Comparison of Hybrid-4DEnVar and Hybrid-4DVar Data Assimilation Methods for Global NWP

ANDREW C. LORENC, NEILL E. BOWLER, ADAM M. CLAYTON, AND STEPHEN R. PRING

Met Office, Exeter, United Kingdom

DAVID FAIRBAIRN

University of Surrey, Guildford, Surrey, and Met Office, Exeter, United Kingdom

ABSTRACT

The Met Office has developed an ensemble-variational data assimilation method (hybrid-4DEnVar) as a potential replacement for the hybrid four-dimensional variational data assimilation (hybrid-4DVar), which is the current operational method for global NWP. Both are four-dimensional variational methods, using a hybrid combination of a fixed climatological model of background error covariances with localized covariances from an ensemble of current forecasts designed to describe the structure of "errors of the day." The fundamental difference between the methods is their modeling of the time evolution of errors within each data assimilation window: 4DVar uses a linear model and its adjoint and 4DEnVar uses a localized linear combination of nonlinear forecasts. Both hybrid-4DVar and hybrid-4DEnVar beat their three-dimensional versions, which are equivalent, in NWP trials. With settings based on the current operational system, hybrid-4DVar performs better than hybrid-4DEnVar. Idealized experiments designed to compare the time evolution of covariances in the methods are described: the basic 4DEnVar represents the evolution of ensemble errors as well as 4DVar. However, 4DVar also represents the evolution of errors from the climatological covariances, whereas 4DEnVar does not. This difference is the main cause of the superiority of hybrid-4DVar. Another difference is that the authors' 4DVar explicitly penalizes rapid variations in the analysis increment trajectory, while the authors' 4DEnVar contains no dynamical constaints on imbalance. The authors describe a four-dimensional incremental analysis update (4DIAU) method that filters out the high-frequency oscillations introduced by the poorly balanced 4DEnVar increments. Possible methods for improving hybrid-4DEnVar are discussed.

1. Introduction

For more than a decade four-dimensional variational data assimilation (4DVar) has been used by most of the main global numerical weather prediction (NWP) centers (Rabier 2005); the Met Office implemented 4DVar in 2004 (Rawlins et al. 2007). A weakness of the basic 4DVar method is the use of a fixed "climatological" model of the error covariance in the background forecast, which does not describe the flow-dependent errors of the day as well as ensemble Kalman filter methods (Lorenc 2003b). To address this Clayton et al. (2013) implemented a hybrid-4DVar method, where the term "hybrid" refers to a

combination of a climatological covariance model with covariances calculated from an ensemble of forecasts, designed to sample the current uncertainty.

Another weakness of 4DVar is its cost, both in the time needed to run sequential iterations of linear and adjoint models for a high-resolution NWP system on a massively parallel computer, and in the effort needed to maintain and develop these models as the NWP system evolves because of advances in science and computing power (Kalnay et al. 2007). These costs can be avoided by using the ensemble trajectories (i.e., the time evolution of each member) to directly estimate a four-dimensional background error covariance. Pioneering attempts such as Tian et al. (2008) did not include localization so were not directly applicable to large NWP systems and relatively small ensembles. With localization the approach has come to be called four-dimensional (4D) ensemble-variational

Corresponding author address: Andrew C. Lorenc, Met Office, Fitzroy Road, Exeter, Devon EX1 3PB, United Kingdom.
E-mail: andrew.lorenc@metoffice.gov.uk

data assimilation (4DEnVar; Lorenc 2013); it was introduced by Liu et al. (2008, 2009) and tested in an NWP model by Buehner et al. (2010a,b). We have coded 4DEnVar within the Met Office global data assimilation software system. Because the operational system is hybrid-4DVar, already using a localized ensemble covariance to represent errors of the day, our tests were able to eliminate this effect and do a clean comparison of the hybrid-4DEnVar and hybrid-4DVar methods. This distinguishes the current paper from most others on hybrid data assimilation, which focused on the benefit of adding hybrid covariances to three-dimensional variational data assimilation (3DVar; Hamill and Snyder 2000; Wang et al. 2008a,b, 2013; Kleist and Ide 2015a), or to 4DVar (Clayton et al. 2013; Zhang and Zhang 2012), or on comparing 4DVar with hybrid-4DEnVar (Buehner et al. 2013; Gustafsson et al. 2014). The studies most related to ours are Wang and Lei (2014) and Kleist and Ide (2015b), who compared 4DEnVar with 3DEnVar.

Our initial plan was for trials of the hybrid-4DEnVar system with as many settings as possible based on those of the operational hybrid-4DVar (Clayton et al. 2013); these showed that although hybrid-4DEnVar performed quite well, improving on the hybrid-3DEnVar method used for example by Hamill and Snyder (2000) and Wang et al. (2013), it did not make as big an improvement as that shown by hybrid-4DVar over hybrid-3DVar. To better understand this difference in the methods we performed some additional diagnostic experiments: of the time evolution of error covariances implicit in the various methods and of the different methods used to control high-frequency oscillations. They indicated that, unless there is strong space localization, the 4DEnVar method can model the time evolution of ensemble covariances as well as 4DVar—in our experiments it is the 3D treatment of the climatological covariance which is the main contribution to the relatively poor performance of hybrid-4DEnVar. The new 4D incremental analysis update (4DIAU) time-filtering scheme used with hybrid-4DEnVar performed well enough compared to the well-established J_c method of 4DVar systems (Gauthier and Thépaut 2001) that this difference makes only a relatively minor contribution.

The structure of this paper is as follows: in section 2 we describe the data assimilation methods; in section 3 we describe realistic trials to evaluate hybrid-4DEnVar (the new method) compared to hybrid-4DVar (the operational method) and their 3D equivalents; in section 4 we describe the additional diagnostic studies of the time evolution of error covariances and in section 5 we present diagnostics of the 4DIAU and J_c schemes used in the main trials, comparing them with the incremental analysis update

(IAU) method used for 3D analysis methods (Bloom et al. 1996). Finally, in section 6, we discuss the conclusions and suggest means of improving the performance of 4DEnVar.

2. Data assimilation methods used

a. Four-dimensional variational methods and notation

All the data assimilation (DA) schemes studied in this paper are based on the advanced hybrid ensemble-variational data assimilation system described by Clayton et al. (2013). It is convenient to describe them using 4D trajectories (i.e., the sequence of 3D states describing the evolution over a time window). We use an underline to extend the standard notation of Ide et al. (1997) to four dimensions, for instance, $\underline{\mathbf{x}}^b$ is the background trajectory. The expected error covariance of $\underline{\mathbf{x}}^b$ is $\underline{\mathbf{P}}$. This defines a Gaussian probability distribution function (pdf) for the 4D increment $\delta\underline{\mathbf{x}}$:

$$\delta\underline{\mathbf{x}} \sim N(0, \underline{\mathbf{P}}), \tag{1}$$

which gives the probability density that $\underline{\mathbf{x}}^b + \delta\underline{\mathbf{x}}$ is the true trajectory. All the methods variationally determine the $\delta\underline{\mathbf{x}}$ that maximizes the posterior Bayesian likelihood by minimizing a penalty function measuring the distances from the background and the observations:

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

where the second term is the observational penalty, measuring the difference of the observations in the time window $\underline{\mathbf{y}}^o$ from their model estimates $\underline{\mathbf{y}}$. The latter are calculated as accurately as possible, using the nonlinear observation operator. For the purposes of this paper we can simplify this to

$$\underline{\mathbf{y}} = \underline{H}(\underline{\mathbf{x}}^b + \delta\underline{\mathbf{x}}), \tag{3}$$

where observation operator $\underline{H}$ includes time, horizontal, and vertical interpolations followed by the potentially nonlinear calculations for each observed parameter. In practice the forecast model providing the background $\underline{\mathbf{x}}^b$ is usually run at higher resolution than the ensemble and the incremental linear model used to define the increment $\delta\underline{\mathbf{x}}$, so the interpolations are performed separately for $\underline{\mathbf{x}}^b$ and $\delta\underline{\mathbf{x}}$ as described by Lorenc et al. (2000). This approach is better than the more usual incremental method (Courtier et al. 1994), which calculates the innovation $\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^b)$ as a preliminary step, then increments it using an observation operator linearized about $\mathbf{x}^b$. Using Eq. (3) enables all the methods in this paper to handle

nonlinear $\underline{H}$ without an expensive outer loop rerunning the forecast model.

For large NWP systems, the background term $\delta\underline{\mathbf{x}}^T\underline{\mathbf{P}}^{-1}\delta\underline{\mathbf{x}}$ in Eq. (2) involves too large a matrix to compute directly. Variational methods overcome this by constucting the increments by a series of transforms, chosen such that the background term is simplified in the transformed space. 4DVar and 4DEnVar differ in the way their transforms define the 4D increment $\delta\underline{\mathbf{x}}$ and, hence, the implicit 4D covariance $\underline{\mathbf{P}}$; 4DVar uses a linear forecast model whereas 4DEnVar uses a linear combination of precalculated ensemble forecasts each made by a full nonlinear model. They are described in sections 2b and 2c. Both methods can use either climatological covariances, ensemble covariances, or a hybrid—a linear combination of the two. Both use the "alpha-control variable" method to localize the ensemble covariances (Lorenc 2003b), and both require methods to control high-frequency noise. The technicalities of doing these things differ in the two methods; they are described with each method below. Both methods can also be extended to account explicitly for error in the forecast model (Desroziers et al. 2014), but a description of this is beyond the scope of this paper.

b. 4DVar—Using a linear forecast model and its adjoint

While both methods studied in this paper are four-dimensional and variational, the term 4DVar has come to be associated with methods that achieve their four-dimensional constraint by running a forecast model and its adjoint at each iteration. In line with the recommendations of Lorenc (2013), in this paper we reserve 4DVar, even with a prefix, to such methods.

In traditional strong-constraint 4DVar, we assume that the evolution of the atmosphere is described by a forecast model and ignore model errors. Then we can "reduce the control variable" (Le Dimet and Talagrand 1986) and define 4D trajectories using the forecast model and its initial conditions: $\underline{\mathbf{x}} = \underline{M}(\mathbf{x})$. Note the unusual definition of $\underline{M}$ as having a 3D input $\mathbf{x}$ and a 4D output $\underline{\mathbf{x}}$. In incremental 4DVar (Courtier et al. 1994) we replace $\mathbf{x}$ by $\mathbf{x}^b + \delta\mathbf{x}$ and make the assumption that

$$\underline{M}(\mathbf{x}^b + \delta\mathbf{x}) = \underline{M}(\mathbf{x}^b) + \delta\underline{\mathbf{x}} \quad \text{and} \quad (4)$$

$$\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\delta\mathbf{x}, \quad (5)$$

where $\underline{\mathbf{M}}$ is a linear approximation to the tangent-linear model of $\underline{M}$; it is run at lower resolution and has simplified parameterizations. Because of these approximations, Rawlins et al. (2007) call it the perturbation forecast (PF) model.

Because of the reduction of the control variable in Eq. (5), the 4D covariance implicit in 4DVar is

$$\underline{\mathbf{P}} = \underline{\mathbf{M}}\mathbf{P}\underline{\mathbf{M}}^T, \quad (6)$$

where $\mathbf{P}$ is the 3D background error covariance at the beginning of the window.

