A Marginal Adjustment Rank Histogram Filter for Non-Gaussian Ensemble Data Assimilation

JEFFREY L. ANDERSON

NCAR Data Assimilation Research Section, Boulder, Colorado

(Manuscript received 13 September 2019, in final form 6 May 2020)

ABSTRACT

An extension to standard ensemble Kalman filter algorithms that can improve performance for non-Gaussian prior distributions, non-Gaussian likelihoods, and bounded state variables is described. The algorithm exploits the capability of the rank histogram filter (RHF) to represent arbitrary prior distributions for observed variables. The rank histogram algorithm can be applied directly to state variables to produce posterior marginal ensembles without the need for regression that is part of standard ensemble filters. These marginals are used to adjust the marginals obtained from a standard ensemble filter that uses regression to update state variables. The final posterior ensemble is obtained by doing an ordered replacement of the posterior marginal ensemble values from a standard ensemble filter with the values obtained from the rank histogram method applied directly to state variables; the algorithm is referred to as the marginal adjustment rank histogram filter (MARHF). Applications to idealized bivariate problems and low-order dynamical systems show that the MARHF can produce better results than standard ensemble methods for priors that are non-Gaussian. Like the original RHF, the MARHF can also make use of arbitrary non-Gaussian observation likelihoods. The MARHF also has advantages for problems with bounded state variables, for instance, the concentration of an atmospheric tracer. Bounds can be automatically respected in the posterior ensembles. With an efficient implementation of the MARHF, the additional cost has better scaling than the standard RHF.

1. Introduction

Two methods for using a set of model forecasts, known as ensemble members or particles, to do sequential data assimilation for geophysical applications were developed in the 1990s. In their most basic form, particle filters ([van Leeuwen 2009](#)) duplicate ensemble members that agree with observations and eliminate members that do not. The assimilation part of the algorithm does not modify the state variables of ensemble members, so important univariate (like boundedness) and multivariate (like approximate geostrophic balance) properties are preserved. Particle methods can represent arbitrary distributions but scale poorly with increasing model size ([Snyder et al. 2008, 2015](#)). The other class of methods, collectively known as ensemble (Kalman) filters ([Houtekamer and Mitchell 1998](#); [Burgers et al. 1998](#); [Pham 2001](#); [Tippett et al. 2003](#)), modifies the state variables of ensemble members during the assimilation. These methods generally scale better than basic particle filters but may not preserve prior

constraints and may not appropriately represent non-Gaussian distributions ([Lei et al. 2010](#)).

For the past two decades, a theme in data assimilation has been trying to develop methods that retain the strengths and avoid the deficiencies of the particle and ensemble filters ([Mandel and Beezley 2009](#); [Stordal et al. 2011](#); [Frei and Künsch 2013](#)). This has meant extending ensemble filters to deal better with non-Gaussian nonlinear priors ([Zupanski 2005](#); [Hodyss 2012](#); [Bocquet and Sakov 2012](#); [El Gharamti et al. 2015](#); [Sakov et al. 2018](#); [Vetra-Carvalho et al. 2018](#); [Nino-Ruiz et al. 2018](#)) and making particle filters more computationally feasible ([Majda et al. 2014](#); [Ades and Van Leeuwen 2015](#); [Poterjoy 2016](#); [Penny and Miyoshi 2016](#); [Van Leeuwen et al. 2019](#)).

One algorithm extending ensemble Kalman filters is the rank histogram filter (RHF; [Anderson 2010](#)). The RHF can represent non-Gaussian prior distributions for observed variables along with arbitrary likelihoods with small incremental cost. [Metref et al. \(2014\)](#) extended the RHF to a multivariate algorithm that allowed non-Gaussian prior information to be used for all state variables and maintained multivariate non-Gaussian structure.

Corresponding author: Jeffrey L. Anderson, jla@ucar.edu

DOI: 10.1175/MWR-D-19-0307.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

However, their algorithm was significantly more expensive than standard ensemble filters.

Here, another approach for extending the RHF to multivariate distributions is described. While not as general as the method of Metref et al. it can be implemented with a similar computational cost scaling to standard ensemble filters and a relatively small constant factor incremental cost in practice. Section 2 reviews the rank histogram filter and describes how it can be used to correct posterior marginal ensembles obtained from a standard ensemble filter. Application of this new method to idealized problems and low-order models is discussed in sections 3 through 5, section 6 discusses computational cost, and section 7 provides conclusions.

2. The marginal adjustment rank histogram filter

Anderson (2010; hereafter A10) described the RHF, an ensemble assimilation method that extended the ensemble Kalman filter to represent non-Gaussian priors and arbitrary likelihoods for observed variables. The last paragraph of A10 noted that a variant of the RHF could compute a posterior ensemble for each state variable directly from the observation likelihood. This algorithm is illustrated in Fig. 1a for an ensemble of size $N = 5$ and proceeds as follows:

- 1) The forward operator is computed and the likelihood evaluated for each ensemble member. Steps 2–9 are done for each state variable separately.
- 2) The prior ensemble of the state variable (green asterisks in Fig. 1a) is sorted and partitions the real line into $N + 1$ regions. Each ensemble member has an associated value of the likelihood (red asterisks in Fig. 1a) computed in step 1.
- 3) Each region is assumed to contain $1/(N + 1)$ of the prior probability similar to the rank histogram (Anderson 1996; Hamill 2001).
- 4) For each of the $N - 1$ bounded regions, the prior probability is assumed to be uniformly distributed as shown by the filled green prior PDF in Fig. 1a.
- 5) In the two unbounded tail regions, the prior probability is assumed to be distributed like the tail of a normal distribution with variance equal to the prior ensemble sample variance. For the left region, the mean is selected so that $1/(N + 1)$ of the probability occurs for values less than the smallest ensemble member. For the right region, the mean is selected so that $1/(N + 1)$ of the probability occurs for values greater than the largest ensemble member (filled green prior PDF in unbounded regions of Fig. 1a).
- 6) A piecewise continuous approximate likelihood is constructed for the regions (red dashed lines in Fig. 1a). For the $N - 1$ bounded regions, the

FIG. 1. Schematic description of the MARHF algorithm. (a) The flat-tail RHF algorithm applied to a generic state variable. A five-member prior ensemble is shown by green asterisks, the relative observation likelihood associated with each ensemble member by red asterisks, and the posterior ensemble by blue asterisks. An approximate piecewise continuous prior density is represented by the green shading, an approximate piecewise continuous likelihood by the red dashed lines, and a piecewise continuous posterior by the blue lines. (b) How a posterior ensemble from a standard ensemble filter is adjusted. The blue asterisks are the posterior ensemble members duplicated from (a). The blue circles are the posterior ensemble members from a standard filter with the ensemble index inside the circle. The arrows indicate how the RHF posterior values replace the values from the standard filter. For example, standard posterior member 4 has the smallest value from the standard filter, and that value is replaced with the smallest value from the RHF.

