H_T M_{0,T} \end{pmatrix},
\]

where \( \mathcal{H}_i(\bullet) \) means use the quantity to the right as an argument and \( \mathcal{H}_i \) and \( H_i \) are defined in section 2. The above serves to define the operators \( \mathcal{H}, \mathcal{M}, \mathcal{H}, \) and \( \mathcal{M} \). Furthermore the linear step \( H_M X^b \) can be avoided by approximating with the nonlinear operations \( \left[ \mathcal{M}(x^b_{t|0}) - \mathcal{M}(x^b) \right]/\sqrt{N-1} \) (approximating the \( i \)th column of \( H_M X^b \)). This EnKS is \( \mathcal{S} \) in Figure 1.

3.3.2. An EnKS with a 4D state vector

An EnKS may be alternatively formulated with 4D state vectors (van Leeuwen and Evensen, 1996), contained in the \((T + 1) n \times N\)-element matrix \( X^b \) (the \( i \)th column being \( \delta x^b_{t|0}/\sqrt{N-1} \), where \( \delta x^b_{t|0} \) is the 4D analogue of (27)):

\[
X^b = \begin{pmatrix} X^b(0) \\ \vdots \\ X^b(T) \end{pmatrix} = MX^b,
\]

and a similar ensemble exists for the analysis, \( \bar{X}^b \). An EnKS can be derived by underlining all quantities (e.g. \( X^b \rightarrow \bar{X}^b \)) in (28):

\[
\bar{x}^b_{t|0} = \bar{x}^b_{0|0} + X^b (H H^b)^T \left[ R + H X^b (H H^b)^T \right]^{-1} \bar{d}^b_{t|0},
\]

where now \( \bar{d}^b_{t|0} = y^b_{t|0} - H (\bar{x}^b_{t|0}) \), and \( H (\bar{x}^b_{t|0}) \) and \( H \) are defined in (30) and (31). Again the linear step \( H X^b \) can be avoided with the approximation \( \left( H (\bar{x}^b_{t|0}) - H (x^b) \right)/\sqrt{N-1} \) (approximating the \( i \)th column of \( H X^b \)). This EnKS is \( \mathcal{S} \) in Figure 1. An EnKS may also be posed in a formulation which considers observations in batches over the window (Evensen and van Leeuwen, 2000).

The EnKS formulations with 3D and 4D state vectors will be important in the next section when they form the basis of methods that we shall call En4DVar and 4DEnVar respectively.

4. Pure EnVar: combining an ensemble with Var

The last decade has seen experimentation with combining the EnKF with Var. Perhaps the simplest means of doing this is to use a Var analysis to recentre the EnKF analysis ensemble. This (one-way) coupling has been used by Zhang et al. (2009) with the simple Lorenz 96 model (Lorenz, 1996), and by Bowler et al. (2008) with MOGREPS (the Met Office Global and Regional Ensemble Prediction System).

However, in this section we look at one-way coupling the other way round and review two (related) methods of using the EnKF ensemble forecasts to define the background-error statistics in Var (collectively known as ‘EnVar’). We focus on 4D methods (3D methods follow in the same way as in section 2.1.3). The two methods differ in the use – or not – of the linear/adjoint models as used in 4D-Var. Following the nomenclature recommended by Lorenc (2013), we call a pure EnVar system that uses a linear model ‘En4DVar’ (the labels ‘E4DVar’, ‘4DVarBen’, and ‘4DVarBenkf’ have also been used by authors). In all such labels, ‘4DVar’ appears, indicating that it uses the same linearized machinery as 4D-Var does. Also following Lorenc (2013), we call a pure EnVar system that does not use a linear model ‘4DEnVar’. En4DVar and 4DEnVar are discussed below (without localization at this stage). Bear in mind that this nomenclature is not consistent with all the literature, especially in articles prior to Lorenc (2013), Liu et al. (2008, 2009) for instance called their method EnD4Var even though it does not use a linear model, but this has been changed to the recommended 4DEnVar in later articles (e.g. Liu and Xiao, 2013).

Note that pure En4DVar and pure 4DEnVar use background errors from one source (here the ensemble), but further coupling can be achieved by merging the ensemble statistics with \( B_\theta \) of Var. This forms another family of methods called ‘hybrid EnVar’ (Lorenc, 2013) (section 5) but again this term is not always used consistently in the literature.

4.1. Pure En4DVar: requiring linear/adjoint models

A deterministic analysis \( x^b \) (essentially the ensemble mean of (29)) may be found using a Var technique. Consider the cost function (Liu et al., 2008):

\[
f_{\text{En4DVar}}(x_{\text{ens}}) = \frac{1}{2} \| x_{\text{ens}} \|^2 + \frac{1}{2} \sum_{t=0}^{T} \| d_t - H_t M_d x^b_{t|0} \|^2_\mathbb{R}^j \] (34)

\[
\frac{1}{2} \| x_{\text{ens}} \|^2 + \frac{1}{2} \| d - H_M X^b x_{\text{ens}} \|^2_\mathbb{R}^{-1},
\]

which is a function of a new \( N \)-element control variable \( x_{\text{ens}} \), and where \( d_t = y^b_{t|0} - \mathcal{H}_i(\left \{ M_{0,T}(x^b) \right \}) \) and \( d = y^b - \mathcal{H}(x^b) \). The latter of the two equations uses the time-compact (underlined)
notation. This control variable comprises coefficients that multiply the ensemble perturbations in the linear combination as follows:

$$\delta \mathbf{x} = \mathbf{X}^b \mathbf{x}_{\text{ens}}$$

(Lorenc, 2003). Here \( \mathbf{X}^b \) acts as a CVT, which is taken from a parallel EnKF system (or even from a population of NMC-style forecast differences (Wang et al., 2014), where the National Meteorological Centre (NMC) method approximates forecast errors with a set of lagged forecasts (Parrish and Derber, 1992)).

To see how (34) works, compare the above equations with those of preconditioned Var (section 2.3). Comparing (34) with (21) (with definition (20) [where in (21) and (20) no model error is considered, \( \delta \mathbf{y}_{\text{var}}(t) = 0 \)). These systems have the same form, except that the CVT \( U \) in (17) becomes \( \mathbf{X}^b \) in (36). The implied \( \mathbf{B} \)-matrix of En4DVar is then akin to (24), i.e. \( \mathbf{X}^b \mathbf{T} \) as consistent with (26). Minimizing (34) is efficient as it is well conditioned, and \( \mathbf{x}_{\text{ens}} \) has only \( N \) elements. At the minimum of (34) \( \mathbf{x}_{\text{ens}} = \mathbf{x}_{\text{ens}}^a \), which leads to the analysis \( \mathbf{x}^a = \mathbf{X}^b + \mathbf{X}^b \mathbf{x}_{\text{ens}}^a \).

The gradient of (34) (Appendix A) is required for the minimization and has the forms:

$$\nabla \mathbf{J}_{\mathbf{x}_{\text{ens}}}^{\text{En4DVar}} = \mathbf{x}_{\text{ens}} - \mathbf{X}^b \sum_{t=0}^{T} \mathbf{M}_{\mathbf{b},t} \mathbf{H}^b_{t}^{-1} (\mathbf{d} - \mathbf{H}_{\mathbf{b},t} \mathbf{X}^b \mathbf{x}_{\text{ens}})$$

(37)

where, as in 4D-Var, the linear and adjoint operators appear. This line with Lorenc (2013)) is summarized in Figure 1 (but note the legend). The (time-compact) cost function and gradient of this model and observation operators are linear is equivalent to the mean EnKS solution in section 3.3.1 (Appendix C).

4.2. Pure 4DEnVar: avoiding linear/adjoint models

Just as the linear and adjoint operators can be avoided in the EnKF and EnKS (section 3), a similar approximation can be applied to En4DVar. For \( p \) observations at time \( t \), the combination \( \mathbf{H}_{\mathbf{b},t} \mathbf{M}_{\mathbf{b},t} \mathbf{X}^b \) can be translated to the explicit \( p \times N \)-element matrix \( \mathbf{Y}^a \) using the nonlinear operators \( \mathcal{H} \) and \( \mathcal{M}_{\mathbf{b},t} \) (Liu et al., 2008):

$$\mathbf{Y}^a = \mathbf{H}_{\mathbf{b},t} \mathbf{M}_{\mathbf{b},t} \mathbf{X}^b \approx \frac{1}{\sqrt{N-1}} \left( \mathcal{H}_{\mathbf{b},t}(\mathcal{M}_{\mathbf{b},t}(\mathbf{x}^b_{\text{ens}})) - \bar{\mathbf{y}}^a, \ldots, \mathcal{H}_{\mathbf{b},t}(\mathcal{M}_{\mathbf{b},t}(\mathbf{x}^b_{\text{ens}})) - \bar{\mathbf{y}}^a \right),$$

(38)

and let \( \bar{\mathbf{y}}^a = \left( \begin{array}{c} \mathbf{y}^a_0 \\ \vdots \\ \mathbf{y}^a_T \end{array} \right) = \mathbf{H}_{\mathbf{b},t} \mathbf{X}^b \),

(39)

where \( \bar{\mathbf{y}}^a \) is the model observation vector at time \( t \) based on the ensemble mean, \( \bar{\mathbf{y}}^a = \mathcal{H}_{\mathbf{b},t}(\mathcal{M}_{\mathbf{b},t}(\mathbf{X}^b_{\text{ens}}))) \) and \( \mathbf{Y}^a \) has dimension \( p \times N \). The (time-compact) cost function and gradient of this formulation are adapted from (35) and (37) using (39):

$$\nabla \mathbf{x}_{\text{ens}}^{\text{4DEnVar}} = \mathbf{x}_{\text{ens}} - \bar{\mathbf{Y}}^a \mathbf{R}^{-1} (\mathbf{d} - \bar{\mathbf{Y}}^a \mathbf{x}_{\text{ens}})$$

(41)

Notice that \( \mathcal{M}_{\mathbf{b},t}(\mathbf{x}^b_{\text{ens}}) = \mathbf{x}^b_{\text{ens}}(t) \) (comprising \( \mathbf{x}^b_{\text{ens}} \) in (32)) may be computed in advance, which means that this method is akin to using an EnKS with a 4D state vector (section 3.3.2). In (40) and (41), the CVT is \( \delta \mathbf{x} = \mathbf{X}^b \mathbf{x}_{\text{ens}} \). This method is formally called 4DEnVar (\( \mathbf{Y}^a \) in Figure 1) (Desroziers et al., 2014; Fairbairn et al., 2014), and is similar to methods of Hunt et al. (2004) and Tian et al. (2008). The 3D counterpart is 3DEnVar (although it is essentially the same as 3DVar) and alternative names are given in sections 4 and 7.5.

4.3. Comments on EnVar methods

For En4DVar and 4DEnVar the implied background-error covariance at \( t = 0 \) is \( \mathbf{X}^b \mathbf{T} \), and between times \( t_1 \) and \( t_2 \) it is \( \mathbf{X}^b(t_1) \mathbf{T} \mathbf{X}^b(t_2) \), where for En4DVar \( \mathbf{X}^b(t) \) comprises columns of \( \mathbf{X}^b \) propagated to time \( t \) by \( \mathbf{M}_{\mathbf{b},t} \). \( \mathbf{X}^b(t) \) for but for 4DEnVar \( \mathbf{X}^b(t) \) comprises columns propagated by the nonlinear model as \( \mathbf{M}_{\mathbf{b},t}(\mathbf{x}^b_{\text{ens}}) - \mathbf{M}_{\mathbf{b},t}(\bar{\mathbf{X}}^a) \) for each column \( i \). En4DVar and 4DEnVar are each typically used to find a single deterministic analysis per analysis cycle, and the perturbed ensemble members, \( \mathbf{X}^b \), are used in these methods only to define the CVT. There are advantages and disadvantages of EnVar over traditional Var:

- **They do not require a background-error covariance model as Var does.** This is important when modelling background errors in \( \mathbf{X}^b \) involving processes that are too complicated, nonlinear or when geophysical balances are not relevant.
- **The control vector \( \mathbf{x}_{\text{ens}} \) has \( N \) elements while the control vector \( \delta \mathbf{x} \) in 3D-Var or strong-constraint 4D-Var has \( n \) elements.** As \( N \ll n \), this represents an efficient problem that has shrunk in size to reflect the low rank of \( \mathbf{P}^b_{\text{var}} \).
- **4DEnVar does not need linearized models, but 4D-Var and En4DVar do.** If NWP centres no longer have to develop, maintain and run linear and adjoint models, this results in a significant cost gain, especially if linearizing highly nonlinear processes with \( \mathbf{M} \) can be avoided (e.g. Xu, 1996; Stiller and Ballard, 2009; Stiller, 2009). As a consequence however, other quantities that rely on the adjoint will no longer be available, such as adjoint sensitivity analysis (e.g. Benedetti et al., 2003) or Hessian singular vector calculations (e.g. Lawrence et al., 2009).
- **As with the EnKF/S, EnVar has a suitability for parallel processing for computing the matrix \( \mathbf{Y}^a \).** Such systems have limitless parallelizability with the number of ensemble members.
- **The low-rank property of the implied background-error covariance matrix in EnVar means that sampling error problems will inevitably arise when \( N \) is small.** This usually requires some kind of mitigation such as localization (section 6).

4.4. Generating an ensemble within EnVar

The EnVar forms given in sections 4.1 and 4.2 produce only a single analysis (the mode of the posterior), while the EnKF/S produce an ensemble of possible analyses. However EnVar permits further constraints to the analysis, such as a tangent linear normal mode constraint (Kleist et al., 2009; Wang et al., 2013) or a variationally-based initialization term applied with an extra term—normally called \( J_c \)—added to the cost function

\[ J_c = \frac{1}{2} \sum \mathbf{x}_{\text{ens}}^2 + \frac{1}{2} \mathbf{d}^2 - \mathbf{x}_{\text{ens}}^2 \mathbf{R}^{-1} \]

\[ \nabla \mathbf{x}_{\text{ens}}^{\text{4DEnVar}} = \mathbf{x}_{\text{ens}} - \bar{\mathbf{Y}}^a \mathbf{R}^{-1} (\mathbf{d} - \bar{\mathbf{Y}}^a \mathbf{x}_{\text{ens}}) \]

(41)
(Clayton et al., 2012; Ge et al., 2012), which can help the analysis to be close to a defined balance.