Traditional 4DVar uses a fixed climatological estimate $\mathbf{B}$, which is itself modeled using a transform $\mathbf{U}$ and a transformed control variable $\mathbf{v}^c$ (Lorenc et al. 2000; Ingleby 2001) such that $\mathbf{P} = \mathbf{B} = \mathbf{U}\mathbf{U}^T$. This gives

$$\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\mathbf{U}\mathbf{v}^c. \quad (7)$$

To get a common penalty function for all our 4DVar options, we use a common control variable $\mathbf{v}$ which in this case is just a copy $\mathbf{v} = \mathbf{v}^c$:

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}\mathbf{v}^T + \frac{1}{2}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o)^T\underline{\mathbf{R}}^{-1}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o) + J_c, \quad (8)$$

where $\underline{\mathbf{y}}$ is given by Eqs. (7) and (3). The implicit 4D covariance is thus

$$\underline{\mathbf{P}} = \underline{\mathbf{M}}\mathbf{U}\mathbf{U}^T\underline{\mathbf{M}}^T. \quad (9)$$

Each iteration in the variational minimization of Eq. (8) requires an integration of the linear model $\underline{\mathbf{M}}$ and its adjoint $\underline{\mathbf{M}}^T$—this is consistent with their appearance in the implicit 4D covariance in Eq. (9).

Control of high-frequency noise in the 4DVar experiments was achieved by adding a digital-filter J_c term to Eq. (8), similar to that of Gauthier and Thépaut (2001). This term penalizes high-frequency oscillations in the PF model trajectory. We discuss its properties in section 5.

Ensemble 4DVar (En-4DVar) implicitly uses a localized ensemble error instead of the climatological covariance $\mathbf{B}$:

$$\mathbf{P} = \mathbf{C}\circ\mathbf{X}\mathbf{X}^T, \quad (10)$$

where $\mathbf{X} = [\mathbf{x}'_1 \quad \cdots \quad \mathbf{x}'_N]$ is constructed using normalized ensemble perturbations valid at the beginning of the window $\mathbf{x}'_k = (\mathbf{x}_k - \overline{\mathbf{x}})/\sqrt{N-1}$. The symbol $\circ$ denotes an element-by-element (Schur) product of two matrices and $\mathbf{C}$ is a correlation matrix designed to localize the ensemble covariance and reduce sampling errors (Hamill et al. 2001). Lorenc (2003b) and Wang et al. (2007) showed the localization could be achieved efficiently in a variational framework by transforming to an alpha-control variable field $\boldsymbol{\alpha}_k$ for each ensemble member. $\boldsymbol{\alpha}_k$ defines the local weight given to each perturbation, so that

$$\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\sum_k \boldsymbol{\alpha}_k \circ \mathbf{x}'_k. \quad (11)$$

We ensure that each $\boldsymbol{\alpha}_k$ is smooth using techniques copied from the climatological covariance transform $\mathbf{U}$, setting $\boldsymbol{\alpha}_k = \mathbf{U}^\alpha \mathbf{v}_k^\alpha$, such that

$$\mathbf{C} = \mathbf{U}^\alpha (\mathbf{U}^\alpha)^{\mathrm{T}}. \tag{12}$$

We concatenate all the $\mathbf{v}_k^\alpha$ into the single control vector $\mathbf{v}$ so that the background penalty term is transformed into a simple inner product:

$$\frac{1}{2} \mathbf{v} \mathbf{v}^{\mathrm{T}} = \frac{1}{2} \sum_k \mathbf{v}_k^\alpha (\mathbf{v}_k^\alpha)^{\mathrm{T}}. \tag{13}$$

The same 4DVar penalty function in Eq. (8) can be used, this time calculating $\underline{\mathbf{y}}$ using Eqs. (11) and (3). The implicit 4D error covariance is still constructed using the model:

$$\underline{\mathbf{P}} = \underline{\mathbf{M}} (\mathbf{C} \circ \mathbf{X} \mathbf{X}^{\mathrm{T}}) \underline{\mathbf{M}}^{\mathrm{T}}. \tag{14}$$

The technical details are the same as in Clayton et al. (2013), in particular our implementation is unusual in localizing transformed variables. There is no reason to regard any particular representation of a model perturbation as fundamental: we can use grid points or spectral coefficients; streamfunction and velocity potential or wind components. So any invertible linear transform $\mathbf{T}$ can be used in

$$\delta \underline{\mathbf{x}} = \underline{\mathbf{M}} \mathbf{T}^{-1} \sum_k \boldsymbol{\alpha}_k \circ \mathbf{T} \mathbf{x}_k' \quad \text{and} \tag{15}$$

$$\underline{\mathbf{P}} = \underline{\mathbf{M}} \mathbf{T}^{-1} (\mathbf{C} \circ \mathbf{T} \mathbf{X} \mathbf{X}^{\mathrm{T}} \mathbf{T}) \mathbf{T}^{-\mathrm{T}} \underline{\mathbf{M}}^{\mathrm{T}}, \tag{16}$$

to give a different but equally valid localization method. In the experiments described here we followed Clayton et al. (2013) and used the transform $\mathbf{T}$ from our 3DVar and 4DVar system (Lorenc et al. 2000). Since localization can affect balance (Mitchell et al. 2002) the choice of which variables to localize can be important; our localization in Eq. (16) is applied to streamfunction, divergence, unbalanced pressure, and unbalanced moisture (Ingleby et al. 2013), causing less imbalance (Kepert 2009).

Hybrid-4DVar is constructed by a weighted combination of traditional 4DVar and ensemble 4DVar:

$$\delta \underline{\mathbf{x}} = \underline{\mathbf{M}} \left(\beta_c \mathbf{U} \mathbf{v} + \beta_e \sum_k \mathbf{U}^\alpha \mathbf{v}_k^\alpha \circ \mathbf{x}_k' \right). \tag{17}$$

We concatenate $\mathbf{v}^c$ and all the $\mathbf{v}_k^\alpha$ into the single control vector $\mathbf{v}$. The same 4DVar penalty function in Eq. (8) can be used, this time calculating $\underline{\mathbf{y}}$ using Eqs. (17) and (3). The implicit covariance has the squares of the weights in Eq. (17):

$$\underline{\mathbf{P}} = \underline{\mathbf{M}} (\beta_c^2 \mathbf{B} + \beta_e^2 \mathbf{C} \circ \mathbf{X} \mathbf{X}^{\mathrm{T}}) \underline{\mathbf{M}}^{\mathrm{T}}. \tag{18}$$

Assuming they both give independent valid estimates of $\mathbf{P}$, we get[1]

$$\beta_c^2 + \beta_e^2 = 1, \tag{19}$$

but, following Clayton et al. (2013), the trials in section 3 used $\beta_c^2 = 0.8$ and $\beta_e^2 = 0.5$.

3DVar versions of all the above can be obtained by replacing the PF model $\underline{\mathbf{M}}$ by a persistence forecast for the increments $\underline{\mathbf{l}}$; that is, copying the increments at one time to all other times in the window:

$$\delta \underline{\mathbf{x}} = \underline{\mathbf{l}} \delta \mathbf{x}. \tag{20}$$

The time interpolation of the background forecast in Eq. (3) is still done correctly; it is only the increments that are persisted. [This is known as the first guess at appropriate time (FGAT) method, e.g., Lorenc and Rawlins (2005).] To minimize phase errors in time, it is better if the increments are considered as being at the middle of the window, and the ensemble covariances are also valid at this time. In 3DVar there is no time-evolving increment so we cannot use our 4DVar J_c term to control imbalance. For future flexibility we wish to investigate methods with no internal model steps, ruling out internal constraints such as the tangent-linear normal mode constraint (TLNMC; Kleist et al. 2009) or the external initialization scheme we used when 3DVar was operational—an incremental version of digital filtering initialization (DFI; Lynch and Huang 1992). The IAU method (Bloom et al. 1996) was used in the experiments in section 3. Its time-filtering properties are discussed in section 5.

c. 4DEnVar—Using a precalculated ensemble of trajectories

This method is based on En-4DVar, but is designed to avoid the use of the linear and adjoint models at each iteration so as to avoid the computational cost on parallel computers and the maintenance cost of designing the models and coding them. Instead we extend to four dimensions all the En-4DVar equations, applying them to the sequence of states in trajectories rather than at the single time at the beginning of the window. So we use an implicit localized ensemble error covariance:

$$\underline{\mathbf{P}} = \underline{\mathbf{C}} \circ \underline{\mathbf{X}} \underline{\mathbf{X}}^{\mathrm{T}}, \tag{21}$$

where $\underline{\mathbf{X}} = [\underline{\mathbf{x}}_1' \cdots \underline{\mathbf{x}}_N']$ is constructed using normalized ensemble perturbations valid through the window, and $\underline{\mathbf{x}}_k' = (\underline{\mathbf{x}}_k - \bar{\underline{\mathbf{x}}})/\sqrt{N-1}$. The method uses four-dimensional alpha-control variable fields $\underline{\boldsymbol{\alpha}}_k$, for each ensemble

---

[1] Lorenc (2003b) incorrectly suggests that $\beta_c + \beta_e = 1$.

member, which define the local weight given to each perturbation trajectory, so that

$$\delta\underline{\mathbf{x}} = \sum_k \underline{\boldsymbol{\alpha}}_k \circ \underline{\mathbf{x}}'_k. \qquad (22)$$

Since each $\underline{\mathbf{x}}'_k$ is a normalized difference of nonlinear forecasts, Eq. (22) is a linear combination of nonlinear forecasts (using localized weights). It is a linear function of the $\underline{\boldsymbol{\alpha}}_k$ and hence a different kind of linear model to that in Eq. (5).

For simplicity in this initial implementation and copying most other work, we have not yet exploited the potential variation of $\underline{\boldsymbol{\alpha}}_k$ in time; instead we have used a persistence forecast $\underline{\boldsymbol{\alpha}}_k = \underline{\mathbf{I}}\,\boldsymbol{\alpha}_k$. Other technical details for $\boldsymbol{\alpha}_k$ are the same as in En-4DVar: we ensure that it is smooth using a transform, setting $\underline{\boldsymbol{\alpha}}_k = \underline{\mathbf{U}}^\alpha_k \mathbf{v}^\alpha_k = \underline{\mathbf{I}}\,\mathbf{U}^\alpha \mathbf{v}^\alpha_k$. The corresponding implicit 4D localization matrix used in Eq. (21) is

$$\underline{\mathbf{C}} = \underline{\mathbf{U}}^\alpha (\underline{\mathbf{U}}^\alpha)^{\mathrm{T}} = \underline{\mathbf{I}}\,\mathbf{U}^\alpha (\mathbf{U}^\alpha)^{\mathrm{T}} \underline{\mathbf{I}}^{\mathrm{T}} = \underline{\mathbf{I}}\,\mathbf{C}\underline{\mathbf{I}}^{\mathrm{T}}. \qquad (23)$$

As in En-4DVar we concatenate all the $\mathbf{v}^\alpha_k$ into the single control vector $\mathbf{v}$. A very similar penalty function to Eq. (8) can be used:

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}\mathbf{v}^{\mathrm{T}} + \frac{1}{2}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o)^{\mathrm{T}}\underline{\mathbf{R}}^{-1}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o), \qquad (24)$$

this time calculating $\underline{\mathbf{y}}$ using Eqs. (22) and (3).

