Hybrid Ensemble–Variational Filter: A Spatially and Temporally Varying Adaptive Algorithm to Estimate Relative Weighting

MOHAMAD EL GHARAMTI^a

^a *National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 30 March 2020, in final form 8 September 2020)

ABSTRACT: Model errors and sampling errors produce inaccurate sample covariances that limit the performance of ensemble Kalman filters. Linearly hybridizing the flow-dependent ensemble-based covariance with a time-invariant background covariance matrix gives a better estimate of the true error covariance. Previous studies have shown this, both in theory and in practice. How to choose the weight for each covariance remains an open question especially in the presence of model biases. This study assumes the weighting coefficient to be a random variable and then introduces a Bayesian scheme to estimate it using the available data. The scheme takes into account the discrepancy between the ensemble mean and the observations, the ensemble variance, the static background variance, and the uncertainties in the observations. The proposed algorithm is first derived for a spatially constant weight and then this assumption is relaxed by estimating a unique scalar weight for each state variable. Using twin experiments with the 40-variable Lorenz 96 system, it is shown that the proposed scheme is able to produce quality forecasts even in the presence of severe sampling errors. The adaptive algorithm allows the hybrid filter to switch between an EnKF and a simple EnOI depending on the statistics of the ensemble. In the presence of model errors, the adaptive scheme demonstrates additional improvements compared with standard enhancements alone, such as inflation and localization. Finally, the potential of the spatially varying variant to accommodate challenging sparse observation networks is demonstrated. The computational efficiency and storage of the proposed scheme, which remain an obstacle, are discussed.

KEYWORDS: Bayesian methods; Filtering techniques; Inverse methods; Kalman filters; Ensembles; Numerical weather prediction/forecasting

1. Introduction

For the past ~2 decades, the pursuit of improving the performance of ensemble Kalman filters (EnKFs) has been the focus of many research studies across various Earth system applications including numerical weather prediction (NWP) and oceanography (Carrassi et al. 2018, and references therein). The EnKF (Evensen 2003) approximates the background error covariance matrix with the sample covariance from an ensemble of model states. For unbiased forecasts and in the limit of a large ensemble, this usually produces robust error covariance that accurately represents the uncertainty in the background ensemble mean (Furrer and Bengtsson 2007; Sacher and Bartello 2008). Owing to the large dimension of Earth system models, using a large ensemble size is not always possible. Moreover even in the most sophisticated atmospheric and oceanic solvers, model errors are known to exist. Because of this, the resulting background error covariance estimate may have large inaccuracies that can degrade the quality of the Kalman update.

Methods to ameliorate the imperfections in the background covariance include inflation (e.g., Pham et al. 1998; Mitchell and Houtekamer 2000; Anderson and Anderson 1999; Whitaker and Hamill 2012; El Gharamti 2018), localization

(e.g., Houtekamer and Mitchell 1998; Bishop and Hodyss 2009a,b; Anderson 2012; Lei et al. 2016) and the use of multiphysics ensembles (e.g., Skamarock et al. 2008; Meng and Zhang 2008; Berner et al. 2011). Inflation increases the spread of the ensemble around the ensemble mean without changing the rank of the covariance. The rank may change if the inflation is designed to vary in space (El Gharamti et al. 2019). Localization, on the other hand, tackles sampling error by reducing spurious correlations in the error covariance usually resulting in a full-rank matrix. The multiphysics, often referred to as multischeme, approach addresses model error and limited ensemble spread by using ensemble members with different model configurations.

Because the ensemble size is limited, certain parts of the state's distribution may be hidden or unknown to the sample covariance matrix. Hamill and Snyder (2000) argued that blending in some static (climatological) information in the ensemble covariance might present the ensemble with new correction directions. This can be achieved through a linear combination of the flow-dependent ensemble covariance and a stationary background covariance, typically used in 3DVAR and Optimal Interpolation (OI) systems. The work of Bishop and Satterfield (2013) and Bishop et al. (2013) confirmed this and proposed an analytical justification for such linear hybridization, under certain assumptions. In practice, many studies have used this hybrid covariance approach (e.g., Etherton and Bishop 2004; Buehner 2005; Wang et al. 2008; Kuhl et al. 2013; Clayton et al. 2013; Penny et al. 2015; Bowler et al. 2017) and proved its effectiveness to enhancing the ensemble filtering procedure.

Denotes content that is immediately available upon publication as open access.

Corresponding author: Moha Gharamti, gharamti@ucar.edu

Linearly combining the background ensemble and static covariance matrices requires tuning a weighting factor. Gharamti et al. (2014) proposed to optimize this factor by maximizing (or minimizing) the information gain between the forecast and the analysis statistics. The authors tested their approach for state and parameter estimation in a subsurface contaminant transport problem. Ménétrier and Auligné (2015) used a variational procedure to find the optimal localization and hybridization parameters simultaneously. Their technique tackles sampling error by minimizing the quadratic error between the localized-hybrid covariance and an asymptotic covariance matrix (obtained with a very large ensemble). The authors validated their method using a 48 h WRF ensemble forecast system in north America. Satterfield et al. (2018) estimated the weighting factor by first, finding the distribution of the true error variance given the ensemble variance, and then computing its mean. Using high-order polynomial regression, the authors could approximately estimate the weighting between the static and the ensemble covariance.

Instead of weighing the ensemble and static background covariance matrices with a deterministic factor, the following study aims to investigate the behavior of the hybrid system when the weighting factor is assumed to be “a random variable.” A Bayesian algorithm is proposed in which the probability density function (pdf) of the weighting factor is updated, at each data assimilation (DA) cycle, using the available data. The derivation is presented for two scenarios: a spatially constant and a spatially varying weighting field. The proposed scheme has a number of attractive features: 1) it is sequential in nature such that the posterior pdf of the weight at the current time becomes the prior at the next DA cycle, 2) it allows for adaptive selection of the covariance weighting without manual tuning, 3) it can mitigate not only sampling errors but also model errors, and 4) its ease of implementation.

The rest of the paper is organized as follows. Section 2 introduces a Bayesian approach to estimate the hybrid weighting factor. The performance of the proposed scheme is compared to the EnKF using twin-experiments in section 3. Sensitivity to ensemble size, model error and observation network is tested and analyzed. A summary of the findings and further discussions are found in section 4.

2. Adaptive hybrid ensemble–variational scheme

In the EnKF, the prior distribution $p(\mathbf{x}_k|\mathbf{Y}_{k-1})$ of state $\mathbf{x} \in \mathbb{R}^{N_x}$, at any time t_k , is assumed Gaussian with the following moments:

$$\bar{\mathbf{x}}_k = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}_k^i, \quad i = 1, 2, \dots, N_e, \quad (1)$$

$$\mathbf{P}_k^e = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (\mathbf{x}_k^i - \bar{\mathbf{x}}_k)(\mathbf{x}_k^i - \bar{\mathbf{x}}_k)^T, \quad (2)$$

$\mathbf{x}_k^i$ are the individual state ensemble members, $\bar{\mathbf{x}}_k$ is the sample ensemble mean, $\mathbf{P}_k^e$ is the sample covariance, N_e is the ensemble size (e refers to “ensemble”), and N_x is the size of the state. Let $\mathbf{Y}_{k-1}$ be the set of all observations up to time $k - 1$.