We also mention a couple of related EnVar schemes that are capable of producing an ensemble. The ‘Maximum Likelihood Ensemble Filter’ (MLEF) (Zupanski, 2003) can also bypass use of the linearized model. MLEF, which preconditions the variational problem on the Hessian, has been demonstrated in a 3D context. This preconditioning replaces (36) with the following CVT:

$$\delta x = X^b (1 + S)^{-T/2} \chi_{\text{MLEF}},$$

(42)

where in 3D $S = (R^{-1/2} H X^b)^T (R^{-1/2} H X^b)$, which is written in a way so that the Hessian (as before) can be approximated with two runs of the nonlinear observation operator. Hessian preconditioning not only allows a very efficient minimization but it also allows calculation of an analysis ensemble, so the method can be used without a separate ensemble generator. The ‘ Ensemble Variational Integrated localized (or Lancezos)’ (EVIL) scheme of Auligné et al. (2016) is another method (discussed further in section 5.5 where it is applied to a hybrid setting).

5. Hybrid methods

According to the definition in Lorenc (2013), a hybrid method refers to one that blends $B_0$ from Var with $P_N^{(0)}$ from an ensemble (recall $B_0$ is full-rank but quasi-static and crudely modelled, and $P_N^{(0)}$ is flow-dependent but rank deficient). Six ways of combining these matrices are discussed here. The first is an ensemble-based recalibration of the $B_0$-matrix in Var, the second and third are (equivalent) ways of averaging $B_0$ and $P_N^{(0)}$; the fourth averages the gain matrices instead of the covariance matrices; the fifth returns to averaging $B_0$ and $P_N^{(0)}$, but provides an ensemble of analyses as part of the variational iterations; and the sixth combines $B_0$ with $P_N^{(0)}$, but not by averaging them.

Hybridization introduces climatological error information into EnsVar (where the hybrid system becomes less sensitive to ensemble size than the pure EnKF; Zhang et al., 2009). The use of hybrid covariances has long been thought to be useful (Gustaffson, 2007; Bishop and Satterfield, 2013). Here we shall consider the former viewpoint (equivalent to hybridization of, for example, 4D-Var with En4DVar). From this perspective, the hybrid background-error covariance matrix, here denoted $B^h$, replaces the usual $B_0$-matrix in Var. Localization is not discussed until section 6 since at this stage it is a distraction from the hybridization procedure.

5.1. Ensemble recalibration of $B_0$ in Var

Some operational centres use a parallel ensemble prediction system to continuously recalibrate the model of $B_0$ in their Var system (this blending of ensemble and climatological information would arguably classify this as a hybrid method, so we should call it a $B^h$-matrix). Météo-France and the ECMWF, for example, use the spread of their ensemble prediction systems (initialized not from an EnKF but from ensembles of perturbed 4D-Var assimilations; Belo Pereira and Berre, 2006; Berre et al., 2007, 2009; Isaksen et al., 2010) to model the ‘variances of the day’ of $B_0$ (Bonavita et al., 2011, 2012; Raynaud et al., 2012). The ensemble is known to have significant sampling errors (and so needs to be filtered) and is known to be under-spread (and so needs to be inflated).

Recently the ECMWF and Météo-France extended their hybrid schemes to include on-line calibration of the spatial correlations in their B-matrices. At the ECMWF, e.g. these spatial structures are described by a ‘wavelet diagonal’ scheme (Fisher, 2003), which can represent different correlation length-scales for different geographical regions. The original implementation of this scheme used a climatological calibration, which implied static background-error length-scales, but these are known in reality to be flow-dependent (e.g. longer length-scales in higher pressure systems). The ECMWF has experimented with two ways of calibrating their wavelet scheme, namely using samples of background errors from either (i) a 12 day window (25 ensemble members, twice per day, leading to 600 samples), or (ii) a combination of current members (25 members within an 8 h window – 200 samples) plus samples representing climatology (400 samples spread over the seasons, leading again to 600 samples in total) (Bonavita et al., 2016). The 600 samples in each case are considered necessary to calibrate the wavelet model without filtering or localizing. Even though the forms of their $B^h$-matrix models (Derber and Bouttier, 1999; Fisher, 2003) are not changed by recalibration, important flow-dependence is added as a consequence of recalibration with either (i) or (ii) above, resulting in modestly improved forecasts right out to 10 days in some quantities (Bonavita et al., 2016).

5.2. The explicit average of error covariance matrices

The classic hybrid background-error covariance matrix, $B^h$, is defined according to Hammill and Snyder (2000) as the weighted average of $B_0$ and $P_N^{(0)}$:

$$B^h = (1 - \beta) B_0 + \beta P_N^{(0)},$$

(43)

where $\beta (0 \leq \beta \leq 1)$ is a tunable factor controlling the weighting of the static and flow-dependent covariances, and $B^h$ replaces $B_0$ in (10). The use of explicit matrices is obviously impractical for large systems and so methods of using $B^h$ implicitly are introduced.

5.3. The implicit average of error covariance matrices

This method merges the CVTs of (17) and (36) to represent $B_0$ and $P_N^{(0)}$ implicitly. We show this in the context of hybridizing 4D-Var with En4DVar:

$$\chi_{\text{4DVar}}(\delta x_{\text{var}}, x_{\text{ens}}) = \frac{1}{2} \| \chi_{\text{var}} \|^2 + \frac{1}{2} \| x_{\text{ens}} \|^2 + \frac{1}{2} \sum_{t=0}^{T} \| \delta d_t \|^2_{P^{-1}_t},$$

(44)

where

$$\delta d_t = d_t - H_x M_{xt} \left( \sqrt{1-\beta} U B_0 \chi_{\text{var}} + \sqrt{\beta} B^h x_{\text{ens}} \right)$$

(45)

and $d_t = y_t - H_x \left( M_{xt}(x^h) \right)$. (cf. (44)–(45) with (34)). In (44) the augmented control vector $(\delta x_{\text{var}}, x_{\text{ens}})$ has $n + N$ elements where $\delta x_{\text{var}}$ is associated with $B_0$ (as $x_{\text{var}}$ in section 2.3), and $x_{\text{ens}}$ associated with $P_N^{(0)}$ (as $x_{\text{ens}}$ in section 4). Recall that $n$ is the size of the state, $N$ is the number of ensemble members, $U$ is the original $n \times n$ CVT of section 2.3, and $x^h$ is the $n \times N$ matrix comprising the ensemble perturbations $(x^h - \bar{x})/\sqrt{N-1}$. The hybrid En4DVar cost function can be compactly written as

$$J_{\text{En4DVar}}(x^h) = \frac{1}{2} \| x^h \|^2 + \frac{1}{2} \sum_{t=0}^{T} \| \delta d_t \|^2_{P^{-1}_t},$$

(46)

where $x^h = \left( \delta x_{\text{var}}, x_{\text{ens}} \right)$. (47)

For conciseness, the formula for the gradient of $J_{\text{En4DVar}}$ with respect to $x^h$ is left to section 7, where localization is also
considered. The CVT for (46) relates $\delta x_{\text{var}}$ and $x_{\text{ens}}$ to $\delta x$ at $t = 0$:

$$\delta x = \sqrt{1 - \beta U x_{\text{var}}} + \sqrt{\beta} x_{\text{ens}},$$

$$= U h, x_{\text{h}} = \left(\sqrt{1 - \beta U} \sqrt{\beta} x_{\text{ens}}^h\right),$$

$$(48)$$

(cf. (17)). The hybrid CVT which consolidates $U$ and $x_{\text{h}}$ is called $U_h$, which is the $n \times (n+N)$ matrix $\left(\sqrt{1 - \beta U} \sqrt{\beta} x_{\text{ens}}^h\right)$. In the $x_{\text{h}}$ representation, background errors have covariance $\left(x_{\text{ens}}^h x_{\text{ens}}^{hT}\right) = I$, which implies the following error covariance in the model representation ($\delta x$):

$$B_h = \left(\delta x \delta x^T\right)_h$$

$$= \left(1 - \beta\right)\left[U U^T + \beta X X^T\right] = \left(1 - \beta\right)B_0 + \beta B_{(N)},$$

$$(49)$$

where $\bullet_h$ performs an expectation over the background population. Equations (49) and (43) are identical, which shows that minimizing (44) with respect to $\delta x_{\text{h}}$ (with CVT (48)), and then transforming the resulting control variable increment into model space with (48), is the same as minimizing the strong-constraint 4D-Var cost function with $B_0 \rightarrow B_h$. This is hybrid En4DVar ($\otimes$ in Figure 1, $\otimes$ for En3DVar).

A benefit of form (44) over (43) is efficiency as it avoids prohibitively large matrices. Wang et al. (2007b) provide an alternative derivation of (49). Although this hybrid method of representing the background-error covariance matrix has been applied to strong-constraint 4D-Var, it may also be used with the weak-constraint formulation (21). When $\beta = 1$ ($\beta = 0$), this system is identical to pure En4DVar (pure Var) (section 4.1).

5.4. Combining gain matrices

The most common type of hybrid schemes average $B_0$ with $P_{(N)}^{\text{ENS}}$ but an alternative is to follow Penny (2014) by averaging the Kalman gains of pure Var (gain $K^{\text{VAR}}$) and of pure EnKF (gain $K^{\text{ENS}}$) to give $K^h$. For 3D schemes, e.g.

$$K^h = \beta_1 K^{\text{VAR}} + \beta_2 K^{\text{VAR}} + \beta_3 K^{\text{VAR}} \text{HK}^{\text{ENS}},$$

$$(50)$$

where $K^{\text{ENS}} = P_{(N)}^{\text{ENS}} \left(H P_{(N)}^{\text{ENS}} H^T + R\right)^{-1},$

$$K^{\text{VAR}} = B_0 H^T \left(H B_0 H^T + R\right)^{-1},$$

and $\beta_1$ are scalars. $K^{\text{VAR}}$ and $K^{\text{ENS}}$ capture the way that the respective schemes work but are not computed explicitly.

Penny (2014) applied this averaging to 3D-Var and an LETKF with the Lorenz 96 model (Lorenz, 1996) and the choices $\beta_1 = 1$, $\beta_2 = \alpha$, and $\beta_3 = -\alpha$. This leads to the hybrid gain:

$$K^h = K^{\text{ENS}} + \alpha K^{\text{VAR}} (I - \text{HK}^{\text{ENS}}),$$

$$(53)$$

equivalent to first running the EnKF to give the ensemble mean $x_{\text{ens}}^h$, running Var with this as the background to give $x_{\text{var}}^h$, and then doing the following weighted average to give the hybrid analysis $x^h = x_{\text{ens}}^h + \alpha x_{\text{var}}^h$. Penny (2014) found that this kind of hybrid improves the analysis over those of the separate pure schemes. However the benefit of this hybrid is that it requires little extra coding to existing EnKF and Var schemes.

5.5. EVIL

Most EnVar systems use a separate EnKF-style system to generate $X^0$ for each cycle. The fact that the EnVar and EnKF are separate can lead to inconsistencies, and hence to sub-optimality in hybrid EnVar. EVIL (Ensemble Variational Integrated Localized (or Lanzcos)) (Auligné et al., 2016) represents a modification to hybrid EnVar which uses information gained in the Var minimization procedure to estimate an analysis ensemble, instead of the usual single analysis.

EVIL is based on the fact that conjugate gradient (CG)-based minimization algorithms are closely related to Lanczos methods (Paige and Saunders, 1975; Fisher and Courtier, 1995; El Akraoui et al., 2013). In outline, the gradient descent vectors from $q$ iterations of the CG procedure form a Krlov sub-space in which the Hessian of the cost function is tridiagonal (q here takes the role of N). Typically $q \sim O(50)$, so such a Hessian can be easily diagonalized. The eigenvectors of the Hessian in control space – stored in the $n \times q$ matrix $Z_h$ - and the eigenvectors - in the $q \times q$ diagonal matrix $\Theta$ – are called Ritz pairs. In general in quadratic systems, the analysis-error covariance matrix, $A$, is the inverse of the Hessian. This leads to the following approximation for an $n \times q$ analysis ensemble, $X^h$ in 3D (Auligné et al., 2016):

$$X^h = X^h + A H^T R^{-1} \left(Y^o - H X^h\right),$$

$$\approx X^h + U_h Z_h \Theta^{-1} Z_h^T U_h^T H^T R^{-1} \left(Y^o - H X^h\right),$$

$$(54)$$

where $Z_h \Theta^{-1} Z_h^T$ is the analysis-error covariance (in control space) found from the CG/Lanczos procedure, $Y^o$ is a $p \times q$ matrix of stochastically perturbed observations, and $U_h$ is the hybrid CVT (48). (Note that the 3D objects like $X^h, H,$ etc. could be extended to 4D with the underline notation of sections 3.3.1 and 3.3.2 as appropriate.) The EnVar minimization would provide the Ritz pair information and (54) would provide the machinery to compute an analysis ensemble – then propagated in time for the next forecast ensemble. EVIL is $\otimes$ in Figure 1.

Equation (54) is a particular implementation of EVIL called stochastic-EVIL (S-EVIL). Other implementations are deterministic EVIL (D-EVIL), which uses the Ritz pairs to compute a square-root of the analysis-error covariance matrix, and a resampling EVIL (R-EVIL) which can yield additional ensemble members (Auligné et al., 2016). It may also be applied to bi-conjugate gradient algorithms giving dual representations of the Ritz pairs.

The number of iterations, $q$, may need to be relatively high to obtain a reasonable representation of the correct eigen-spectrum of $A$, and Auligné et al. (2016) suggest that several hundred may be needed in their cut-down (single-level) system. This may not yet be affordable to run routinely, but is an option for future systems. The ability of $Z_h \Theta^{-1} Z_h^T$ to represent the analysis-error covariance matrix will need to be checked thoroughly, as will the performance of the method with nonlinear operators where the link between the Hessian and the analysis-error covariance matrix is no longer strictly valid.