Hybrid-4DEnVar is an obvious extension, given the benefit of hybrid methods listed by Clayton et al. (2013). However, since the objective of 4DEnVar is to avoid the use of models inside the iteration, the climatological covariance component has to be implemented using a persistence forecast as in 3DVar Eq. (20):

$$\delta\underline{\mathbf{x}} = \beta_c\,\underline{\mathbf{I}}\,\mathbf{U}\mathbf{v}^c + \beta_e \sum_k \underline{\mathbf{I}}\,\mathbf{U}^\alpha \mathbf{v}^\alpha_k \circ \underline{\mathbf{x}}'_k. \qquad (25)$$

As for hybrid-4DVar we concatenate $\mathbf{v}^c$ and all the $\mathbf{v}^\alpha_k$ into the single control vector $\mathbf{v}$. The same penalty function in Eq. (24) can be used, this time calculating $\underline{\mathbf{y}}$ using Eqs. (25) and (3). The implicit covariance has the squares of the weights in Eq. (25):

$$\underline{\mathbf{P}} = \beta_c^2\,\underline{\mathbf{I}}\,\mathbf{B}\underline{\mathbf{I}}^{\mathrm{T}} + \beta_e^2\,\underline{\mathbf{C}} \circ \underline{\mathbf{X}}\,\underline{\mathbf{X}}^{\mathrm{T}}. \qquad (26)$$

Notice that, apart from the J_c term, Eq. (24) is identical to the hybrid-4DVar penalty in Eq. (8), and both use the same method as in Eq. (3) for predicting the observed values. It is the different calculations of the 4D increments in Eqs. (25) and (17), which distinguish the methods. 4DEnVar does not require $\underline{\mathbf{M}}$ and so Eqs. (21) and (26) do not require $\underline{\mathbf{M}}$ and $\underline{\mathbf{M}}^{\mathrm{T}}$: there is no need for linear and adjoint model integrations at each iteration.

Time filtering using a 4DIAU is used in our 4DEnVar experiments, since the digital-filter J_c term is not expected

TABLE 1. Experiments compared in section 3.

Name in text	DA method	Initialization
4DVar	Hybrid-4DVar	J_c
4DEnVar	Hybrid-4DEnVar	4DIAU
3DVar	Hybrid-3DVar	IAU
3DEnVar	Hybrid-3DEnVar	IAU
4DVar4DIAU	Hybrid-4DVar	4DIAU

to be effective without a model. The four-dimensional IAU is like that of Bloom et al. (1996), but with an important difference: while the IAU adds the same 3D analysis increment gradually over a period during a model forecast, the 4DIAU adds a different 3D field valid at each time step of the forecast,[2] taken from the four-dimensional $\delta\underline{\mathbf{x}}$.

The differences between the J_c initialization used in 4DVar, this 4DIAU method, and the IAU used in the 3D analysis experiments are discussed further in section 5.

3DEnVar versions of the above can be obtained by making the 4D trajectories such as $\underline{\mathbf{x}}'_k$ only contain one state, in the middle of the time window. The variable $\delta\underline{\mathbf{x}}$ then correspondingly only contains one state, and its time interpolation is inactivated when evaluating Eq. (3). The 4DIAU reverts to the original IAU.

Apart from the different methods for controlling high-frequency noise, the only difference between the methods is in their handling of the time evolution of increments in each window. Remove this by taking the three-dimensional versions and the methods are the same: En-3DVar is equivalent to 3DEnVar and hybrid-3DVar is equivalent to hybrid-3DEnVar. This helps in understanding the results in section 3. [For technical reasons in our implementation, there are other slight differences: $\mathbf{T}$ in Eq. (15) is linearized about the background $\mathbf{x}^b$ in En-4DVar and about the ensemble mean $\bar{\mathbf{x}}$ in 4DEnVar; they should not affect the quality of the results.]

3. NWP trials

The object of this first set of experiments was to compare the hybrid-4DEnVar method with the hybrid-4DVar system documented by Clayton et al. (2013) and used for operational global NWP at the Met Office. We, therefore, based the settings for hybrid-4DEnVar (e.g., for localization and hybridization of the ensemble covariances) on that paper, with no attempt to tune them

---

[2] In our experiments, $\delta\underline{\mathbf{x}}$ was stored as seven 3D states spanning the 6-h window. The nearest state in time was used without time interpolation.

(a) 4DVar vs 4DEnVar: 3.138%

(b) 3DVar vs 3DEnVar: 0.007%

(c) 4DVar vs 3DVar: 3.620%

(d) 4DEnVar vs 3DEnVar: 0.474%

(e) 4DVar4DIAU vs 4DEnVar: 2.594%
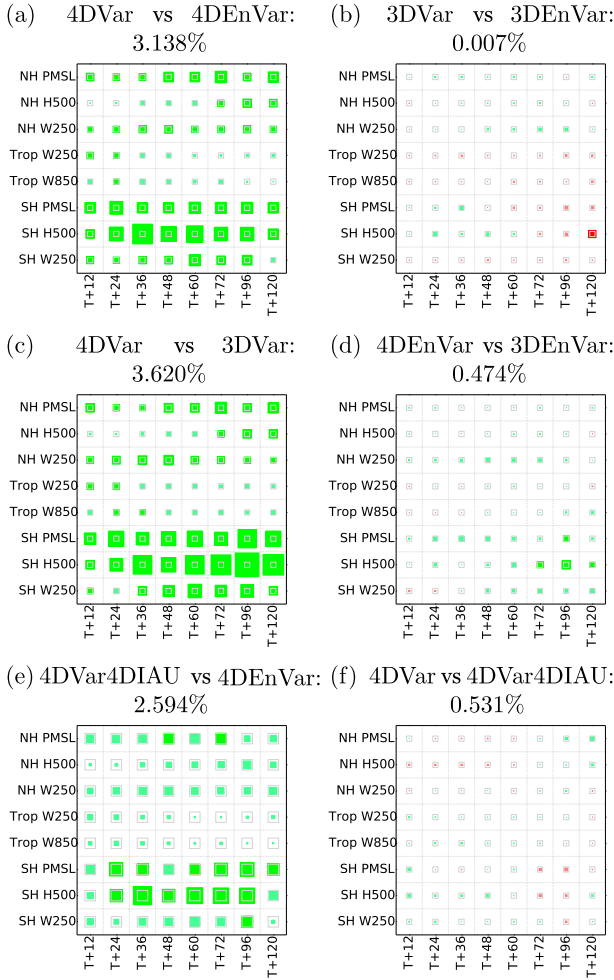
(f) 4DVar vs 4DVar4DIAU: 0.531%

FIG. 1. Pairwise comparison of the RMS misfit of forecasts to observations, for a range of fields and ranges. The relative improvement of the first experiment compared to the second is proportional to the areas of the filled squares plotted. The hollow gray square shows the size for a 2% difference and the biggest difference that can be plotted is 10%. The squares are colored green if the first experiment was better and red if the second experiment was better. Each caption shows the average of the values plotted: in each pair the first experiment performed better.

for the new method. We also ran hybrid-3DEnVar and hybrid-3DVar experiments, with the expectation that they would be very similar. By using these as controls, we are able to measure how much adding the time dimension improves each method.

The other important difference is the J_c term used in hybrid-4DVar. This penalizes high-frequency oscillations revealed by applying a digital time filter to $\delta\underline{\mathbf{x}}$ from Eq. (17). To eliminate this difference we also ran the hybrid-4DVar method with the J_c term turned off, using the 4DIAU initialization used for the hybrid-4DEnVar experiment to control high-frequency noise. The experiments are summarized in Table 1.
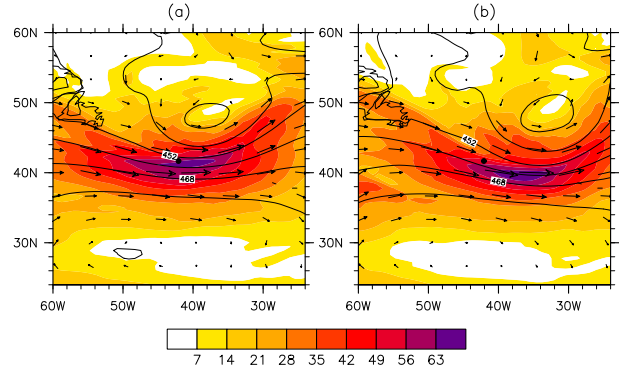

FIG. 2. Plot of the background wind fields for the jet stream example at height level 29, at times (a) 0300 and (b) 0900 UTC 1 Nov 2011. The wind magnitude (m s^{-1}) is given by the colors and the wind vectors are given by the arrows. The pressure contours at level 29 are given by the black lines. The observation is located at the black dot, but is only present at 0300 UTC.

The trials were run at slightly lower horizontal resolution than the Met Office operational global NWP system, which for the experimental period in 2013 ran its deterministic forecast with a grid spacing of about 26 km: the grids we used were 640 × 481 (about 42 km) for the deterministic forecast model and 432 × 325 (about 62 km) for the ensemble forecasts and the perturbation forecast model in 4DVar. All used the same 70 levels, with the top level at 80 km. The ensemble was precalculated by running a version of the Met Office Global and Regional Ensemble Prediction System (MOGREPS; Bowler et al. 2008; Flowerdew and Bowler 2011). Note that MOGREPS includes no external initialization or time filtering of the initial ensemble states or perturbations. Experiments showed that such measures were unnecessary, and indeed the ensemble perturbation trajectories are found to be relatively well-balanced after $T + 3$ h—the time they are first required for use in 4DVar or 4DEnVar. All the trials used the same 44-member ensemble trajectories to generate ensemble covariances using Eqs. (10) or (21) as appropriate. They assimilated the operational observations in a 6-h cycle for the month of July 2013. Each experiment ran forecasts to 5 days every 12 h. These were verified against observations for a range of fields and time ranges as indicated on the two axes on each panel in Fig. 1. The first three rows in each panel are measured north of 20°N, the next two rows are between 20°N and 20°S and the last three rows are south of 20°S. Length of forecast increases along each row.

The experiments were then compared in pairs by calculating the relative difference between their RMS errors for each verified field. These are shown graphically for each pair of experiments by the shaded squares in Fig. 1, whose areas are proportional to the percentage
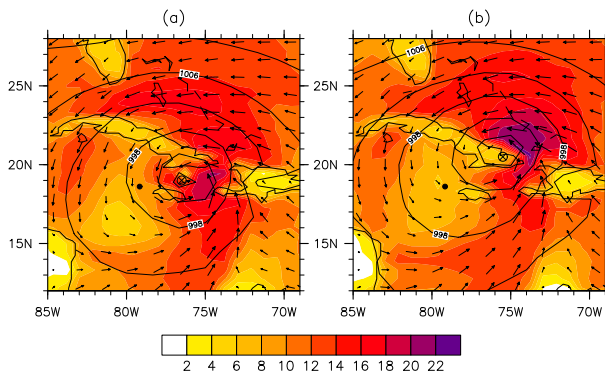
FIG. 3. Plot of the background surface wind fields for the Hurricane Sandy example at times (a) 0300 and (b) 0900 UTC 25 Oct 2012. Format is equivalent to Fig. 2, but that the pressure contours are plotted at mean sea level. The eye of the hurricane is approximately located at the ⊗ symbol. The observation is only present at 0300 UTC.

change in RMS error. The caption for each panel also shows the average of the values plotted.