The EnKF is able to provide reliable background error covariances as long as the forecast model is not extremely biased and in the case of large ensemble sizes. Random and systematic model errors, over successive DA cycles, can cause filter divergence. The use of small ensembles, on the other hand, causes the EnKF to be rank-deficient and the true variance to be underestimated.

Hybridizing the ensemble covariance with static (time-invariant) background covariance $\mathbf{B}$ can be done as follows:

$$\hat{\mathbf{P}}_k = \alpha \mathbf{P}_k^e + (1 - \alpha) \mathbf{B}, \quad (3)$$

where α is a weighting factor chosen between 0 and 1 and $\hat{\mathbf{P}}_k$ is the hybrid form of the background covariance that can be used during the update; $\mathbf{B}$ is a stationary error covariance matrix typically used in 3–4D variational and optimal interpolation systems. It can be thought of as a climatological covariance; often formed from a large inventory of historical forecasts that are sampled over large time windows. One technique to construct $\mathbf{B}$ comes from NWP and uses the so-called National Meteorological Center (NMC; Parrish and Derber 1992) method. Matrix $\mathbf{B}$ can also be constructed through a series of transform (e.g., balance, horizontal, vertical) operators (Wu et al. 2002). A detailed procedure on how to model $\mathbf{B}$ (for cloud and chemical DA systems) can be found in Descombes et al. (2015). In general, storing the full entries of $\mathbf{B}$ may not be possible and thus smart decomposition and storage tools are necessary.

Setting $\alpha = 1$, Eq. (3) results in an EnKF with a purely flow-dependent ensemble-based covariance. For $\alpha = 0$, the system morphs into an ensemble optimal interpolation (EnOI) scheme. Some studies impose two different weighting factors on the ensemble and the static covariances such that the sum of the weights does not add to 1. This is usually attributed to under or overdispersion in the spread of certain variables in $\mathbf{B}$. Assuming that both $\mathbf{P}_k^e$ and $\mathbf{B}$ give independent estimates of the true background covariance then one would expect the weights to add to 1 (Lorenc et al. 2015). Accordingly, in this study we adhere to this argument and use a single factor α .

a. Bayesian approach to choose α

Here, the weighting factor α is assumed to be a random variable. The motivation for this is the fact that the underlying conditions of the ensemble and the observations change over time. For instance, model errors cause the ensemble spread to evolve at different times and seasons. Furthermore, the number and uncertainty of observations may also change dramatically (e.g., in a convection event). A single deterministic value for α may not be able to efficiently respond to such changes. Similar arguments have led to development of various adaptive inflation techniques. To this end, we propose to estimate α , at any analysis time t_k , following Bayes’s rule as follows:

$$p(\alpha|\mathbf{d}_k) \approx p(\alpha) \cdot p(\mathbf{d}_k|\alpha), \quad (4)$$

where $\mathbf{d}_k = \mathbf{y}_k - \overline{h(\mathbf{x}_k^i)}$ is the background innovation. The observations at t_k are given by $\mathbf{y}_k$ and h is a nonlinear forward operator relating the model state variables to the observed quantities. The overbar denotes averaging over all ensemble

members. The first term on the right hand side of Eq. (4) represents the prior marginal distribution for α . The other term is the likelihood, assumed Gaussian, with mean $\mathbf{0}$ and variance:

$$\mathbb{E}(\mathbf{d}_k \mathbf{d}_k^T) = \mathbf{R}_k + \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T, \quad (5)$$

where $\mathbf{R}_k$ is the observation error covariance matrix and $\mathbf{H}_k$ is a linearized forward operator.¹ $\mathbf{P}_k = \mathbb{E}(\varepsilon \varepsilon^T)$ is the true (theoretical) prior covariance matrix and ε denotes the unbiased background errors. Equation (5) assumes that the background and observation errors are uncorrelated (Desroziers et al. 2005) and that $\overline{h(\mathbf{x}_k^i)} = h(\overline{\mathbf{x}_k})$. Assuming that the hybrid covariance $\hat{\mathbf{P}}_k$ in Eq. (3) is a good approximation of $\mathbf{P}_k$, we can substitute its form to obtain the following:

$$\mathbb{E}(\mathbf{d}_k \mathbf{d}_k^T) = \mathbf{R}_k + \alpha \mathbf{H}_k \mathbf{P}_k^e \mathbf{H}_k^T + (1 - \alpha) \mathbf{H}_k \mathbf{B} \mathbf{H}_k^T. \quad (6)$$

From Eq. (6), the likelihood can be written as

$$p(\mathbf{d}_k | \alpha_k) = \frac{1}{\sqrt{2\pi}\theta} \exp\left(-\frac{\mathbf{d}_k^T \mathbf{d}_k}{2\theta^2}\right), \quad (7)$$

in which $\theta^2(\alpha) = \text{trace}(\mathbf{R}_k) + \alpha \text{trace}(\mathbf{H}_k \mathbf{P}_k^e \mathbf{H}_k^T) + (1 - \alpha) \text{trace}(\mathbf{H}_k \mathbf{B} \mathbf{H}_k^T)$. We note that this likelihood is only Gaussian with respect to the innovation and not α . This will be clear in the following 1D example.

Initially, the prior pdf for α needs to be specified. A simple choice could be a Gaussian. Since α only varies between 0 and 1, a beta distribution may be a more suitable choice. The algorithm proceeds at each assimilation step, by taking the product of the pdfs as shown in Eq. (4). After finding the posterior pdf for α , the mode can be computed and used in Eq. (3) in the next forecast cycle. The subsequent prior is set to be the same as the posterior.

ILLUSTRATION

We would like to test the Bayesian scheme outlined above in a simple 1D example. We consider a single variable ensemble with variance $\sigma_e^2 = 0.9$, a static background variance of $\sigma_s^2 = 0.2$ (s refers to “static”), observation error variance $\sigma_o^2 = 0.1$ and an innovation $d = 2.5$. For the prior distribution of α , we present two examples: 1) a Gaussian distribution centered at 0.5 with variance equal to 0.05 and 2) a beta distribution with scale and shape parameters both equal to 2. The parameters of the beta distribution are chosen such that the mode and the variance match those of the Gaussian pdf.

The prior pdfs and the likelihood are plotted in Fig. 1. For visual purposes, the likelihood (red curve) has been scaled so it can be viewed properly on the same axis with the other pdfs. Multiplying the priors and the likelihood produces the two blue posterior pdfs. Analytically, the posteriors are not exactly beta and Gaussian like their prior counterparts (although visually

FIG. 1. A single variable example illustrating the Bayesian update for α . Two choices for the marginal prior distribution of α are considered: beta (thick black line) and Gaussian (thin black line). The likelihood function estimated from Eq. (7) is represented by the dot-dashed red line. The thick and thin blue lines are the posterior distributions for the beta and Gaussian priors, respectively. “M” in the legend denotes the mode of the pdf, while “V” is the variance.

they are similar). Before the update, equal weight is placed on the ensemble and static background variance. After the update, the modes of the Gaussian-like and beta-like posteriors are: $\alpha = 0.66$ and $\alpha = 0.76$, respectively. Because of the relatively large innovation (prior mean minus observation), the update places more weight on the larger variance (the ensemble here) thus increasing the background spread. Roughly speaking, this increases the “apparent” consistency between the prior and the observation. As can be seen, the variance of the prior pdfs shrunk by almost 42% after the update.