5.6. The ensemble reduced-rank Kalman filter

In (43), $B_0$ is able to influence how the DA affects all degrees of freedom, including the subspace spanned by the ensemble. In Petrie and Bannister (2011), another hybrid scheme is proposed which, instead of averaging, essentially uses $P_{(N)}^{\text{ENS}}$ in the ensemble subspace and uses $B_0$ elsewhere. The general method is based on the ‘reduced-rank Kalman filter’ (or the ‘simplified Kalman filter’) developed at ECMWF in the 1990s for use with Hessian singular vectors (Fisher, 1998; Fisher and Andersson, 2001; Beck and Ehrendorfer, 2005), but Petrie and Bannister (2011) adapted it for use with an ensemble and called it the ‘Ensemble Reduced-Rank Kalman Filter’ (EnRRKF).
The cost function for the EnRRKF is a function of the \( n \)-element control vector \( \chi_{\text{EnRRKF}} \):

\[
J_{\text{EnRRKF}}(\chi_{\text{EnRRKF}}) = \| \chi_{\text{EnRRKF}} \|^2_{P^{-1}} \chi_{\text{EnRRKF}}
\]

\[
= \frac{1}{2} \| \chi_{\text{EnRRKF}} \|_{P^{-1}}^2 + \bar{\chi}_{\text{EnRRKF}} P^{-1} \chi_{\text{EnRRKF}}
\]

\[
+ \frac{1}{2} \| \bar{\chi}_{\text{EnRRKF}} \|_{P^{-1}}^2 + \frac{1}{2} \sum_{i=0}^{T} \| \delta_{i} \|_{G^{-1}}^2,
\]

where \( \chi_{\text{EnRRKF}} = \chi_{e} + \bar{\chi}_{\text{EnRRKF}} \),

\[
\delta_{i} = d_{i} - H_{i} M_{i} u X_{\text{EnRRKF}}
\]

and \( d_{i} = y_{i} - H_{i} \chi_{\text{EnRRKF}} \).

The first part of \( \chi_{\text{EnRRKF}} \) in (56) is \( \chi_{e} \) (\( n \)-elements) which represents the subspace (the part that has flow-dependent covariances) and is non-zero only in the first \( N \) elements. The second part is \( \bar{\chi}_{\text{EnRRKF}} \) (\( n \)-elements) which is the remainder (the part that has climatological covariances) and is non-zero only in the last \( n - N \) elements. Due to the zero terms, summing these vectors in (56) does not lose any information. The first line of (55) has three terms due to errors that lie

(i) exclusively inside the \( N \)-dimensional subspace,
(ii) inside and outside this subspace, and
(iii) exclusively outside this subspace respectively.

The CVT for the EnRRKF is:

\[
\delta_{X} = UX_{\text{EnRRKF}}
\]

where \( U \) is the same CVT used in pure Var (17) and \( X \) is a special \( n \times n \) matrix \((XX^{T}) = I \) and \( X^{T}X = I \) (do not confuse \( X \) with \( X^{T} \)). \( X \) has the following special properties: when \( \delta_{X} \) lies in the ensemble subspace, \( X^{T}U^{-1} \delta_{X} \) is a vector which can be non-zero only in the first \( N \) elements (the last \( n - N \) elements are zero), and when \( \delta_{X} \) lies outside the ensemble subspace, \( X^{T}U^{-1} \delta_{X} \) is a vector which can be non-zero only in the last \( n - N \) elements (the first \( N \) elements are zero). This reflects the structures of \( \chi_{e} \) and \( \bar{\chi}_{\text{EnRRKF}} \) and can be achieved using Householder transforms (Fisher, 1998). Although \( P_{\chi}^{-1} \) appears in (55), it acts only on vectors that are always zero in the last \( n - N \) elements. This means that only part of the inverse matrix \( P_{\chi}^{-1} \) (the left-most \( n \times N \) part) needs to be known, which is achievable for \( N \sim O(10) - O(10^2) \). This reduced \( P_{\chi}^{-1} \) matrix is derived in Petrie and Bannister (2011) and has a complicated, but calculable form.

To our knowledge, the EnRRKF has not yet been tested, so it is impossible to assess it. Since the flow-dependent part of the error covariance is nearly full-rank in the \( N \)-dimensional space (or at least a \( K \)-dimensional subspace (\( K < N \)) can be defined where \( P_{\chi}^{(K)} \) is full rank), there is reason to suppose that localization in the EnRRKF is not essential. Unlike for the other hybrid methods, the rank of \( P_{\chi}^{(N)} \) computed with (26) is \( \leq (N - 1) \), which leads to the introduction of analysis noise (Houtekamer and Mitchell, 1998; van Leeuwen, 1999; Hamill et al., 2001; Houtekamer and Mitchell, 2005; Ehrendorfer, 2007; Meng and Zhang, 2012; Houtekamer and Zhang, 2016). In normal situations when \( N \ll n \), filter divergence can be mitigated with inflation (Anderson and Anderson, 1999) and rank deficiency can be mitigated with localization (Hamill et al., 2001; Lorenc, 2003).

Localization plays a fundamental role in EnKF-based systems, but before this section, for simplicity, it has been missing from the equations. This section is about how localization is included. Different means of reducing far-field sampling errors by localization are used. One common method modifies the sample covariances by a Schur product in model space (Lorenc, 2003; Hamill et al., 2001). This is the type of localization that is reviewed below. Another method works in observation space by limiting the observations used to update a particular point to those within a prescribed radius, and/or by gradually reducing the influence of the observations as a function of distance from the point (Houtekamer and Mitchell, 2005). The latter method is used for instance in the local ensemble Kalman filter (Ott et al., 2004; Szunyogh et al., 2005) and with other ensemble DA techniques (Bishop et al., 2001; Pham, 2001; Szunyogh et al., 2008), but presents problems with interpreting observations with non-local observation operators (Fertig et al., 2008), such as satellite radiance measurements.

6.1. Schur product localization applied in model space

Sampling noise due to small \( N \) is most significant when the true correlation values are small. Since true correlation values are expected to be small between points that are separated by a large distance, a common approach is to filter computed correlation values according to the distance separating the points. Consider a single field and let the sample covariance between points \( p \) and \( q \) (at positions \( r_{p} \) and \( r_{q} \)) be the matrix element \( \hat{P}_{(N)}^{b} \). This is multiplied by a moderation function \( c(r_{p}, r_{q}) \) which is unity when \( r_{p} = r_{q} \) and goes to zero when \( |r_{p} - r_{q}| \rightarrow \infty \) (Gaspari and Cohn, 1999). This is localization. In the context of (26), the localized covariance function is

\[
c(r_{p}, r_{q}) \left[ \hat{P}_{(N)}^{b} \right] = c(r_{p}, r_{q}) \left( \frac{N}{N - 1} \sum_{i=0}^{T} \delta_{i}^{b}(r_{p}) \delta_{i}^{b}(r_{q}) \right),
\]

where \( \delta_{i}^{b}(r_{p}) \) is the \( i \)-th background perturbation (27) at \( r_{p} \). Define an \( n \times n \) localization matrix \( C \) to have elements \( c_{pq} = c(r_{p}, r_{q}) \) and define the localized background-error covariance matrix to be \( \hat{P}_{(N)}^{b} \), whose matrix elements are

\[
c(r_{p}, r_{q}) \left[ \hat{P}_{(N)}^{b} \right] = c(r_{p}, r_{q}) \left[ \hat{P}_{(N)}^{b} \right] \in C \cdot \hat{P}_{(N)}^{b} \in C^{T} \cdot \hat{P}_{(N)}^{b} \in C^{T}.
\]

This can be written in the following compact way:

\[
\hat{P}_{(N)}^{b} = C \cdot \hat{P}_{(N)}^{b} \in C^{T}.
\]

where the \( \circ \) symbol is the Schur (element-by-element or Hadamard) product, and where a covariance with a hat indicates its localized form. (This hat notation will apply to any quantity: covariance matrices and cost functions, etc.) This localization can be generalized to the multivariate situation (Bishop and Hodysy, 2009a, 2009b, 2011; Bannister, 2015). Matrix \( C \) is mathematically a correlation matrix, and should be chosen such that \( \hat{P}_{(N)}^{b} \) remains a true covariance (Gaspari and Cohn, 1999).

The Schur product localization discussed above may be written explicitly, which allows localization to be used in DA without the need for the explicit covariance matrices in (59). The two known forms are labelled ‘B05’ and ‘L03’ after Buehner (2005) and Lorenc (2003) respectively (as is done in Wang et al., 2007b).

6.2. The ‘B05’ representation of the localization Schur product

In a similar way to the decomposition of \( \hat{P}_{(N)}^{b} \) as \( \hat{P}_{(N)}^{b} = X^{T}X^{T} \) (26), suppose similarly that \( C \) can be decomposed as \( UU^{T} \) (26).
(where the $n \times M$ matrix $U^C$ is a kind of ‘square-root’ of $C$), and $P^{N}_{(N)}$ can be decomposed as $P^{N}_{(N)} = \hat{X}^b \hat{X}^b^T$ (where the $n \times NM$ matrix $\hat{X}^b$ is defined below). Appendix D shows that $P^{N}_{(N)}$ can be written as:

\[
\hat{P}^{N}_{(N)} = \hat{X}^b \hat{X}^b^T, \quad (60)
\]

\[
\hat{X} = \frac{1}{\sqrt{N-1}} \left( \delta x^b_{(1)} \circ \mathbf{u}^C_{(1)}, \ldots, \delta x^b_{(1)} \circ \mathbf{u}^C_{(N)}, \ldots, \delta x^b_{(N)} \circ \mathbf{u}^C_{(1)}, \ldots, \delta x^b_{(N)} \circ \mathbf{u}^C_{(N)} \right) = \frac{1}{\sqrt{N-1}} \left( \text{diag}(\delta x^b_{(1)} U^C, \ldots, \text{diag}(\delta x^b_{(N)} U^C) \right) \quad (61)
\]

(Buehner, 2005), where $\delta x^b_{(i)}/\sqrt{N-1}$ is the vector occupying the ith column of $\hat{X}^b$, $\mathbf{u}^C_{(i)}$ is the vector occupying the ith column of $U^C$, and the diag operator evaluates to an $n \times n$ diagonal matrix whose diagonal comprises its n-element vector argument. In other words, the columns of $\hat{X}^b$ are formed in (61) from every possible Schur-product-pair of columns of $\mathbf{X}^b$ and $U^C$, and $\hat{X}^b$ can be thought of as a matrix of $NM$ scaled effective ensemble members whose covariance $\hat{X}^b \hat{X}^b^T$ is the localized background-error covariance $\hat{P}^{N}_{(N)}$. As $NM > N$, $\hat{P}^{N}_{(N)}$ would be expected to have a higher rank (maximum rank $\sim NM$) than $P^{N}_{(N)}$ (maximum rank $\sim N$). This approach may be used to localize any of the pure EnVar or hybrid schemes discussed in sections 4 and 5 by changing (36) to

\[
\delta x = \hat{X}^b \chi_{\text{ens}}, \quad (62)
\]

where $\chi_{\text{ens}}$ is the corresponding $NM$-element control vector.

6.3. The ‘L03’ representation of the localization Schur product

Another compact representation of a localized ensemble background-error covariance matrix is often used for the variational formulations of sections 4 and 5 (Lorenc, 2003). This representation is explained with reference to localizing the En4DVar system (34) (37 in Figure 1), which involves $N$ new control vectors:

\[
\mathbf{P}^{4D}_{\text{En4DVar}}(\hat{V}) = \frac{1}{2} \sum_{i=1}^{N} \left\| \mathbf{\hat{a}}_{(i)} \right\|_2^2, \quad (63)
\]

\[
\mathbf{\hat{a}}_{(i)} = \left( \mathbf{\hat{a}}_{(1)} \ldots \mathbf{\hat{a}}_{(N)} \right), \quad (64)
\]

\[
\delta x = \frac{1}{\sqrt{N-1}} \sum_{i=1}^{N} \delta x^b_{(i)} \circ \mathbf{\hat{a}}_{(i)} = \left[ \mathbf{X}^b \circ \mathbf{\hat{A}} \right] \chi_{\text{ens}}, \quad (65)
\]

and $\mathbf{\hat{a}}_{(i)}$ is the n-element control vector associated with the ith ensemble member, which are assembled into the $n \times N$ matrix $\mathbf{\hat{A}}$, and $1_N$ is the column vector whose $N$ elements all contain 1. Equation (65) is an unpreconditioned form of L03 localization. A preconditioned form can be devised with $N$ alternative vectors

$\mathbf{\hat{V}}$.

(\text{where the $n \times M$ matrix $U^C$ is a kind of ‘square-root’ of $C$, and $P^{N}_{(N)}$ can be decomposed as $P^{N}_{(N)} = \hat{X}^b \hat{X}^b^T$ (where the $n \times NM$ matrix $\hat{X}^b$ is defined below). Appendix D shows that $P^{N}_{(N)}$ can be written as:)

\[
\hat{P}^{N}_{(N)} = \hat{X}^b \hat{X}^b^T, \quad (60)
\]

\[
\hat{X} = \frac{1}{\sqrt{N-1}} \left( \delta x^b_{(1)} \circ \mathbf{u}^C_{(1)}, \ldots, \delta x^b_{(1)} \circ \mathbf{u}^C_{(N)}, \ldots, \delta x^b_{(N)} \circ \mathbf{u}^C_{(1)}, \ldots, \delta x^b_{(N)} \circ \mathbf{u}^C_{(N)} \right) = \frac{1}{\sqrt{N-1}} \left( \text{diag}(\delta x^b_{(1)} U^C, \ldots, \text{diag}(\delta x^b_{(N)} U^C) \right) \quad (61)
\]

(Buehner, 2005), where $\delta x^b_{(i)}/\sqrt{N-1}$ is the vector occupying the ith column of $\hat{X}^b$, $\mathbf{u}^C_{(i)}$ is the vector occupying the ith column of $U^C$, and the diag operator evaluates to an $n \times n$ diagonal matrix whose diagonal comprises its n-element vector argument. In other words, the columns of $\hat{X}^b$ are formed in (61) from every possible Schur-product-pair of columns of $\mathbf{X}^b$ and $U^C$, and $\hat{X}^b$ can be thought of as a matrix of $NM$ scaled effective ensemble members whose covariance $\hat{X}^b \hat{X}^b^T$ is the localized background-error covariance $\hat{P}^{N}_{(N)}$. As $NM > N$, $\hat{P}^{N}_{(N)}$ would be expected to have a higher rank (maximum rank $\sim NM$) than $P^{N}_{(N)}$ (maximum rank $\sim N$). This approach may be used to localize any of the pure EnVar or hybrid schemes discussed in sections 4 and 5 by changing (36) to

\[
\delta x = \hat{X}^b \chi_{\text{ens}}, \quad (62)
\]

where $\chi_{\text{ens}}$ is the corresponding $NM$-element control vector.