The comparison of hybrid-4DVar with hybrid-4DEnVar, which is the objective of this paper, is shown in Fig. 1a. It shows that hybrid-4DEnVar performs on average over 3% worse, with the biggest degradations occurring in 500-hPa height forecasts in the Southern Hemisphere. The rest of this paper focusses on understanding the causes of this result.

We start by considering Fig. 1b, which confirms that, as expected, the hybrid-3DVar and hybrid-3DEnVar systems perform equally well. These experiments are used as reference in Figs. 1c,d. Figure 1c shows that adding the time dimension in hybrid-4DVar gives a big benefit, especially in the Southern Hemisphere. This is usual for 4DVar (e.g., it was seen in Rawlins et al. 2007). In comparison, Fig. 1d shows a much smaller improvement due to the time dimension in hybrid-4DEnVar.

We can eliminate the J_c term as the major cause of the improvement seen in the hybrid-4DVar experiment by considering the last two panels. Figure 1e shows that, using the same 4DIAU method to control noise, hybrid-4DVar still beats hybrid-4DEnVar, by nearly as much as in Fig. 1a. Figure 1f shows that there is some advantage from the J_c term over 4DIAU in 4DVar—we study this more in section 5.

So we can conclude that the time dimension is less well handled in hybrid-4DEnVar than in hybrid-4DVar. Experiments to show why are described in the next section.

4. Diagnosing the 4D covariances

a. Experimental setup

We diagnose the behavior of the 4D covariances using single-observation experiments. We chose two different cases:

TABLE 2. The two observation types for the single-observation experiments in section 4.

Case	Jet stream	Hurricane Sandy
Direction	Westerly	Northerly
$\mathbf{y} - \mathbf{y}^o$ (m s^{-1})	10	10
Location	41°N, 41°W at level 29 (~450 hPa)	18°N, 79°W at level 1 (surface)
Time	0300 UTC 1 Nov 2011	0300 UTC 25 Oct 2012

1) a strong midlatitude jet stream—a good test of strong advection, and
2) a hurricane—a good test of complex nonlinear physics, including moist processes.

The same configurations are used for each case, except that a 44-member ensemble is used for the jet stream, whereas only a 22-member ensemble is available for Hurricane Sandy—this was the operational ensemble size at that time. The difference in ensemble size is likely to change the structure of the analysis increments. However, it is unlikely to change the conclusions of the experiments, since it does not change the ability of the DA methods to propagate the increment through the assimilation window. As for the trials described in section 3, the ensemble and 4DVar perturbation forecast resolutions were both 432 × 325, 70 levels. However, the deterministic forecast model resolution was 1024 × 769, 70 levels rather than 640 × 481, 70 levels.

For the first case we chose a region in the polar jet stream, located in the North Atlantic Ocean about 1500 km southeast of Newfoundland, Canada. The observation is located at level 29, which corresponds to a height of 5796 m and a pressure of about 450 hPa. The wind speed for this example is about 60 m s^{-1}. Figure 2 shows the background $\mathbf{x}^b$ wind fields for the jet stream at 0300 (Fig. 2a) and 0900 UTC (Fig. 2b) 1 November 2011. The assimilation window is defined between these times. The observation is located at the beginning of the window.

For the second case we chose Hurricane Sandy as it was tracking northward through the Caribbean. Figure 3 shows the background $\mathbf{x}^b$ surface wind fields at times 0300 (Fig. 3a) and 0900 UTC (Fig. 3b) 25 October 2012. The assimilation window is defined between these times. The eye of the hurricane is located near southern Cuba (given by the ⊗ symbol), and the strongest winds are on the eastern flank of the hurricane. The observation is located southwest of the hurricane and is only present at the beginning of the window.

The pseudo-observations are specified as an increment to the background. The two observations are described in Table 2.

The single-observation experiments are designed to explore the 4D representation of the climatological and

TABLE 3. Covariance (Cov) models used for the single-observation experiments.

Cov model	i	ii	iii	iv
β_c^2	1.0	0.0	0.0	0.5
β_e^2	0.0	1.0	1.0	0.5
L (km)	—	500	1200	1200

ensemble background-error covariances for hybrid-4DVar and hybrid-4DEnVar. They are also designed to explore the effect of horizontal localization on the ensemble covariance. Bearing this in mind, four covariance models with differing β and Gaussian localization scales[3] L are investigated, given in Table 3.

The single observations are located at the beginning of the window, which, since we turned off the 4DVar J_c term, implies that the hybrid-4DVar and hybrid-4DEnVar analysis increments should be identical at t_0.[4] Note that our climatological covariance is built using transforms designed to give appropriately balanced increments (Lorenc et al. 2000) and our ensemble localization in Eqs. (15) and (16) uses the same transforms, so we did not expect these single-observation experiments to exhibit substantial imbalance. This was confirmed by repeating the 4DVar experiments with the J_c term switched on (not shown); results did not change significantly.

The differences between 4DVar and 4DEnVar at later times are caused by the way these methods generate the 4D analysis increment. All the methods were derived assuming model errors were negligible, so the analysis should be a trajectory of the nonlinear model. In this paper we refer to deviations from this as "strong-constraint errors" (E_{SC}). They are calculated by

$$E_{SC} = \{M_{0\to6}[\mathbf{x}^b(t_0) + \delta\mathbf{x}^a(t_0)]\} \\ - \{M_{0\to6}[\mathbf{x}^b(t_0)] + \delta\mathbf{x}^a(t_6)\}, \quad (27)$$

where $M_{0\to6}$ is the model propagation from the beginning to the end of the assimilation window.

For hybrid-4DVar, Eq. (27) measures the errors due to the tangent linear hypothesis, and the resolution difference between the deterministic forecast $M_{0\to6}$ and the perturbation forecast $\mathbf{M}_{0\to6}$ used to compute $\delta\mathbf{x}^a(t_6)$. For hybrid-4DEnVar, Eq. (27) measures errors from four sources:

- the resolution difference between the deterministic and ensemble forecasts;

- the time invariance of the analysis increment generated by the climatological part of the background error covariance (as in 3DVar FGAT);
- the linearization errors introduced by the alpha-control variable—a linear combination of trajectories need not itself be a trajectory; and
- the errors introduced by the space localization (of the ensemble covariance between the initial and final times) not moving with the flow.

The relative stong-constraint error is calculated by dividing the area-integrated strong-constraint error by the corresponding integrated nonlinearly propagated analysis increment:

$$RE_{SC} = 100 \times \frac{\|E_{SC}\|}{\|M_{0\to6}[\mathbf{x}^b(t_0) + \delta\mathbf{x}^a(t_0)] - M_{0\to6}[\mathbf{x}^b(t_0)]\|}, \quad (28)$$

where $\|\cdot\|$ is the square root of the integral, over the appropriate level, of the wind error squared.

b. Jet stream results

ANALYSIS INCREMENTS

The hybrid-4DVar and hybrid-4DEnVar analysis increments are compared for the four Cov models. Since the hybrid-4DVar and hybrid-4DEnVar increments are theoretically identical at the time of the observation (the start of the window), only the hybrid-4DEnVar increments are shown at this time. The increments at the end of the window are shown for both methods.

Firstly, hybrid-4DEnVar and hybrid-4DVar are compared for Cov model i, where the pure climatological covariance is used ($\beta_c^2 = 1.0$, $\beta_e^2 = 0.0$). The plots of the increments are shown in Figs. 4a–c. For this Cov model, hybrid-4DVar is equivalent to 4DVar and hybrid-4DEnVar is equivalent to 3DVar (FGAT). At the start of the window (Fig. 4a), the analysis increments exhibit a homogeneous structure, which is generated by the 3D climatological covariance. As expected from the time-constant design of 3DVar, the analysis increment does not change by the end of the assimilation window (Fig. 4b). On the other hand, the analysis increment for 4DVar moves downstream (Fig. 4c), since $\mathbf{B}$ is implicitly propagated by the PF and adjoint models ($\underline{\mathbf{M}}\,\mathbf{B}\underline{\mathbf{M}}^T$). This experiment clearly shows the implicit flow dependence generated by the PF and adjoint models.

Next Cov model ii is examined, where hybrid-4DEnVar and hybrid-4DVar use a pure ensemble covariance ($\beta_c^2 = 0.0$, $\beta_e^2 = 1.0$) and severe localization ($L = 500$ km). Cov model ii is a comparison between 4DEnVar and En-4DVar and is shown in Figs. 4d–f. It is evident that the analysis increment at the beginning of the window (Fig. 4d)

---

[3] For a discussion of the relationship between L and other localization functions see Clayton et al. (2013, their section 2.5).

[4] Very small differences exist in practice due to the differences in linearization noted at the end of section 2.
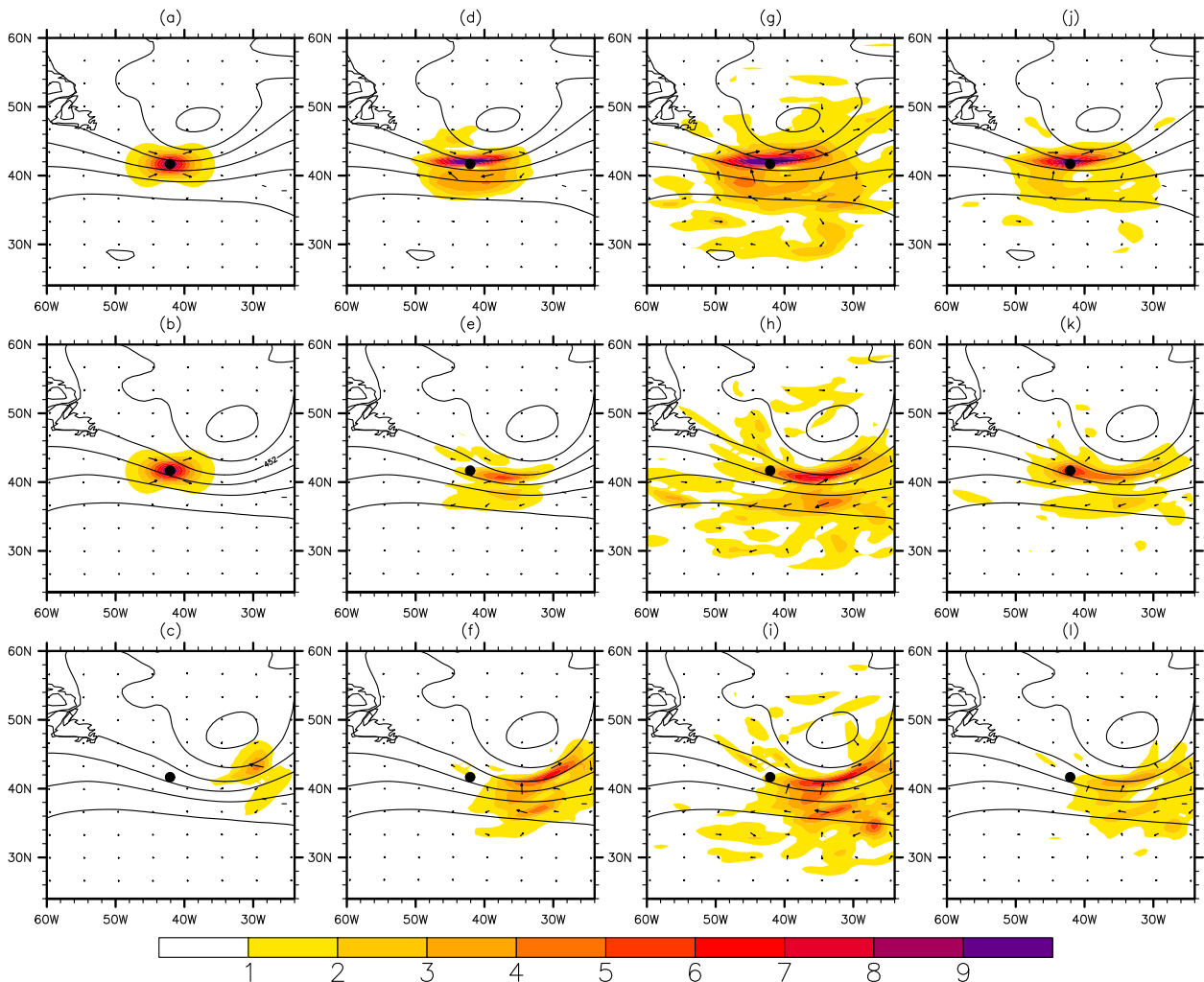
FIG. 4. Wind analysis increments for the jet stream case. Using Cov model i: (a) hybrid-4DEnVar increment at the start of the window (like the hybrid-4DVar increment at this time), (b) at the end of the window, and (c) hybrid-4DVar increment at the end of the window. (d)–(f),(g)–(i),(j)–(l) Equivalent methods to (a)–(c) and correspond to Cov models ii, iii, and iv, respectively. As in Fig. 2 the dot shows the observation position.