To study the impact of the innovation on the proposed Bayesian update scheme, we consider 3 cases: small bias ($d = 0.01$), moderate bias ($d = 1$) and large bias ($d = 3$). We assess the behavior for a wide range of values and different combinations for σ_e^2 and σ_s^2 . Both variances are set to vary between 0.01 and 3 with a step of 0.01. The observation error variance is fixed and equal to 0.2 and we only look at the Gaussian prior case² in which $p(\alpha) \sim \mathcal{N}(0.5, 0.05)$. The resulting posterior estimates for α are shown in Fig. 2. As can be seen from the 3 scenarios, when the background variances match (i.e., $\sigma_e^2 = \sigma_s^2$), the algorithm weights them equally such that $\alpha = 0.5$. When the bias is large, as we saw in Fig. 1, the larger variance gets assigned bigger weight to better fit the observation. For small biases, the algorithm behaves in an opposite fashion in the sense that the smaller variance gets the bigger weight. At first this might seem surprising, however, it actually makes sense. Since the estimate has little bias, the algorithm has more confidence with the less uncertain quantity (or the one with

¹ The linearization of the forward operator is usually not required in practical applications. The term $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T$ can be obtained by applying the nonlinear operator and then computing the sample statistics.

² The beta case provided similar conclusions, so we dropped it for brevity.

FIG. 2. Bayesian update for α in the single-variable example for different bias scenarios: (left) small bias, (center) moderate bias, and (right) large bias. The results are shown for a range of values for the ensemble variance σ_e^2 and the static variance σ_s^2 .

smaller variance). Finally, with moderate biases the algorithm seems to alternate behavior depending on the magnitude of the different ensemble and the static variance combinations.

b. Spatially varying α

Is it appropriate to use a single value for α for the entire state domain? The answer in most models is generally no. This is because both model and sampling errors may not be homogeneous in space. For instance, ocean models have long exhibited strong warm SST biases in the Southern Ocean and cold biases in the North Pacific and Atlantic Oceans. Another reason is that we often work with complex observation networks where some regions are densely observed and others are only sparsely observed. This generally implies that the ensemble spread in different regions can be vastly different. As an example, one could look at in situ radiosonde observations. Densely observed troposphere over Europe and North America might have smaller ensemble variance compared to other less constrained areas in the Southern Hemisphere. Based on this, it is nearly impossible to address different issues in different parts of the domain if α is spatially homogenous. The problem may be alleviated if each state variable is assigned a different weighting factor.

1) THE ALGORITHM

A spatially varying weighting factor consists of N_x elements; $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_{N_x}]^T$. In vector form, the hybridized covariance matrix given in Eq. (3) becomes

$$\hat{\mathbf{P}} = \mathbf{W}^e \circ \mathbf{P}^e + \mathbf{W}^s \circ \mathbf{B}, \quad (8)$$

$$\hat{\mathbf{P}}_{ij} = \sqrt{\alpha_i} \sqrt{\alpha_j} \mathbf{P}_{ij}^e + \sqrt{1 - \alpha_i} \sqrt{1 - \alpha_j} \mathbf{B}_{ij}, \quad i, j = 1, 2, \dots, N_x, \quad (9)$$

where $\mathbf{W}^e$ and $\mathbf{W}^s$ are $N_x \times N_x$ weighting matrices with elements: $\mathbf{W}_{ij}^e = \sqrt{\alpha_i \alpha_j}$ and $\mathbf{W}_{ij}^s = \sqrt{(1 - \alpha_i)(1 - \alpha_j)}$, respectively. The $\circ$ notation indicates element-wise or Hadamard product of matrices. For clarity, time indexing has been dropped. Unlike Eq. (3), each element of the hybrid covariance matrix will be modulated nonuniformly by α . It is further easy to see that $\hat{\mathbf{P}}$ in Eq. (8) is still symmetric and semipositive definite.

Since both $\mathbf{W}^e \circ \mathbf{P}^e$ and $\mathbf{W}^s \circ \mathbf{B}$ are semipositive definite, their sum also yields a semipositive definite matrix.

At each DA cycle, the prior probability density function for α is a multivariate normal. The goal is to use the data and update α following Bayes's rule as in Eq. (4). Each component in the prior weighting factor vector can be described by $\alpha_j \sim \mathcal{N}(\bar{\alpha}_{p,j}, \sigma_{p,\alpha_j}^2)$ with mean $\bar{\alpha}_{p,j}$ and variance σ_{p,α_j}^2 (subscript p refers to “prior”). Choosing Gaussians as marginals pdfs for α_j leads to computationally efficient approximation of the Bayesian problem. Unlike the spatially homogenous case, the beta distribution here yields very complex equations that do not provide explicit solutions for α_j . To estimate the multivariate posterior distribution for α , the observations are assimilated serially. The mean and the variance of the weighting at every grid cell is updated after processing each observation. We follow Anderson (2009) and for each observation, the sample correlations between the observed prior ensemble and the rest of the state variables are used to describe the correlations between the individual weighting factors in space:

$$\ell_j = \frac{\sum_{i=1}^{N_e} (x_{p,j}^i - \bar{x}_{p,j}) [h(x_{p,j}^i) - \bar{h}(x_{p,j}^i)]}{\sqrt{\sum_{i=1}^{N_e} (x_{p,j}^i - \bar{x}_{p,j})^2} \sqrt{\sum_{i=1}^{N_e} [h(x_{p,j}^i) - \bar{h}(x_{p,j}^i)]^2}}, \quad (10)$$

$$\rho_j = \phi|\ell_j|, \quad j = 1, 2, \dots, N_x, \quad (11)$$

where $x_{p,j}^i$ is the i th prior ensemble member of state variable j . The sample correlation coefficient is given by ℓ_j and ϕ is a localization factor. Similar to the update of the state, increments to the prior values of α are limited to neighboring grid cells only. Replacing α by $\rho_j \alpha_j$ in Eq. (6), the variance of the background innovation can be obtained:

$$\theta_j^2 = \sigma_o^2 + \rho_j \alpha_j \sigma_e^2 + (1 - \rho_j \alpha_j) \sigma_s^2, \quad (12)$$

where σ_e^2 and σ_s^2 denote the ensemble and static background variances of the j th element in the state vector, respectively. Compared to Eq. (6), which was only defined in observation space, θ_j^2 now enables the representation of the individual weighting factors in the state space. Note that setting $\rho_j = 1$

reduces this algorithm to the spatially constant variant. The posterior distribution for the j th element of α is written as follows:

$$p(\alpha_j|d) = \frac{1}{2\pi\theta_j\sigma_{p,\alpha_j}} \exp \left[-\frac{d^2}{2\theta_j^2} - \frac{(\alpha_j - \bar{\alpha}_{p,j})^2}{2\sigma_{p,\alpha_j}^2} \right]. \quad (13)$$

The product of the prior and the likelihood is approximated with a Gaussian and later used in the next cycle as the Gaussian prior. The procedure is repeated until all the observations have been assimilated.