6.3. The ‘L03’ representation of the localization Schur product

Another compact representation of a localized ensemble background-error covariance matrix is often used for the variational formulations of sections 4 and 5 (Lorenc, 2003). This representation is explained with reference to localizing the En4DVar system (34) (37 in Figure 1), which involves $N$ new control vectors:

\[
\mathbf{P}^{4D}_{\text{En4DVar}}(\hat{V}) = \frac{1}{2} \sum_{i=1}^{N} \left\| \mathbf{\hat{a}}_{(i)} \right\|_2^2, \quad (63)
\]

\[
\mathbf{\hat{a}}_{(i)} = \left( \mathbf{\hat{a}}_{(1)} \ldots \mathbf{\hat{a}}_{(N)} \right), \quad (64)
\]

\[
\delta x = \frac{1}{\sqrt{N-1}} \sum_{i=1}^{N} \delta x^b_{(i)} \circ \mathbf{\hat{a}}_{(i)} = \left[ \mathbf{X}^b \circ \mathbf{\hat{A}} \right] \chi_{\text{ens}}, \quad (65)
\]

and $\mathbf{\hat{a}}_{(i)}$ is the n-element control vector associated with the ith ensemble member, which are assembled into the $n \times N$ matrix $\mathbf{\hat{A}}$, and $1_N$ is the column vector whose $N$ elements all contain 1. Equation (65) is an unpreconditioned form of L03 localization. A preconditioned form can be devised with $N$ alternative vectors

$\mathbf{\hat{V}}$.

(\text{where the $n \times M$ matrix $U^C$ is a kind of ‘square-root’ of $C$, and $P^{N}_{(N)}$ can be decomposed as $P^{N}_{(N)} = \hat{X}^b \hat{X}^b^T$ (where the $n \times NM$ matrix $\hat{X}^b$ is defined below). Appendix D shows that $P^{N}_{(N)}$ can be written as:)}
member $i$ and waveband $j$ this gives $\delta \hat{x}_{ij} = S^{-1} \hat{x} S \delta \hat{x}_{ij}$, where $S$ is a spectral transform, and $\Psi$ is the diagonal bandpass filter associated with band $j$. Under this scheme, the alternative version of (61) is

$$\tilde{X}^b = \frac{1}{\sqrt{N-1}} \left( \text{diag}(\delta \hat{x}_{1j}), \ldots, \text{diag}(\delta \hat{x}_{Nj}) \right),$$

where $U^C$ is a square-root of a localization matrix associated with band $j$, i.e., $U^C U^C^T = C$. This process allows band (i.e. scale-selective) localization, which is useful because less localization (i.e. larger localization length-scales) is needed for lower wavebands (i.e. larger error scales) than for higher wavebands (Buehner, 2012). This method effectively increases the potential rank of the background-error covariances by a factor of $J$, which results in the need for fewer ensemble members for similar noise level (Buehner, 2012). This method is extended in (Buehner and Shlyaeva, 2015).

B05 and L03 localizations are equivalent, but their efficiencies depend upon the chosen rank of $U^C$. Consider for instance an implementation of localization in pure En4DVar.

- The total number of control vector elements required for B05 is $NM$, making up $\tilde{X}_{\text{ens}}$ in (62). Elements of $\tilde{X}_{\text{ens}}$ are associated with $NM$ perturbation vectors in $\tilde{X}^b$. For lower-rank implementations of $U^C$ (where $M < N$), B05 is the more efficient approach. B05 is used, for example, in the Environment Canada global and regional deterministic DA systems (Buehner et al., 2015a; Caron et al., 2015), and also in the recent work of Liu and Xue (2016).
- The total number of control vector elements required for L03 is $N\delta$ (comprising $\tilde{V}$ in (66)–(68)), which are associated with the $N$ perturbation vectors in $\tilde{X}^b$. For full-rank implementations of $U^C$, L03 is the more efficient approach (since $\tilde{X}^b$ does not have to be stored) as long as $U^C$ can be modelled efficiently. L03 is used, for example, in the Met Office’s global hybrid system (Clayton et al., 2012).

We note that it is arguably more straightforward to account for model error in 4DEnVar schemes that use B05 localization (section 8).

7. Summary of the EnKF, pure EnVar, hybrid EnVar

We have presented many options for approximating the DA problem which incorporate an ensemble in one way or another. These have been classified as either pure ensemble (EnKF/S, section 3), pure EnVar (essentially Var approaches to finding the mean state of the pure ensemble system, section 4) and hybrid EnVar (extending EnVar with added information from the B05-matrix of Var, section 5). The B05 (section 6.2) and L03 (section 6.3) localization formulations are now applied to these DA systems to give the final forms of the equations that operational DA can deal with. It is possible (within bounds) to mix-and-match a DA method with a localization method, leaving us with a relatively large number of combinations.

7.1. Localization in the EnKF

The most common way of localizing $P_b^{(N)}$ in an EnKF (3) in Figure 1) is by restricting the observations that are allowed to affect each analysis point (discussion in section 6). However, to localize in model space, B05 is the appropriate implicit method. As described in section 6.2, a localized version of (28) is found by replacing $P_b \rightarrow \hat{P}_b$ in the first line, which is equivalent to replacing $\tilde{X}^b \rightarrow \tilde{X}$ in the last line, where $\tilde{X}$ is defined in (61) (not shown).

To avoid linearized operators in the resulting expression, $H\tilde{X}^b$ can be approximated by the nonlinear operations $H(x^b + \delta \hat{x}_{ij}) = H(x^b)$ (approximating the $i$th column of $H\tilde{X}^b$, where $\delta \hat{x}_{ij}$ is the $i$th column of $\hat{x}^b$, and $x^b$ is the ensemble mean).

7.2. Model space localization in the EnKS

7.2.1. Localization in the EnKS (3D state vector)

The EnKS with a 3D state vector (where observations are predicted by combining the forecast model with the observation operator (29), (5) in Figure 1) may also be localized with B05; in fact it would be difficult to use the observation space localization mentioned in section 6 with this EnKS, as the presence of the forecast model means that even point observations for $t > 0$ cannot be associated with points in the model domain at $t = 0$. This localized EnKS is found by a similar procedure to that described in section 7.1, applied to (29) (not shown).

To avoid linearized operators in the resulting expression, $H_b \tilde{X}^b$ can be approximated by $H_b \hat{X}^b = H_b \hat{X}^b$ (approximating the ith column of $H_b \tilde{X}^b$).

7.2.2. Localization in the EnKS (4D state vector)

The EnKS with a 4D state vector (5) in Figure 1) uses $\tilde{X}^b$ instead of $X^b$ (32), which eliminates the forecast model from within the DA (33). The localization must also be extended to the time dimension, $C \rightarrow C$, so $U^C \rightarrow U^C$, where $C = U^C U^C^T$. This version of the localized EnKS is found by a similar procedure to that described in sections 7.1 and 7.2.1, applied to (33), by replacing $\tilde{X}^b \rightarrow \tilde{X}$ (not shown).

As above, each column of $H_b \tilde{X}^b$ can be approximated by the nonlinear operations $H_b(x^b + \delta \hat{x}_{ij}) = H_b(x^b)$ (approximating the ith column of $H_b \tilde{X}^b$).

7.2.3. Comments on the implied localized background-error covariances in these systems

Due to the way that the different localizations work in the above 3D and 4D versions of the EnKS, they do not necessarily share the same implied localized background-error covariances, even when the forecast model is linear. In order to examine how they differ, consider the implied error covariance matrix between fields at time $t$. This is studied by looking at the analysis increment (propagated to time $t$) due to observations at (only) $t$. Under these conditions the analysis increment of the 3D state vector form of the EnKS (29) $(X^b \rightarrow \tilde{X}^b)$ is:

$$M_0(x_{ij} - \tilde{x}_{ij}) = \sum x_{ij} \left( H M_0 \tilde{X}^b \right)^T$$

$$\times \left\{ R H M_0 \tilde{X}^b (H M_0 \tilde{X}^b)^T \right\}^{-1}$$

$$\times \left\{ y_{ij} - H \tilde{M}_0 (x_{ij}) \right\}$$

$$= \tilde{P}_{(N)}(t) H^T R H \tilde{P}_{(N)}(t) H^T \left\{ y_{ij} - H \tilde{M}_0 (x_{ij}) \right\},$$

where $\tilde{P}_{(N)}(t) = M_0(x_{ij} - \tilde{x}_{ij})^T M_0(x_{ij} - \tilde{x}_{ij})$ and $\tilde{y}_{ij} = \tilde{P}_{(N)}(0) = C \tilde{P}_{(N)}$. As usual, the hat (no hat) means "localized" ("unlocalized"). This analysis allows us to see that $\tilde{P}_{(N)}(t)$ is the propagated localized background-error covariance implied in the localized EnKS with 3D state vector, which is just the localized covariance at $t = 0$ propagated with the linear model.
The analysis increment of the 4D state vector form of the EnKS (33) \((\mathbf{X}^b \rightarrow \hat{\mathbf{X}}^b)\) due to observations at (only) t is now studied. It exists at all times in the window, but the increment at t can be isolated with the operator comprising \(T + 1\) blocks, each \(n \times n\) \(\mathbf{I}_t = \{0 \cdots 1 \cdots 0\}\), where all blocks are 0, except for the one corresponding to time t, which is I. Noting that for this case \(H\hat{\mathbf{X}}^b = H\hat{\mathbf{X}}^b(t)\), and \(X^b = \hat{\mathbf{X}}^b(t)\), the time t analysis increment from (33) \((\mathbf{X}^b \rightarrow \hat{\mathbf{X}}^b)\) is

\[
\mathbf{X}^b(t) - \hat{\mathbf{X}}^b(t) = \mathbf{I}_t \left[ \left( \mathbf{X}^b(t) \hat{\mathbf{X}}^b(t) \right) \mathbf{R}^{-1} + \mathbf{H} \hat{\mathbf{X}}^b(t) \left( \mathbf{H}^T \mathbf{R}^{-1} \right) \right]^{-1} \left( \mathbf{X}^b(t) - \mathbf{H} \hat{\mathbf{X}}^b(t) \right).
\]

(71)

This analysis allows us to see that the localized covariance matrix implied here at time t is \(\hat{\mathbf{X}}^b(t) \hat{\mathbf{X}}^b(t)^T = \mathbf{I}_t \left( \mathbf{C} \mathbf{P}^b \mathbf{C}^T \right) \mathbf{I}_t = \mathbf{C}(t) \odot \mathbf{P}^b(t)\). The differences between this covariance and the one for the 3D state vector are that (i) here no linearized model is used and (ii) the localization at time t, \(\mathbf{C}(t)\) here needs to be specified (via \(\mathbf{U}^b\)), rather than being propagated from \(\mathbf{C}(0)\). This difference also separates the localized En4DVar and 4DEnVar schemes (below).

7.3. Localization and hybridization in En4DVar

In this section, En4DVar (the method in section 4.1 of finding the mean of the EnKS variationally, making use of the linear models of 4D-Var) is localized with B05 and with L03 (Ω in Figure 1, and Ω is for hybrid En3DVar). The equations shown are also hybridized with the climatological \(\mathbf{B}_0\).

7.3.1. B05 localization in hybrid En4DVar

The 4DEnVar equations hybridized and localized with B05 emerge from a straightforward application of concepts in sections 5.3 and 6.2 to (35). The cost function, CVT and gradient formulae for this case are given as:

\[
\hat{\mathbf{H}}^{\text{En4DVar}} \left( \delta \hat{\mathbf{X}}_{\text{var}} , \hat{\mathbf{X}}_{\text{ens}} \right) = \frac{1}{2} \left\| \delta \hat{\mathbf{X}}_{\text{var}} \right\|^2 + \frac{1}{2} \left\| \hat{\mathbf{X}}_{\text{ens}} \right\|^2 + \frac{1}{2} \left\| \mathbf{d} - \mathbf{H} \delta \hat{\mathbf{x}} \right\|^2 \mathbf{R}^{-1},
\]

(72)

\[
\delta \mathbf{x} = \sqrt{1 - \beta} \mathbf{U} \delta \mathbf{x}_{\text{var}} + \sqrt{\beta} \hat{\mathbf{X}}_{\text{ens}} - \mathbf{H} \delta \hat{\mathbf{x}}.
\]

(73)

\[
\nabla_{\hat{\mathbf{x}}} \hat{\mathbf{H}}^{\text{En4DVar}} = \left( \nabla_{\hat{\mathbf{x}}} \hat{\mathbf{H}}^{\text{En4DVar}} \right) \left( \nabla_{\hat{\mathbf{x}}} \hat{\mathbf{H}}^{\text{En4DVar}} \right) = \left( \hat{\mathbf{x}}_{\text{var}} - \sqrt{1 - \beta} \mathbf{U} \mathbf{H}^{-1} \mathbf{R}^{-1} \left( \mathbf{d} - \mathbf{H} \delta \hat{\mathbf{x}} \right) \right).
\]

(74)

where \(\mathbf{H}_k\) is defined in (31), \(d\) is defined after (33), and we define the augmented (hybridized) control vector as \(\hat{\mathbf{x}}_{\text{var}} = (\hat{\mathbf{x}}_{\text{var}} \hat{\mathbf{x}}_{\text{ens}})\), which has \(n + NM\) elements in total. Here \(\hat{\mathbf{x}}_{\text{ens}}\) is the \(N\)-element vector associated with the \(N\) ensemble members in \(\mathbf{X}^c\) combined with the \(M\) localization members in \(\mathbf{U}^c\) (61). The differences with the original form (35) due to localization (75) include the extended length of the control vector \((\mathbf{x}^c \rightarrow \hat{\mathbf{x}}_{\text{ens}})\), and \(X^b \rightarrow \hat{X}^b\).

