MDPI

Article

A quantile-conserving ensemble filter based on kernel density estimation

Ian Grooms [1]* **and Christopher Riedel** [2]

[1] Department of Applied Mathematics, University of Colorado, Boulder, CO, 80309, USA; ian.grooms@colorado.edu
[2] University Corporation for Atmospheric Research, Boulder, CO, 80309; criedel@ucar.edu
* Correspondence: ian.grooms@colorado.edu

Abstract: Ensemble Kalman filters are an efficient class of algorithms for large-scale ensemble data assimilation, but their performance is limited by their underlying Gaussian approximation. A two-step framework for ensemble data assimilation allows this approximation to be relaxed: The first step updates the ensemble in observation space, while the second step regresses the observation state update back to the state variables. This paper develops a new quantile-conserving ensemble filter based on kernel density estimation and quadrature for the scalar first step of the two-step framework. It is shown to perform well in idealized non-Gaussian problems, as well as in an idealized model of assimilating observations of sea ice concentration.

Keywords: data assimilation; ensemble filter; sea ice concentration

1. Introduction

Data assimilation combines observations, both in situ and remotely sensed, with a forecast model to estimate the state of a dynamical system. Ensemble Kalman Filters (EnKFs; [1,2]) are a class of ensemble data assimilation methods that are widely used in large-scale Earth science applications. Ensemble filters operate sequentially in time: Each time that observations become available the ensemble forecast is adjusted, and the resulting analysis ensemble is then forecast to the next time that observations become available. The accuracy of EnKFs is limited (inter alia) by their underlying assumption of joint Gaussianity, and there is a wide range of approaches to relaxing or removing this assumption. The class of non-Gaussian extensions of the EnKF that is of interest here derives from the two-step algorithm for implementing an EnKF developed by Anderson [3]. In the first step an EnKF is used to update an ensemble of forecast observations; in the second step the ensemble of observation increments is converted to an ensemble of state-variable increments using linear regression. Both of these steps can be generalized, leading to non-Gaussian (i.e. non-Kalman) ensemble filters [4].

The focus here is on developing a non-parametric, non-Gaussian ensemble filter for the first step of this two-step process, where we assume that the first step (the update in observation space) is a scalar Bayesian estimation problem. The restriction to scalar problems is not very restrictive in light of the well-known property of Bayesian estimation that observations with independent errors can be assimilated serially, in any order; the problem of assimilating many observations can therefore often be reduced to a sequence of problems assimilating scalar observations. The approach adopted here is based on the probability integral transform

$$H^a = F_a^{-1}\Big(F_f\Big(H^f\Big)\Big). \tag{1}$$

In the above equation H^f and H^a are random variables distributed according to the prior (forecast) and posterior (analysis) in observation space, respectively, and F_f and F_a are

Citation: Grooms, I.; Riedel, C. KQCEF. *Remote Sens.* **2024**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

the cumulative distribution functions (cdfs) of H^f and H^a. The letter y is often used to denote the value of an observation, and the standard observation model that relates the observation y to the state vector $\boldsymbol{x}$ is

$$y = h(\boldsymbol{x}) + \epsilon \tag{2}$$

where h is the observation operator (sometimes called the 'forward' operator) and ϵ is the observation error. The letter H in Equation (1) is deliberately chosen here to emphasize the difference between the observation including observation error and the observation without including observation error. If $\boldsymbol{X}^f$ is a random variable distributed according to the prior, then $H^f = h\left(\boldsymbol{X}^f\right)$ is a random variable distributed according to the prior in observation space. As described in [4], two-step ensemble filters do not need to assume a standard observation model, and even when observations are related to the state by the standard model one does not need to choose to estimate $h(x)$ in the first step, but we retain this formalism here for simplicity of exposition.

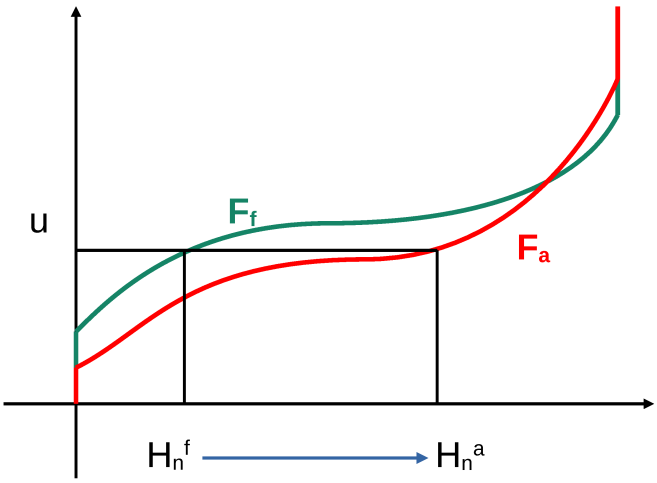

Figure 1. An illustration of a QCEF. The prior ensemble member H_n^f is mapped by the prior cdf F_f to the value $u \in [0,1]$. This value is then mapped to the analysis ensemble member H_n^a by the analysis quantile function F_a^{-1}. The cdfs illustrated here correspond to the case of mixed distributions on a bounded domain as discussed in Section 2.

As described in more detail in Appendix A, the transform Equation (1) operates by mapping the forecast distribution for H^f to a uniform distribution on $[0,1]$, and then mapping from the uniform distribution to the posterior distribution of H^a; it relates to ensemble filtering as follows. Suppose that one has a forecast ensemble in observation space h_n^f, $n = 1, \ldots, N_e$, and one wishes to generate an analysis ensemble in observation space h_n^a, $n = 1, \ldots, N_e$. If F_f and F_a^{-1} were known, then one could generate an analysis ensemble by simply applying the transform to the forecast ensemble:

$$h_n^a = F_a^{-1}\left(F_f\left(h_n^f\right)\right). \tag{3}$$

Typically neither F_f nor F_a are known a priori, so they need to be estimated from the ensemble and, in the case of F_a, from the observation value. Note that F_a^{-1} is the quantile function for the posterior distribution, and that the ensemble members retain their quantiles when moving from the forecast to the analysis. Ensemble filters that operate in this way have therefore recently been called 'Quantile-Conserving Ensemble Filters' (QCEFs [5]). The

action of a QCEF on a single ensemble member is illustrated in Figure 1. Figure 1 illustrates a QCEF where the prior and posterior cdfs have jump discontinuities; the appropriate definitions of these functions and their inverses in such a case is discussed in Appendix A.

This paper develops a new method for estimating the QCEF transform. The method uses kernel density estimation to approximate the prior probability density function (pdf). Extra care is taken with the kernel density estimation to handle random variables with bounded ranges, as well as to handle situations with a nonzero total probability on one or both boundaries. This estimate of the prior pdf is then integrated approximately using quadrature to produce an approximation of the prior cdf F_f. Bayes Theorem states that, up to a normalizing constant, the posterior pdf equals the prior pdf multiplied by the likelihood function. The new method therefore again uses quadrature to approximate the posterior cdf F_a; in addition to the approximation of the prior pdf, this only requires the ability to evaluate the likelihood function. Finally, a rootfinding method is applied to the equation

$$F_a(h_n^a) - F_f\left(h_n^f\right) = 0 \tag{4}$$

to solve for the analysis ensemble members h_n^a.

The details of the new, non-parametric QCEF method are described in Section 2. Section 3 compares the new method to the Bounded Normal Rank Histogram Filter (BNRHF; [6–8]) and to an EnKF in synthetic non-Gaussian DA problems where the posterior distribution is exactly known. Section 4 then compares these methods in the context of a assimilating observations of sea ice concentration (SIC) into a single-column sea ice model (Icepack). Conclusions are offered in Section 5.

2. Mixed delta-kernel distributions

This section describes the methods used to estimate the parameters in a mixed distribution of the form

$$p(h) = p_l\delta(h - h_l) + p_m p_{\text{int}}(h) + p_u\delta(h - h_u) \tag{5}$$

where δ denotes the Dirac delta distribution. Here the random variable H is assumed to take values in the range $h \in [h_l, h_u]$. The constants $p_{\{l,m,u\}}$ are the total probabilities at the lower bound h_l, in the interior (h_l, h_u), and at the upper bound h_u. As probabilities they are non-negative and sum to one. If $h_l = -\infty$ then $p_l = 0$, and if $h_u = \infty$ then $p_u = 0$. The subscript l, m, and u stand for lower, middle, and upper, respectively, and the subscript int on p_{int} stands for interior. We assume that p_{int} is a probability density function in its own right, in particular that it integrates to one.