- likelihood is the mean of the likelihood values of the ensemble members bounding the region.
- 7) The likelihood for the two unbounded tail regions is also constant and has the value of the likelihood of the ensemble member that bounds one side of the region. This is referred to in A10 as the “flat tail” RHF and differs from the method described in detail there, which uses a Gaussian distribution for the likelihood in the unbounded regions.
- 8) The continuous posterior distribution is the product of the continuous prior and likelihood, normalized so that the result is a probability distribution (piecewise continuous blue curve in Fig. 1a). The posterior is Gaussian in the unbounded regions and uniform in each bounded region.
- 9) The N posterior ensemble members (blue asterisks in Fig. 1a) are chosen so that they partition the real line into regions with $1/(N + 1)$ of the posterior probability

in each. These are referred to here as the RHF posterior ensemble for a state variable.

Here, this RHF posterior is used to adjust the posterior marginal ensemble for a state variable computed by a standard ensemble Kalman filter-type algorithm (Anderson 2003). Figure 1b shows how this marginal adjustment rank histogram filter (MARHF) is computed. The blue asterisks are the posterior ensemble computed with the RHF applied directly to a state variable from Fig. 1a. The blue circles are the posterior ensemble members computed with a standard method and the number inside is the ensemble index, the order in which the ensemble members are indexed in the standard filter code. The marginal adjustment is indicated by the arrows in Fig. 1b. In this example, standard posterior member 4 has the smallest value, and this value is replaced by the smallest RHF posterior value. More generally, the value of the n th smallest standard posterior ensemble member is replaced with the n th smallest value from the RHF. In other words, the RHF posterior for a state variable is assigned to ensemble members so that it has the same rank statistics as the marginal from the standard method. The computation of the RHF posterior (Fig. 1a) and the marginal adjustment (Fig. 1b) is done independently for each state variable.

The standard posterior for a state variable can be generated using any traditional method, for example, the perturbed observations EnKF (Burgers et al. 1998; Evensen 2003), the ensemble adjustment Kalman filter (EAKF; Anderson 2001), or the multivariate RHF as described in A10. All results shown here use the RHF as the standard method so that the treatment of observed variables is similar to the way in which the MARHF adjusts state variables.

3. Application to bivariate Gaussian prior

As a first test, the various methods are applied to bivariate Gaussian prior distributions with unit variance and a specified correlation. For each correlation and each ensemble size tested, 100 000 trials are evaluated. Each trial has a prior ensemble that is a random draw from the prior. One of the two variables is observed, and the observation is a randomly selected prior ensemble member perturbed by a random draw from $N(0, 1)$ to simulate observation error. The continuous Kalman filter is the best linear unbiased estimate (BLUE) for this problem.

Correlation values from the set $\{0, 0.1, 0.2, \dots, 1.0\}$ are explored. The standard EAKF and RHF using regression to update state variables and the MARHF are tested for ensemble sizes of 40, 80, 160, and 1280. The standard EAKF is simply an algorithm for computing the Kalman filter for this problem and provides the BLUE for a given sample size. Errors of the ensemble methods compared to

FIG. 2. Time-mean RMSE of posterior ensemble mean for the unobserved variable in an ensemble assimilation problem with a prior drawn from a bivariate Gaussian with unit variance. The horizontal axis indicates the correlation of the prior Gaussian. Results are shown for the standard EAKF with regression (dashed lines) and the MARHF (solid lines) for ensemble sizes of 40, 80, 160, and 1280.

the analytic Kalman filter results (the same as an EAKF with an infinite ensemble size) are examined.

Figure 2 shows the posterior RMSE of the ensemble mean of the unobserved variable for the EAKF and the MARHF averaged over the 100 000 trials. As expected, the EAKF produces smaller RMSE for all prior correlations. The RMSE reduces for both methods as ensemble size increases suggesting that the MARHF may be converging to the BLUE solution with large ensemble size. Figure 3 shows the RMSE of the unobserved variable posterior variance averaged over the trials. The MARHF has larger RMSE than the EAKF for a given ensemble size and prior correlation, but errors decrease with ensemble size suggesting that the MARHF may be converging to the correct solution. Figure 4 shows the RMSE of the posterior correlation between the observed and unobserved variable averaged over the trials. The MARHF posterior correlation RMSE is not much larger than for the EAKF and the values become closer as the ensemble size increases.

For problems with Gaussian priors, the MARHF is not competitive with the optimal EAKF. However, it may be converging to the correct solution with increasing ensemble size.

4. Application to low-order dynamical systems

a. Experimental design

Ensemble filters available in the Manhattan release of DART (Anderson et al. 2009; <https://www.image.ucar.edu/DAReS/DART/>) have been applied to idealized

FIG. 3. As in Fig. 2, except that RMSE of posterior ensemble variance for the unobserved variable is plotted.

problems in the Lorenz-63 (Lorenz 1963; hereafter L63) and 40-variable Lorenz-96 (Lorenz and Emanuel 1998; hereafter L96) models. For both models, an initial truth integration of 1.1 million time steps was generated starting from a preliminary initial condition of 1 for the first state variable (x in L63) and 0 for all others. States at time steps that were multiples of 100 000 were saved as a set of 10 initial conditions for subsequent assimilation experiments. A number of different cases were explored for each model. A “case” is distinguished by the assimilation period, the observation error variance, and (for L96) the forward operator.