7.3.2. L03 localization in hybrid 4DEnVar

The cost function for 4DEnVar with L03 localization (66) was used as the example in section 6.3. Incorporating the hybrid scheme from section 5.3 leads to the cost function, CVT and gradient formulae for this case given as:

\[
\hat{\mathbf{H}}^{\text{4DEnVar}} \left( \delta \hat{\mathbf{X}}_{\text{var}} , \hat{\mathbf{X}}_{\text{ens}} \right) = \frac{1}{2} \left\| \delta \hat{\mathbf{X}}_{\text{var}} \right\|^2 + \frac{1}{2} \left\| \hat{\mathbf{X}}_{\text{ens}} \right\|^2 + \frac{1}{2} \left\| \mathbf{d} - \mathbf{H} \delta \hat{\mathbf{x}} \right\|^2 \mathbf{R}^{-1},
\]

(78)

\[
\delta \mathbf{x} = \sqrt{1 - \beta} \mathbf{U} \delta \mathbf{x}_{\text{var}} + \sqrt{\beta} \hat{\mathbf{X}}_{\text{ens}} - \mathbf{H} \delta \hat{\mathbf{x}}.
\]

(79)

\[
\nabla_{\hat{\mathbf{x}}} \hat{\mathbf{H}}^{\text{4DEnVar}} = \left( \nabla_{\hat{\mathbf{x}}} \hat{\mathbf{H}}^{\text{4DEnVar}} \right) \left( \nabla_{\hat{\mathbf{x}}} \hat{\mathbf{H}}^{\text{4DEnVar}} \right) = \left( \hat{\mathbf{x}}_{\text{var}} - \sqrt{1 - \beta} \mathbf{U} \mathbf{H}^{-1} \mathbf{R}^{-1} \left( \mathbf{d} - \mathbf{H} \delta \hat{\mathbf{x}} \right) \right).
\]

(80)

where we define the augmented (hybridized) control vector for L03 as \(\hat{\mathbf{x}}_{\text{var}} = (\delta \hat{\mathbf{x}}_{\text{var}} \hat{\mathbf{x}}_{\text{ens}})\), which has \(n + N\) elements in total. Note that the ensemble part, \(\hat{\mathbf{x}}_{\text{ens}}\), comprises \(N\) \(n\)-element vectors (\(\hat{\mathbf{x}}_{\text{ens}}\) assembled here as columns of matrix \(\hat{\mathbf{X}}_{\text{ens}}\), so it should strictly be called a control matrix). The gradient of (75) with respect to \(\hat{\mathbf{X}}_{\text{ens}}\) (forming the \(n \times N\) sub-matrix) gives the lower part of (77), which has matrix elements

\[
\left[ \nabla_{\hat{\mathbf{x}}} \hat{\mathbf{H}}^{\text{4DEnVar}} \right]_{ij} = \hat{\mathbf{H}}^{\text{4DEnVar}} / \partial (\hat{\mathbf{x}}_{ij}).
\]
where $H$ is defined in (31), $d$ is defined after (33), and the hybridized control vector has the same structure as in section 7.3.1.

Apart from the 4D nature of the ensemble-related matrices like $X^b$, notice that $U$ (now underlined) now needs to be a 4D operator to make up for the lack of the linearized forecast model to propagate perturbations from $t = 0$. This is an important difference between hybrid En4DVar and hybrid 4DEnVar. The are possible ways to structure $U$ and some are listed below.

- Let $U$ be a persistence model of climatological perturbations (Desroziers et al., 2014; Lorenc et al., 2015):
  \[
  U = \begin{pmatrix} I & \vdots & I \end{pmatrix},
  \] (81)
  where $U$ is the $t = 0$ CVT as used in (17) in section 2.3. The preceding matrix contains $T + 1$ blocks, each comprising the $n \times n$ identity matrix. The contribution from $\sqrt{1-\beta U} \delta x_{t\text{ var}}$ in (79) then behaves as 3D-FGAT (section 2.1.3 and Poterjoy and Zhang, 2015) and the climatological component of the implied background-error covariance between any two state-vector components is independent of time. Even though this is classified as a 4D method, it does not exploit the climatological covariance propagation property of 4DVar. This is the simplest option which is used with most, if not all, applications of hybrid 4DEnVar at present.

- A more sophisticated form of $U$ is to extend a 3D covariance model used to represent $U$ to 4D. A strategy for modelling 3D background-error covariances is to exploit the normal modes of the system (e.g. Žagar et al., 2004), where the control vector represents coefficients that weight perturbations as linear combinations of 3D normal modes. This could in principle be extended to 4D (where columns of $U$ would be related to the evolving normal modes of the system) although would be limited by linearity of the forecast model.

- A further option is to replace the climatological term of 4DEnVar with that of En4DVar, but use a highly simplified model — e.g. with advection only, $M^{\text{adv}}$. The computational cost issues associated with use of linear models (and adjoints) would remain, but such a simple model would require little maintenance.

It might also be possible to propagate the localization square-root matrix in 4DEnVar (either with the full, linearized or simplified model). Column $i$ of the square-root matrix $U^C$ (section 6.2, call $u_i^C$) would then, e.g. comprise the pre-computed column vector $\chi^b_\delta = [\chi^b(0), M^\text{adv}_0 \chi^b(0), \ldots, M^\text{adv}_{T-1} \chi^b(T-1)]$. This, and ideas of P. Arbogast (2016; personal communication) are possible ways of dealing with the 4D aspects of localization.

### 7.4.2. L03 localization in hybrid 4DEnVar

The hybrid 4DEnVar equations with L03 localization emerge from an application of concepts in sections 5.3 and 6.3 to (40). The cost function, CVT and gradient formulae for this case are given as:

\[
\begin{align*}
\hat{p}^{\text{hybrid4DVar}}(\delta \chi_{t\text{ var}}, \hat{\mathbf{y}}) &= \frac{1}{2} \| \delta \chi_{t\text{ var}} \|^2 + \frac{1}{2} \| \mathbf{d} - H \delta \mathbf{x}_{t\text{ var}} \|^2, \\
\delta \mathbf{x} &= \sqrt{1 - \beta U} \delta \chi_{t\text{ var}} + \sqrt{\beta} \left( U \delta \mathbf{v} + U \mathbf{v}_{t\text{ var}} \right) N, \\
\nabla_{\mathbf{y}} \hat{p}^{\text{hybrid4DVar}} &= \left( \nabla \chi_{t\text{ var}}^{\text{hybrid4DVar}} \right) \hat{p}^{\text{hybrid4DVar}}, \\
\nabla_{\delta \mathbf{x}} \hat{p}^{\text{hybrid4DVar}} &= \left( \nabla \chi_{t\text{ var}}^{\text{hybrid4DVar}} \right) \hat{p}^{\text{hybrid4DVar}}.
\end{align*}
\] (82)

where the total control ‘vector’ $\hat{x}_b^c = (\delta \chi_{t\text{ var}}, \mathbf{c})$ has $n + n(T+1)$ elements in total. In these forms, virtually all objects are underlined, indicating that they are 4D. This is the way that localized hybrid 4DEnVar avoids the Jacobian of the forecast model (i.e. the forecast part of the problem is dealt with in the preparation of $\chi^b$). The gradient $\nabla_{\delta \mathbf{x}} \hat{p}^{\text{hybrid4DVar}}$ that appears (as $N$ repeated columns) in (84) is not found from (15); that equation gives $\nabla_{\delta \mathbf{x}} \hat{p}^{\text{hybrid4DVar}}$, the gradient with respect to the initial perturbation in 4D-Var. Instead $\nabla_{\delta \mathbf{x}}^{\text{hybrid4DVar}}$ is the following $n(T+1)$-element vector that does not require the model adjoint:

\[
\nabla_{\delta \mathbf{x}}^{\text{hybrid4DVar}} = \begin{pmatrix} \nabla_{\delta \mathbf{x}} \hat{p}^{\text{En4DVar}} \\
\vdots \\
\nabla_{\delta \mathbf{x}} \hat{p}^{\text{hybrid4DVar}} \\
\end{pmatrix},
\] (83)

where $d_i$ is defined by (12).

Note that $\hat{\mathbf{y}}$ is a much larger matrix than $\hat{\mathbf{y}}$ is in hybrid En4DVar with L03 localization. The other difficulty over En4DVar is that a 4D localization must be modelled via $U^C$ (discuss section 7.2.3). This, and the ‘U problem’ are important outstanding issues in the development of hybrid 4DEnVar (the discussion in section 7.4.1 still stands for L03 localization).

### 7.4.3. Comments on the Jacobian of the observation operator and on the implied localized background-error covariances in these systems

Hybrid 4DEnVar in (78) and (82) retains the Jacobian $H$. This could be avoided whenever $H \chi^b$ appears where it can be approximated with differences between runs of the nonlinear observation operator; e.g. for perturbations at time $t$ at column $i$ of $X^b$ the approximation would be

\[
\mathcal{H}_t \left[ \chi^b(t) + \delta \mathbf{x}^b(t) \right] = \mathcal{H}_t \left[ \chi^b(t) \right].
\]

There is no obvious way of avoiding $H$ elsewhere.

The ensemble part of the implied localized background-error covariances between localized En4DVar and localized 4DEnVar differ in the same way as those between the 3D and 4D versions of the EnKS (section 7.2.3). Having a 3D state vector (where future states are generated by the model within the DA), En4DVar has the implied covariance at time $t$: $\mathbf{P}^{\text{En4DVar}}_{t|t} = M^{\text{adv}}_t \mathbf{P}^{\text{En4DVar}}_{t|t}$ (quantities in the brackets are valid at $t = 0$), and having a 4D state vector, 4DEnVar has the implied covariance at time $t$:

$\mathbf{P}^{\text{hybrid4DVar}}_{t|t} = \mathbf{C}(t) \mathbf{P}^{\text{En4DVar}}_{t|t}$. This result is independent of whether
Figure 2. How the background-error covariance matrix evolves throughout the assimilation window for various 4D methods. (a) 4D-Var, (b) En4DVar (or the EnKS with 3D state vector), (c) 4DEnVar (or the EnKS with 4D state vector), (d) hybrid En4DVar and (e) hybrid 4DEnVar. For simplicity, model error is not considered, and $M$ is short for $M_0$, $t$. In hybrid 4DEnVar (e), $B$ is not evolved (as is current practice). The ellipses represent the background-error covariance ‘bubbles’ at times 0 and $t$, and the curves represent background ensemble members, where the thick black curve in each panel is the ensemble mean. In all Var techniques, only a single analysis trajectory is conventionally found.

the B05 or the L03 formulation is used. Figure 2 gives a pictorial representation of how the background-error covariances evolve throughout the assimilation window in each 4D scheme discussed in this article.

7.5. Summary of nomenclature

Figure 1 serves to summarize the nomenclature, including the names recommended by (Lorenc, 2013) and in this article, and alternative names. It also summarizes the basic features of the localized schemes including the number of control variables (the Figure caption gives more information). As is the case throughout this article, we recognise, but do not elaborate on, the spectrum of methods that count as an EnKF (e.g. ETKF, etc.).

8. Model error with (hybrid) En4DVar and 4DEnVar

No assimilation is optimum until it accounts correctly for all sources of error. Model error is an important source, which is a particular concern when the assimilation window is ‘long’. We have already considered model error with weak-constraint 4D-Var (section 2.1), and here we consider how model error might be represented in En4DVar and 4DEnVar.

8.1. Model error in (hybrid) En4DVar

Accounting for model error in methods that use the linearized model, e.g. hybrid En4DVar, may be achieved by augmenting the control variable with the variables associated with model error, $\delta \eta(1), \ldots, \delta \eta(T)$, adding a model error term as in (10), and modifying the prediction operator of the observations at time $t$. For instance hybrid En4DVar with B05 localization (F in Figure 1) has the modified model observations (based on (8)):

$$y_t^\tau = \mathcal{H}_t[M_0, t(x^\tau)] + H_t \left[ M_0, \left( \sqrt{1-\beta} U \delta \chi_{\text{var}} + \sqrt{\beta} \hat{x}_{\text{ens}} \right) \right] + \sum_{t=1}^T M_{t, \tau} \delta \eta(\tau).$$

The new terms accumulate the estimated model error contributions up to and including time $t$. A similar procedure is valid for L03 localization.**

8.2. Model error in 4DEnVar

Accounting for model error in methods that do not use the linearized model, like hybrid 4DEnVar, is more in line with the requirements of many operational centres, but is arguably harder to do and there are few examples in the literature. The key is to build flexibility into a scheme to relax the need for the solution to be synthesized from model trajectories over the window. One possibility, based around the B05 localization is to allow a different linear combination of members at each time (Amezcua et al., 2017; Goodliff et al., 2017). Instead of one control vector associated with the ensemble, $\overline{x}_{\text{ens}}$ in (79), there would be $T+1$

**In practice the preconditioned versions of the model error variables would be used (((18) of section 2.3)).
such vectors. The versions of (78) and (79) become (87) and (88) respectively,
\[
\mathcal{J}_{4D\text{EnVar}} \left[ \delta \chi_{\text{var}} \tilde{\chi}_{\text{ens}}(0), \ldots, \tilde{\chi}_{\text{ens}}(T) \right] = \frac{1}{2} \delta \chi_{\text{var}}^2 \mathbf{I} + \frac{1}{2} \sum_{t=0}^{T} \left[ \tilde{\chi}_{\text{ens}}(t)^2 \right] + \left\| \mathbf{y} - \mathbf{H} \mathbf{X}^b \right\|_2^2 + \frac{1}{2} \left( \sum_{t=0}^{T} \delta \chi_{\text{var}}^2 \right)^{\frac{1}{2}}, \tag{87}
\]
\[
\delta \mathbf{x} = \sqrt{1-\beta} \mathbf{U} \delta \mathbf{x}_{\text{var}} + \sqrt{\beta} \left( \begin{array}{c}
\tilde{\mathbf{X}}^b(0) \tilde{\mathbf{X}}_{\text{ens}}(0) \\
\vdots \\
\tilde{\mathbf{X}}^b(T) \tilde{\mathbf{X}}_{\text{ens}}(T)
\end{array} \right), \tag{88}
\]
where there are \( n+(T+1)NM \) elements to the control vector in total (or \( (T+1)NM \) for the non-hybrid version). The last term of (87) penalizes model trajectory misfits, which have covariance \( \mathbf{Q}_2 \).

There is an inconsistency in system (87) and (88), namely that model error has been accounted for in the part associated with \( \tilde{\chi}_{\text{ens}}(t) \), but not in the part associated with \( \delta \chi_{\text{var}} \) (the latter is akin to 3DFGAT). Clearly accounting for model error effectively in hybrid 4DEnVar will require a great deal of research.