Bounded variables appear in many applications. For example, rainfall is never negative, and concentration mixing ratios are between zero and one. Bounded variables also appear in the context of parameter estimation; for example microphysical parameters in numerical weather prediction models or albedo parameters in snow and ice models are bounded. Mixed distributions are also relevant to many applications; for example, one might make a forecast with a nonzero probability of having no rain, while the remaining probability is distributed across a continuous range of possible rainfall rates. These distributions are also relevant to sea ice data assimilation, as seen in Section 4.

We assume that the likelihood function is available, up to a positive multiplicative constant, i.e.

$$\ell(h; y) \propto p_{Y|H=h}(y) \tag{6}$$

where $p_{Y|H=h}(y)$ denotes the probability density of observing value $Y = y$ given that $H = h$, and the proportionality constant is positive and independent of h.

In the present context, Bayes' theorem takes the following form

$$p^a(h) = p_{H|Y=y}(h) \propto \ell(h; y)p(h) \tag{7}$$

where the proportionality constant is independent of h and serves only to guarantee that the total integrated probability is one. When the prior $p(h)$ is a mixed distribution of the form Equation (5), Bayes theorem implies that the posterior distribution will also be a mixed distribution, with posterior density

$$p^a(h) = p_l^a \delta(h - h_l) + p_m^a p_{\text{int}}^a(h) + p_u^a \delta(h - h_u). \tag{8}$$

The posterior class weights are given by

$$p_l^a = \frac{\tilde{p}_l^a}{\tilde{p}_l^a + \tilde{p}_m^a + \tilde{p}_u^a}, \quad \tilde{p}_l^a = p_l \ell(h_l; y) \tag{9a}$$

$$p_m^a = \frac{\tilde{p}_m^a}{\tilde{p}_l^a + \tilde{p}_m^a + \tilde{p}_u^a}, \quad \tilde{p}_m^a = p_m \mathbb{E}_{\text{int}}[\ell(h; y)] \tag{9b}$$

$$p_u^a = \frac{\tilde{p}_u^a}{\tilde{p}_l^a + \tilde{p}_m^a + \tilde{p}_u^a}, \quad \tilde{p}_u^a = p_u \ell(h_u; y) \tag{9c}$$

where $\mathbb{E}_{\text{int}}$ denotes expectation with respect to p_{int}, i.e.

$$\mathbb{E}_{\text{int}}[\ell(h; y)] = \int_{h_l}^{h_u} \ell(h; y) p_{\text{int}}(h) \mathrm{d}h. \tag{10}$$

The posterior probability density in the interior is

$$p_{\text{int}}^a(h) = \frac{\ell(h; y) p_{\text{int}}(h)}{\mathbb{E}_{\text{int}}[\ell(h; y)]}. \tag{11}$$

These expressions follow from Bayes' theorem Equation (7) where the proportionality constant, which is the denominator in Equation (9), is set to ensure unit total probability.

The remaining subsections of this section describe methods that use an ensemble h_n^f, $n = 1, \ldots, N_e$ to estimate the prior and posterior class weights and interior pdf, to approximate the application of the prior cumulative distribution function (cdf) to the ensemble, and finally to approximate the application of the posterior quantile function to produce an analysis ensemble h_n^a, $n = 1, \ldots, N_e$.

2.1. Class weights

The prior class weights are easy to estimate from the ensemble. Suppose that there are N_l ensemble members on the left boundary ($h_n^f = h_l$), N_m in the middle, and N_u on the upper boundary. Then the prior class weights are estimated as

$$p_{\{l,m,u\}} \approx \frac{N_{\{l,m,u\}}}{N_e}. \tag{12}$$

The only subtlety in the estimation of the posterior class weights lies in the approximation of the expected value of the likelihood function $\ell(h; y)$ over the interior distribution. We approximate this using a simple Monte-Carlo approximation. Suppose that $h_{n_j}^f$ $j = 1, \ldots, N_m$ are in the interior; then

$$\mathbb{E}_{\text{int}}[\ell(h; y)] \approx \frac{1}{N_m} \sum_{j=1}^{N_m} \ell\left(h_{n_j}^f; y\right). \tag{13}$$

The approximations given by Equations (12) and (13) are inserted into the definition Equation (9) to obtain approximations for the posterior class weights.

2.2. Interior probability density using kernel density estimation

Kernel density estimation (KDE) is a classic way of using an ensemble of realizations of a random variable to approximate the probability density function of the random variable [9]. Suppose that K is a function with non-negative output that integrates to unity, and that $h^f_{n_j}$, $j = 1, \ldots, N_m$ are independent realizations of a random variable H with pdf p_{int}. Then

$$p_{\text{int}}(h) \approx p^{\text{KDE}}_{\text{int}}(h) \frac{1}{N_m} \sum_{j=1}^{N_m} \frac{1}{w_j} K\left(\frac{h - h^f_{n_j}}{w_j} \right) \tag{14}$$

where $w_j > 0$ are called the bandwidths and K is the kernel. Kernel density estimation theory addresses the choice of kernel function K and bandwidths w_j that make the approximation accurate.

We here choose to use the Epanechnikov kernel

$$K(h) = \frac{3}{4}\left(1 - h^2\right)_+ \tag{15}$$

where $(\cdot)_+$ is the so-called Rectified Linear Unit that returns zero for negative arguments, and returns the argument for non-negative arguments. We choose this kernel primarily because it has compact support and is computationally inexpensive to evaluate, though it has some other optimality properties in the context of kernel density estimation [10].

We set the bandwidths w_j using the adaptive method described in [9, §5.3]. This method proceeds as follows. First we use the standard rule of thumb to set a non-adaptive bandwidth parameter

$$w_0 = \frac{2\sigma}{N_m^{1/5}} \tag{16}$$

where σ is the standard deviation of the interior ensemble [9, §3.4]. The value 2 in this equation is specific to the Epanechnikov kernel; other kernels have different constants (cf. [4]).

Next a 'pilot' estimate of the density is obtained using the nearest-neighbor method [9, §2.5]

$$\tilde{p}_{\text{int}}(h) = \frac{k}{2N_m d_k(h)} \tag{17}$$

where $d_k(h)$ is the distance from h to the k^{th} nearest neighbor in the ensemble. We set $k = \lfloor \sqrt{N_m} \rfloor$ where $\lfloor \cdot \rfloor$ rounds down to the nearest integer.

This pilot estimate of the density is then used to construct local bandwidth factors

$$\lambda_j = \left[\frac{\tilde{p}_{\text{int}}\left(h^f_{n_j}\right)}{g} \right]^{-1/2} \tag{18}$$

where g is the geometric mean of $\tilde{p}_{\text{int}}(h^f_{n_j})$. These local bandwidth factors are finally used to modulate the non-adaptive bandwidth w_0, producing adaptive bandwidths

$$w_j = w_0 \lambda_j. \tag{19}$$

2.3. Boundary corrections on the interior probability density

The foregoing kernel density estimate of the probability density function is a consistent approximation even when the data come from a bounded distribution, but the order of accuracy is significantly reduced near boundaries. There are many approaches to updating the kernel density approximation for bounded distributions. Two directions not pursued here include reflecting the data around the boundaries, and transforming the data from a

bounded to an unbounded distribution before performing kernel density estimation in the unbounded domain.

The approach taken here is to locally modify the kernel near the boundary [11,12]. We define the following two functions

$$l(t) = \frac{-64(-2+t(4+3t(t-2)))}{(1+t)^4(19+3t(t-6))}, \tag{20}$$

$$m(t) = \frac{240(t-1)^2}{(1+t)^4(19+3t(t-6))}. \tag{21}$$

(The argument t here is simply a dummy variable; it does not denote time.) The form of these functions depends on the choice of kernel function as described in [12, §1]; the expressions given above correspond to the Epanechnikov kernel. These functions are used to modify the j^{th} kernel in the approximation Equation (14) near the boundary. Specifically, if the j^{th} term is evaluated at a point h that is within a distance w_j of the lower boundary, then the kernel is multiplied by

$$l\left(\frac{h-h_l}{w_j}\right) + \left(\frac{h-h^f_{n_j}}{w_j}\right) m\left(\frac{h-h_l}{w_j}\right). \tag{22}$$

If the j^{th} term is evaluated at a point h that is within a distance w_j of the upper boundary, then the kernel is multiplied by

$$l\left(\frac{h_u-h}{w_j}\right) - \left(\frac{h-h^f_{n_j}}{w_j}\right) m\left(\frac{h_u-h}{w_j}\right). \tag{23}$$

This correction returns the accuracy of the approximation near the boundary to the same order of accuracy enjoyed by the approximation away from boundaries. In the absence of boundaries the approximation Equation (14) integrates to unity, as required for a probability density function. The correction breaks this property, so that the corrected approximation no longer integrates to unity. It must therefore be scaled by an appropriate factor, equal to the integral of the corrected approximation over the support of the pdf. This scaling factor (inter alia) is approximated using quadrature as described in the next subsection.