The updated estimate $\bar{\alpha}_{u,j}$ is found by maximizing (13). Steps and computations required to find the mode of $p(\alpha_j|d)$ are provided in the [appendix](#). The variance of the posterior distribution, σ_{u,α_j}^2 , can be numerically approximated ([Anderson 2007](#)) by taking the ratio r of the posterior density evaluated at $\bar{\alpha}_{u,j} + \sigma_{p,\alpha_j}$ to the value at $\bar{\alpha}_{u,j}$:

$$\sigma_{u,\alpha_j}^2 = -\frac{\sigma_{p,\alpha_j}^2}{2\log(r)}. \quad (14)$$

Other more robust ways to estimate the updated variance can be used instead. We argue that the focus should be directed on how to correctly and efficiently estimate the mode of (13), rather than its variance. Similar adaptive techniques used in the literature ([El Gharamti et al. 2019](#)) to estimate the covariance inflation factor often fix the variance in time and only allow the inflation value to change. This limits the possibility of losing spread in the estimated parameter.

2) IMPLEMENTATION AND COST

Implementing the adaptive hybrid ensemble–variational scheme requires the knowledge of a static background covariance matrix $\mathbf{B}$ in advance. If $\mathbf{B}$ is available and stored on the machine, then one can simply proceed with the steps below. If not, then it is necessary to perform an offline procedure to compute $\mathbf{B}$. This can be achieved using the methods discussed in [section 2](#). In this study, $\mathbf{B}$ is considered a climatological covariance and uses a set of long model runs:

$$\mathbf{B} = \frac{1}{N_s} \sum_{i=1}^{N_s} (\mathbf{x}^{s,i} - \bar{\mathbf{x}})(\mathbf{x}^{s,i} - \bar{\mathbf{x}})^T, \quad (15)$$

where $\mathbf{x}^{s,i}$ denote model states resulting from running the model over large forecast cycles, $\bar{\mathbf{x}}$ is the average background state, and N_s is the number of the climatological states (often very large; $N_s \gg N_e$). Note that as long as we have access to $\mathbf{x}^{s,i}$, forming and storing $\mathbf{B}$ is not necessary.

The update procedure follows the two-step DA approach ([Anderson 2003](#)), commonly used in the Data Assimilation Research Testbed (DART; [Anderson et al. 2009](#)). The update of the weighting factors happens before updating the state as follows:

1) Forecast step:

Starting from the most recent analysis state, the ensemble members are propagated using the model until the next observation time.

2) Analysis step:

Loop over the observations: $o = 1 \rightarrow N_o$.

2.1) Update α :

Loop over each state variable: $j = 1 \rightarrow N_x$.

Find: $d, \phi, \ell_j, \rho_j, \sigma_{e_j}^2, \sigma_{s_j}^2$.

For each pair of $\{\bar{\alpha}_{p,j}, \sigma_{p,\alpha_{p,j}}\}$, compute $\{\bar{\alpha}_{u,j}, \sigma_{p,\alpha_{u,j}}\}$ using (A2) and (14).

2.2) Compute the observation increments:

Evaluate the forward operators on the background ensemble and the static states.

Compute the hybridized observation space variance:

$$\sigma_{h_o}^2 = \bar{\alpha}_{u,o}\sigma_{e_o}^2 + (1 - \bar{\alpha}_{u,o})\sigma_{s_o}^2.$$

Using the observation value and its variance, compute the increments: δy .

2.3) Update $\mathbf{x}$:

Loop over each state variable: $j = 1 \rightarrow N_x$.

Compute the hybridized state-observation covariance:

$$\sigma_{h_{j,o}} = \sqrt{\bar{\alpha}_{u,j}}\sqrt{\bar{\alpha}_{u,o}} \cdot \sigma_{e_{j,o}} + \sqrt{1 - \bar{\alpha}_{u,j}}\sqrt{1 - \bar{\alpha}_{u,o}} \cdot \sigma_{s_{j,o}}.$$

Add the observation increments to the state: $x_{u,j}^i = x_{p,j}^i + \phi\sigma_{h_{j,o}}\sigma_{h_o}^{-2}\delta y$.

The covariance between the j th variable and the observed ensemble (climatology) is denoted by $\sigma_{e_{j,o}}$ ($\sigma_{s_{j,o}}$). The posterior weighting factor at the current observation location is denoted by $\bar{\alpha}_{u,o}$. The observation increments can be computed in various ways depending on the choice of the filtering algorithm (e.g., EnKF and EAKF). Note that the proposed implementation scheme does not support observations that are not defined on the same grid as the state. For instance, when assimilating radiances (typically nonlocal), one cannot compute $\bar{\alpha}_{u,o}$ in steps 2.2 and 2.3 and thus $\sigma_{h_o}^2$ and $\sigma_{h_{j,o}}$ will not be available.

In terms of computational cost, updating the weighting factors in step 2.1 is relatively small compared to the other operations in steps 2.2 and 2.3. To find the posterior distribution of α , one first needs to solve the cubic formula in (A2) and then evaluate (14). Such a computation is dominated by evaluating $2N_x$ cubic roots and N_x natural logarithms. For large ensemble sizes (i.e., $N_e > 20$), the added computational cost by step 2.1 was found to be smaller than 10% of the total computation time. This was demonstrated in a wide range of toy models such as Lorenz 63 and Lorenz 96. There is no reason to believe that this overhead cost would be considerably different in large systems. In addition to step 2.1, computing the forward operators on the background climatological states, in step 2.2, every assimilation cycle can be quite costly. This issue diminishes when the observation networks do not—immensely—change between assimilation cycles. In other words, if one assumes the observation network to be fixed in time then computing $\mathbf{H}_k\mathbf{B}\mathbf{H}_k^T$ to find $\sigma_{s_o}^2$ and $\sigma_{s_{j,o}}$ is done only once.

Apart from complexity, the proposed hybrid technique requires storing $\mathbf{B}$ (often $\mathbf{B}^{1/2}$ is enough given the symmetry) or the climatological states that form $\mathbf{B}$ in addition to storing the weighting factors themselves. To this end, if the storage volume required for a single state vector is $\mathcal{S}_x$, then data storage required by the proposed spatially varying hybrid scheme is $\mathcal{S}_x(N_e + N_s + 1)$ compared to $\mathcal{S}_xN_e$ for nonhybrid ensemble filters. This makes the method quite prohibitive in terms of

storage especially when working with a large atmosphere or ocean model. This is, however, an issue for every hybrid scheme and not only the proposed one. More insightful discussion on possible ways to mitigate this can be found in section 4.