Another point of interest is the nature of the 4D ensemble comprising \( \mathbf{X}^b \) (and hence the ‘localized’ version \( \tilde{\mathbf{X}}^b \)). Changing only the linear combination of members at each time step does not allow significant deviations from the ‘model attractor’, which would be required when the model is in error. An obvious way of dealing with this is to modify the way that the ensemble is prepared by allowing the trajectories to deviate stochastically. There are many well-documented ways of doing this (Palmer and Williams, 2010), e.g. using multiple physical parametrization schemes (Stensrud et al., 2000; Berner et al., 2011), using the stochastic kinetic energy backscatter scheme (Shutts, 2005), perturbing physics tendencies (Buiizza et al., 1999; Bouttier et al., 2012; Frenay et al., 2012), or perturbing physics parameters (Bowler et al., 2008; Hacker et al., 2011; Gebhardt et al., 2011; Vie et al., 2012; Baker et al., 2014). Timing errors may be represented by adding ensemble members with a slightly different validity time to the background state (Gustafsson et al., 2014). Including additive model error during model integration has been done by many authors in the context of the EnKF to good effect (e.g. Houtekamer et al., 2005; Hamill and Whitaker, 2005; Kalnay et al., 2007).

9. Which method to use, and how do the methods compare in use?

There is now a variety of known DA methods, so here we compare the merits and the relative performance of each as reported in the literature.

9.1. Factors to be considered

The decision on which DA method to use depends upon many factors, some of which are listed here.

- Parallelizability: Methods that do not require time integration within the DA algorithm have the edge on parallelizability, since the forward and adjoint integrations need to be done serially. Suitable methods include the EnKS (4D state vector), 4DEnVar, and (hybrid) 4DEnVar. Although the EnKS (3D state vector) and (hybrid) En4DVar do time integration within the DA algorithm this can be allocated to separate processors.

- Existing 4D-Var infrastructure. If linearized and adjoint versions of the forecast model are already available (e.g. used already in 4D-Var), then upgrading to (hybrid) En4DVar may well be beneficial (section 9.2).

- Systems with highly nonlinear processes or frequently upgraded forecast models. It is well known that highly nonlinear models (e.g. with moist processes, switches, etc.) present difficulties with linearizing. There are also complications over linearizing models that have, for example, semi-Lagrangian dynamical cores. Anyway most models are time consuming to linearize and are expensive to run, so methods that avoid the need for linearized and adjoint models are desirable. Methods like the EnKS (4D state vector) and (hybrid) 4DEnVar are desirable in this case.

- Known background-error statistics. In systems that have background-error covariance statistics that are easily modelled and whose variability is negligible then 4D-Var (strong- and weak-constraint) remains useful. This is not normally the case for weather forecasting, even though 4D-Var is historically successful.

- Model error is important. If the model error accumulated over the required assimilation time window is significant compared to other sources of error that are accounted for already, then a method with model error should be considered. Of the methods shown in Figure 1, weak-constraint 4D-Var (section 2.1.2), weak-constraint (hybrid) EnVar (section 8), and the EnKF account for model error. There are other possibilities, e.g. methods that use a perfect model, but absorb model error into the \( \mathbf{R}_c \)-matrix (Howes, 2016).

- Deterministic or probabilistic. Probabilistic forecasts are increasingly in demand (e.g. Palmer, 2012). If the information contained in a single forecast is not adequate, and measures of analysis and forecast uncertainty are required, then the EnKF, EnKS (and other flavours) are useful. Of course (hybrid) En4DVar and 4DEnVar need an ensemble system running alongside them anyway. It is possible to run these methods many times, by perturbing inputs like the background state and the observations (Liu and Xiao, 2013) or use a scheme like EVIL (section 5.5), which does not need a separate ensemble.

- Non-instantaneous or ‘rate-of-change’ observations. If observations are to be assimilated that are a function of the state at multiple times (e.g. precipitation accumulation) then methods that have 4D state vectors are the natural choice, e.g. 4D-Var, EnKS and (hybrid) En4DVar/4DEnVar. The equations describing these systems will need to be modified to allow for such observations.

9.2. Systems studied in literature and their relative performance

There are now a large number of studies testing and comparing the DA methods discussed in this article. In this section we summarize the main findings. We consider three kinds of study:

(i) investigations in simple models (Table 2),
(ii) investigations with operational-scale models, but with simulated observations (Table 3), and
(iii) as (ii), but with real data (Table 4).

The nomenclature used here is consistent with Figure 1, even though some articles use their own.

The main outcomes are that the 4D methods tend to be better than 3D methods, hybrid methods tend to be better than pure Var, ensemble, or EnVar methods, (and hybrids are more robust when \( N \) is small) but 4DEnVar still needs some development effort if it is to become better than En4DVar in all systems; only one study found hybrid 4DEnVar gave better performance than hybrid En4DVar (Gustafsson and Bojorava, 2016).
2014). However 4DEnVar is much cheaper to run than En4DVar so, if the performance of 4DEnVar could be improved through further work, it could then allow the cost savings in computation to be shifted to models with higher resolution and with more ensemble members. The results from the simplified systems are in general in agreement with those from the full systems, confirming the value of the simplified studies, and all studies have helped ultimately guide centres to decide on which methods to use operationally.

10. Summary of variational/ensemble methods used in NWP centres

It is often reported that most, if not all, NWP centres have the capability to combine their ensemble and variational systems. Table 5 lists the capabilities of a number of leading (inter)national NWP centres and research groups. Obviously this list is a snapshot and the true capabilities will change as developments progress.

11. Summary and outlook

Data assimilation needs to work well to serve NWP. State-of-the-art Var systems assume that background information obeys Gaussian statistics, with a covariance, \( P^b \), estimated from a quasi-static representation, from an ensemble, or a combination of both. The use of ensemble information in Var systems using the raft of methods discussed in this article to give appropriate flow dependence of \( P^b \), has been challenging, but largely successful.

This article is intended to be a pedagogical review of Var methods and the practical ways of introducing ensemble information into them to take advantage of the efficiency of Var and the flow-dependence of an ensemble, while minimizing their drawbacks (Buehner et al., 2015b, and section 1). Broadly speaking, this is presently done in three ways.

1. Use the ensemble to recalibrate the (co)variances of the \( B_b \)-matrix used as an approximation for \( P^b \) in 4D-Var (section 5.1).
2. Use the ensemble to define \( P^b \) by preconditioning the control vector on the ensemble itself (so-called pure EnVar methods). This describes the analysis increment as a linear combination of ensemble perturbations (where the control vector contains the coefficients) and implies a background-error covariance matrix equal to the sample covariance of the ensemble, \( P^b_{(N)} \). This can be done using methods that exploit (En4DVar) or avoid (4DEnVar) the linear/adjoint forecast model (section 4).
3. Average \( B_b \) with \( P^b_{(N)} \) (so-called hybrid EnVar). This is usually done with an augmented control vector comprising a part that is associated with the static part (as used in traditional Var) and a part that is associated with the ensemble (the same as that used in 1. here). Alternative methods are reviewed such as the a hybrid gain method, the EVIL method, and an ensemble reduced-rank Kalman filter (section 5.3).

These methods use the Var machinery to solve a DA problem and most require a separate ensemble system, which many NWP centres have anyway for ensemble prediction purposes. Practical methods that use ensemble information rely on localization to mitigate some of the effects of sampling error. The mathematical complexities that localization brings can mask the workings of the equations and so the methods are shown first without localization, but it is introduced later with two Schur product methods documented by Buehner (2005) (B05) and Lorenc (2003) (L03).

EnVar using 2. and 3. here has been shown to be fruitful and the key reasons are summarized as:

- (Hybrid) En4DVar and (hybrid) 4DEnVar are efficient ways of finding the deterministic (single) analysis.
- The methods provide flow-dependency to the \( P^b \) used in the DA.

Table 2. Studies made of some of the techniques in this article applied to simplified models.

| Article | Methods tested | Model | \( N \) | Best performance | Comments |
|---------|----------------|-------|-------|-----------------|---------|
| Etherton and Bishop (2004) | 3D-Var, hybrid En3DVar | 2D turbulent barotropic vorticity | 16, 64 | Hybrid En3DVar | Model error included, and using sub-optimal ensemble |
| Hunt et al. (2004) | EnKF, 4DEnVar | Lorenz 96 (Lorenz, 1996) | 40 | 4DEnVar | -- |
| Wang et al. (2007a) | Hybrid OI, EnKF | Two-layer primitive equation | 5, 20, 50 | hybrid OI for small \( N \) | -- |
| Liu et al. (2008) | 3/4D-Var, EnKF, 4DEnVar | 1D shallow-water | 50 | 4DEnVar comparable to 4D-Var | 4DEnVar cheaper than 4D-Var |
| Tian et al. (2008) | 4D-Var, EnKF, 4DEnVar | NCAR Land (Oleson et al., 2004), Lorenz 63 (Lorenz et al., 1963) | 60 | 4DEnVar | Based on ‘Proper Orthogonal Decomposition’ |
| Wang et al. (2009) | EnSRF, hybrid En3DVar | Two-layer primitive equation | 50, 200 | Hybrid En3DVar | Includes model error due to unresolved scales |
| Penny (2014) | LETKF, hybrid covariance, hybrid gain | Lorenz 96 | 10, 40 | Hybrid En4DVar | Hybrid less sensitive to window length and \( N \) |
| Goodliff et al. (2015) | ETKF, ETKS, En4DVar, 4DVar, hybrid En4DVar | Lorenz 63 | 5, 10, 20, 50, 100 | ETKS/ hybrid gain | Hybrids allow smaller \( N \) |
| Poterjoy and Zhang (2015) | Hybrid En4DVar, hybrid 4DVar | Lorenz 96 | 5, 10, 20, 40 | Hybrid En4DVar | Methods similar in data-dense areas |
Table 3. Studies made of some of the techniques in this article applied to operational-scale models, but with simulated or severely reduced observations.

| Article | Methods tested | Model | N  | Best performance | Comments |
|---------|----------------|-------|----|------------------|----------|
| Wang et al. (2008a) | 3D-Var, hybrid En3DVar | WRF (Skamarock et al., 2005) | 50 | Hybrid En3DVar | Benefit of hybrid especially over data-sparse regions |
| Liu et al. (2009) | 3/4DEnVar | WRF | 37 | 4DEnVar | Humidity particularly sensitive to sampling error |
| Buehner et al. (2010a) | 3D-FGAT, 4D-Var, En3/4DVar, 4DEnVar | GEM (EC) (Cote et al., 1998) | 72, 96 | – | Single obs./column expts. |
| Kleist and Ide (2015a) | 3D-Var, hybrid 3DEnVar | NCEP GFS (NCEP, 2015) | 80 | Hybrid 3DEnVar | – |
| Kleist and Ide (2015b) | 4D-Var, 4DEnVar, hybrid 4DEnVar | NCEP GFS | 40, 80 | Hybrid 4DEnVar | Improvement from 3D to 4D less than improvement in (2015a) |

Table 4. Studies made of some of the techniques in this article applied to operational-scale models with real observations.

| Article | Methods tested | Model | N  | Best performance | Comments |
|---------|----------------|-------|----|------------------|----------|
| Wang et al. (2008b) | 3D-Var, hybrid En3DVar | WRF | 50 | Hybrid En3DVar | – |
| Buehner et al. (2010b) | 3D-FGAT, 4D-Var, En3/4DVar, 4DEnVar | GEM (EC) | 96 | En4DVar | One month experiment |
| Hamill et al. (2011b) | 3D-Var, EnKF, hybrid En3DVar | NCEP GFS | ? | EnKF/hybrid En3DVar | – |
| Wang (2011) | 3D-Var, hybrid 3DEnVar | WRF | 32 | Hybrid 3DEnVar | Hybrid was able to adjust hurricane position |
| Clayton et al. (2012) | 3/4D-Var, hybrid En3/4DVar | Global UM (Davies et al., 2005) | 23 | Hybrid En4DVar | 4D has marginal improvement over 3D in hybrid |
| Zhang and Zhang (2012) | 4D-Var, EnKF, hybrid En4DVar | WRF | 10, 40 | Hybrid En4DVar | Period of summer convection |
| Buehner et al. (2013) | 3/4D-Var, hybrid 3DEnVar, hybrid 4DEnVar | GEM (EC) | 192 | Hybrid 4DEnVar | Hybrid 4DEnVar is viable method |
| Kuhl et al. (2013) | 4D-Var, hybrid En4DVar | NOGAPS (Hogan et al., 1991) | 80 | Hybrid En4DVar | – |
| Liu and Xiao (2013) | 3D-Var, 3D-FGAT, 3/4DEnVar | WRF | 38 | 4DEnVar | Includes ensemble of En4DVar |
| Schwartz et al. (2013) | 3D-Var, hybrid En3DVar | WRF | 32 | Hybrid En3DVar | Hybrid best for one outer loop, 3D-Var often best for three outer loops |
| Wang et al. (2013) | 3D-Var, EnKF, En3DVar, hybrid En3DVar | NCEP GFS | 80 | En3DVar | Tested one-way and two-way coupling |
| Zhang et al. (2013) | 3D-Var, EnKF, hybrid En3DVar, hybrid En4DVar | WRF | 10, 20, 40, 80 | Hybrid En4DVar | Hybrid En3DVar with half N comparable to EnKF |
| Pan et al. (2014) | 3D-Var, hybrid En3DVar | WRF | 40 | Hybrid En3DVar | Tested one-way and two-way coupling |
| Poterjoy and Zhang (2014) | 4D-Var, EnKF, hybrid En3DVar | WRF | 60 | Hybrid En4DVar | Tested period of tropical cyclogenesis and hurricane path |
| Gustafsson et al. (2014) | 3D-Var, 4D-Var, hybrid En3DVar, hybrid En4DVar | HIRLAM Unden et al. (2002) | 20, 40 | Hybrid En4DVar | Marginal gain only |
| Gustafsson and Bojarova (2014) | 4D-Var, hybrid En4DVar, hybrid 4DEnVar | HIRLAM | 20 | Hybrid 4DEnVar | – |
| Schwartz and Liu (2014) | 3D-Var, EnSRF, hybrid En3DVar | WRF | 50 | Hybrid En3DVar | Hybrid best for precip. positions, but methods otherwise comparable |
| Schwartz et al. (2014) | 3D-Var, EnSRF, hybrid En3DVar | WRF-Chem (Schwartz et al., 2012) | 50 | Hybrid En3DVar | Aerosol assimilation |
| Wang and Lei (2014) | En3/4DVar | NCEP GFS | 80 | 4DEnVar | Cost savings allow higher resolution |
| Buehner et al. (2015a) | 4D-Var, hybrid 4DEnVar | 4DEnVar | 256 | Hybrid 4DEnVar | Hybrid 4DEnVar an order of magnitude cheaper than 4D-Var |
| Caron et al. (2015) | 4D-Var, hybrid 4DEnVar | Regional version of GEM (EC) | 192, 256 | Hybrid 4DEnVar | – |
| Lorenc et al. (2015) | hybrid En3/4DVar, hybrid 3/4DEnVar | Global UM | 44 | Hybrid 4DEnVar | – |

- Although the En4DVar and 4DEnVar solutions are nominally equivalent to the ensemble mean of the analogous EnKF system (these methods and their hybrid counterparts allow an extension to the 4D domain), they allow access to the full range of observation operators, and allow other on-line features to be facilitated, namely variational bias correction and variational initialization.