2.4. Cumulative distribution functions using quadrature and sampling

Applying the QCEF transform requires evaluating the approximations to the prior and posterior cumulative distribution functions (cdfs) described in the foregoing sections. For values in the interior, i.e. not on the boundaries, this is achieved using quadrature. Before describing the quadrature, we discuss the application of the prior cdf F_f to values on the boundary.

2.4.1. Boundary sampling

The formal definition of F_f as the cdf of the prior given by Equation (5) implies that

$$F_f(h_l) = p_l.$$

But with this formal definition the QCEF transform Equation (1) would not produce an analysis ensemble distributed according to the true posterior distribution. As detailed in Appendix A, in order for Equation (1) to work correctly for any pair of prior and posterior distributions of the form Equation (5) and Equation (8), F_f should map values on the lower boundary to a uniform distribution on $[0, p_l]$, and it should map values on the upper boundary to a uniform distribution on $[1-p_u, 1]$.

We achieve this as follows. First draw a single random number u from a uniform distribution on $[0,1]$, then set

$$v = \frac{u}{N_e}. \tag{24}$$

Suppose that we have N_l prior ensemble members on the lower boundary $h_l = h^f_{n_s}$, $s = 1, \ldots, N_l$. Then we set

$$F_f\left(h^f_{n_s}\right) = v + \frac{s-1}{N_e}. \tag{25}$$

Similarly, suppose that we have N_u prior ensemble members on the upper boundary $h_u = h^f_{n_t}$, $t = 1, \ldots, N_u$. Then we set

$$F_f\left(h^f_{n_t}\right) = 1 - \left(v + \frac{t-1}{N_e}\right). \tag{26}$$

Strictly speaking, $F_f(h)$ is an abuse of notation, because the map defined above is multi-valued, and is a function of an entire ensemble rather than a single argument, and is not deterministic. Ultimately what matters is that we have a map from the forecast ensemble h^f_n to a set of values that are uniformly distributed on $[0,1]$.

The question of how to define F_f for boundary values was addressed in [8] in the context of the bounded normal rank histogram filter (BNRHF). They chose to map all members on the boundary to the middle of the corresponding interval; e.g. ensemble members on the lower boundary map to the value $p_l/2$. On the one hand this preserves the property that identical ensemble members are all treated identically, where the approach taken here arbitrarily separates identical ensemble members. On the other hand, as will be seen in Section 3, this choice leads to an inability to correctly represent the posterior in certain cases.

2.4.2. Quadrature

For any $h \in [h_l, h_u]$, the kernel-density approximation to the interior prior cdf is

$$\int_{h_l}^{h} p^{\text{KDE}}_{\text{int}}(t)\mathrm{d}t. \tag{27}$$

The kernel approximation to the interior pdf $p^{\text{KDE}}_{\text{int}}$ is piecewise-smooth, with discontinuous derivatives at the points $h^f_{n_j} \pm w_j$, where w_j is the radius of kernel that is centered at $h^f_{n_j}$. We begin by defining breakpoints h_k that consist of the points $h^f_{n_j} \pm w_j$ and the points h_l and h_u (if they are finite). These points are sorted, and any points outside the range $[h_l, h_u]$ are discarded. The remaining points define a set of subintervals $I_k = [h_k, h_{k+1}]$. If $h \in I_{k+1}$ then we have

$$\int_{h_l}^{h} p^{\text{KDE}}_{\text{int}}(t)\mathrm{d}t = \sum_{i=1}^{k} \int_{I_i} p^{\text{KDE}}_{\text{int}}(h)\mathrm{d}h + \int_{h_k}^{h} p^{\text{KDE}}_{\text{int}}(t)\mathrm{d}t. \tag{28}$$

In the absence of boundary corrections these integrals can be evaluated exactly by integrating the Epanechnikov kernel. In the presence of boundary corrections it is possible, but tedious, to compute the integrals analytically; for the posterior interior cdf the integrand is multiplied by the likelihood function $\ell(h; y)$, so none of the integrals can be computed analytically. In these cases we approximate each of these integrals using Gauss-Legendre quadrature with five-points, which is exact for polynomials up to degree nine [13, §5.3]. The integrals pre-computed on each sub-interval using Gauss-Legendre quadrature to save computational cost, so that a single evaluation of the cdf requires only computing the final integral

$$\int_{h_k}^{h} p^{\text{KDE}}_{\text{int}}(t)\mathrm{d}t \tag{29}$$

and adding it to the sum of the pre-computed integrals. New integrals must be pre-computed for each new forecast, so the savings are not as great as they would be if it were possible to pre-compute values once and use them for all forecasts.

2.5. Application of the quantile function using rootfinding

The QCEF probability integral transform Equation (1) requires inverting the approximation to the posterior cdf. This is accomplished via rootfinding methods. We define the function

$$f(h) = F_a(h) - F_f\left(h_n^f\right). \tag{30}$$

If $f(h_n^a) = 0$ then $h_n^a = F_a^{-1}(F_f(h_n^f))$, as desired.

To obtain a first guess for the solution, we use the pre-computed values discussed in the previous section to form a piecewise-linear approximation to $f(h)$ between the points h_k. The root of this piecewise-linear approximation can be found analytically, and serves as the first guess supplied to the rootfinding algorithm.

The rootfinding method starts with the secant method [13, §2.3], which produces a sequence of approximations to the root. This continues until either (i) the tolerance (2^{-39}) is met, (ii) the max number of iterations (50) is reached, or (iii) two consecutive approximations bracket the root. If (iii) occurs, i.e. if two approximations of h_n^a are found that produce values of f with opposite sign, then the algorithm switches to an improved bisection method from [14] called the 'Interpolation, Truncation, and Projection (ITP) method,' which terminates either when the tolerance is met, or the max number of iterations is reached. The rootfinding problems that need to be solved to update the ensemble are independent of each other and can therefore be carried out in parallel, but the implementation used here does not make use of parallelism.

The worst-case performance of the combined secant-ITP method is that of the bisection method, i.e. the error reduces by a factor of 2 at each iteration (linear convergence). The method will therefore reach the tolerance before hitting the max number of iterations as long as the initial error is less than 2^{11}. The best performance comes from the case where the method never switches from secant to ITP; in this case the order of the method is between linear and quadratic, with exponent $(1+\sqrt{5})/2 \approx 1.62$ [13, Theorem 2.3]. Using the configuration recommended in [14], the best-case performance of the ITP method is convergence with order $\sqrt{2}$.

The computational bottleneck in a single iteration of the rootfinding method is the evaluation of the cdf F_a. If the integrals on the sub-intervals were not pre-computed as described in the foregoing section, then these integrals would have to be re-computed at every step of the rootfinding iteration, and for every ensemble member; the pre-computation thus leads to significant savings. With the pre-computation, the most expensive part of a single evaluation of the cdf is the computation of the Gauss-Legendre approximation to the integral on a single sub-interval Equation (29), which requires evaluating $p_{\text{int}}^{\text{KDE}}$ five times.

3. Non-cycling tests

This section compares three ensemble filters on three scalar problems where the answer is known analytically. The three algorithms are

- The Ensemble Adjustment Kalman Filter [EAKF; 15],
- The Bounded Normal Rank Histogram Filter [BNRHF; 7,8],
- The Kernel-based QCEF (KQCEF) developed in the preceding section.

The ensemble size is varied through $N_e = 20, 40, 60, 80$, and at each ensemble size the test is repeated 100 times. In all three problems the likelihood is normal

$$\ell(h; y) = e^{-\frac{1}{2}\left(\frac{h-y}{\gamma}\right)^2}. \tag{31}$$

The tests and their results are presented in the following subsections.