3. Observing system simulation experiments

The proposed adaptive hybrid ensemble–variational algorithm is tested using the 40-variable Lorenz system (L96; Lorenz 1996):

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad i = 1, 2, \dots, 40. \quad (16)$$

A fourth-order Runge–Kutta scheme is used to integrate the system with a time step $dt = 0.05$ ($\equiv 6$ h in real time). The variables are assumed equally spaced in a $[0, 1]$ cyclic domain. A “truth” run is first generated using $F = 8$, to obtain a set of reference states. The truth starts from a point on the attractor and consists of 100 000 model time steps. Synthetic observations are available every 5 time steps and they are contaminated with Gaussian noise of mean 0 and variance 1. All results ignore the first 10 000 DA cycles to avoid any transient behavior. The DA scheme used in all runs is the ensemble adjustment Kalman filter (EAKF). Hereafter, the proposed spatially constant and varying hybrid schemes are referred to as: EnKF-OI-c and EnKF-OI-v, respectively.

The initial ensemble members are randomly drawn from a normal distribution centered at ξ_0 with a unit variance. The initial ensemble mean ξ_0 is obtained by integrating the model for 5 years in real-world time starting from the last forecast of the truth run. Climatology states $\mathbf{x}^{s,i}$ are sampled from a very long model run every 5000 time steps for a total of $N_s = 1000$. The choice of 1000 static background states was made through offline single cycle DA experiments (not shown) where a wide range of values for α were tested. N_s could be different for other models depending on the details of the dynamics. Figure 3 shows the static background covariance matrix and an initial ensemble covariance constructed using 10 members. As can be seen, $\mathbf{B}$ has a clear banded structure and generally less noise compared to $\mathbf{P}^e$. The percentage of the variance explained by the leading empirical orthogonal functions (EOF modes) suggests that $\mathbf{B}$ is a full rank matrix while $\mathbf{P}^e$ has a rank of 9 (i.e., $N_e - 1$). The first 3 EOF modes associated with $\mathbf{P}^e$ account for more than 50% of the total variance of the system. Spatially, the large-scale patterns dominate the first few EOFs in $\mathbf{B}$ and small-scale inaccurate ones take on the last EOFs, which contribute very little to the total variance.

Each data assimilation run below was repeated 20 times, each with randomly drawn initial ensemble and observation errors. The presented metrics, e.g., RMSE, are averaged over all 20 runs.

a. Ensemble size

In the first set of experiments, we assume perfect model conditions to assess the performance of the proposed adaptive hybrid scheme for different ensemble sizes. The forecast model and the climatology both use a forcing of 8 like the truth run. The

FIG. 3. (top) Initial ensemble covariance $\mathbf{P}^e$ constructed using 10 members. (bottom) A static covariance matrix $\mathbf{B}$ obtained using 1000 climatological states. Also shown, according to the right y axis on both plots, the increase in the explained variance (%) captured by the entire EOF modes. The EOFs are found by computing the eigenvalues and eigenvectors of $\mathbf{B}$. The derived eigenvalues provide a measure of the percent variance explained by each mode.

observation network consists of the odd numbered variables for a total of 20 observations. We compare five different filtering schemes: EnKF with no enhancements (i.e., no inflation or localization), EnOI, EnKF-OI with a fixed weighting factor $\alpha = 0.5$ (nonadaptive), EnKF-OI-c, and EnKF-OI-v. For the adaptive schemes, the initial distribution of the weighting factor is assumed normal with mean $\bar{\alpha}_{p,j} = 0.5$ and variance $\sigma_{p,\alpha_j}^2 = 0.1$. The initial variance was selected through sensitivity experiments by varying it between 0.001 and 1. Overall, it was found that as the weighting variance decreases, changes to α over time also decrease and vice versa. For many tested ensemble members, starting with weighting variance of 0.1 yielded the best performance. σ_{p,α_j}^2 is kept constant both in time and space throughout all experiments.

Figure 4 plots the prior and the posterior root-mean-square errors (RMSE) for different ensemble sizes ranging from 3 to 200. The RMSE has been averaged both in space and time. For very small ensemble sizes ($N_e < 20$, the EnKF failed to produce

FIG. 4. Prior (solid) and posterior (dashed) spatially and temporally averaged RMSE resulting from EnKF, EnOI, EnKF-OI (α fixed to 0.5), EnKF-OI-c, and EnKF-OI-v for different ensemble sizes. For clarity, the y axis is shown in log scale.

useful forecasts and suffered from ensemble divergence. This is not surprising and it confirms the fact that implementing the EnKF with limited ensemble size and no other statistical enhancements is not reliable and will not yield meaningful results. For $N_e \leq 45$, the EnOI outperformed the EnKF producing better quality forecasts and analyses. Note that the EnOI is independent of the ensemble size because it uses static background perturbations. The EnKF starts becoming competitive with other filters when the ensemble size is at least 80. Among all tested DA schemes, the EnKF-OI-v is the most accurate for all ensemble sizes. The EnKF-OI-c matches the performance of the EnKF-OI-v for $N_e < 50$ but for larger ensemble sizes EnKF-OI-v is the superior scheme. The hybrid EnKF-OI with a fixed weighting factor performs better than the EnKF for all tested ensemble sizes. It is better than the EnOI only for $N_e > 20$.

Figure 4 indicates that for very small ensemble sizes, the proposed EnKF-OI-v is able to match the accuracy of the EnOI. As the ensemble size increases, the performance of EnKF-OI-v improves until it matches and slightly outperforms that of the EnKF for large ensemble sizes. This illustrates the robust functionality of the proposed scheme in the presence of sampling errors only. As can be seen in Fig. 5, the EnKF-OI-v decreases the weight on the ensemble covariance to 0.1 for $N_e \leq 5$. By doing so, the filter can retrieve perturbations with larger magnitude from the static background covariance in order to combat severe sampling errors in the ensemble. As the ensemble size increases, sampling errors diminish and α is shown to converge to 1. Since it is spatially varying, EnKF-OI-v is more responsive to changes in the statistics of the ensemble as compared to EnKF-OI-c. For example, with an ensemble

FIG. 5. Weighting factor obtained using both adaptive hybrid schemes as function of the ensemble size. Each asterisk denotes an average over all assimilation cycles. For EnKF-OI-c, averaging is performed over time. The result of the EnKF-OI-v involves averaging both in space and time.

size of 200 one would expect the ensemble to be the main source for the background uncertainties and that is indeed the case for EnKF-OI-v (α averaged in time and space is 0.96). EnKF-OI-c, on the other hand, still partially relies on the static background covariance matrix assigning it a $\sim 20\%$ weight.

The ratio between the average ensemble spread and the RMSE is found close to unity using the EnKF-OI-v. Additional discussion on this will be presented, using imperfect model conditions, in the next section.

b. Model error

To simulate model errors, we vary the forcing term in the forecast model. 11 DA experiments are performed, using 20 ensemble members, in which F is set to 3, 4, ..., 13. In each experiment, $\mathbf{B}$ is computed using the biased forecast model assuming that we do not have access to the true parameters of the model. Small F values produce a stable system with very little to almost no error growth. Large F values, on the other hand, produce extremely chaotic dynamics. The goal is to compare the performance of the EnKF-OI-v to the EnKF in the presence of (i) inflation and (ii) localization. Inflation is performed by multiplying the prior covariance with a constant factor, λ , greater than 1. Localization is a number chosen between 0 and 1 and used to multiply the state regression coefficients, as shown in the algorithm (section 2b). This number is a function of the distance between the state and the observation (here, cutoff distance), computed using a Gaspari and Cohn correlation function (Gaspari and Cohn 1999). Similar to the previous experiments, every other variable is observed.