Ten independent sequences of observations were generated for each case by starting a truth run from each of the initial conditions and integrating for 5500 assimilation steps, generating observations at each step. Initial ensembles for a given ensemble size were generated for each of the 10 initial conditions by adding a random draw from $N(0, 1)$ to the truth for each ensemble copy of each state variable. Experiments were performed for 5500 assimilation steps with the first 500 steps discarded. The RMSE of the ensemble mean averaged over the 5000 steps is the primary evaluation metric. Ensemble sizes of 20, 40, 80, and 160 were explored.

Four different data assimilation (DA) methods were tested for each case: the EnKF, the perturbed observation EnKF with sorted observation increments (Anderson 2019, hereafter A19), the RHF, and the MARHF. For each triplet of case, ensemble size, and method, a set of fixed multiplicative prior inflation values and Gaspari–Cohn localization halfwidths was tested in experiments using the first of the 10 initial conditions. The values of inflation and localization that resulted in the smallest

FIG. 4. As in Fig. 2, except that the RMSE of posterior correlation is plotted. The additional dotted curves are for the RMRHF, a data assimilation method that uses the RHF to update state variables directly but does not make use of a prior from a standard method.

time-mean RMSE were then used for 9 additional experiments using the remaining 9 initial conditions resulting in a set of 10 experiments of 5500 assimilation steps for each triplet. The adaptive inflation techniques in DART (Anderson 2009; El Gharamti 2018) were not used as they may be challenged by highly non-Gaussian applications.

This experimental procedure leads to a small number of cases where an assimilation diverges, failing to track the truth and producing large RMSE (see results for L63 and L96 in subsequent subsections). This occurs because the “optimal” values of inflation and localization for the first of the 10 initial conditions are problematic for one of the other 9 initial conditions. It was found that a larger inflation value would prevent filter divergence in these cases. This points out the challenges of tuning these parameters with a single case.

Localization is applied in the same way as in standard DART applications. Increments for state variables are computed by regression in the EnKF, EAKF, and RHF, then these increments are multiplied by a scalar α that is a function of the distance between the observation and the state variable. For the MARHF, localizing the state space increments in this way turns out to be identical to the localization obtained if the relative likelihoods are damped by the same factor and the RHF algorithm applied [see Eq. (1) section 6].

b. Lorenz-63 results

The three-variable L63 model was used with the standard parameter settings of $\sigma = 10$, $r = 28$, and $b = 8/3$ with a time step of 0.01 using two-step Runge–Kutta

FIG. 5. Time-mean RMSE for Lorenz-63 assimilation cases with 80-member ensembles. Each panel is for a different assimilation period and contains results for each of the six observation error variances. Blue bars are the RMSE for each of the 10 individual experiments for a given case and the red bar is the mean of these 10 values. The three columns of bars for each observation variance are for the EnKF (left), the RHF (middle), and the MARHF (right).

time differencing. For plotting results, a single time step is defined as being equivalent to 1 h of idealized time to be consistent with previous studies. All three state variables were observed at each assimilation time. For localization, the three state variables were treated as if they were equally spaced on a cyclic domain of length 1 (the distance between any pair of state variables is $1/3$). Thirty-six cases were tested with each characterized by an assimilation period from the set $\{3600\text{ s}, 10800\text{ s}, 21600\text{ s}, 1\text{ day}, 2\text{ days}, 4\text{ days}\}$ and an observation error variance from the set $\{4, 8, 16, 32, 64, 128\}$. These cases were selected so that the posterior RMSE varied from a few percent of the climatological RMSE up to approximately the climatological RMSE. Twenty-four pairs of inflation and localization values were evaluated to tune each case with inflation selected from $\{1.0, 1.01, 1.02, 1.04, 1.08, 1.16\}$ and Gaspari–Cohn (Gaspari and Cohn 1999) localization half-width selected from $\{0.4, 0.6, 0.8, \infty\}$.

Figure 5 shows 80-member ensemble results for the EnKF, RHF, and MARHF for the 36 cases. For each method, the RMSE from each of the 10 initial conditions is displayed along with the mean of the 10. For almost all cases, the MARHF produces the smallest RMSE, followed by the RHF and the EnKF. Exceptions are found for three cases with short periods and small observation error variance for which the EnKF has smaller mean RMSE. The relative difference between the MARHF and the other methods tends to be larger for larger posterior errors that occur for cases with long periods and/or large observation error variance; these cases are expected to have a less linear relation between observed and state variables and less Gaussian priors. The EAKF (not plotted) was worse than all three of these methods for almost all cases.

Figure 6 shows the mean RMSE from the 10 initial conditions for 40-, 80-, and 160-member ensembles for all cases with the RMSE normalized by the RMSE of the

FIG. 6. Normalized time-mean RMSE for Lorenz-63 assimilation cases with 40 (green), 80 (blue), and 160 (red) ensemble members; values plotted are the mean of the 10 individual experiments for each ensemble size. Each panel is for a different assimilation period and contains results for each of the six observation error variances. The three columns of bars for each case are for the EnKF (left), the RHF (middle), and the MARHF (right). The values for each observation error variance on a panel are normalized by the smallest RMSE value for that case so that the best experiment has a value of 1.

best experiment from all methods for each case (a value of 1 is plotted for the best single experiment). For almost all cases and methods the RMSE reduces as the ensemble size increases as expected. Exceptions occur in cases when one or more of the experiments for a given method diverged leading to very large mean RMSE. In a few of these cases, the mean error is so large that it is off the plots in Fig. 6. For cases with large error variance and/or long period, the MARHF is better for all ensemble sizes. In fact, there are several cases for which the 40-member MARHF is better than even the 160-member EnKF or RHF, and many cases for which the 80-member MARHF is better than the 160-member EnKF or RHF. For shorter periods and smaller error variances, the 40-member MARHF is generally among the worst methods, while the 80- and 160-member MARHF are more competitive. For 20-member ensembles (not shown in figures), the

MARHF performed worse than the other methods for almost all cases.