In each test we generate an ensemble drawn from the prior, apply the three filters above to the same prior ensemble, and use the Kolmogorov-Smirnov [KS; 16] test on the null hypothesis that the analysis ensemble is drawn from the known posterior distribution. A drawback of this (or any) test of the null hypothesis that the analysis ensemble is drawn from the true posterior distribution is that the null hypothesis can be rejected even when the analysis ensemble is drawn from a distribution that is close to the true posterior distribution. (Naturally, if the analysis distribution is close to the true posterior distribution, then it takes a large ensemble to confidently reject the null hypothesis.) The KS test could, for example, reject the null hypothesis in a situation where the analysis ensemble mean has very small error, but the shape of the analysis ensemble clearly does not reflect the true posterior distribution.

An alternative would be to directly estimate a distance (or divergence) between the analysis and true posterior distributions, e.g. using the Kullback-Leibler (KL) divergence. A strength of the KL divergence as compared to the KS test is that the KL divergence provides a number representing the error in the analysis distribution (whereas the KS test simple accepts or rejects the null hypothesis). A weakness of the KL divergence is that it can be hard to interpret, whereas the KS test has a very clear meaning. Like the KS test, sample-based estimates of the KL divergence are sensitive to ensemble size [17,18].

3.1. Normal Prior

In this problem the prior is Gaussian, with mean zero and unit standard deviation. The posterior is therefore also Gaussian, with known mean and variance. In addition to varying the ensemble size, we also vary the value y of the observation and the observation error standard deviation γ. If y is far from 0 then the observation is in the tail of the prior distribution. If $\gamma \ll 1$ then the observation error is small compared to the prior.

Each panel in Figure 2 shows the fraction of times that the KS test rejected the null hypothesis at the 5% significance level. The horizontal axis shows the value of y, while the vertical axis shows the logarithm of the observation error variance. Low values indicate that the analysis ensemble produced by the ensemble filter was frequently consistent with the true posterior distribution. The top row shows the EAKF results, the middle row shows the BNRHF results, and the bottom row shows the KQCEF results. The columns show the results for $N_e = 20, 40, 60$, and 80, from left to right.

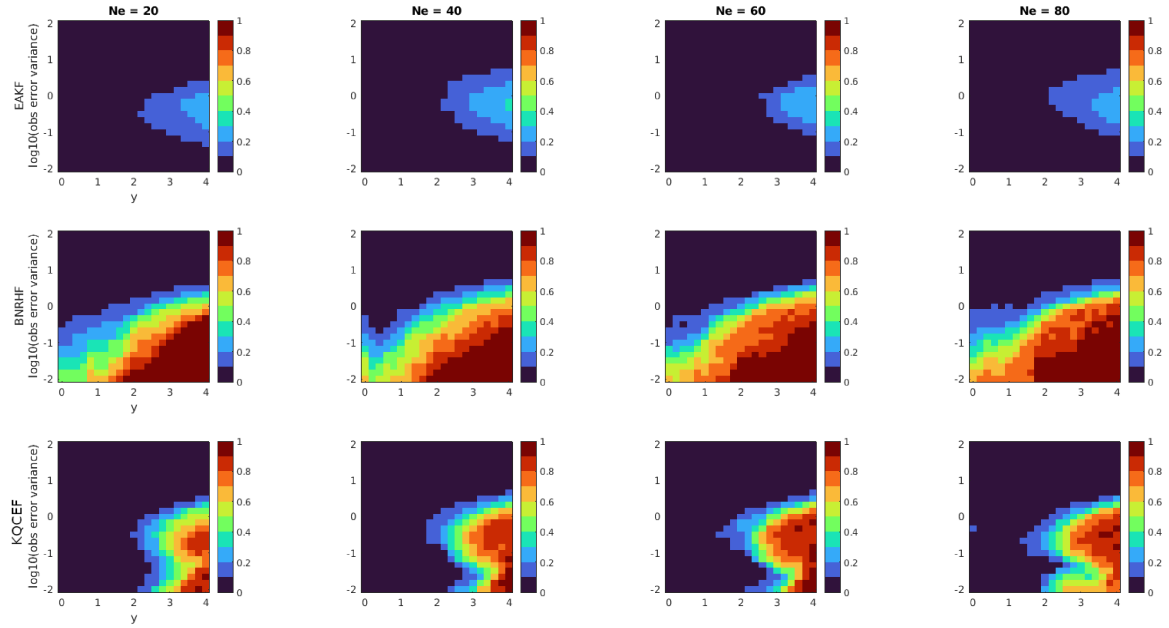

Figure 2. Each panel shows the fraction of 100 experiments where the null hypothesis that the analysis ensemble was drawn from the known true posterior distribution was rejected by the Kolmogorov-Smirnov test at the 5% significance level. The prior is a standard normal. The likelihood is normal with mean y that varies along the horizontal axis and variance γ^2 that varies along the vertical axis. The EAKF, BNRHF, and KQCEF results are in the upper, middle, and lower rows. The ensemble size varies across the columns from 20 to 80.

It is seldom the case in practice that the forecast comes from a known family of parametric distributions, and that the likelihood is conjugate to the prior. But in those cases, or in cases where the forecast distribution is well-approximated by a parametric distribution that is conjugate to the likelihood, one expects that a filter based on conjugate parametric distributions (cf. the methods discussed in [19] and [5]) will outperform non-parametric approaches like the BNRHF and KQCEF. That is indeed the case here: the prior and likelihood are both normal, which precisely matches the assumptions of the EAKF, and we see that the EAKF performs best of the three algorithms in this test. The only source of errors in the EAKF is sampling errors in estimating the prior mean and variance. These sampling errors decrease as the ensemble size increases, but so does the sensitivity of the KS test; combined, these effects lead to minimal changes in the accuracy across ensemble sizes. The EAKF performs well overall, but slightly less well when the observation is in the tail of the prior, and when the observation error variance is comparable to the prior variance.

All the methods perform well when the observation error variance is large compared to the prior variance, regardless of the value of y. This is reassuring, because the analysis ensemble should be only minimally changed compared to the forecast ensemble when the observation error variance is large.

When the observation value y is in the tail of the prior, neither BNRHF nor KQCEF performs well. Again, this is not surprising because the representation of the tails of the prior and posterior in the BNRHF and KQCEF relies on ensemble members in the tails, and there are, by definition, not many ensemble members in the tails.

The only benefit of the KQCEF over the BNRHF in this problem comes when the observation error variance is small compared to the prior variance, and the observation value is not too far into the tails of the prior. In such cases the KQCEF significantly outperforms the BNRHF. This is because the BNRHF approximates the likelihood by evaluating it at ensemble members and linearly interpolating between ensemble members; when the observation error is small compared to the prior spread, the gaps between

ensemble members can be so large that the piecewise-linear approximation of the likelihood is inaccurate. See Section 5.1 for further discussion of this issue.

3.2. Bi-normal Prior

In this problem the prior is a Gaussian mixture with two equally-weighted components, each of which has unit standard deviation. One component is centered at -2 while the other is centered at 2. The posterior is therefore also a Gaussian mixture, where the class weights, means, and variances are known functions of the observation value and observation error variance. As in the foregoing problem, in addition to varying the ensemble size, we also vary the value y of the observation and the observation error standard deviation γ.