- In (hybrid) En4DVar, the linearized forecast model (and its adjoint) are used (as in 4D-Var), but this is avoided in (hybrid) 4DEnVar. This is the purpose of developing the latter as the linearization of the forecast model is time-consuming to construct and maintain, is costly to run, and is limited by the nature of the physical processes, especially parametrizations with switches.
As these methods use information from an ensemble, there is still sensitivity to the ensemble size, \( N \), especially when it is small, but hybrid methods have been shown to reduce the sensitivity to \( N \).

(2) Hybrid EnVar methods are competitive with, or outperform, the pure Var or ensemble counterparts.

There are outstanding issues that still need to be solved. For instance in hybrid 4D-EnVar the 4D ensemble is averaged with 3D-Var, yet may be further apart, and so the localization of the 3D-Var analysis may be wrongly damped by localization. Such a problem would be wrongly damped by localization. Such a problem can lead to better forecasts of moisture-related quantities like precipitable water and equitable threat score for precipitation (Caron et al., 2013). It also should be noted that all of the methods may be suitable for application to other geophysical problems like coupled systems and estimating surface fluxes of chemical species.

Finally a note on nomenclature. As the number of methods grows, it becomes important to label each succinctly and accurately. Lorenz (2013) has attempted to do this, and authors have tended to follow. We support the view that the ‘hybrid’ label should be used only for methods that combine ensemble and static aspects in their background errors, but we do think that it should be necessary to specify this when a method does use information from both sources (unlike point 5. in Lorenc, 2013); see Figure 1. It is also useful to distinguish between 3D and 4D pure ensemble methods. In this case it is sensible to specify 4D ensemble methods as smoothers (EnKS) rather than as filters (EnKF).

Furthermore hybrid 3D-EnVar (unlike 3D-Var) can improve the position of a hurricane (Wang, 2011). However, there are some less encouraging studies showing that, for example, humidity is particularly sensitive to sampling error and is not improved with 4DEnVar over 3DEnVar (Liu et al., 2009). Operational centres like the Met Office use EnVar systems for their global models, but have yet to develop similar systems for their regional models. It remains an important question how much these techniques can improve analyses and forecasts in their high-resolution regional models. Environment Canada though has already shown that 4DEnVar can lead to better forecasts of moisture-related quantities like precipitable water and equitable threat score for precipitation (Caron et al., 2013). It should also be noted that all of the methods may be suitable for application to other geophysical problems like coupled systems and estimating surface fluxes of chemical species.

There is scope to test the methods covered in this review to situations where pure Var can struggle. One application is quantitative precipitation forecasting at the convective scale. There are already some encouraging results in this area: e.g. hybrid En4DVar can outperform pure ensemble and Var methods in summer convection (Zhang and Liu, 2012), and hybrid En3DVar is better than pure ensemble and Var methods for analysing precipitation positions (Schwartz and Liu, 2014).
Acknowledgements

The author is supported by NERC via the NCEO. I am constantly learning new things about DA from my colleagues, but I would especially like to thank people who, over the years, have given inspiring talks or I have had interesting conversations with concerning the subject matter of this article, namely Javier Amiezca, Jeff Anderson, Dale Barker, Craig Bishop, Neil Bowler, Mark Buehner, Jean-Francois Caron, Mike Cullen, Martin Ehrendorfer, David Fairbairn, Jonathan Flowerdew, Mike Fisher, Michael Goodliff, Gordon Inverarity, Daryl Kleist, Amos Lawless, David Livings, Andrew Lorenz, Stefano Migliorini, Chris Snyder, Peter Jan van Leeuwen, Marek Wlasak and Fuqing Zhang. Thanks also to the anonymous reviewers of this article for providing important comments.

Appendices

Appendix A: The vector derivative

Consider a quadratic function of the n-element vector x:

\[ f(x) = \frac{1}{2} (\mathbf{H}x + y)^T S^{-1} (\mathbf{H}x + y), \]

where y (p-elements), \( \mathbf{H} \) (p \times n-elements) and the symmetric matrix \( S \) (p \times p-elements). Derivatives of such a function appear frequently, and so we derive a general result. The vector derivative is the following n-element vector:

\[ \nabla_f = \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right), \]

where \( x_i \) is the \( i \)th component of \( x \). First decompose (A1) into a function of its components \( x_i \):

\[ f(x) = \frac{1}{2} \sum_{i=1}^{p} \left( \sum_{j=1}^{n} \mathbf{H}_{ij} x_j + y_j \right) \left( \sum_{k=1}^{n} \mathbf{H}_{ki} x_k + y_k \right). \]  

(A3)

Differentiating this mechanically with respect to one arbitrary element \( x_{ij} \) gives, by the product rule,

\[ \frac{\partial f}{\partial x_{ij}} = \frac{1}{2} \left[ \sum_{j=1}^{p} \sum_{k=1}^{n} \mathbf{H}_{ij} \frac{\partial x_j}{\partial x_{ij}} \mathbf{H}_{kj} \right] + \frac{1}{2} \left[ \sum_{j=1}^{p} \sum_{k=1}^{n} \mathbf{H}_{ij} \frac{\partial x_k}{\partial x_{ij}} \mathbf{H}_{kj} \right] \]

\[ = \frac{1}{2} \left[ \sum_{j=1}^{p} \mathbf{H}_{ij} \sum_{k=1}^{n} \mathbf{H}_{kj} x_k + y_k \right] + \frac{1}{2} \left[ \sum_{j=1}^{p} \mathbf{H}_{ij} \sum_{k=1}^{n} \mathbf{H}_{kj} x_j + y_j \right]. \]  

(A4)

In the second line of the last expression, re-index the summations as \( i \leftrightarrow j \) and \( j \rightarrow k \). Noting that \( (S^{-1})_{ij} = (S^{-1})_{ji} \), this shows that the two lines in the last expression are identical, so:

\[ \frac{\partial f}{\partial x_{ij}} = \frac{1}{2} \left[ \sum_{j=1}^{p} \mathbf{H}_{ij} \sum_{k=1}^{n} (S^{-1})_{ji} \mathbf{H}_{kj} x_k + y_k \right] + \frac{1}{2} \left[ \sum_{j=1}^{p} \mathbf{H}_{ij} \sum_{k=1}^{n} (S^{-1})_{ij} \mathbf{H}_{kj} x_j + y_j \right]. \]  

(A5)

This can be confirmed to be the decomposition of the \( n \)th element of the matrix expression:

\[ \nabla_f = \mathbf{H}^T S^{-1} (\mathbf{H}x + y). \]  

(A6)

Appendix B: The matrix derivative

Consider a quadratic function of the \( n \times N \) matrix V:

\[ f(V) = \frac{1}{2} \left( \mathbf{H} (\mathbf{X} \circ [\mathbf{V}]) \mathbf{1}_N + y \right)^T S^{-1} \left( \mathbf{H} (\mathbf{X} \circ [\mathbf{V}]) \mathbf{1}_N + y \right), \]  

(B1)

where \( y \) (p-elements), \( \mathbf{H} \) (p \times n-elements), \( \mathbf{X} \) (n \times N-elements), \( \mathbf{U} \) (n \times n-elements), the symmetric matrix \( S \) (p \times p-elements), and \( \mathbf{1}_N \) (N-elements, each of value 1). This is the structure of cost functions that use the L03 formulation of localizing ensemble-derived background-error covariances ((66) in section 6.3). The matrix derivative is defined as

\[ \nabla_V f = \left( \frac{\partial f}{\partial \mathbf{V}_{i1}}, \ldots, \frac{\partial f}{\partial \mathbf{V}_{iN}} \right), \]  

(B2)

where \( \mathbf{V}_{ij} \) is the \( ij \)th element of \( V \). In order to derive a matrix expression for this gradient, we use the chain rule. Let \( \mathbf{v} = \left[ \mathbf{X} \circ [\mathbf{V}] \right] \mathbf{1}_N \), where \( \mathbf{v} \) has \( N \) elements. \( \nabla_V f \) can be found using the vector derivative result in Appendix A:

\[ \nabla_V f = \mathbf{H}^T S^{-1} \left( \mathbf{H} \mathbf{v} + y \right). \]  

(B3)

The chain rule allows the derivative with respect to \( \mathbf{Z} \) to be found:

\[ \frac{\partial f}{\partial \mathbf{Z}_{ij}} = \sum_{i=1}^{N} \frac{\partial \mathbf{v}_i}{\partial \mathbf{Z}_{ij}} \frac{\partial f}{\partial \mathbf{v}_i} \]  

(B4)

where \( \frac{\partial \mathbf{v}_i}{\partial \mathbf{Z}_{ij}} \) appears in (B4) is the \( ij \)th component of (B3), and \( \frac{\partial \mathbf{v}_i}{\partial \mathbf{Z}_{ij}} \) can be found by first expanding out the definition of \( \mathbf{v} \) in terms of \( \mathbf{V} \) given above:

\[ \mathbf{v}_i = \sum_{j'=1}^{N} \mathbf{X}_{ij'} \mathbf{U}_{ij'} \mathbf{v}_{ij'}, \]  

(B5)

so

\[ \frac{\partial \mathbf{v}_i}{\partial \mathbf{V}_{ij}} = \sum_{j'=1}^{N} \mathbf{X}_{ij'} \mathbf{U}_{ij'} \frac{\partial \mathbf{v}_{ij'}}{\partial \mathbf{V}_{ij}} = \sum_{j'=1}^{N} \mathbf{X}_{ij'} \mathbf{U}_{ij'} \delta_{ij'}. \]

(B6)

Substituting this into the chain rule gives:

\[ \frac{\partial f}{\partial \mathbf{Z}_{ij}} = \sum_{i=1}^{N} \mathbf{X}_{ij} \mathbf{U}_{ij} \frac{\partial f}{\partial \mathbf{v}_i} = \mathbf{X}_i \mathbf{U}_i, \]

(B7)

which can be confirmed to be the \( ij \)th element of the matrix expression:

\[ \nabla_V f = \mathbf{U}^T \left[ \mathbf{X} \circ [\nabla_V f \cdots \nabla_V f] \right], \]  

(B8)

where \( \nabla_V f \) is the \( n \times N \) matrix of repeated columns of \( \nabla_V f \).
Appendix C: The variational form of the Ensemble Kalman Smoother

The correspondence between the EnKS formula (29) and the minimizing state of (35) is demonstrated here. The gradient of (35) with respect to \( x_{\text{ens}} \), is given in (37). Setting this to zero for the minimum gives the following solution:

\[
\mathbf{x}_{\text{ens}} = \left[ I + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{R}^{-1} \mathbf{H}_m \mathbf{X}^b \right]^{-1} \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{R}^{-1} \mathbf{d} \tag{C1}
\]

(Recall the definitions \( \mathcal{H}_m \) and \( \mathbf{H}_m \) in (31), \( \mathbf{d} \) is defined after (35), \( \mathbf{R} \) is the time-ordered observation-error covariance matrix, and \( \mathbf{X}^b \) is the matrix whose columns comprise the normalized ensemble of 3D perturbations, \( (\mathbf{x}^b_{(i)} - \bar{\mathbf{x}}^b) / \sqrt{N-1} \)). The model space increment is found from (36):

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ I + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{R}^{-1} \mathbf{H}_m \mathbf{X}^b \right]^{-1} \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{R}^{-1} \mathbf{d} \tag{C2}
\]

The standard Sherman–Morison–Woodbury identity (e.g. Appendix A of Ehrendorfer, 2007):

\[
\left( \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right) \mathbf{B} = \mathbf{H}^T \mathbf{R}^{-1} \left( \mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R} \right) \tag{D3}
\]

applied to (C2) with \( \mathbf{B} \rightarrow \mathbf{I}, \mathbf{H} \rightarrow \mathbf{H}_m \mathbf{X}^b \), and \( \mathbf{R} \rightarrow \mathbf{R} \) gives:

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{C3}
\]

The EnKS (29) produces an ensemble of analyses. Taking the mean gives:

\[
\bar{\mathbf{x}} = \bar{\mathbf{X}}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{D4}
\]

This shows that, as long as the ensemble mean is close to the real observation mean, the ‘localization’ perturbations – see below – are the same. The Covariance matrix of \( \mathbf{x}^b \) as defined from (D3) gives:

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{D5}
\]

This shows that, as long as the ensemble mean is close to the real observation mean, the ‘localization’ perturbations – see below – are the same. The Covariance matrix of \( \mathbf{x}^b \) as defined from (D3) gives:

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{D5}
\]

This shows that, as long as the ensemble mean is close to the real observation mean, the ‘localization’ perturbations – see below – are the same. The Covariance matrix of \( \mathbf{x}^b \) as defined from (D3) gives:

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{D5}
\]

This shows that, as long as the ensemble mean is close to the real observation mean, the ‘localization’ perturbations – see below – are the same. The Covariance matrix of \( \mathbf{x}^b \) as defined from (D3) gives:

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{D5}
\]

This shows that, as long as the ensemble mean is close to the real observation mean, the ‘localization’ perturbations – see below – are the same. The Covariance matrix of \( \mathbf{x}^b \) as defined from (D3) gives:

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{D5}
\]

This shows that, as long as the ensemble mean is close to the real observation mean, the ‘localization’ perturbations – see below – are the same. The Covariance matrix of \( \mathbf{x}^b \) as defined from (D3) gives:

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{D5}
\]

This shows that, as long as the ensemble mean is close to the real observation mean, the ‘localization’ perturbations – see below – are the same. The Covariance matrix of \( \mathbf{x}^b \) as defined from (D3) gives:

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{D5}
\]

This shows that, as long as the ensemble mean is close to the real observation mean, the ‘localization’ perturbations – see below – are the same. The Covariance matrix of \( \mathbf{x}^b \) as defined from (D3) gives:

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{D5}
\]

This shows that, as long as the ensemble mean is close to the real observation mean, the ‘localization’ perturbations – see below – are the same. The Covariance matrix of \( \mathbf{x}^b \) as defined from (D3) gives:

\[
\delta \mathbf{x} = \mathbf{X}^b \left[ \mathbf{H}_m \mathbf{X}^b + \left( \mathbf{H}_m \mathbf{X}^b \right)^T \mathbf{H}_m \right]^{-1} \mathbf{d} \tag{D5}
\]

This shows that, as long as the ensemble mean is close to the real observation mean, the ‘localization’ perturbations – see below – are the same. The Covariance matrix of \( \mathbf{x}^b \) as defined from (D3) gives:
The left-hand side is matrix element $k, k'$ of $\langle \delta x \delta x^T \rangle$ and the right-hand side of the last line is matrix element $k, k'$ of $P_{(N)} \cdot C$. This shows that the error covariance matrix implied by the CVT is the error covariance matrix sampled from the forecast ensemble, $P_{(N)}$ localized with $C$.