The L63 model provides good illustrations of the differences in posterior ensembles generated by the different methods. As an example, an 80-member prior ensemble from day 4600 of the L63 RHF applied to the case with assimilation period 1 day and observation error variance 16 was used as the prior for all of the methods. Figure 7 shows the projection of the prior and posterior ensembles onto the x - z plane for the EnKF, the RHF, and the MARHF. The posterior is the result of assimilating a single observation of x that is marked by the red line in the figure panels with observation error standard deviation of 4 indicated by the red horizontal region. The prior spans a significant portion of the attractor (shown in gray shading) and is clearly not close to Gaussian. The EnKF and RHF use regression to update the unobserved z variable so the increments all have the

FIG. 7. Example of prior and posterior 80-member ensembles after assimilating an observation of the x variable for Lorenz-63 projected on an x - z plane. The prior is indicated by large green dots and the posterior by smaller blue dots with thin lines showing the increments between prior and posterior. The observed value of x is indicated by the vertical red line and the observation error standard deviation by the horizontal red region. The three panels show the results of applying an (left) EnKF, (center) RHF, and (right) MARHF. An approximation of the model attractor is depicted in gray by 100 000 samples from an integration of the model with no data assimilation.

same slope for both methods. Both posteriors contain many ensemble members that are far from the attractor and any of the initial ensemble members. The MARHF adjusts the increments from the RHF so increment vectors do not have to be parallel. For the MARHF, the posterior ensemble members are closer to prior ensemble members and more consistent with the attractor.

Figure 8 shows the projection of the prior and posterior on the y - z plane; neither of these variables is observed. Again, the EnKF and RHF increments all have the same slope and the posterior is clearly far from the attractor. However, the MARHF appears to be more consistent with the attractor. There are

nevertheless several ensemble members that are well off the attractor.

c. Lorenz-96 results

The L96 model with 40 variables was used as a larger test case. The same configuration was used as in A19 with $F = 8.0$, fourth-order Runge–Kutta time differencing and a 0.05-unit time step. Several previous DA studies like Sakov et al. (2012) and Anderson (2016) and references therein also used this configuration. As in A19, state variables are assumed to be uniformly spaced on a cyclic $[0, 1]$ domain. A subset of the cases explored in A19 are used here. Three of the forward operators from A19 are examined but only for the configuration

FIG. 8. Example of prior and posterior 80-member ensembles after assimilating an observation of the x -variable for Lorenz-63 projected on a y - z plane. The prior is indicated by large green dots and the posterior by smaller blue dots with thin lines showing the increments between prior and posterior. The observed value of x is indicated on the panels of Fig. 7. The three panels show the results of applying an (left) EnKF, (center) RHF, and (right) MARHF. An approximation of the model attractor is depicted in gray by 100 000 samples from an integration of the model with no data assimilation.

TABLE 1. Observation error variances used for cases with different forward operators for the Lorenz-96 model.

Forward operator	Obs error variance
Identity	1, 2, 4, 16
Square root	0.25, 0.5, 1, 2
Square	16, 64, 256

where observations are taken at 40 uniformly randomly distributed “observing stations.” The station locations are fixed in time and the same for all experiments. Let x_o be the value of the state linearly interpolated to one of the station locations. The observation types are identity: $y_o = x_o$; square root: $y_o = \text{sgn}(x_o)\sqrt{|x_o|}$; and square: $y_o = x_o^2$.

For L96, a case is defined by its choice of assimilation period selected from the set $\{1, 2, 3, 6, 12\}$ time steps, the observation error variance, and the forward operator. A total of 20 cases were explored for the identity and

square root forward operators and 15 for the square forward operator. Table 1 shows the set of error variances that were used for each forward operator.

As for L63, the localization and inflation for each method were optimized for each case. A total of 70 pairs of inflation/deflation and localization were tested for each case and method. Inflation/deflation was selected from the set $\{0.95, 0.98, 0.99, 1, 1.02, 1.04, 1.08, 1.16, 1.32, 1.64\}$ and localization half-width from the set $\{0.125, 0.15, 0.175, 0.2, 0.25, 0.4, \infty\}$. Note that the deflation values were not tested when optimizing in A19 and there were cases in the current study for which deflation did lead to the smallest RMSE for some methods.

Figure 9 shows the RMSE for the 10 experiments and their mean for each of the 20 cases with identity forward operator for the EnKF, RHF, and MARHF. As for L63, the MARHF tends to be better for larger observation error variance or longer assimilation period. As noted in A19, it is difficult to do better than standard regression

FIG. 9. Time-mean RMSE for Lorenz-96 assimilation cases with identity forward operator and 80-member ensembles. Each panel is for a different observation error variance and contains results for each of the five assimilation periods. Blue bars are the RMSE for each of the 10 individual experiments for a given case and the red bar is the mean of these 10 values. The three columns of bars for each assimilation period are for the EnKF (left), the RHF (middle), and the MARHF (right).

FIG. 10. Normalized time-mean RMSE for Lorenz-96 assimilation cases with identity forward operators and 40 (green), 80 (blue), and 160 (red) ensemble members; values plotted are the mean of the 10 individual experiments for each ensemble size. Each panel is for a different observation error variance and contains results for each of the five assimilation periods. The three columns of bars for each assimilation period are for the EnKF (left), the RHF (middle), and the MARHF (right). The values for each observation error variance on a panel are normalized by the smallest RMSE value for that observation error variance so that the best method has a value of 1. Note the different vertical scale in (d).

methods for identity observations in L96. For most of the cases, there is relatively little variation in the RMSE for the 10 experiments so that even small differences in mean RMSE appear to be significant.

Figure 10 shows the normalized mean RMSE from the 10 experiments for 40-, 80-, and 160-member ensembles for identity observations. For 40 members, the MARHF has the smallest RMSE only for cases with the largest error variance, while the EnKF is best for most cases. As the ensemble size gets larger, the MARHF improves more rapidly so that for 160 members, the MARHF has the smallest RMSE for 13 of the 20 cases. For some of the cases with large error variance, the 80-member MARHF has smaller RMSE than the other methods with 160 members.