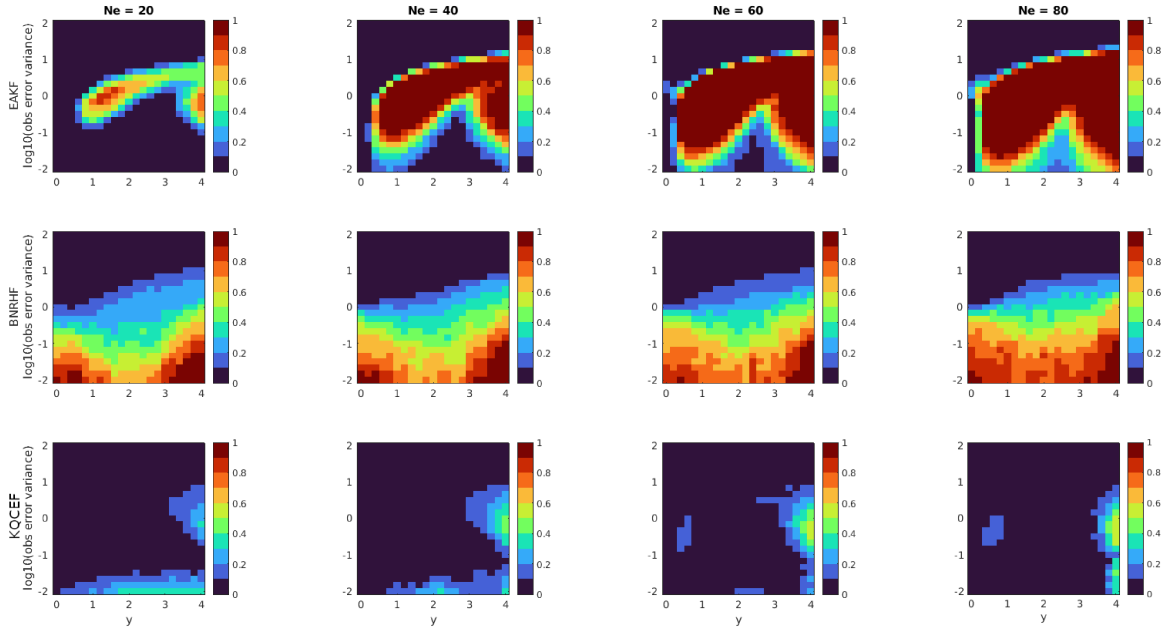

Figure 3. Each panel shows the fraction of 100 experiments where the null hypothesis that the analysis ensemble was drawn from the known true posterior distribution was rejected by the Kolmogorov-Smirnov test at the 5% significance level. The prior is an equal mixture of two normals, each with unit variance, with means at -2 and 2. The likelihood is normal with mean y that varies along the horizontal axis and variance γ^2 that varies along the vertical axis. The EAKF, BNRHF, and KQCEF results are in the upper, middle, and lower rows. The ensemble size varies across the columns from 20 to 80.

Each panel in Figure 3 shows the fraction of times that the KS test rejected the null hypothesis at the 5% significance level. The arrangement of the panels is the same as Figure 2. The KQCEF performs best of the three algorithms in this test. As in the normal problem, the errors are largest (but still not too large) when the observation is in the tail of the prior, and the observation error variance is close to one. The errors are smaller here than in the normal problem because an observation value of $y = 4$ is only two standard deviations from one of the Gaussian modes of the prior, whereas in the normal problem an observation value of $y = 4$ is four standard deviations from the mean of the prior. There are also slightly elevated error levels when the observation error variance is small and the ensemble size is small. These appear presumably because the ensemble members are being split between the two modes of the prior, and each mode is poorly resolved.

The BNRHF performs well on this problem only when the observation error variance is large, in which case all the methods perform well. When the observation error variance is small the BNRHF performs worse than the EAKF. In this case the problem exhibits what some authors have called 'medium nonlinearity' [20,21]: The prior is very non-Gaussian but the posterior is approximately Gaussian. In such cases the EAKF effectively ignores the

prior and moves all of the ensemble members close to the observed value y. The BNRHF uses an approximation of the likelihood function and posterior pdf that is piecewise linear between the ensemble members. If the likelihood is so tight that the structure of its peak falls between ensemble members, then the BNRHF cannot correctly represent the posterior pdf, which leads to errors. See Section 5.1 for further discussion of this issue.

When the observation error variance is close to the prior variance, the EAKF performs very poorly. But more importantly, the EAKF errors *increase* as the ensemble size increases. This is because the EAKF uses an affine transformation to update the ensemble members, and this affine transformation preserves the equally-weighted bimodal structure of the prior. As the ensemble size increases the KS test becomes more sensitive and can tell the difference between the unequally-weighted bimodal structure of the true posterior and the equally-weighted bimodal structure produced by the EAKF.

3.3. Mixed prior

In this problem the prior is a mixed distribution on $[0, 1]$ consisting of delta distributions at $h = 0$ and $h = 1$ each with probability p_0 and a truncated normal in the interior with mean $1/2$ and variance (of the un-truncated normal) $1/16$. The observation error variance is fixed at $\gamma^2 = 1/64$, and the observation value y is varied between zero and one. The posterior is therefore also a mixed distribution with delta distributions on the boundaries and a truncated normal in the interior, and all of the class weights and truncated-normal parameters can be computed in terms of the parameters of the prior and likelihood.

The EAKF can produce analysis ensemble members outside the range $[0, 1]$. It is possible to update the EAKF so that it respects the bounds, e.g. by truncating values outside the boundary back to the boundary. The ad-hoc updates amount to an unfair comparison with the more sophisticated BNRHF and KQCEF though, so we refrain from testing the EAKF on this problem, and compare only the BNRHF and KQCEF.

The value of p_0 controls the initial number of ensemble members on the boundaries. For $p_0 = 0$ there are no forecast ensemble members on the boundaries, and neither BNRHF nor KQCEF will move ensemble members onto the boundary when there are none there initially, regardless of the value of y. The test with $p_0 = 0$ simply tests the ability of BNRHF and KQCEF to handle cases with bounded variables.

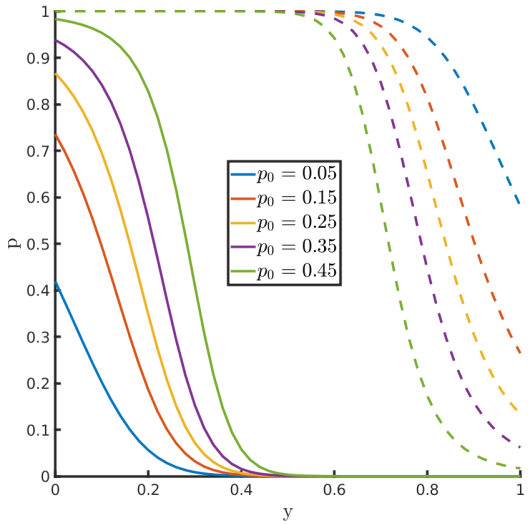

Figure 4. Division of the total posterior probability between the left boundary, interior, and right boundary as a function of the observation value y for the test with mixed and bounded prior. Each color corresponds to the prior probability on each boundary. The solid line indicates the posterior probability on the left boundary; the distance between the solid and dashed lines indicates the posterior probability in the interior; and the distance from the dashed line to 1 indicates the posterior probability on the right boundary.

As p_0 increases from zero, the number of forecast ensemble members on the boundaries increases. For a fixed positive value of p_0, the value of y controls how many ensemble members move off of or onto each boundary. This is illustrated in Figure 4. The value y varies along the horizontal axis. At a fixed value of y, the posterior probability for each class (lower boundary, interior, upper boundary) is indicated by the lines, which divide the vertical axis between the three classes. Each color corresponds to a different value of p_0; the distance below the solid lines indicates the posterior probability on the left boundary; the distance between the solid and dashed lines indicates the posterior probability in the interior; and the distance above the dashed lines indicates the posterior probability on the right boundary. For example, when $y = 1/2$ we see that the posterior probability is almost entirely in the interior for all values of p_0, meaning that all ensemble members on the left and right boundaries should all move to the interior when $y = 1/2$.

As another example, suppose that $p_0 = 0.05$, meaning that when $N_e = 20$ the prior ensemble has one member on each boundary. If $y = 0$ then the posterior probability on the left boundary is approximately 0.4, so we expect to have approximately 8 posterior ensemble members on the left boundary. In this case the posterior probability on the upper boundary is near zero, so the remaining 12 posterior ensemble members should be in the interior.

In this problem we test two aspects of the posterior separately. The analysis ensemble produced by each filter allocates a certain number of ensemble members to each class (left boundary, interior, right boundary). This distribution of ensemble members between classes should be statistically indistinguishable from a draw from a multinomial distribution, so we begin by testing the null hypothesis that the analysis ensemble, condensed to just numbers-per-class, is a draw from the multinomial distribution whose parameters correspond to the true posterior.

Figure 5 shows the fraction of times that the multinomial test rejected the null hypothesis at the 5% significance level. The horizontal axis shows the value of y, while the vertical axis shows the value of p_0. The KQCEF performs better than the BNRHF in this test. At each ensemble size there are combinations of p_0 and y where BNRHF fails the multinomial test 100% of the time. The reason for these failures is that the BNRHF moves ensemble members off the boundary all at once or not at all. For some combinations of p_0 and y, the true posterior requires some of the ensemble members on the boundary to move off the boundary and into the interior. For other combinations of p_0 and y, the true posterior requires some of the ensemble members on the boundary to move into the interior and some onto the other boundary. For these combinations of p_0 and y the analysis ensemble produced by the BNRHF either moves all or none of the members off the boundary, which is different enough from the true posterior that it fails the test, especially when the prior ensemble members on the boundary should be split in half. As the ensemble size grows the difference between half and all (or between half and none) grows, making the error more egregious. The KQCEF has some difficulty in similar regions of parameter space, but is still very accurate.