1) EFFECTS OF INFLATION

Average prior RMSE results for 9 different inflation values are shown in Fig. 6. Both EnKF and EnKF-OI-v diverge when using $\lambda = 2$ for all tested forcing values. When modeling biases are large and the model is very chaotic ($F = 12$ and 13), the EnKF becomes numerically unstable and inflation does not seem to help. In contrast, EnKF-OI-v does not blow up in these

FIG. 6. Average prior RMSE resulting from (top) the EnKF and (middle) the EnKF-OI-v for different forcing and inflation configurations. Shaded regions in gray indicate filter divergence. (bottom) The temporally and spatially averaged weighting factor from EnKF-OI-v. The ensemble size is set to 20.

biased conditions. Both schemes behave equally well for small forcing values; e.g., $F = 3, 4$. As F increases, the EnKF-OI-v estimates are significantly more accurate than those of the EnKF. The smallest RMSE is obtained using EnKF-OI-v for $F = 8$ (no model errors) and $\lambda = 1.08$. We note that the hybrid scheme with no inflation (i.e., $\lambda = 1$) is able to outperform the EnKF with any inflation for all tested bias scenarios. It is not shown here, but we reached a similar assessment for the posterior RMSE estimates.

To understand the interaction between the adaptive hybrid algorithm and inflation, we plot the resulting spatially and temporally averaged weighting factor in the bottom panel of

Fig. 6. It is evident that the change in α is proportional to that of λ . As inflation increases, biases and sampling errors in $\mathbf{P}^e$ decrease and this allows the adaptive algorithm to give it a higher weight. For instance, in the case of $F = 5$ increasing λ from 1 to 1.5 led to an increase in α from 0.3 to 0.9. Under stable conditions and relatively small inflation ($\leq 4\%$), more weight is placed on $\mathbf{B}$ in order to increase the prior variance and make the ensemble statistics more consistent with the observations.

2) EFFECTS OF LOCALIZATION

Instead of inflating the background variance, we now limit the size of the update for each state variable through localization. 10 different cutoff radii ranging from 0.1 to 100 are tested. Average prior RMSE resulting from the EnKF and the EnKF-OI-v are shown in Fig. 7. Both EnKF and EnKF-OI-v perform best when the localization radius is relatively small given sampling errors and strong model biases. Similar to the previous set of experiments, the performance of the EnKF strongly degrades for large values of F . The EnKF-OI-v systematically outperforms the EnKF for all tested localization and forcing cases. Moreover, with no localization (cutoff is 100) and in the presence of model errors the EnKF-OI-v can match the forecast accuracy of a localized EnKF with 0.2 cutoff.

The impact of localizing the flow-dependent background ensemble covariance on the update of the weighting factor is described in Fig. 8. Visibly, one can notice two opposing patterns (shown by the dashed-dotted line). Under rapidly changing and chaotic conditions (i.e., $F \geq 8$), α tends to increase as the impact of localization increases. Given the strong error growth in the forecast model, the spread of the ensemble stays large enough throughout the experiment. Consequently, the adaptive algorithm pushes α closer to 1 retaining the localized flow-dependent perturbations. On the contrary, when the conditions are more stable ($F < 8$), α shrinks as the algorithm attempts to increase the variability of the underdispersed ensemble using the static background perturbations. To confirm this analysis, we plot in the left panel of Fig. 8 the ratio of the average ensemble spread (AES) obtained using the EnKF to that of the EnKF-OI-v. The AES of the EnKF for $F < 8$ is very small compared to the EnKF-OI-v. In the case of $F = 3$, the plotted ratio is ~ 0.1 indicating that the prior spread of the EnKF is only 10% the spread of the EnKF-OI-v. The variance of both schemes is approximately the same for $F \geq 8$ with an AES ratio roughly equal to 1.

c. Observation network

The spatially adaptive hybrid scheme is further tested using four sparse observation networks. Observed locations in these networks are as follows; Data void I (DV-I): the first 20 variables, DV-II: the first and last 5 variables, DV-III: 10 variables in the center of the domain and DV-IV: only 5 center variables. These experiments assume a perfect forecast model and use an ensemble size of 20. Localization cutoff radius is 0.1 and inflation is turned off.

The time-averaged spatial distribution of the weighting factor resulting from each experiment is plotted in Fig. 9. Because the ensemble covariance is heavily localized, the

FIG. 7. Average prior RMSE resulting from (left) the EnKF and (right) the EnKF-OI-v for different forcing and localization configurations ($N_e = 20$). Shaded regions in gray indicate filter divergence.

weighting factor values are close to 1. Similar behavior was observed and discussed in Fig. 8. Spatially, observed locations are assigned smaller weight than the nonobserved ones because they are constantly updated and thus they have smaller ensemble spread. Partially relying on **B** in these regions helps increase the prior spread and improve the overall stability of the filter. The convergence of the individual weighting factors in space happens rather quickly. In the case of DV-III, for example, starting from 0.5 the optimal weights were reached after ~ 20 DA cycles (Fig. 10). Beyond that, no major changes are detected for the weights of the unobserved variables. The weights at the observed locations, on the other hand, can be seen varying between 0.81 and 0.98.

Combined with model errors ($F = 10$), DV-I is used in a final assessment to compare the performance of all filtering schemes. The best performance (obtained after tuning inflation and localization) of the EnKF, EnKF-OI ($\alpha = 1/2$), EnKF-OI-c,

and EnKF-OI-v are compared for different ensemble sizes ranging from 15 to 200. As can be shown in Fig. 11, the proposed EnKF-OI-c and EnKF-OI-v are able to outperform the EnKF for all tested ensemble members, especially for small ensemble sizes. In fact, the adaptive hybrid schemes using 20 members are shown to produce more accurate prior estimates compared to a well-tuned 120-member EnKF. Unlike the perfect modeling conditions (Fig. 4) in which we saw EnKF-OI-v adapts to $\alpha \approx 1$ for large ensemble size, here α only reached 0.7 for $N_e = 200$. This indicates that relying only on ensemble perturbations in case of sparse observations and in the presence of model errors may yield unsatisfactory state estimates. The nonadaptive hybrid scheme performs well for small ensemble sizes, however, it is unable to match the accuracy of the EnKF for large ensemble sizes. The improvements of the EnKF-OI-v over the EnKF-OI-c are minimal ($\sim 3\%$). This may be attributed to the nature of the spatial

FIG. 8. (left) Ratio of the average ensemble spread resulting from the EnKF to that of EnKF-OI-v. (right) Spatially and temporally averaged weighting factor from the EnKF-OI-v runs. Results are shown for different forcing and localization configurations. The ensemble size is set to 20. Shaded regions in gray indicate filter divergence (i.e., hybrid filter never diverged).