References

Amezcu J, Goodliffe M, van Leeuwen PJ. 2017. A weak-constraint 4DEnsembleVar. Part I: formulation and simple model experiments. Tellus 69A: 1271564, doi: 10.1002/2016600076.1271564.

Anderson JL. 2001. An ensemble adjustment filter for data assimilation. Mon. Weather Rev. 129: 2884–2903.

Anderson JL, Anderson SL. 1999. A Monte Carlo implementation of the ensemble Kalman filter for on-line Bayesian estimation. Mon. Weather Rev. 129: 2884–2903.

Bonavita M, Torrisi L, Marcucci F. 2008. The ensemble Kalman filter in an operational regional NWP system: Preliminary results with real observations. Q. J. R. Meteorol. Soc. 134: 1733–1744.

Buehner M, Charron M. 2007. Spectral and spatial localization of background-error correlations using a wavelet formulation in 4D-Var. Mon. Weather Rev. 135: 1414–1436.

Buehner M, He B. 2010a. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. Mon. Weather Rev. 138: 1550–1566.

Buehner M, Houtekamer PL, Charette C, Mitchell HL, He B. 2010b. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. Mon. Weather Rev. 138: 1567–1586, doi: 10.1175/2010MWR3158.1.

Buehner M, Morneau I, Charette C. 2013. Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. Nonlin. Proc. Geophys. 20: 669–682.

Burgers M, McIegget-Cowan R, Beaulne A, Charette C, Garand L, Heidtette L, Grime L, larochelle S, Morneau SR, Morneau J, Zadra A. 2015a. Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. Mon. Weather Rev. 143: 2352–2359, doi: 10.1175/MWR-D-14-00354.1.

Burgers G, van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126: 1908–1925.

Burgers G, van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126: 1908–1925.

Buehner M, Morneau I, Charette C. 2013. Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. Nonlin. Proc. Geophys. 20: 669–682.

Buehner M, McIegget-Cowan R, Beaulne A, Charette C, Garand L, Heidtette L, Grime L, larochelle S, Morneau SR, Morneau J, Zadra A. 2015a. Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. Mon. Weather Rev. 143: 2352–2359, doi: 10.1175/MWR-D-14-00354.1.

Burgers G, van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126: 1908–1925.

Buehner M, Morneau I, Charette C. 2013. Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. Nonlin. Proc. Geophys. 20: 669–682.

Buehner M, McIegget-Cowan R, Beaulne A, Charette C, Garand L, Heidtette L, Grime L, larochelle S, Morneau SR, Morneau J, Zadra A. 2015a. Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. Mon. Weather Rev. 143: 2352–2359, doi: 10.1175/MWR-D-14-00354.1.

Burgers G, van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126: 1908–1925.

Buehner M, Morneau I, Charette C. 2013. Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. Nonlin. Proc. Geophys. 20: 669–682.

Buehner M, McIegget-Cowan R, Beaulne A, Charette C, Garand L, Heidtette L, Grime L, larochelle S, Morneau SR, Morneau J, Zadra A. 2015a. Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. Mon. Weather Rev. 143: 2352–2359, doi: 10.1175/MWR-D-14-00354.1.

Burgers G, van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126: 1908–1925.

Buehner M, Morneau I, Charette C. 2013. Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. Nonlin. Proc. Geophys. 20: 669–682.

Buehner M, McIegget-Cowan R, Beaulne A, Charette C, Garand L, Heidtette L, Grime L, larochelle S, Morneau SR, Morneau J, Zadra A. 2015a. Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. Mon. Weather Rev. 143: 2352–2359, doi: 10.1175/MWR-D-14-00354.1.
CLAYTON AM, LORENC AC, BARKER DM. 2012. Operational implementation of 4D-Var using an incremental approach. Q. J. R. Meteorol. Soc. 138: 1279–1294.

DEE DP. 2002. ‘An adaptive scheme for cycling background-error variances using 4D-Var’. Q. J. R. Meteorol. Soc. 128: 2097–2110, doi:10.1002/qj.475.

GUSTAFSSON N, BOJAROVA J. 2014. Four-dimensional ensemble variational (4D-En-Var) data assimilation for the High Resolution Limited Area Model (HIRLAM). Nonlin. Proc. Geophys. 21: 745–762, doi:10.5194/npg-21-745-2014.

HABEN S, LAWLESS AS, NICHOLS NK. 2011. Conditioning of incremental variational data assimilation, with application to the Met Office system. Tellus 63A: 782–792.

HACKER JP, SNYDER C, HA S-Y, POCERNICH M. 2011. Linear and non-linear response to parameter variations in a mesoscale model. Tellus 63A: 429–444.

HAMILT DM. 2006. Ensemble-based atmospheric data assimilation. In Predictability of Weather and Climate, Palmer TN, Hagedorn R. (eds.). Cambridge University Press, Cambridge, UK.

HAMILT DM, SNYDER C. 2000. A hybrid ensemble Kalman filter-3D variational assimilation scheme. Mon. Weather Rev. 128: 2959–2971.

HAMILT DM, WHITAKER JS. 2005. Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. Mon. Weather Rev. 133: 3132–3147.

HAMILT DM, WHITAKER JS, SYNDER C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Weather Rev. 129: 2776–2790.

JEPSEN PL, MITCHELL HL. 1998. Global ensemble predictions of 2009’s tropical cyclones initialized with an ensemble Kalman filter. Mon. Weather Rev. 126: 1143–1162.

JEPSEN PL, MITCHELL HL. 2001. Ensemble Kalman filter. Q. J. R. Meteorol. Soc. 127: 1443–1462.

JEPSEN PL, MITCHELL HL. 2005. Ensemble Kalman filter. Q. J. R. Meteorol. Soc. 131: 3209–3235.

JEPSEN PL, ZHANG F. 2016. Review of the ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 144: 4489–4532.

JEPSEN PL, LEFRAVR E, DEROE J, RITCHIE H, MITCHELL HL. 1995. A system simulation approach to ensemble prediction. Mon. Weather Rev. 123: 1225–1242.

JEPSEN PL, PIIKINEN G, BUHNER M, CHARRON M, SPACEK L, HANSEN B. 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. Mon. Weather Rev. 133: 604–620.

JEPSEN PL, DENG X, MITCHELL HL, BAEK S-J, GAGNON N. 2014. Higher resolution in an operational ensemble Kalman filter. Mon. Weather Rev. 142: 1143–1162.

JEPSEN PL, PIIKINEN G, BUHNER M, CHARRON M, SPACEK L, HANSEN B. 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. Mon. Weather Rev. 133: 604–620.

JEPSEN PL, DENG X, MITCHELL HL, BAEK S-J, GAGNON N. 2014. Higher resolution in an operational ensemble Kalman filter. Mon. Weather Rev. 142: 1143–1162.

JEPSEN PL, PIIKINEN G, BUHNER M, CHARRON M, SPACEK L, HANSEN B. 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. Mon. Weather Rev. 133: 604–620.

JEPSEN PL, DENG X, MITCHELL HL, BAEK S-J, GAGNON N. 2014. Higher resolution in an operational ensemble Kalman filter. Mon. Weather Rev. 142: 1143–1162.

JEPSEN PL, PIIKINEN G, BUHNER M, CHARRON M, SPACEK L, HANSEN B. 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. Mon. Weather Rev. 133: 604–620.

JEPSEN PL, DENG X, MITCHELL HL, BAEK S-J, GAGNON N. 2014. Higher resolution in an operational ensemble Kalman filter. Mon. Weather Rev. 142: 1143–1162.

JEPSEN PL, PIIKINEN G, BUHNER M, CHARRON M, SPACEK L, HANSEN B. 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. Mon. Weather Rev. 133: 604–620.
Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray L. 2005. Designing new models. J. Atmos. Sci. 62: 1574–1587.

Lorenc EN. 1963. Deterministic non-periodic flow. J. Atmos. Sci. 20: 130–148.

Ménétrier M, Auligné T. 2015. An overlooked issue of variational data assimilation. Mon. Weather Rev. 143: 3925–3930, doi: 10.1175/MWR-D-14-00404.1.

Meng Z, Zhang F. 2007. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. Mon. Weather Rev. 135: 1403–1423.

Meng Z, Zhang F. 2008a. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVar in a realtime data case study. Mon. Weather Rev. 136: 522–540.

Meng Z, Zhang F. 2008b. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVar in a month-long experiment. Mon. Weather Rev. 136: 3671–3682.

Meng Z, Zhang F. 2012. Ensemble-based data assimilation. In Encyclopedia of Atmospheric Sciences (2nd edn), North GD, Polv JAZF (eds.). Vol. 2: 241–247. Elsevier: Amsterdam, Netherlands.

Moirs RE, Emanuel KA. 2002. Influence of added observations on analysis and forecast errors: Results from idealized systems. J. R. Meteorol. Soc. 128: 285–312.

Navon IM, Daescu DN, Liu Z. 2005. The impact of background error on incomplete observations for 4D-Var data assimilation with the FSU GSM. NCEP Environmental Modeling Center. The Global Forecast System (GFS) – Global Spectral Model (GSM). NOAA, NCEP: College Park, MD. http://www.emc.ncep.noaa.gov/GFS/doc.php (accessed 13 January 2017).

Oleson KW, Dai Y, Bonan G, Bosilovich M, Dickinson R, Dirmeyer P, Hoffman F, Houser P, Levis S, Niu G-Y, Thornton P, Verizon T, Yang Z-L, Zeng X. 2004. "Technical description of the Community Land Model (CLM)", Technical Note NCAR/TN-461+STR. NCAR, Boulder, CO. http://www.cgd.ucar.edu/tts/clm/distribution/clm3.0/TechnicalNote/CLM_TechNote.pdf (accessed 13 January 2017).

Ott E, Hunt BR, Szunyogh I, Zimin AV, Kostelich EJ, Corazza M, Kalnay E, Patil DJ, Yorke JA. 2004. A local ensemble Kalman filter for atmospheric data assimilation. Tellus 56A: 413–428.

Paige CC, Saunders MA. 1975. Solution of sparse indeﬁnite systems of linear equations. SIAM J. Numer. Anal. 12: 617–629.

Palmer TN. 2012. Towards the probabilistic Earth system simulator: A vision for the future of climate and weather prediction. Q. J. R. Meteorol. Soc. 138: 841–861.

Palmer TN, Williams P. 2010. Stochastic Physics and Climate Modelling. Cambridge University Press: Cambridge, UK.

Pan Y, Zhu K, Xue M, Wang X, Hu M, Benjamin SG, Werngard SS, Whitaker JS. 2014. A GSM-based coupled EnSRF–En3DVar hybrid data assimilation system for the operational rapid refresh model: Tests at a reduced resolution. Mon. Weather Rev. 142: 3756–3780.

Park SK, Zapanski D. 2003. Four-dimensional variational data assimilation for mesoscale and storm scale applications. Met. Atmos. Phys. 82: 173–208.

Parrish DF, Derber JC. 1992. The National Meteorological Center’s spectral statistical-interpolation analysis system. Mon. Weather Rev. 120: 1747–1793.

Penny SG. 2014. The hybrid local ensemble transform Kalman filter. Mon. Weather Rev. 142: 2319–2349, doi: 10.1175/mwr-d-13-00131.1.

Petrie RE, Bannister RN. 2011. A method for merging flow-dependent forecast error statistics from an ensemble with static statistics for use in high-resolution variational data assimilation. Comput. Fluids 46: 387–391.

Pham DT. 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. Mon. Weather Rev. 129: 1194–1207.

Poterjio J, Zhang F. 2014. Intercomparison and coupling of ensemble and four-dimensional variational data assimilation methods for the analysis and forecasting of hurricane Karl (2010). Mon. Weather Rev. 142: 3347–3364.

Poterjio J, Zhang F. 2015. Systematic comparison for four-dimensional data assimilation methods with and without the tangent linear model using hybrid background-error covariance: 4D-Var versus 4DVar. Mon. Weather Rev. 143: 1601–1621.

Rabier F. 2005. Overview of global data assimilation developments in numerical weather prediction centres. Q. J. R. Meteorol. Soc. 131: 3215–3233.

Rabier F, Jarnainen H, Klinker E, Mahfouf J-F, Simmons A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. Q. J. R. Meteorol. Soc. 126: 1143–1170.

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne T. 2007. The Met Office global four-dimensional variational data assimilation system. Mon. Weather Rev. 135: 3347–3362.

Raynaud L, Berre L, Desroziers G. 2011. An extended specification of flow-dependent background-error variances in the Météo-France global 4D-Var system. Q. J. R. Meteorol. Soc. 137: 607–619, doi: 10.1002/qj.795.

Raynaud L, Berre L, Desroziers G. 2012. Accounting for model error in the Météo-France ensemble data assimilation system. Q. J. R. Meteorol. Soc. 138: 249–262.

Rhodin A. 2015. ‘Multi Scale VarEnKF Localization using Wavelets’. In Presentation at International Symposium on Data Assimilation, 2016 The Authors. Quarterly Journal of the Royal Meteorological Society. published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society