Figure 11 shows all 10 experiments and their mean for 80-member ensembles for the square root forward operator, while Fig. 12 shows the mean results for

ensembles with 40, 80, and 160 members. The enhanced nonlinearity between observations and state variables leads to the MARHF having smaller RMSE for many cases with larger error variance and/or period. There can be quite a bit of variability across the 10 experiments for cases with short period, but the differences in mean appear to be significant for most cases. For 160-member ensembles, Fig. 12 shows that 17 out of 20 cases have smallest RMSE with the MARHF. Only for a few of the largest error variance cases is the 40-member MARHF best. For some 160-member cases, the EnKF had one or more experiments for which the ensemble diverged so that the RMSE was larger than for smaller ensembles. This appears to be related to the outlier behavior noted in A10 that motivated the original development of the RHF. Neither the RHF nor the MARHF demonstrated this type of behavior.

FIG. 11. As in Fig. 9, except for with the square-root forward operator.

Figure 13 shows the normalized mean RMSE for the square forward operator. Many more experiments diverged for this forward operator, especially for small error variance and short period. This problem was more common for the EnKF, which had RMSE outside the plotted range for a number of cases. Otherwise, the results for the square forward operator are similar to the other cases, with the MARHF having smaller RMSE for larger observation error variance and longer periods.

5. Application to bounded variables

Many state variables in Earth system models are bounded; for instance, variables that represent the amount or concentration of a substance are nonnegative. Less commonly, one finds variables that are bounded above or confined to a finite range, for instance, the amount of water in a fixed capacity soil bucket or sea ice fraction. Many ensemble assimilation techniques can sometimes generate posterior ensemble members that lie outside of the bounds. This can be problematic for subsequent model forecasts. As pointed out in A10, it is straightforward to include

additional constraints on the prior in the standard RHF to enforce bounds. This prevents posterior ensemble members for *observed* quantities from violating bounds, but regressing the ensemble increments for observed variables onto state variables can still generate posterior ensemble members that violate bounds.

The MARHF can enforce bounds on any state variable posterior by bounding the continuous prior. Figure 14 is a modified version of Fig. 1a for a variable that is nonnegative like a tracer concentration. Consistent with the rank histogram (Anderson 1996), 1/6 of the probability density in the continuous prior is assumed to lie below the smallest ensemble member, but that region is now bounded below by zero. The rest of the flat tail RHF algorithm is unmodified. The continuous posterior has zero probability of negative values and the posterior ensemble has no negative members, unlike the case in Fig. 1 where the prior was not bounded.

An idealized exploration of applying an MARHF to bounded quantities makes use of a lognormal bivariate prior distribution. Prior ensembles are generated by taking the antilog of random draws from the same distribution described in section 3. One of the variables is

FIG. 12. As in Fig. 10, except for the square-root forward operator.

observed with a likelihood that has a gamma distribution with shape parameter equal to the value of a randomly chosen ensemble member and scale parameter 1.0. The gamma distribution likelihood has zero probability for negative values, consistent with the bounded nature of the variables.

It is clearly unrealistic and suboptimal for measurements of bounded quantities to have nonzero likelihood for impossible values (Bishop 2016; Posselt and Bishop 2018). Nevertheless, many assimilation techniques use Gaussian likelihoods that are unbounded. To simulate this situation, the mean and variance of the gamma likelihood are used to generate a suboptimal Gaussian likelihood for comparison with methods that use the actual gamma.

Five assimilation methods are applied: an EAKF using the approximate Gaussian likelihood, an RHF using the Gaussian likelihood, an MARHF using the Gaussian likelihood, an RHF using the gamma likelihood, and an MARHF using the gamma likelihood. As in section 3, 11 different correlations are explored for the log space prior bivariate distribution with 100 000 trials for each ensemble size from the set {40, 80, 160, 1280}. In this

case, there is no comparison to an analytic result. Instead, a particle filter is applied to get the statistics of the posterior for each ensemble. This is done by computing the likelihood weighted average, variance, and correlation of the prior ensembles. The ensemble filter posteriors are compared to the particle filter statistics. Note that this does not indicate that a particle filter with these sizes would be the benchmark in a *cycling* assimilation because of the challenges noted in Snyder et al. (2008) and Slivinski and Snyder (2016).

Figure 15 shows the RMSE of the posterior ensemble mean averaged over 100 000 trials for the unobserved variable as a function of the prior log space correlation. The EAKF always has largest RMSE and the MARHF with the gamma likelihood smallest RMSE for all ensemble sizes. While the RMSE for the MARHF with the gamma is uniformly decreasing as a function of ensemble size, this does not appear to be the case for the other methods for most prior correlations.

Figure 16 shows the RMSE of the posterior variance of the unobserved variable. Again, the EAKF is always largest and the MARHF with the gamma is always smallest. However, in this case all of the methods seem

FIG. 13. As in Figs. 10 and 12, except for the square forward operator.

to have RMSE that is decreasing with ensemble size. The RHF with the gamma likelihood is nearly the same as the RHF with the Gaussian likelihood, while the MARHF with the Gaussian is much closer to the MARHF with the gamma than in Fig. 15.

Figure 17 shows the RMSE of the posterior correlation. In general, errors for all methods are quite similar. There is no indication that the errors of any of the methods are decreasing with increasing ensemble size. For ensemble sizes 160 and 1280, the MARHF with Gaussian observation error has the smallest error, while the MARHF with the gamma likelihood is quite similar. This is consistent with expectations that the marginal statistics of the MARHF will be nearly correct, but bivariate statistics that come from correcting the regression prior of an RHF continue to be approximate.

Figure 18 shows the fraction of posterior ensemble members for the unobserved variable that are negative, violating the bound. Both MARHF methods have no negative members, while the other three methods have greater than 4% negative for some value of prior correlation for all ensemble sizes. It does not appear that the fraction of negative members is decreasing as the ensemble size increases for these other methods.