exhibits a heterogenous structure, since the ensemble co-variance captures the "errors of the day." Note that such severe localization is not used in operational practice, but it clearly demonstrates an important effect. Both increments have moved downstream by the end of the assimilation window, but the En-4DVar increment (Fig. 4f) has moved farther downstream than the 4DEnVar increment (Fig. 4e). The localization between different times in 4DEnVar makes no allowance for the flow, which explains why the increment is smaller in magnitude and has not moved as far downstream as the En-4DVar increment. Fairbairn et al. (2014) used a simple model to demonstrate that this effect degrades the 4DEnVar analysis.

Cov model iii is shown in Figs. 4g–i. This is the same as Cov model ii but uses the longer 1200-km localization length scale used operationally at the Met Office. Thus,

the analysis increments are spread over a larger area. The analysis increments at the end of the window for En-4DVar and 4DEnVar are similar, since the localization length scale is large enough to not significantly degrade the time correlations of the 4DEnVar ensemble covariance.

Cov model iv is shown in Figs. 4j–l, where the hybrid background-error covariance matrix is used. The analysis increments for Cov model iv are effectively a linear combination of the increments for Cov model i and Cov model iii. It is clear that the main difference between the two methods is the way they use the climatological covariance, rather than the ensemble covariance. The hybrid-4DEnVar method uses a time-invariant climatological co-variance, so part of the analysis increment is permanently centered around the observation. The 4D ensemble co-variance then propagates part of the analysis increment
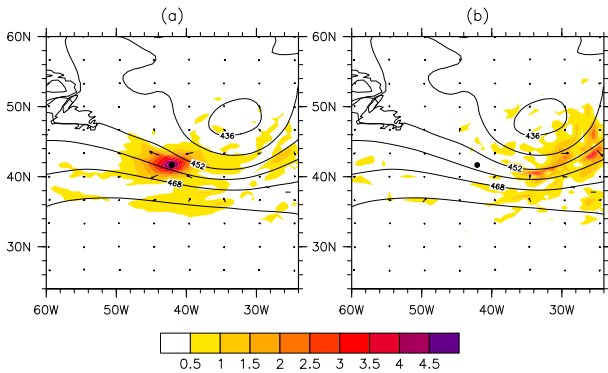
FIG. 5. Strong-constraint wind errors for the jet stream case and Cov model iv, for which the increments were plotted in Figs. 4k,l: (a) Hybrid-4DEnVar and (b) hybrid-4DVar. The dot shows the observation position.

downstream. On the other hand, the hybrid-4DVar climatological and ensemble covariances are both propagated by the PF and adjoint models. Hence all of the hybrid-4DVar analysis increment is propagated downstream.

c. Strong-constraint errors

The strong-constraint errors for Cov model iv are shown in Fig. 5. The location of the largest errors for hybrid-4DEnVar agrees with Fig. 4k, where part of the analysis increment was not propagated downstream from the observation. These large errors are not present in hybrid-4DVar. Instead, hybrid-4DVar has some errors downstream, which are probably caused mainly by the simplified physics in the PF model, though as for hybrid-4DEnVar there will also be errors due to the higher resolution used for the deterministic forecast. The strong-constraint errors for hybrid-4DEnVar appear to be larger overall than the strong-constraint errors for hybrid-4DVar, which is clarified below.

d. Relative errors

The relative strong-constraint errors for Cov model iii and Cov model iv are shown in Table 4. For Cov model iii, 4DEnVar performs slightly better than En-4DVar. This suggests that the linear assumption made by using the 4DVar PF and adjoint models is less accurate than the linear assumption made by using the 4DEnVar alpha-control variable. However, for Cov model iv, hybrid-4DVar performs significantly better than hybrid-4DEnVar. The results for Cov model iv agree with the trials of section 3, where hybrid-4DVar performed better than hybrid-4DEnVar. The results suggest that the inferior performance of hybrid-4DEnVar in the trials is related to the inferior 4D representation of the climatological background-error covariance.

e. Hurricane Sandy results

1) ANALYSIS INCREMENTS

The experiments for the Hurricane Sandy case are performed in exactly the same way as the experiments for the jet stream case. The analysis increments for Cov model iii and Cov model iv are shown in Fig. 6. It is evident that the ensemble covariance spreads the increment around the hurricane. Because of the high winds and sharp gradients near a hurricane, any variations in position cause an unusually large ensemble covariance; in comparison the climatological covariance is almost negligible. This explains why the increments for Cov model iv (Figs. 6d–f) are approximately half the size of the increments for Cov model iii (Figs. 6a–c).

2) STRONG-CONSTRAINT ERRORS

Figure 7 shows the strong-constraint errors for Cov model iv. The strong-constraint errors for both hybrid-4DEnVar and hybrid-4DVar are largest near the hurricane eye. The strong-constraint errors for hybrid-4DVar are significantly larger than the strong-constraint errors for hybrid-4DEnVar. This may be related to the simplified physics of the PF model. In particular, its moist processes such as precipitation (Stiller and Ballard 2009; Stiller 2009), which are important in hurricane development, use very simple parameterizations with global coefficients almost certainly inappropriate for hurricanes.

3) RELATIVE ERRORS

Table 4 shows the relative strong-constraint errors for Cov models iii and iv. In Cov model iii, 4DEnVar performs significantly better than En-4DVar. As discussed above, this is probably due to the simplified PF model physics. In Cov model iv, hybrid-4DEnVar also performs better than hybrid-4DVar, since the ensemble covariances dominate the hybrid near hurricanes.

5. J_c, IAU, and 4DIAU

The Met Office 4DVar scheme includes a penalty J_c on high-frequency oscillations, as revealed by applying a digital time filter to the increment trajectory $\delta\mathbf{x}$. The penalty term is similar to that of Gauthier and Thépaut (2001) but uses a different norm to measure the size of the filter increments.[5] This constraint is effective enough that no further measures are needed to deal with imbalances—we just add

---

[5] While Gauthier and Thépaut (2001) use the full energy norm, the Met Office scheme uses only the elastic term that depends on the pressure increment, and applies this to both pressure and pressure tendency.

TABLE 4. The relative strong-constraint errors (RE) for Cov models iii and iv with the jet stream case and the Hurricane Sandy case.

Methods	Cov model	Jet RE (%)	Sandy RE (%)
4DEnVar	iii	51	57
En-4DVar	iii	54	69
Hybrid-4DEnVar	iv	78	66
Hybrid-4DVar	iv	66	75

the analysis increment valid at the start of the window directly to the background $\mathbf{x}^b$. Since there is no model within 4DEnVar, we cannot expect the 4DVar J_c term to be effective, so it was omitted and there is currently no dynamical penalty on imbalance within 4DEnVar. Thus, the increments produced by hybrid-4DEnVar are poorly balanced, making it necessary to deal with the high-frequency gravity–inertia waves that are generated when the increments are added to the model.[6]

A simple method of filtering high-frequency oscillations is the IAU scheme (Bloom et al. 1996). Here a fraction of the same 3D analysis increment is added each model time step over a time window. This weighted displacement has a time-filtering effect on the increments, similar to that of the time filter in the J_c term (Polavarapu et al. 2004). The IAU method was used for our 3DVar and 3DEnVar experiments, inserting the analysis increment over a 6-h span with uniform weights.

Since 4DEnVar directly produces a 4D analysis increment, it is possible to modify the IAU scheme to add increments valid at the correct time to each model time step; we call this 4DIAU. Although superficially similar, 4DIAU has different time-filtering properties to IAU because there is no displacement in time.[7] In theory, waves that evolve the same way in $\delta\mathbf{x}$ as in $\underline{M}$ (when added to $\mathbf{x}^b$) are not time filtered as they would be in the J_c term or the IAU. However, as long as the ensemble trajectories themselves contain relatively little noise, most noise will come from the manipulations of each 3D state in $\delta\underline{\mathbf{x}}$, and will not be dynamically consistent in time. Thus, with 4DEnVar we can expect 4DIAU to filter gravity wave noise to some extent while not damping real signals correctly represented in the ensemble.

To investigate the degree of high-frequency gravity wave activity in the various systems, we obtained near-surface pressure p_1 increment time series produced by adding

hybrid-4DVar and hybrid-4DEnVar analysis increments to a test forecast, using a variety of insertion strategies:

1) instantaneous at $T - 3$—the start of the analysis window;
2) instantaneous at $T + 0$;
3) instantaneous at $T + 3$;
4) uniform IAU from $T - 3$ to $T + 3$, using the increment valid at $T + 0$; and
5) uniform 4DIAU from $T - 3$ to $T + 3$, using the nearest hourly increment to each time.

The part of the p_1 evolution due to the analysis increment was obtained by subtracting the time series from the background forecast trajectory, and two balance diagnostics were calculated. The mean absolute tendency plots on the left of Fig. 8 tend to emphasize very high-frequency oscillations, which are dissipated rapidly by the filtering built into the forecast model's dynamics. The mean absolute high-pass p_1 plots on the right measure the average amplitude of the high-frequency oscillations revealed by applying the same time filter as used in the 4DVar J_c term—a more relevant measure for data assimilation. For the high-frequency oscillations captured by the filter, this diagnostic gives more weight to relatively slow oscillations than the tendency diagnostic. Since these slower oscillations tend to persist longer in the model, the value of this diagnostic reduces more slowly than that of the tendency diagnostic. The frequency response of the J_c filter, and that of the standard 6-h uniform-weights IAU scheme used with 3DVar and in these balance experiments is shown in Fig. 9.