FIG. 9. Average weighting factor in space obtained using the proposed EnKF-OI-v for $N_e = 20$. Different curves represent four different observation networks.

correlations in the L96, which are not as significant as those presented by a large GCM.

4. Summary and discussion

A new spatially and temporally varying adaptive hybrid ensemble and variational filtering algorithm is introduced. The individual weights assigned to the ensemble and the static background covariance matrices are assumed to be random variables. A Bayesian approach is used to estimate these weights conditioned on available data. The resulting data assimilation scheme can be decomposed into: 1) a filter to estimate the weighting factors α and 2) another filter to estimate the state $\mathbf{x}$. Using the Lorenz 96 nonlinear model, the proposed scheme was tested against the ensemble Kalman filter in various settings.

Under perfect modeling conditions, EnKF-OI-v (spatially varying form of the scheme) was found to be more accurate

FIG. 10. Weighting factor in space obtained using the proposed EnKF-OI-v for the first 800 DA cycles. The ensemble size is set to 20.

FIG. 11. Spatially and temporally averaged prior RMSE obtained for different ensemble sizes using EnKF, EnKF-OI ($\alpha = 1/2$), EnKF-OI-c, and EnKF-OI-v. The schemes are tuned for best performance using inflation and localization. Model errors are simulated using $F = 10$ in the forecast model and the observation network is set to data void I (DV-I).

than the EnKF for a wide range of ensemble sizes. For very small ensemble sizes, both EnKF-OI-v and EnKF-OI-c (spatially constant variant) are less prone to sampling errors as they morph into an EnOI system fully utilizing the static background perturbations. For large ensemble sizes, the adaptive scheme makes the hybrid filter behave just like an EnKF with pure flow-dependent prior statistics. With model imperfections, the EnKF-OI-v was found to be distinctly more stable than the EnKF, producing quality forecasts even in highly biased conditions. The adaptive scheme appropriately selected the weighting coefficients given the amount of inflation and localization imposed on the ensemble. To illustrate, with enough variability in the system highly localized ensembles got assigned larger weight. Spatially, EnKF-OI-v was able to detect densely observed regions and counteract ensemble spread reduction by increasing the weight on the static background covariance.

One of the drawbacks of the proposed adaptive algorithm is its computational cost. As discussed in section 2, a large inventory of climatological model states need to be stored on the machine if $\mathbf{B}$ is not available. Applying the forward operator on these time-invariant model states is an additional burden. To make this algorithm feasible, one should explore ways to reduce the computational and storage cost. A possible strategy would be to approximate $\mathbf{B}$ with a surrogate that uses a climatological run obtained with a lower-resolution model. During the update, interpolating the static background perturbations on the high-resolution grid can be performed. Storing only the leading few modes of $\mathbf{B}$ could be another way to reduce storage issues. Assuming storage issues are resolved, one can counter the computational complexity of the proposed scheme by limiting the size of the ensemble. For instance, in the presence of model errors ($F = 10$) the adaptive hybrid system with 20 members could outperform the forecast accuracy of a typical 80-member tuned EnKF, eliminating more than 60% of the customary EnKF cost. These savings, which may not be the same in other models, could compensate for the additional costs incurred by the adaptive hybrid scheme.

This work proposed the idea of using a different weighting coefficient for every grid cell. One of the advantages of such an approach is that it removes some of the sensitivity of ensemble filters to choices of localization (Fig. 7). The hybrid covariances, apart from being full rank, are generally less noisy and have better identifiable spatial structures. This study, however, did not provide a complete assessment of the properties of the resulting hybrid covariances when a spatially constant or a spatially varying weighting field is used. This is an interesting line of research that could be explored going forward. Furthermore, proposing non-Gaussian distributions for the weighting factors should be contemplated. Allowing the variance of α to change in time is also something this study did not focus on and would be worth investigating. Testing the performance of the EnKF-OI-v in large Earth system models using real data will be considered in future research. This includes assimilating spatially heterogeneous datasets and nonlocal observations.

The proposed spatially varying adaptive hybrid scheme requires the observations to be processed serially. If the observations are to be assimilated all at once, as in many NWP systems, one may be restricted to using the spatially constant adaptive form. In the current L96 framework, the difference between the performance of EnKF-OI-v and EnKF-OI-c was minimal. However, this might not be the case in other systems where spatial correlations have different patterns and significance. In general, estimating even a spatially constant weighting factor is far more superior than using a time-invariant one making this work quite relevant to NWP systems. In fact, one possible application could be the state-of-the-art hybrid ensemble-variational systems such as hybrid 4DEnVar (Lorenc et al. 2015). Without using an adjoint or a tangent linear model, having the right weighting between the ensemble and the static background covariance can be crucial.

Acknowledgments. The author thanks three anonymous reviewers for their useful comments and suggestions. The author would also like to thank Jeff Anderson for intriguing discussions. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

Data availability statement. Data sharing is not applicable to this article as no new data were created or analyzed in this study.

APPENDIX

Posterior Estimate of the Weighting Factor

To find the maximum of (13), we first take its derivative with respect to α_j :

$$\frac{\partial p(\alpha_j|d)}{\partial \alpha_j} = \omega \left[\frac{\rho_j(\sigma_{e_j}^2 - \sigma_{s_j}^2)}{2\sigma_{p,\alpha_j}\theta_j^3} \left(\frac{d^2}{\theta_j^2} - 1 \right) - \frac{\alpha_j - \bar{\alpha}_{p,j}}{\sigma_{p,\alpha_j}^3 \theta_j} \right], \quad (\text{A1})$$

where ω is the exponential term of (13) divided by 2π . Setting (A1) to zero and rearranging the terms yields a cubic polynomial:

$$c_1 \alpha_j^3 + c_2 \alpha_j^2 + c_3 \alpha_j + c_4 = 0, \quad (\text{A2})$$

and the coefficient are as follows:

$$c_1 = -2\gamma^2, \quad (\text{A3})$$

$$c_2 = 2\bar{\alpha}_{p,j}\gamma^2 - 4\gamma\beta, \quad (\text{A4})$$

$$c_3 = 4\bar{\alpha}_{p,j}\gamma\beta - 2\beta^2 - \rho_j\sigma_{p,\alpha_j}^2\gamma(\sigma_{e_j}^2 - \sigma_{s_j}^2), \quad (\text{A5})$$

$$c_4 = 2\bar{\alpha}_{p,j}\beta^2 + \rho_j\sigma_{p,\alpha_j}^2(\sigma_{e_j}^2 - \sigma_{s_j}^2)(d^2 - \beta), \quad (\text{A6})$$

where $\gamma = \rho_j(\sigma_{e_j}^2 - \sigma_{s_j}^2)$ and $\beta = \sigma_o^2 + \sigma_{s_j}^2$. The roots of (A2) can be found using the cubic formula. If multiple real roots exit, the closest to the prior $\bar{\alpha}_{p,j}$ is selected as the mode, $\bar{\alpha}_{u,j}$ of the posterior.