It is beyond the scope of this manuscript to explore applying the MARHF to realistic models with bounded quantities, but these bivariate results suggest that there could be significant advantages that will be explored in future work. Initial applications of the MARHF with 80 members for global NWP at

FIG. 14. Schematic of the flat-tail rank histogram filter algorithm applied to a positive-definite quantity. A five-member prior ensemble is shown by green asterisks, the relative observation likelihood associated with each ensemble member by red asterisks, and the posterior ensemble by blue asterisks. An approximate piecewise continuous prior density is represented by the green shading, an approximate piecewise continuous likelihood by the red dashed lines, and a piecewise continuous posterior by the blue lines.

FIG. 15. Time-mean RMSE of posterior ensemble mean for the unobserved variable in an ensemble assimilation problem with a prior drawn from a bivariate lognormal distribution with unit variance and an observation likelihood with a gamma distribution. The horizontal axis indicates the correlation of the prior bivariate distribution in log space. Results are shown for the standard EAKF with an approximate Gaussian likelihood (black dashed), an RHF with Gaussian likelihood (red dashed), an MARHF with Gaussian likelihood (blue dashed), an RHF with gamma likelihood (red solid), and an MARHF with gamma likelihood (blue solid) for ensemble sizes of (top left) 40, (top right) 80, (bottom left) 160, and (bottom right) 1280.

approximately 1° resolution have been competitive with an EAKF.

6. Computational details and cost

The MARHF is an adjustment to the results of a standard ensemble filter update. The additional computations for the MARHF for assimilating a single observation that impacts M state variables for an N -member ensemble include the following (with cost measured by number of floating point operations):

- 1) Sorting each of the impacted state variable ensembles to generate the rank histogram prior [cost $O(MN \log N)$ for a reasonable sort].
- 2) Computing the likelihood for each ensemble member [cost $O(N)$].
- 3) Multiplying the prior value times the observation likelihood for each ensemble member [cost $O(MN)$].

- 4) Identifying the position of the new ensemble members in the posterior [cost is $O(MN)$ since the posterior ensemble is already sorted].

The total cost for assimilating P observations is then $O(PMN \log N)$ compared to the $O(PMN)$ for standard ensemble filters. These costs might be further reduced in order when used with localization, but the reduction factor would be the same for both the MARHF and the standard filter.

As pointed out in A19, it is possible to reduce the cost of ensemble filter algorithms like the MARHF that require sorting by taking advantage of the fact that prior ensembles of state variables are generally not completely scrambled after the assimilation of a single observation. Instead, the posterior is often nearly sorted and applying an algorithm like insertion sort can reduce the expected cost of sorting a state variable ensemble from $O(N \log N)$ to $O(N)$.

FIG. 16. As in Fig. 15, but for the RMSE of posterior ensemble variance for the unobserved variable. Note that the RHF and RHF with gamma likelihood are nearly identical.

All results for this study used this approach and the cost of the sorts was found to scale as $O(N)$ for ensemble sizes of 10–1280.

An additional computational cost reduction is possible by modifying the MARHF algorithm so that the impact of all observations is computed in a single step at each assimilation time. To do this, the parallel ensemble Kalman filter algorithm described in Anderson and Collins (2007) that is used in DART was applied. Suppose that there are P observations to be assimilated at a given time. The assimilation step begins by computing the ensemble of forward operators for all of the P observations and defining an extended state vector that includes both the model state variables and the observations. Observations are assimilated serially and impact the entire extended state. A cumulative likelihood is kept for each state variable as the observations are assimilated so that

$$L_c^n = \sum_{p=1}^P [\alpha_p L_p^n + (1 - \alpha_p) \bar{L}_p], \quad n = 1, \dots, N, \quad (1)$$

where L is a likelihood, superscript n indexes the N ensemble members, subscript p indexes the P observations, subscript c is the cumulative likelihood, α_p is the localization factor for the p th observation, and the overbar is an ensemble mean. Once all observations have been assimilated by the standard filter, a single RHF update can be done using the cumulative ensemble likelihoods to compute the posterior marginal ensemble for each state variable and adjust the posterior from the standard method as in Fig. 1. The total additional cost for the MARHF is now due to

- 1) computation of the cumulative likelihood for the joint state, $O(PMN) + O(P^2N)$;
- 2) a single RHF update for each joint state variable, $O(MN \log N) + O(PN \log N)$, which can be reduced to $O(MN) + O(PN)$ for nearly sorted priors.

The cost of the Anderson and Collins method is $O(PMN) + O(P^2N)$ due to the extra updates required for observed variables in the extended state. The final incremental cost for the MARHF has the same scaling.

FIG. 17. As in Fig. 15, but for the RMSE of posterior correlation.

Despite its additional cost, the Anderson and Collins method has been demonstrated to be an effective way to use parallel computing efficiently. The additional cost is small if the number of observations is small compared to the number of state variables. If this is not the case, the incremental cost can be made small by assimilating observations in subsets so that the number of observations in each subset is small compared to the number of state variables. In summary, it is possible to implement the MARHF so that the incremental cost does not change the scaling of the standard ensemble filter cost and may be smaller with efficient sorting.

7. Discussion

Not surprisingly, the MARHF does not produce results that are as good as ensemble Kalman filters that implicitly assume Gaussianity when applied to nearly Gaussian priors. However, for a limited additional cost, the MARHF can produce significantly better results for priors that are not Gaussian. Like the standard RHF, the MARHF can handle arbitrary non-Gaussian priors but for all state variables. The RHF is also able to enforce

additional constraints on ensemble members, for instance, boundedness, and the MARHF extends this capability to all state variables.

It is possible to modify the RHF assumptions about the relationship between the ensemble and an associated continuous probability distribution. As an example, it is sometimes advantageous to make assumptions about priors having fat tails to account for errors that are not represented in model forecast priors. More than $1/(N + 1)$ of the probability mass can then be assumed to lie in the tails when creating the prior continuous probability. This assumption could be maintained, or discarded, when creating the posterior ensemble sample from the continuous posterior depending on the desired outcome. Prior ensemble inflation could also be implemented by modifying the way in which the ensemble members define the continuous ensemble. This approach was not explored here but could allow inflation methods that are more appropriate than the standard multiplicative or additive methods, especially for highly non-Gaussian priors.