The basic findings are as follows:[8]

- The degree of imbalance produced by a hybrid-4DEnVar increment is similar to that produced by the $T - 3$ increment from hybrid-4DVar without a J_c term (cf. the black, red, and pink lines in Figs. 8e,f with the black lines in Figs. 8c,d).
- The hybrid-4DEnVar with 4DIAU appears to be well balanced—better than hybrid-4DVar with J_c (cf. the blue lines in Figs. 8e,f with the black lines in Figs. 8a,b).
- For hybrid-4DEnVar, 4DIAU is almost as effective as the 6-h IAU scheme (cf. the blue and green lines in Figs. 8e,f.).
- For hybrid-4DVar without J_c, 4DIAU (blue lines in Figs. 8c,d) is less effective than both IAU (green lines in the same figures) and J_c (black lines in Figs. 8a,b). In fact, 4DIAU is not much better than just applying the

---

[6] We ran an experiment without such measures: the scores (not shown) were much worse than for any experiment in section 3.

[7] Because our $\delta\underline{\mathbf{x}}$ is stored at hourly intervals and we take the nearest state rather than interpolating, there is actually an hour-long IAU of each substate. This filters out oscillations with periods less than about an hour: a much weaker filter than in the 6-h IAU, as shown in Fig. 9.

[8] The apparent increases in imbalance after $T + 12$ are due to diurnal variations in grid-scale noise produced by the model's convection scheme, and are unrelated to imbalances produced by the analysis increments.
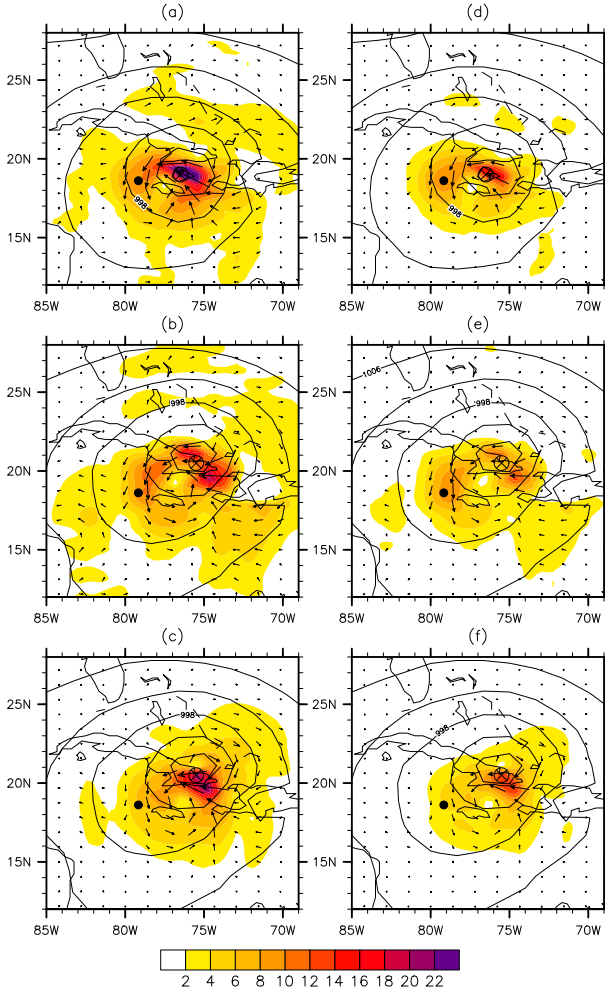
FIG. 6. Wind analysis increments for the Hurricane Sandy case. (left) Using Cov model iii: (a) hybrid-4DEnVar increment at the start of the window (approximately equivalent to hybrid-4DVar at this time), (b) at the end of the window, and (c) hybrid-4DVar increment at the end of the window. (right) Using Cov model iv: (d)–(f) equivalent methods to (a)–(c). As in Fig. 3 the dot and ⊗ symbol show the observation and hurricane positions.
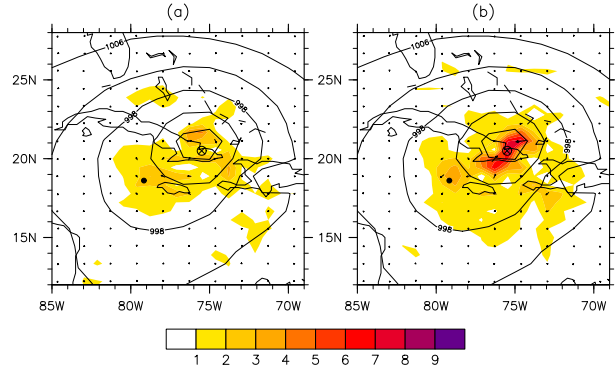

FIG. 7. Strong-constraint wind errors for the Hurricane Sandy case and Cov model iv shown in Figs. 6e,f: (a) Hybrid-4DEnVar and (b) hybrid-4DVar.

analysis increment at $T - 3$ and allowing the filtering built into the forecast model to deal with the noise (cf. the black and blue lines in Figs. 8c,d).

The time-filtering properties of IAU and 4DIAU are a result of destructive interference between increments added at different times, so the ineffectiveness of 4DIAU when used with 4DVar is due to the dynamical consistency of the increments. The essence of this effect is illustrated in Fig. 10, which shows (first seven rows) high-frequency components of the p_1 field at $T + 6$ produced by adding analysis increments (climatological component alone, corresponding to $\mathbf{v}$) at $T - 3$, $T - 2$, ..., $T + 3$. This insertion process is similar to what is done in IAU or 4DIAU, but here we add full analysis

increments, and add the increments every hour rather than every time step. Figures 10a,c are for 4DIAU-like insertions, where the increment times match the insertion times. Figures 10b,d are for IAU-like insertions where the same $T + 0$ increment is added at all times. For hybrid-4DEnVar (Figs. 10a,b), the climatological components of the analysis increments are the same at all times, so 4DIAU is equivalent to IAU. The (same) increment added at different times produces much different high-frequency structures at $T + 6$, which tend to cancel each other when averaged together (bottom panels). For hybrid-4DVar without a J_c term, the effect of the IAU-like insertions (Fig. 10d) is similar to what we see with hybrid-4DEnVar. However, because of the dynamical consistency between the hourly analysis increments, the 4DIAU-like insertions (Fig. 10c) produce very similar high-frequency patterns at $T + 6$: what differences we do see are the result of mismatches between the evolution of analysis increments in the 4DVar PF model and in the full model, but these are relatively small. Thus, the average increment shown on the bottom row of Fig. 10c is similar to the individual increments above, with very little cancellation evident.

Corresponding results for the ensemble components of the analysis increments (corresponding to the $\mathbf{v}_k^\alpha$) are shown in Fig. 11. For hybrid-4DEnVar the ensemble components are now time dependent due to the time dependence of the ensemble perturbations. The individual ensemble perturbation trajectories are well balanced, so most imbalance is generated by their Schur multiplication with the 4D alpha-control variable fields $\underline{\boldsymbol{\alpha}_k}$ in Eq. (22). There is no reason to expect gravity waves generated by the Schur product to be dynamically consistent in time, and indeed the structures produced by the 4DIAU-like insertion (Fig. 11a) differ between insertion times. Thus, the column average
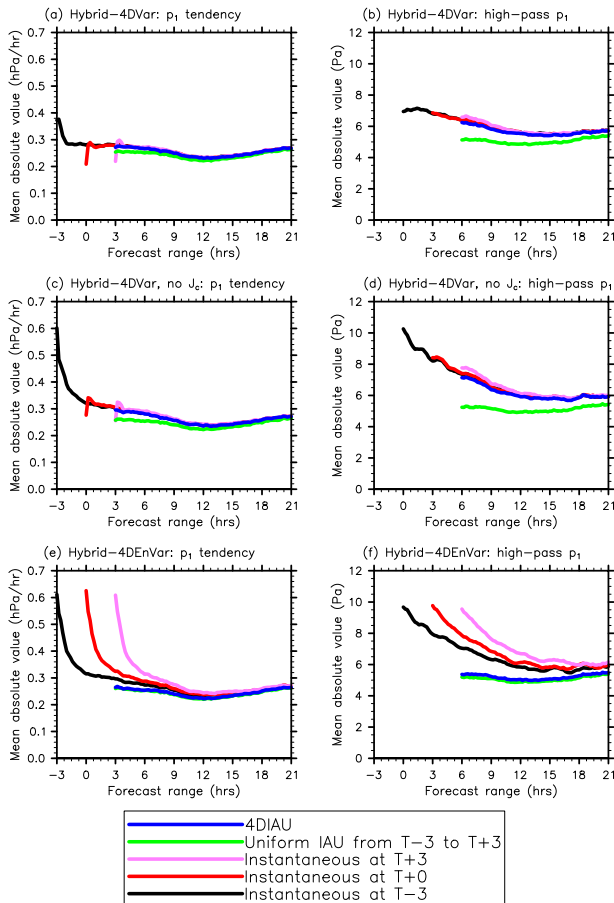
FIG. 8. Mean absolute values of (a),(c),(e) the p_1 increment tendency fields, and (b),(d),(f) the fields obtained by applying a high-pass time filter to the p_1 increment time series. The first row is for increments from a standard hybrid-4DVar run, the second is for a hybrid-4DVar run without the J_c term, and the third is for a standard hybrid-4DEnVar run.
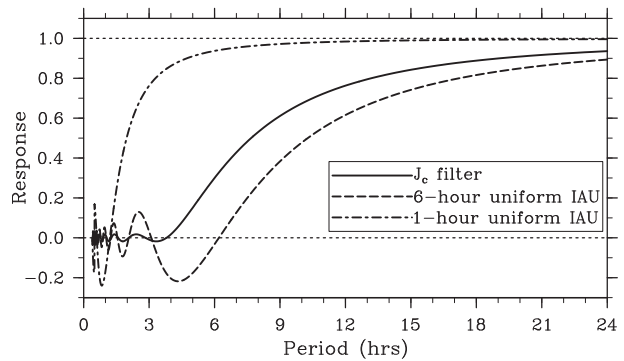

FIG. 9. Response functions for the filter used in the 4DVar J_c term, for the standard 6-h IAU scheme, and for the 1-h IAU scheme built into 4DIAU.

overestimate of the effect of not having a J_c term in 4DEnVar.

6. Discussion and conclusions

We have set out a consistent documentation of the use of climatological, ensemble, and hybrid covariances in 4DVar and 4DEnVar.

We have described an appealingly simple 4DIAU method, for adding the 4D increment from 4DEnVar to the background in a way that reduces imbalance in the subsequent forecast, while not damping real moving features seen in the ensemble.

We have *not* in this paper studied the benefit from hybrid covariances—using an ensemble to add "errors of the day" information to the climatological covariance traditionally used in 3DVar and 4DVar. Instead we used the same 3D hybrid covariances and studied the effect of differences in the way they were extended to four-dimensions in hybrid-4DEnVar compared to hybrid-4DVar. The main conclusion was that, with the same settings (tuned for hybrid-4DVar), the new hybrid-4DEnVar system performed worse than the existing hybrid-4DVar. In detail:

1) The main flaw in our hybrid-4DEnVar, compared to hybrid-4DVar, is that the climatological component of the covariance is not evolved in time: it is used as in 3DVar. [In retrospect it is not surprising that this effect dominates, since we followed Clayton et al. (2013) and gave the climatological covariances a weight of $\beta_c^2 = 0.8$ in the hybrid.]