REFERENCES

- Anderson, J. L., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642, [https://doi.org/10.1175/1520-0493\(2003\)131<0634:ALLSFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2).
- , 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, **59A**, 210–224, <https://doi.org/10.1111/j.1600-0870.2006.00216.x>.
- , 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83, <https://doi.org/10.1111/j.1600-0870.2008.00361.x>.
- , 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **140**, 2359–2371, <https://doi.org/10.1175/MWR-D-11-00013.1>.
- , and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758, [https://doi.org/10.1175/1520-0493\(1999\)127<2741:AMCIOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2).
- , T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, <https://doi.org/10.1175/2009BAMS2618.1>.
- Berner, J., S.-Y. Ha, J. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995, <https://doi.org/10.1175/2010MWR3595.1>.
- Bishop, C. H., and D. Hodyss, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part I: Tests on simple error models. *Tellus*, **61A**, 84–96, <https://doi.org/10.1111/j.1600-0870.2008.00371.x>.
- , and —, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part II: A strategy for the atmosphere. *Tellus*, **61A**, 97–111, <https://doi.org/10.1111/j.1600-0870.2008.00372.x>.
- , and E. A. Satterfield, 2013: Hidden error variance theory. Part I: Exposition and analytic model. *Mon. Wea. Rev.*, **141**, 1454–1468, <https://doi.org/10.1175/MWR-D-12-00118.1>.
- , —, and K. T. Shanley, 2013: Hidden error variance theory. Part II: An instrument that reveals hidden error variance distributions from ensemble forecasts and observations. *Mon.*

- Wea. Rev.*, **141**, 1469–1483, <https://doi.org/10.1175/MWR-D-12-00119.1>.
- Bowler, N., and Coauthors, 2017: The effect of improved ensemble covariances on hybrid variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 785–797, <https://doi.org/10.1002/QJ.2964>.
- Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043, <https://doi.org/10.1256/qj.04.15>.
- Carrassi, A., M. Bocquet, L. Bertino, and G. Evensen, 2018: Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdiscip. Rev.: Climate Change*, **9**, e535, <https://doi.org/10.1002/wcc.535>.
- 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, <https://doi.org/10.1002/QJ.2054>.
- Descombes, G., T. Auligné, F. Vandenbergh, D. Barker, and J. Barre, 2015: Generalized background error covariance matrix model (gen_be v2. 0). *Geosci. Model Dev.*, **8**, 669–696, <https://doi.org/10.5194/GMD-8-669-2015>.
- Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, <https://doi.org/10.1256/qj.05.108>.
- El Gharamti, M., 2018: Enhanced adaptive inflation algorithm for ensemble filters. *Mon. Wea. Rev.*, **146**, 623–640, <https://doi.org/10.1175/MWR-D-17-0187.1>.
- , K. Raeder, J. Anderson, and X. Wang, 2019: Comparing adaptive prior and posterior inflation for ensemble filters using an atmospheric general circulation model. *Mon. Wea. Rev.*, **147**, 2535–2553, <https://doi.org/10.1175/MWR-D-18-0389.1>.
- 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, [https://doi.org/10.1175/1520-0493\(2004\)132<1065:ROHDAS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1065:ROHDAS>2.0.CO;2).
- Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, <https://doi.org/10.1007/s10236-003-0036-9>.
- Furrer, R., and T. Bengtsson, 2007: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivar. Anal.*, **98**, 227–255, <https://doi.org/10.1016/j.jmva.2006.08.003>.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, <https://doi.org/10.1002/QJ.49712555417>.
- Gharamti, M. E., J. Valstar, and I. Hoteit, 2014: An adaptive hybrid EnKF-OI scheme for efficient state-parameter estimation of reactive contaminant transport models. *Adv. Water Resour.*, **71**, 1–15, <https://doi.org/10.1016/j.advwatres.2014.05.001>.
- Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, [https://doi.org/10.1175/1520-0493\(2000\)128<2905: AHEKFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2905: AHEKFV>2.0.CO;2).
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, [https://doi.org/10.1175/1520-0493\(1998\)126<0796: DAUAKE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0796: DAUAKE>2.0.CO;2).
- Kuhl, D. D., T. E. Rosmond, C. H. Bishop, J. McLay, and N. L. Baker, 2013: Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework. *Mon. Wea. Rev.*, **141**, 2740–2758, <https://doi.org/10.1175/MWR-D-12-00182.1>.
- Lei, L., J. L. Anderson, and J. S. Whitaker, 2016: Localizing the impact of satellite radiance observations using a global group ensemble filter. *J. Adv. Model. Earth Syst.*, **8**, 719–734, <https://doi.org/10.1002/2016MS000627>.
- Lorenc, A. C., N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, 2015: Comparison of hybrid-4DVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Wea. Rev.*, **143**, 212–229, <https://doi.org/10.1175/MWR-D-14-00195.1>.
- Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Vol 1, Reading, United Kingdom, ECMWF, 1–18, <https://www.ecmwf.int/node/10829>.
- Ménétrier, B., and T. Auligné, 2015: Optimized localization and hybridization to filter ensemble-based covariances. *Mon. Wea. Rev.*, **143**, 3931–3947, <https://doi.org/10.1175/MWR-D-15-0057.1>.
- Meng, Z., and F. Zhang, 2008: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Mon. Wea. Rev.*, **136**, 522–540, <https://doi.org/10.1175/MWR3352.1>.
- Mitchell, H. L., and P. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416–433, [https://doi.org/10.1175/1520-0493\(2000\)128<416:AAEFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<416:AAEFV>2.0.CO;2).
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763, [https://doi.org/10.1175/1520-0493\(1992\)120<1747:TNCSS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<1747:TNCSS>2.0.CO;2).
- Penny, S. G., D. W. Behringer, J. A. Carton, and E. Kalnay, 2015: A hybrid global ocean data assimilation system at NCEP. *Mon. Wea. Rev.*, **143**, 4660–4677, <https://doi.org/10.1175/MWR-D-14-00376.1>.
- Pham, D. T., J. Verron, and M. C. Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.*, **16**, 323–340, [https://doi.org/10.1016/S0924-7963\(97\)00109-7](https://doi.org/10.1016/S0924-7963(97)00109-7).
- Sacher, W., and P. Bartello, 2008: Sampling errors in ensemble Kalman filtering. Part I: Theory. *Mon. Wea. Rev.*, **136**, 3035–3049, <https://doi.org/10.1175/2007MWR2323.1>.
- Satterfield, E. A., D. Hodyss, D. D. Kuhl, and C. H. Bishop, 2018: Observation-informed generalized hybrid error covariance models. *Mon. Wea. Rev.*, **146**, 3605–3622, <https://doi.org/10.1175/MWR-D-18-0016.1>.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008: A hybrid ETKF–3DVAR data assimilation scheme for the WRF Model. Part II: Real observation experiments. *Mon. Wea. Rev.*, **136**, 5132–5147, <https://doi.org/10.1175/2008MWR2445.1>.
- Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, <https://doi.org/10.1175/MWR-D-11-00276.1>.
- Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, [https://doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2).