A number of ensemble data assimilation techniques apply transformations to the prior ensemble distribution

FIG. 18. As in Fig. 15, but for the fraction of posterior ensemble members that are negative. Both the MARHF and the MARHF with gamma likelihood produce no negative ensemble members.

(Beezley and Mandel 2008), often in an attempt to make the problem more Gaussian or linear or to enforce bounds. A log transform is one example (Lien et al. 2016) that can be used for a bounded quantity. Another class of transforms is called anamorphosis (Simon and Bertino 2009; Schöniger et al. 2012; Amezcua and Van Leeuwen 2014; El Gharamti et al. 2015) and uses either functional or empirically derived transforms. When these transforms are used with standard ensemble filtering methods, there is a need to define a Gaussian likelihood in the transformed space. The RHF and the MARHF remove this requirement and allow an arbitrary likelihood defined in the original space to be used in the transformed space.

As noted, the MARHF does not do as well as standard ensemble Kalman filters for problems that are approximately Gaussian. It is trivial to implement filters that use the MARHF to update certain state variables but use the standard algorithm for other state variables. For example, in an atmospheric chemistry model, it might be useful to apply the MARHF only to tracer concentration variables that are bounded below and may display non-Gaussian distributions. It might also be effective to

apply the MARHF to pairs of observations and state variables that are known a priori to have more nonlinear relations while using a standard filter for other pairs.

One could skip the second step of the MARHF algorithm as described in Fig. 1, simply using an RHF to adjust each state variable but not pairing the posteriors by rank with the standard method posterior. In this algorithm, referred to here as the raw marginal rank histogram filter (RMRHF), the rank order of the prior ensemble for every state variable is unchanged in the posterior. If ensemble member 4 has the smallest value for a given state variable prior, ensemble member 4 is also the smallest in the posterior. The RMRHF was tested with the Gaussian bivariate example of section 3. For marginal statistics, like the RMSE of the posterior mean or variance, the RMRHF is identical to the MARHF. However, results are very different for the posterior correlation, a bivariate statistic. The RMRHF has much larger correlation RMSE that increases as ensemble size increases for correlations greater than 0.5 (Fig. 4, dashed lines). Much of this error is due to a systematic tendency for the RMRHF to have correlations with absolute values that are too large. While the

MARHF may be converging to the correct solution, the RMRHF is clearly not and it is generally not competitive for the applications detailed above.

The success of the MARHF compared to the RMRHF is related to the way in which the values from the posterior marginal distribution are paired with ensemble members. Treating the marginal ensemble members as an adjustment to the standard ensemble Kalman filter in the MARHF (Fig. 1b) resulted in a competitive algorithm. However, the method is still suboptimal as reflected, for example, by the posterior points that are off the attractor in Figs. 7 and 8. Other approaches for pairing marginal posterior values with ensemble members could lead to different filter performance. As an example, one could pair the posterior values for the observed and state variables in ways to minimize the size of the increments. In general, such methods seem to be quite expensive but initial exploration has shown that they can produce better results than the MARHF. These methods are related to others that explore modifying particle filters to move prior members to minimize certain norms. The method of Metref et al. (2014) is also closely related to some more general ways of pairing values with ensembles. Obtaining implementations with the power of the Metref et al. algorithms but with greatly reduced cost may be possible. Combining the strengths of methods that make ensemble Kalman filters more general and particle filters less expensive continues to be a compelling direction for future research.

Acknowledgments. The author is grateful for support from the DART team, Nancy Collins, Moha El Gharamti, Tim Hoar, and Kevin Raeder and to Michael Ying for comments on an earlier version. Three anonymous reviewers were a great help in improving earlier drafts. The National Center for Atmospheric Research is sponsored by the National Science Foundation. 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: All software required to repeat the experiments described in this work are publicly available in the Data Assimilation Research Testbed (<https://dart.ucar.edu/>).

REFERENCES

- Ades, M., and P. J. Van Leeuwen, 2015: The effect of the equivalent-weights particle filter on dynamical balance in a primitive equation model. *Mon. Wea. Rev.*, **143**, 581–596, <https://doi.org/10.1175/MWR-D-14-00050.1>.
- Amezcu, J., and P. J. Van Leeuwen, 2014: Gaussian anamorphosis in the analysis step of the EnKF: A joint state-variable/observation approach. *Tellus*, **66A**, 23493, <https://doi.org/10.3402/TELLUSA.V66.23493>.
- Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530, [https://doi.org/10.1175/1520-0442\(1996\)009<1518:AMFPAE>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2).
- Anderson, J., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, [https://doi.org/10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2).
- , 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).
- , 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>.
- , 2010: A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.*, **138**, 4186–4198, <https://doi.org/10.1175/2010MWR3253.1>.
- , 2016: Reducing correlation sampling error in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **144**, 913–925, <https://doi.org/10.1175/MWR-D-15-0052.1>.
- , 2019: A nonlinear rank regression method for ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **147**, 2847–2860, <https://doi.org/10.1175/mwr-d-18-0448.1>.
- , and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.*, **24**, 1452–1463, <https://doi.org/10.1175/JTECH2049.1>.
- , T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, <https://doi.org/10.1175/2009BAMS2618.1>.
- Beezley, J. D., and J. Mandel, 2008: Morphing ensemble Kalman filters. *Tellus*, **60A**, 131–140, <https://doi.org/10.1111/j.1600-0870.2007.00275.x>.
- Bishop, C. H., 2016: The GIGG-EnKF: Ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. *Quart. J. Roy. Meteor. Soc.*, **142**, 1395–1412, <https://doi.org/10.1002/qj.2742>.
- Bocquet, M., and P. Sakov, 2012: Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems. *Nonlinear Processes Geophys.*, **19**, 383–399, <https://doi.org/10.5194/npg-19-383-2012>.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724, [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2).
- 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>.
- , B. Ait-El-Fquih, and I. Hoteit, 2015: An iterative ensemble Kalman filter with one-step-ahead smoothing for state-parameters estimation of contaminant transport models. *J. Hydrol.*, **527**, 442–457, <https://doi.org/10.1016/j.jhydrol.2015.05.004>.
- 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>.
- Frei, M., and H. R. Künsch, 2013: Bridging the ensemble Kalman and particle filters. *Biometrika*, **100**, 781–800, <https://doi.org/10.1093/biomet/ast020>.
- 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>.

- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, [https://doi.org/10.1175/1520-0493\(2001\)129<0550:IORHFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2).
- Hodyss, D., 2012: Accounting for skewness in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 2346–2358, <https://doi.org/10.1175/MWR-D-11-00198.1>.
- 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:DAUAEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2).
- Lei, J., P. Bickel, and C. Snyder, 2010: Comparison of ensemble Kalman filters under non-Gaussianity. *Mon. Wea. Rev.*, **138**, 1293–1306, <https://doi.org/10.1175/2009MWR3133.1>.
- Lien, G., T. Miyoshi, and E. Kalnay, 2016: Assimilation of TRMM multisatellite precipitation analysis with a low-resolution NCEP global forecast system. *Mon. Wea. Rev.*, **144**, 643–661, <https://doi.org/10.1175/MWR-D-15-0149.1>.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
- , and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414, [https://doi.org/10.1175/1520-0469\(1998\)055<0399:OSFSWO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2).
- Majda, A. J., D. Qi, and T. P. Sapsis, 2014: Blended particle filters for chaotic systems. *Proc. Natl. Acad. Sci. USA*, **111**, 7511–7516, <https://doi.org/10.1073/pnas.1405675111>.
- Mandel, J., and J. D. Beezley, 2009: An ensemble Kalman-particle predictor-corrector filter for non-Gaussian data assimilation. *Computational Science, ICCS 2009*, G. Allen et al., Eds., Lecture Notes in Computer Science, Vol. **5545**, Springer, 470–478.
- Metref, S., E. Cosme, C. Snyder, and P. Brasseur, 2014: A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation. *Nonlinear Processes Geophys.*, **21**, 869–885, <https://doi.org/10.5194/npg-21-869-2014>.
- Nino-Ruiz, E. D., H. Cheng, and R. Beltran, 2018: A robust non-Gaussian data assimilation method for highly non-linear models. *Atmosphere*, **9**, 126, <https://doi.org/10.3390/atmos9040126>.
- Penny, S. G., and T. Miyoshi, 2016: A local particle filter for high-dimensional geophysical systems. *Nonlinear Processes Geophys.*, **23**, 391–405, <https://doi.org/10.5194/npg-23-391-2016>.
- Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207, [https://doi.org/10.1175/1520-0493\(2001\)129<1194:SMFSDA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2).
- Posselt, D. J., and C. H. Bishop, 2018: Nonlinear data assimilation for clouds and precipitation using a gamma inverse-gamma ensemble filter. *Quart. J. Roy. Meteor. Soc.*, **144**, 2331–2349, <https://doi.org/10.1002/qj.3374>.
- Poterjoy, J., 2016: A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev.*, **144**, 59–76, <https://doi.org/10.1175/MWR-D-15-0163.1>.
- Sakov, P., D. S. Oliver, and L. Bertino, 2012: An iterative EnKF for strongly nonlinear systems. *Mon. Wea. Rev.*, **140**, 1988–2004, <https://doi.org/10.1175/MWR-D-11-00176.1>.
- , J.-M. Haussaire, and M. Bocquet, 2018: An iterative ensemble Kalman filter in presence of additive model error. *Quart. J. Roy. Meteor. Soc.*, **144**, 1297–1309, <https://doi.org/10.1002/qj.3213>.
- Schöniger, A., W. Nowak, and H.-J. Hendricks Franssen, 2012: Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography. *Water Resour. Res.*, **48**, W04502, <https://doi.org/10.1029/2011WR010462>.
- Simon, E., and L. Bertino, 2009: Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: A twin experiment. *Ocean Sci.*, **5**, 495–510, <https://doi.org/10.5194/os-5-495-2009>.
- Slivinski, L., and C. Snyder, 2016: Exploring practical estimates of the ensemble size necessary for particle filters. *Mon. Wea. Rev.*, **144**, 861–875, <https://doi.org/10.1175/MWR-D-14-00303.1>.
- Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**, 4629–4640, <https://doi.org/10.1175/2008MWR2529.1>.
- , —, and M. Morzfeld, 2015: Performance bounds for particle filters using the optimal proposal. *Mon. Wea. Rev.*, **143**, 4750–4761, <https://doi.org/10.1175/MWR-D-15-0144.1>.
- Stordal, A. S., H. A. Karlsen, G. Nævdal, H. J. Skaug, and B. B. Valles, 2011: Bridging the ensemble Kalman filter and particle filters: The adaptive Gaussian mixture filter. *Comput. Geosci.*, **15**, 293–305, <https://doi.org/10.1007/s10596-010-9207-1>.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490, [https://doi.org/10.1175/1520-0493\(2003\)131<1485:ESRF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2).
- van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137**, 4089–4114, <https://doi.org/10.1175/2009MWR2835.1>.
- , H. R. Künsch, L. Nerger, R. Potthast, and S. Reich, 2019: Particle filters for high-dimensional geoscience applications: A review. *Quart. J. Roy. Meteor. Soc.*, **145**, 2335–2365, <https://doi.org/10.1002/qj.3551>.
- Vetra-Carvalho, S., P. Jan van Leeuwen, L. Nerger, A. Barth, J. U. Altaf, P. Brasseur, P. Kirchgeßner, and J.-M. Beckers, 2018: State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus*, **70**, 1–43, <https://doi.org/10.1080/16000870.2018.1445364>.
- Zupanski, M., 2005: Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.*, **133**, 1710–1726, <https://doi.org/10.1175/MWR2946.1>.