2) Time-covariance errors resulting from the localization and model not commuting (Fairbairn et al. 2014), and the 3D localization not following the flow, are not important for our 1200-km localization scale and 6-h window, but they do become important for a 500-km scale.

(bottom panel) shows significant destructive interference, comparable with what we see for the IAU-like insertion (Fig. 11b). The relatively small-scale features that remain are likely due to the (balanced) movement and evolution of weather systems represented in the ensemble, which also project onto high frequencies. Comparison of the bottom panels in Figs. 11a,b suggests that this balanced evolution is preserved better by the 4DIAU-like insertion—an advantage for 4DIAU over the standard IAU scheme. For the hybrid-4DVar increments, the behavior is qualitatively similar to what we saw for the climatological components, so again we see that 4DIAU is ineffective.

In light of these results, we see that the 4DVar4DIAU trial described in section 3 included much poorer noise control than the other trials, which suggests that its half-percent deficit against 4DVar (Fig. 1f) is likely an
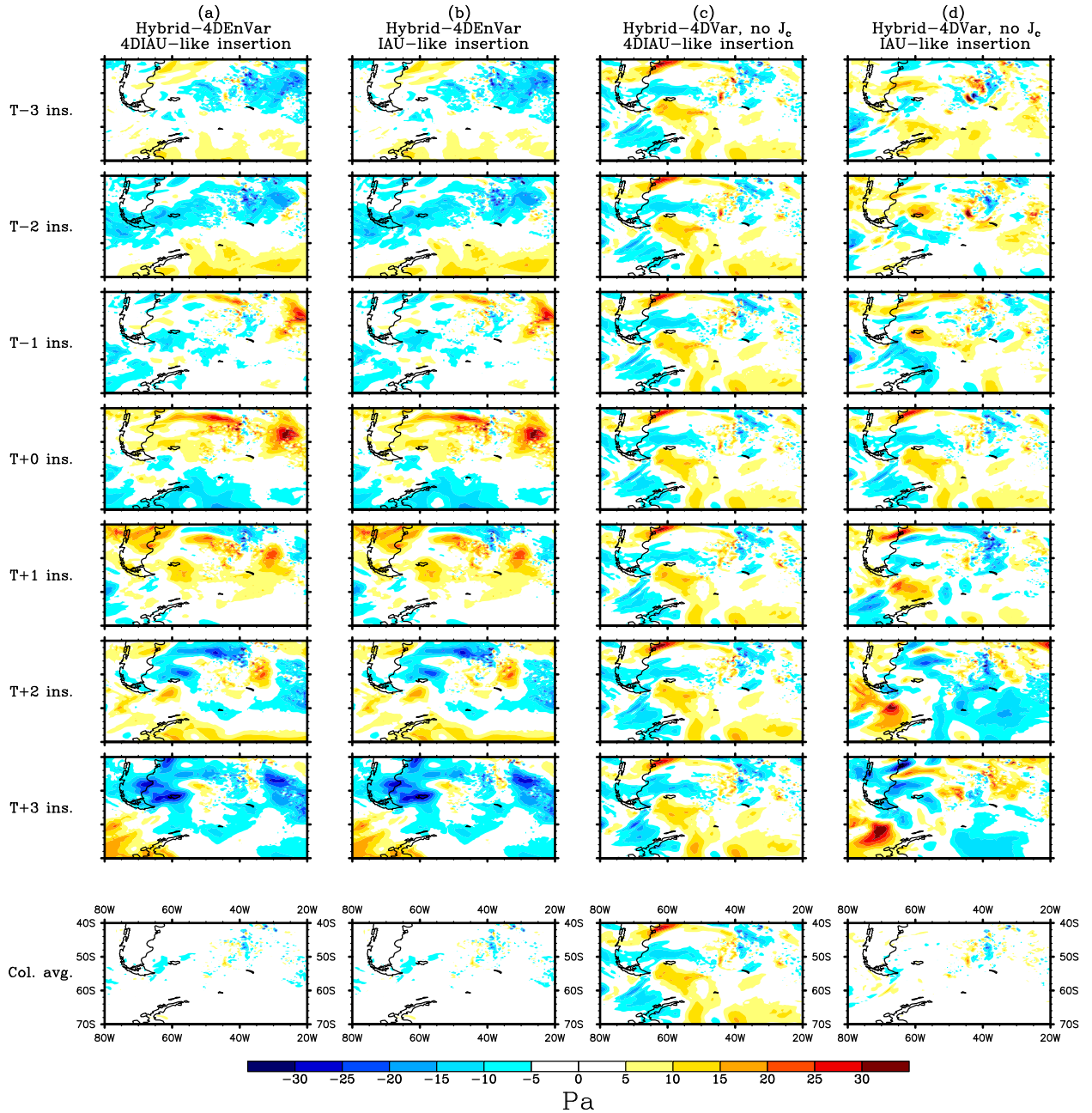
FIG. 10. High-frequency components of the p_1 increment field at $T + 6$ (obtained using the J_c term filter) for analysis increments (climatological components only) added at $T - 3, T - 2, \ldots, T + 3$, and (bottom) the column averages. (a),(b) Hybrid-4DEnVar analysis increments and (c),(d) increments from hybrid-4DVar without J_c. (a),(c) A 4DIAU-like and (b),(d) an IAU-like insertion.

3) The 4DEnVar method was able to use the ensemble component of the covariances often more accurately than in En-4DVar. In the jet stream case evolution errors induced by localization in 4DEnVar were smaller than 4DVar errors, except when the localization was severe. In the hurricane case approximations in the perturbation forecast model induced errors in 4DVar evolutions.

Improving hybrid-4DEnVar

The maintenance and running costs of hybrid-4DVar are larger, so there is an incentive to improve hybrid-4DEnVar. Our results show that to do this we need to reduce the weight on climatological **B** relative to the ensemble covariance. But these weights are usually determined by experiment; both components provide
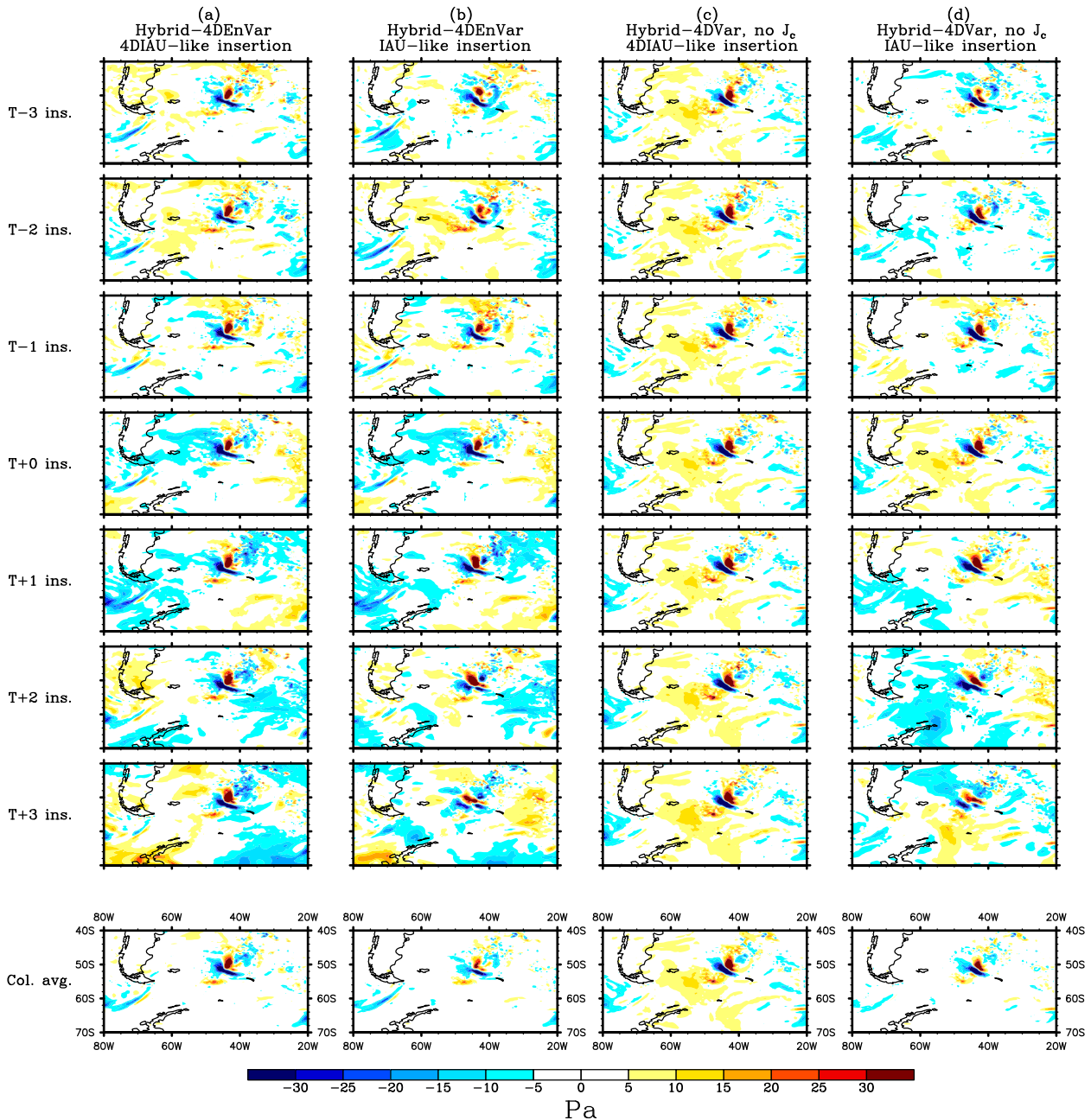
FIG. 11. As in Fig. 10, but for the ensemble components of the analysis increments.

some benefit (Etherton and Bishop 2004; Clayton et al. 2013). Increasing the ensemble weight requires us to first improve the covariances derived from the ensemble by a bigger ensemble, better ensemble generation, and better localization.

The Met Office has increased its operational ensemble size from 23 to 44; further increases are expected as computer power allows. This is guided by Houtekamer et al. (2014), who indicated possible benefits for the Canadian system in increasing the size beyond 192.

The MOGREPS ensemble system (Bowler et al. 2008; Flowerdew and Bowler 2011) was designed and tested to quantify uncertainty in short-period forecasts. Attention in the past focused on having the correct ensemble spread, with less attention on the implied covariances used in data assimilation. Experiments are under way to see whether alternative ensemble generation methods, such as an ensemble of data assimilations (EDA; Bonavita et al. 2012) of 4DEnVar assimilations, give better results.