We also use the KS test to determine the accuracy of the distribution of the interior ensemble members. The true posterior distribution in the interior is a truncated normal with known parameters, and we use the KS test to test the null hypothesis that the analysis ensemble members in the interior (regardless of how many there are in the interior) are drawn from the true posterior distribution in the interior. The results are shown in Figure 6. The problem is particularly difficult when p_0 is large, because in this case there are many prior ensemble members on the boundaries, and few in the interior. BNRHF and KQCEF both use the interior ensemble members to estimate the interior distribution, and when there are few interior ensemble members the approximation of the interior distribution is inaccurate. This is reflected in the results, particularly near $y = 1/2$, because when y is near the middle of the interval the posterior probability is almost entirely in the interior (cf. Figure 4). On the other hand, for values of p_0 and y where the posterior has very few

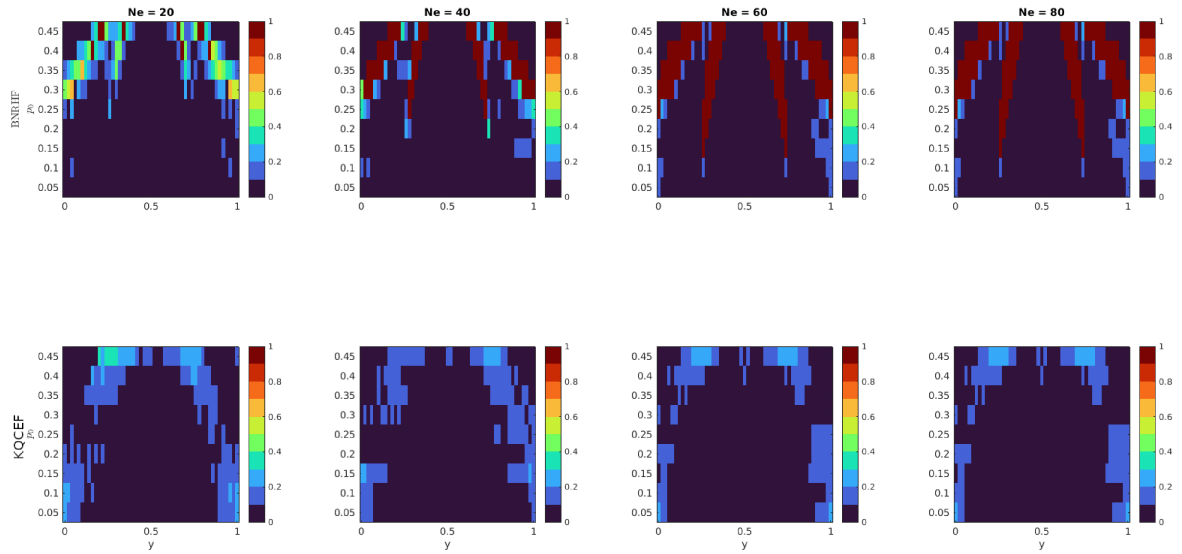

Figure 5. Each panel shows the fraction of 100 experiments where the null hypothesis was rejected at the 5% significance level. The null hypothesis is that the number of posterior ensemble members on the left and right boundaries and in the interior is consistent with a trinomial distribution whose parameters are the known true probabilities on the boundaries and in the interior. The prior is a mixed distribution with probability p_0 on both the left and right boundaries and a truncated normal prior in the interior. The value of p_0 varies along the vertical axis. The likelihood is normal with mean y that varies along the horizontal axis and variance $\gamma^2 = 1/64$. The BNRHF and KQCEF results are in the upper and lower rows, respectively. The ensemble size varies across the columns from 20 to 80.

ensemble members in the interior (e.g. for large p_0 and for y near one of the boundaries), both filters perform well according to this test. The reason for this is that the sensitivity of the KS test depends on the sample size; in situations where there are very few analysis ensemble members in the interior it is hard to tell whether they are distributed correctly or not.

Overall the KQCEF performs significantly better than the BNRHF in this test. This can again be traced back, at least in part, to the fact that the BNRHF moves all the boundary ensemble members together or not at all. When p_0 is large there are many prior ensemble members on the boundaries. If y is near $1/2$, then BNRHF correctly moves the boundary members into the interior, but it moves all the members from the left boundary to one location in the interior, and all the members from the right boundary to another location in the interior. The prior distribution had delta distributions on the boundaries, and the analysis produced by BNRHF incorrectly moves these delta distributions into the interior. The KS test notices these jumps in the interior, and flags the BNRHF analysis distribution as being significantly different from the true posterior. That said, the KQCEF also outperforms the BNRHF even when $p_0 = 0$, i.e. there are no ensemble members on the boundary, but in this case the BNRHF performs better than it does when $p_0 > 0$. The EAKF was not tested on this problem, but it also moves all members from the boundary to the same location in the analysis, and therefore suffers the same problem (in addition to the problem of not respecting the bounds).

4. Application to sea ice concentration

Assimilation of satellite observations of sea ice provides an excellent example of a data assimilation problem with bounded, non-Gaussian distributions. We follow [22] in using the single-column version of the CICE sea-ice model, called Icepack [23], to explore how advances in nonlinear, non-Gaussian data assimilation algorithms might lead to improvements in sea ice data assimilation.

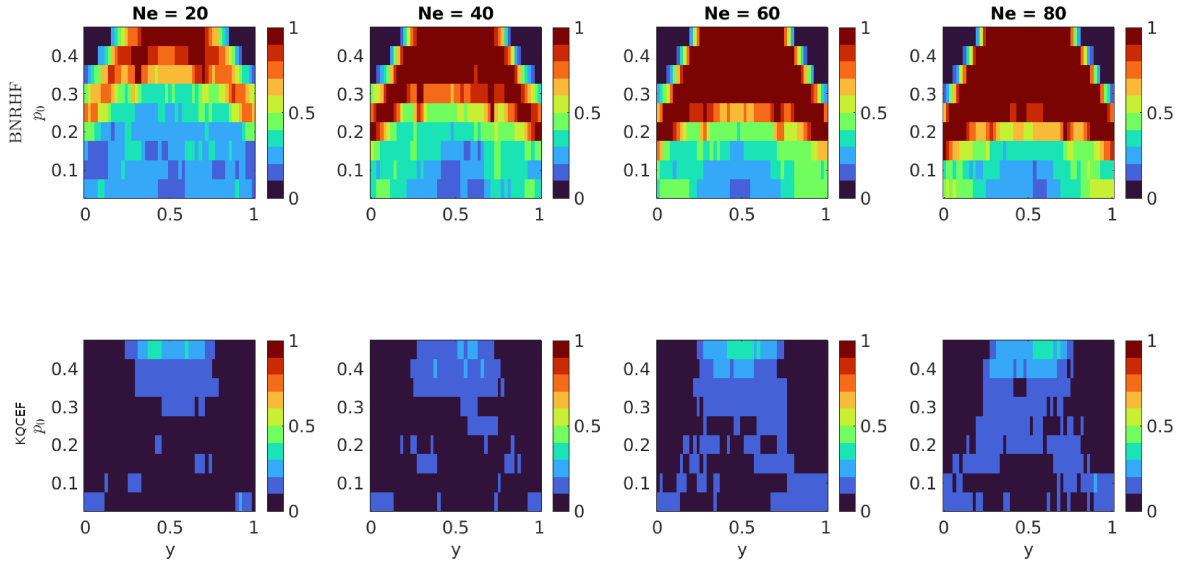

Figure 6. Each panel shows the fraction of 100 experiments where the null hypothesis that the analysis ensemble in the interior was drawn from the known true posterior distribution in the interior was rejected by the Kolmogorov-Smirnov test at the 5% significance level. The prior is a mixed distribution with probability p_0 on both the left and right boundaries and a truncated normal prior in the interior. The value of p_0 varies along the vertical axis. The likelihood is normal with mean y that varies along the horizontal axis and variance $\gamma^2 = 1/64$. The BNRHF and KQCEF results are in the upper and lower rows, respectively. The ensemble size varies across the columns from 20 to 80.