Localization need not be simply a function of distance: see the discussion around Eq. (16). For instance Buehner and Charron (2007) showed that spectral localization is a spatial smoother. The wave band implementation of Buehner (2012) allows us to use different spatial localization scales for each wave band. Multivariate localization (i.e., making the transformed balanced and unbalanced variables independent) would allow us to impose the balance assumed by the climatological covariance model on the ensemble covariances, at the risk of removing real flow-dependent correlations like those seen by Montmerle and Berre (2010). We have implemented these methods and plan to test their benefit.

Wang et al. (2013) and Wang and Lei (2014) found that when the ensemble resolution matched the assimilation model, the weight β_c^2 given to the climatological covariance could be reduced to zero without reducing skill, whereas when the assimilation model resolution was increased the hybrid was necessary. This might be because they used a larger ensemble (80 members) and also the TLNMC; in the experiments of Kleist and Ide (2015b), hybridization only improved results in the absence of the TLNMC. Nevertheless, when the ensemble resolution is less than that desired in the analysis, Kleist and Ide (2015b) suggest using scale-dependent β_c and β_e—something easy to achieve in conjunction with the wave band filter.

Flow-following localization might be done using the ensemble itself (Bishop and Hodyss 2007, 2009a,b), or it might be done by replacing **l** by something more dynamically based, in Eq. (23). We are planning to investigate a simple advection of the alpha-control variable fields [similar to that suggested by Ota et al. (2013)] to reduced the deleterious effect of strong horizontal localization seen in Fig. 4e. Note that time localization can introduce extra degrees of freedom, analogous to the model error control variables in weak-constraint 4DVar (Lorenc 2003a; Fisher and Auvinen 2012; Desroziers et al. 2014). It should be used in conjunction with allowing for model error in the ensemble (Hamill and Whitaker 2005); Fairbairn et al. (2014) showed that stochastic additive inflation in the ensemble improved 4DEnVar in the presence of model error.

REFERENCES

Bishop, C. H., and D. Hodyss, 2007: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quart. J. Roy. Meteor. Soc.,* **133,** 2029–2044, doi:10.1002/qj.169.

——, and ——, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus,* **61A,** 84–96, doi:10.1111/j.1600-0870.2008.00371.x.

——, and ——, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus,* **61A,** 97–111, doi:10.1111/j.1600-0870.2008.00372.x.

Bloom, S. C., L. L. Takacs, A. M. D. Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.,* **124,** 1256–1271, doi:10.1175/1520-0493(1996)124<1256: DAUIAU>2.0.CO;2.

Bonavita, M., L. Isaksen, and E. Holm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. *Quart. J. Roy. Meteor. Soc.,* **138,** 1540–1559, doi:10.1002/qj.1899.

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.,* **134,** 703–722, doi:10.1002/qj.234.

Buehner, M., 2012: Evaluation of a spatial/spectral covariance localization approach for atmospheric data assimilation. *Mon. Wea. Rev.,* **140,** 617–636, doi:10.1175/MWR-D-10-05052.1.

——, and M. Charron, 2007: Spectral and spatial localization of background-error correlations for data assimilation. *Quart. J. Roy. Meteor. Soc.,* **133,** 615–630, doi:10.1002/qj.50.

——, P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.,* **138,** 1550–1566, doi:10.1175/2009MWR3157.1.

——, ——, ——, ——, and ——, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Wea. Rev.,* **138,** 1567–1586, doi:10.1175/2009MWR3158.1.

——, J. Morneau, and C. Charette, 2013: Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlinear Processes Geophys.,* **20,** 669–682, doi:10.5194/npg-20-669-2013.

Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.,* **139,** 1445–1461, doi:10.1002/qj.2054.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120,** 1367–1387, doi:10.1002/qj.49712051912.

Desroziers, G., J.-T. Camino, and L. Berre, 2014: 4DEnVar: Link with 4D state formulation of variational data assimilation and different possible implementations. *Quart. J. Roy. Meteor. Soc.,* **140,** 2097–2110, doi:10.1002/qj.2325.

Etherton, B. J., and C. H. Bishop, 2004: Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev.,* **132,** 1065–1080, doi:10.1175/1520-0493(2004)132<1065:ROHDAS>2.0.CO;2.

Fairbairn, D., S. R. Pring, A. C. Lorenc, and I. Roulstone, 2014: A comparison of 4D-Var with ensemble data assimilation methods. *Quart. J. Roy. Meteor. Soc.,* **140,** 281–294, doi:10.1002/qj.2135.

Fisher, M., and H. Auvinen, 2012: Long window weak-constraint 4D-Var. *ECMWF Seminar on Data Assimilation for Atmosphere and Ocean,* Shinfield Park, Reading, ECMWF, 189–202.

Flowerdew, J., and N. E. Bowler, 2011: Improving the use of observations to calibrate ensemble spread. *Quart. J. Roy. Meteor. Soc.,* **137,** 467–482, doi:10.1002/qj.744.

Gauthier, P., and J.-N. Thépaut, 2001: Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Wea. Rev.,* **129,** 2089–2102, doi:10.1175/1520-0493(2001)129<2089:IOTDFA>2.0.CO;2.

Gustafsson, N., J. Bojarova, and O. Vignes, 2014: A hybrid variational ensemble data assimilation for the high resolution limited area model (HIRLAM). *Nonlinear Processes Geophys.,* **21,** 303–323, doi:10.5194/npg-21-303-2014.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.,* **128,** 2905–2919, doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

——, and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.,* **133,** 3132–3147, doi:10.1175/MWR3020.1.

——, ——, and C. Snyder, 2001: Distance dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.,* **129,** 2776–2790, doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

Houtekamer, P., X. Deng, H. L. Mitchell, S.-J. Baek, and N. Gagnon, 2014: Higher resolution in an operational ensemble Kalman filter. *Mon. Wea. Rev.,* **142,** 1143–1162, doi:10.1175/MWR-D-13-00138.1.

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

Ingleby, N. B., 2001: The statistical structure of forecast errors and its representation in the Met Office global 3D variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.,* **127,** 209–231, doi:10.1002/qj.49712757112.

——, A. C. Lorenc, K. Ngan, F. R. Rawlins, and D. R. Jackson, 2013: Improved variational analyses using a nonlinear humidity control variable. *Quart. J. Roy. Meteor. Soc.,* **139,** 1875–1887, doi:10.1002/qj.2073.

Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy, 2007: 4-D-Var or ensemble Kalman filter? *Tellus,* **59A,** 758–773, doi:10.1111/j.1600-0870.2007.00261.x.

Kepert, J., 2009: Covariance localisation and balance in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.,* **135,** 1157–1176, doi:10.1002/qj.443.

Kleist, D. T., and K. Ide, 2015a: An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-Hybrid results. *Mon. Wea. Rev.,* in press.

——, and ——, 2015b: An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Wea. Rev.,* in press.

——, D. F. Parrish, J. C. Derber, R. Treadon, R. M. Errico, and R. Yang, 2009: Improving incremental balance in the GSI 3DVAR analysis system. *Mon. Wea. Rev.,* **137,** 1046–1060, doi:10.1175/2008MWR2623.1.

Le Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus,* **38A,** 97–110, doi:10.1111/j.1600-0870.1986.tb00459.x.

Liu, C., Q. Xiao, and B. Wang, 2008: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.,* **136,** 3363–3373, doi:10.1175/2008MWR2312.1.

——, ——, and ——, 2009: An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing System Simulation Experiments with advanced research WRF (ARW). *Mon. Wea. Rev.,* **137,** 1687–1704, doi:10.1175/2008MWR2699.1.

Lorenc, A. C., 2003a: Modelling of error covariances by four-dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.,* **129,** 3167–3182, doi:10.1256/qj.02.131.

——, 2003b: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.,* **129,** 3183–3203, doi:10.1256/qj.02.132.

——, 2013: Recommended nomenclature for EnVar data assimilation methods. *Research Activities in Atmospheric and Oceanic Modelling,* WGNE, 2 pp. [Available online at http://www.wcrp-climate.org/WGNE/BlueBook/2013/individual-articles/01_Lorenc_Andrew_EnVar_nomenclature.pdf.]

——, and F. Rawlins, 2005: Why does 4D-Var beat 3D-Var? *Quart. J. Roy. Meteor. Soc.,* **131,** 3247–3257, doi:10.1256/qj.05.85.

——, and Coauthors, 2000: The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.,* **126,** 2991–3012, doi:10.1002/qj.49712657002.

Lynch, P., and X.-Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.,* **120,** 1019–1034, doi:10.1175/1520-0493(1992)120<1019:IOTHMU>2.0.CO;2.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.,* **130,** 2791–2808, doi:10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2.

Montmerle, T., and L. Berre, 2010: Diagnosis and formulation of heterogeneous background-error covariances at the mesoscale. *Quart. J. Roy. Meteor. Soc.,* **136,** 1408–1420, doi:10.1002/qj.655.

Ota, Y., J. C. Derber, E. Kalnay, and T. Miyoshi, 2013: Ensemble-based observation impact estimates using the NCEP GFS. *Tellus,* **65A,** 20038, doi:10.3402/tellusa.v65i0.20038.

Polavarapu, S., S. Ren, A. Clayton, D. Sankey, and Y. Rochon, 2004: On the relationship between incremental analysis updating and incremental digital filtering. *Mon. Wea. Rev.,* **132,** 2495–2502, doi:10.1175/1520-0493(2004)132<2495:OTRBIA>2.0.CO;2.

Rabier, F., 2005: Overview of global data assimilation developments in numerical weather-prediction centres. *Quart. J. Roy. Meteor. Soc.,* **131,** 3215–3233, doi:10.1256/qj.05.129.

Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and T. J. Payne, 2007: The Met Office global 4-dimensional variational data assimilation system. *Quart. J. Roy. Meteor. Soc.,* **133,** 347–362, doi:10.1002/qj.32.

Stiller, O., 2009: Efficient moist physics schemes for data assimilation. II: Deep convection. *Quart. J. Roy. Meteor. Soc.,* **135,** 721–738, doi:10.1002/qj.362.

——, and S. P. Ballard, 2009: Efficient moist physics schemes for data assimilation. I: Large-scale clouds and condensation. *Quart. J. Roy. Meteor. Soc.,* **135,** 707–720, doi:10.1002/qj.400.

Tian, X., Z. Xie, and A. Dai, 2008: An ensemble-based explicit four-dimensional variational assimilation method. *J. Geophys. Res.,* **113,** D21124, doi:10.1029/2008JD010358.

Wang, X., and T. Lei, 2014: GSI-based four dimensional ensemble-variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP global forecast system. *Mon. Wea. Rev.,* **142,** 3303–3325, doi:10.1175/MWR-D-13-00303.1.

——, C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble–3DVAR hybrid analysis schemes. *Mon. Wea. Rev.,* **135,** 222–227, doi:10.1175/MWR3282.1.

——, D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.,* **136,** 5116–5131, doi:10.1175/2008MWR2444.1.

——, ——, ——, and ——, 2008b: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiments. *Mon. Wea. Rev.,* **136,** 5132–5147, doi:10.1175/2008MWR2445.1.

——, D. Parrish, D. Kleist, and J. S. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP global forecast system: Single-resolution experiments. *Mon. Wea. Rev.,* **141,** 4098–4117, doi:10.1175/MWR-D-12-00141.1.

Zhang, M., and F. Zhang, 2012: E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. *Mon. Wea. Rev.,* **140,** 587–600, doi:10.1175/MWR-D-11-00023.1.