The model represents sea ice in a single grid cell. Within this grid cell there are five categories of ice thickness, not counting open water (zero thickness). For each category the model tracks a range of variables, notably including the fraction of the grid cell area that is covered by ice within that category, as well as the total volume of ice in that category. The mean thickness within a category is given by the volume divided by the area fraction, and the mean thickness lies between the thickness bounds that define the category. Each ice category is covered by some amount of snow, and the model tracks the total volume of snow covering each category; the snow thickness covering an ice category is equal to the snow volume covering that category divided by the area fraction of the category. The variable of interest here is the total sea ice concentration (SIC), equal to the sum of the area fractions of the five ice categories; it is also equal to one minus the fractional area of the grid cell that has no ice cover (i.e. open water). Although we investigate the method in the context of the CICE model, the approach developed here could be applied to any model that has SIC as a state variable, e.g. the SIS [24] and neXtSIM [25] models.

4.1. SIC Likelihood

There are several satellites whose sensors return data relevant to SIC, and observations of SIC have been assimilated into sea ice models for decades [26–34]. Assimilating data from any particular satellite requires an observing model appropriate to that satellite; we introduce such a model here. We assume that as the satellite tracks across a single grid cell (corresponding to the Icepack model) it makes N_o measurements at points within the grid cell. At each point it observes whether there is ice or open water; k of the N_o points have open water. This model is relevant to, for example, the laser altimeters aboard the ICESat-2 platform [35]. ICESat-2 provides more information than just ice vs no-ice; assimilation of sea ice thickness observations is an important topic, but out of scope for the present investigation.

We model the observations as a sample from a binomial distribution with parameters a_0 and N_o, where a_0 is the open water area fraction within the grid cell, i.e. one minus the

SIC. The likelihood function is therefore proportional to a beta distribution with parameters $k+1$ and $N_o - k + 1$:

$$\ell(a_0; N_o, k) \propto a_0^k (1 - a_0)^{N_o - k}. \tag{32}$$

The likelihood function is all that the BNRHF and KQCEF filters need to assimilate the observational data; the EAKF requires a Gaussian approximation of the likelihood. The likelihood, when viewed as a function of a_0 has mean

$$y = \frac{k+1}{N_o + 2} \tag{33}$$

and variance [36, §25.3]

$$\gamma^2 = \frac{(k+1)(N_o - k + 1)}{(3 + N_o)(2 + N_o)^2}. \tag{34}$$

We provide these values to the EAKF as the observation value and observation error variance. Note that in the limit $N_o \to \infty$ the observation error variance goes to zero. As described below, we use either $N_o = 10$ or $N_o = 100$. The former produces a larger observation error variance, while the latter produces a smaller observation error variance. The Gaussian approximation of the likelihood is accurate in the configuration with $N_o = 100$. (The red line in the right panel of Figure 10 shows the likelihood function when $N_0 = 100$ and $k = 30$.)

4.2. Experimental Design

We perform data assimilation over a single year (2011) at 75.54°N, 174.45°E. This location is between the East Siberian and Chuckchi Seas; unlike areas nearer the center of the Arctic icepack, this area experiences large changes in SIC over the course of the year. Each ensemble member is forced by atmospheric conditions from distinct members of the CAM6 + DART reanalysis [37]. Each ensemble member includes a slab ocean within Icepack. The initial conditions and lateral heat flux convergence forcing are the same for all ensemble members; they are derived from the ocean component output of a fully-coupled historical simulation of the Community Earth System Model [38].

Even with each member being forced by a different atmosphere, the ensemble spread is small, as the dynamics are not chaotic, and variability associated with lateral processes is missing in the single-column model. To introduce greater ensemble spread, three snow parameters of the Icepack model are perturbed:

- The `R_snw` nondimensional snow-albedo parameter is varied between -2 and 0
- The `rsnw_mlt` melting snow-grain radius parameter is varied between 1.3×10^{-3} and 2.3×10^{-3} m
- The `ksno` parameter determining the thermal conductivity of snow is varied between 0.2 and 0.35 W/m/degree.

The left panel of Figure 7 shows the SIC for free-running (no DA) ensemble from July through October. Outside of this window the SIC is near 100% for all ensemble members. The icepack begins to melt in late July, and the freeze-up occurs between late September and early October, depending on the ensemble member. The combination of perturbed snow parameters and different atmospheric forcing produces a wide range of minimum values for SIC, from less than 10% to more than 90%.

There are 80 ensemble members total because there 80 distinct atmospheric forcing conditions in the CAM6 + DART reanalysis. The black line denotes the reference member that is used to produce synthetic observations in an observing system simulation experiment (OSSE). We run data-assimilating experiments with two ensemble sizes: 20 and 79; the latter is the largest ensemble possible when reserving one of the 80 total members to be the reference case. To reassure the reader that the 20 members of the smaller ensemble span the full range of behaviors as seen in the full ensemble, the members of the ensemble of size 20 are shown with slightly thicker lines in the left panel of Figure 7.

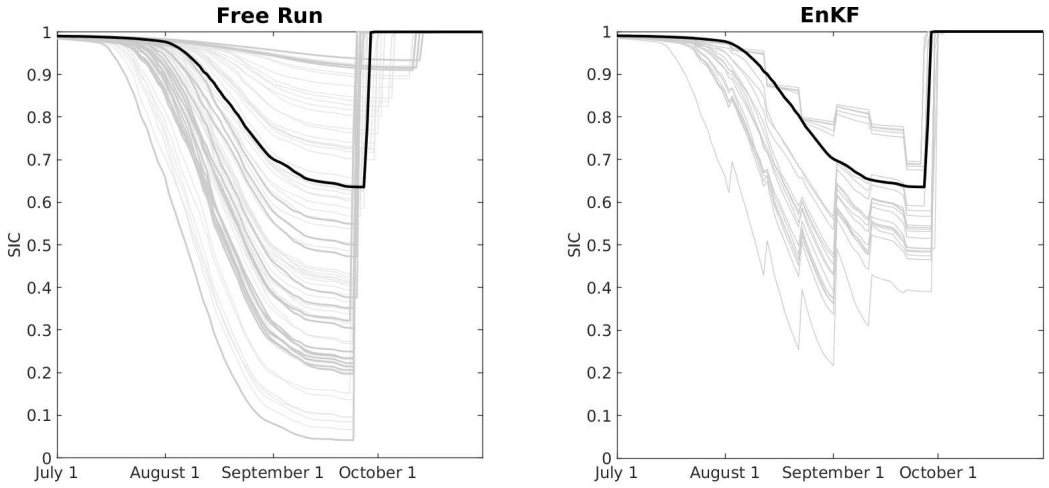

Figure 7. Left panel: SIC as a function of time for a free-running ensemble of 79 members (grey) and a reference member (black). Right panel: SIC as a function of time for an ensemble of size 20 using the EAKF to assimilate observations of SIC every 10 days.

The right panel shows the results from an EnKF run with a 20 member ensemble assimilating $N_o = 10$ observations every 10 days. When all of the ensemble members have SIC close to 1, i.e. at the beginning and end of the plot, as well as in the time period excluded from the plot, the analysis has very little impact on the ensemble. This is because the forecast ensemble spread is very small compared to the uncertainty in the observations, so the observations are effectively ignored. Ignoring the observations in this context is not necessarily a bad thing, because the ensemble forecast is very close to the reference value anyways.

Ice-observing satellites have a range of repeat periods; for example, ICESat-2 has a 91-day repeat period, and CryoSat-2 has a 369 day repeat period, but provides uniform coverage of the Arctic approximately every 30 days. With a 30 day observation interval we would have only 2 or 3 assimilation cycles within the window of time where SIC is dynamically changing, and it would be hard, on the basis of only 2 or 3 assimilation cycles, to discover statistically robust differences between the performance of different filters. We use a 10 day window as a compromise. Although no single satellite is expected to pass over a grid cell every 10 days, there are multiple satellites observing, and in a full model of the Arctic there would be observations of nearby grid cells that would affect the grid cell that is modeled here using Icepack.

With a 10 day window there are approximately 7 analysis cycles in the period between the beginning of the summer melt and the fall freeze-up. Since this is still too few cycles to obtain robust statistics, for each experimental configuration (described below) we run 10 experiments, where each experiment shifts the day of the first assimilation cycle by one day. In this way we end up with data from assimilation cycles on every day of the year while maintaining a 10 day window between cycles.

4.2.1. Experimental Configurations

We use two sets of ensemble sizes: $N_e = 20$ and $N_e = 79$. The 79-member ensemble results from the fact that we only have 80 different atmospheric forcings from the CAM6+DART reanalysis, and one of them serves as the reference. We assimilate either $N_o = 10$ or $N_o = 100$ observations at each cycle. All experiments with $N_o = 10$ use a single set of observations, and all experiments with $N_o = 100$ use a single set of observations.

We use a two-step assimilation cycle. In the first step we use EAKF, BRNHF, or KQCEF to assimilate the open water fraction a_0. Whenever EAKF produces an a_0 outside of $[0, 1]$ we map the offending value back to the nearest edge of the allowable interval.

In the second step we update the area fraction and ice volume for each ice category in the same, simple way: If a_0^- is the open water fraction in the forecast and a_0^+ is the open water fraction after assimilation, then the area fraction for each category is multiplied by $(1 - a_0^+)/(1 - a_0^-)$. I.e. if the open water fraction increases then the area fractions for all of the ice thickness categories decrease by a fixed factor, and vice versa. The ice thickness and snow thickness for each category are not changed by the analysis. More sophisticated methods for updating the ice and snow thicknesses are important research topics (cf. [22]); the method used here is designed to put all of the models on an even footing, and to focus attention on the first step - the update in observation space.

4.3. Results

With three filters, two observing systems ($N_o = 10$ or 100), two ensemble sizes ($N_e = 20$ and 79), and 10 experiments for each configuration, we eventually run 120 years of cycling data assimilation. To assess the performance of the filters, we compute the continuous ranked probability score [CRPS 39,40] using the forecast and analysis ensembles at each assimilation cycle. The CRPS measures the accuracy of the probabilistic estimate provided by the ensemble; lower values are better. Specifically, the CRPS is the L^2 norm of the difference between the empirical cdf associated with the ensemble and a Heaviside step function centered on the reference value. The best possible CRPS is therefore zero, which can only be attained if all of the ensemble members are equal to the reference value. It bears noting that the true Bayesian posterior would not have a CRPS of zero except in the very rare situation where the true posterior is a Dirac delta centered on the reference value. CRPS values have the same units as the quantity of interest, in this case SIC. Thus, a CRPS of 0.001 corresponds to a 0.1% SIC error, which is relatively small.

For a single experimental configuration (filter, N_o, N_e) we compute 58 CRPS values from August 1 through September 28. (Outside this window there is little difference in performance between the filters.) These 58 values come from different experiments: One experiment has assimilation cycles on August 1, August 11, etc., while another experiment has assimilation cycles on August 2, August 12, etc.

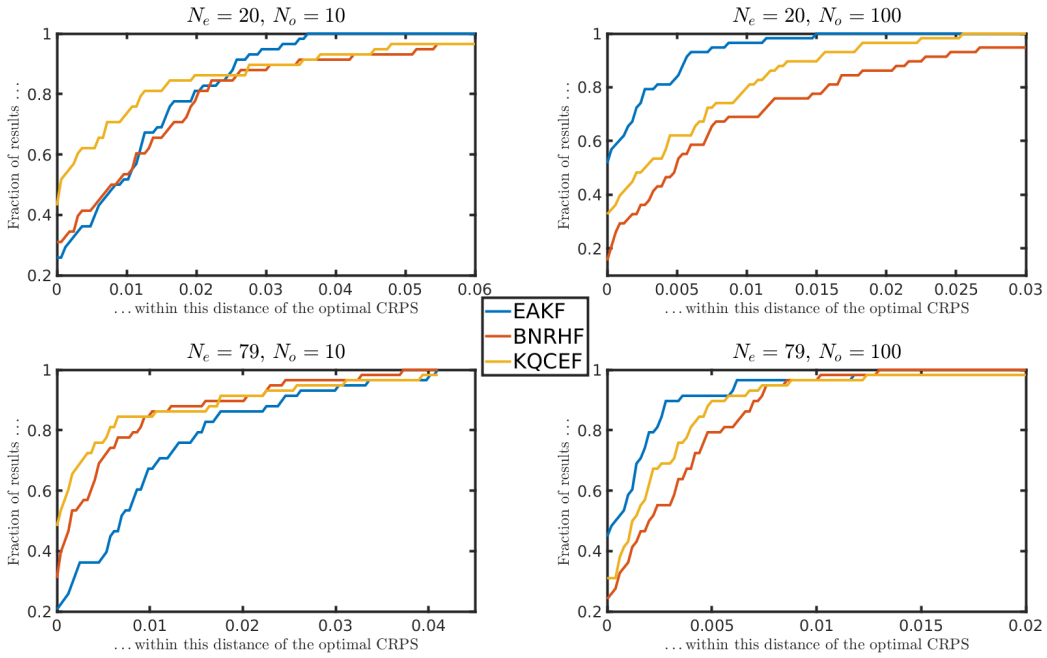

Figure 8. Analysis CRPS performance profiles, as described in the text.

For each experimental configuration we effectively have assimilation cycles on 58 different days. On some days EAKF outperforms the other two filters, on other days BNRHF outperforms the other two, etc. To meaningfully compare these filters we use a

performance profile plot. Each panel of Figure 8 corresponds to a different combination of ensemble size N_e and observation number N_o. The left edge of the panel shows, for each filter, the fraction of the time that it achieved the best CRPS out of the three filters. For example, in the upper left panel ($N_e = 20$, $N_o = 10$), the KQCEF produced the best analysis CRPS about 40% of the time, while the EAKF and BNRHF each produced the best CRPS about 30% of the time. Naturally it doesn't mean much to be second-best (or third-best) if the difference between the results is small. The performance plot takes this into account by adding a tolerance. In the same panel ($N_e = 20$, $N_o = 10$), if we look at the value 0.01 on the horizontal axis, we see that the analysis CRPS produced by the KQCEF was within 0.01 of the best result about 70% of the time, while the EAKF and BNRHF were both within 0.01 of the best CRPS about 50% of the time. The left side of each panel shows how often each filter is close to the optimal filter. The right side of the panel shows how often each filter is far from optimal.

4.3.1. Analysis

The results of Figure 8 can be summarized as follows. The upper left panel shows that when the observation error variance is relatively large ($N_o = 10$) and the ensemble is small ($N_e = 20$), the KQCEF analysis is usually close to optimal. EAKF is slightly more robust in the sense that the analysis CRPS produced by EAKF is always within about 0.35 of the best overall, whereas the analysis CRPS produced by KQCEF is sometimes farther from optimal. In this panel BNRHF performs worst of the three filters.

The lower left panel shows that when the observation error variance is relatively large ($N_o = 10$) and the ensemble is large ($N_e = 79$), the BNRHF and KQCEF are similar to each other, and both better than EAKF. Both of the right panels show that when the observation error variance is relatively small ($N_o = 100$), the EAKF is better than the KQCEF, which is better than the BNRHF. In these right panels the overall CRPS values are small though, so all three filters are performing well overall.

There is an apparent contradiction between the results shown here for small observation error variance and the results from section Section 3. Here the EAKF outperforms the other filters when the observation error variance is small (large N_o) and the ensemble size is large, whereas in Section 3.2 the ranking was KQCEF (best), EAKF, BNRHF (worst). These results can be reconciled by noting the different way in which performance is measured in these experiments. In Section 3.2 the null hypothesis might be rejected because the analysis ensemble has the wrong shape despite being closely centered on the true value, whereas in this section performance is measured using CRPS, which is small when the analysis ensemble is tightly clustered about the true state.

4.3.2. Forecast

Figure 9 is the same as Figure 8 except that Figure 9 shows the CRPS for the ensemble forecast on the 10^{th} day, i.e. at the longest lead time, when the errors are largest. The left panels of Figure 9 shows that when the observation error variance is relatively large ($N_o = 10$), the BNRHF and KQCEF both perform similarly, and are both significantly better than the EAKF. The right panels show that the KQCEF is the best filter for small observation error variance ($N_o = 100$), though the difference compared to the other two filters is reduced at larger ensemble sizes.

These results tell a slightly different story than the analysis CRPS results; in particular, the EAKF is never the best filter overall when considering forecast CRPS. In particular, although the EAKF produces the best analysis when the observation error variance is small, the analysis ensemble error grows faster for the EAKF than for the other filters, so that KQCEF produces a better forecast.

5. Discussion

We have introduced a new approach, based on kernel density estimation and quadrature, to approximating the cdf and quantile function needed to implement a QCEF. Previous

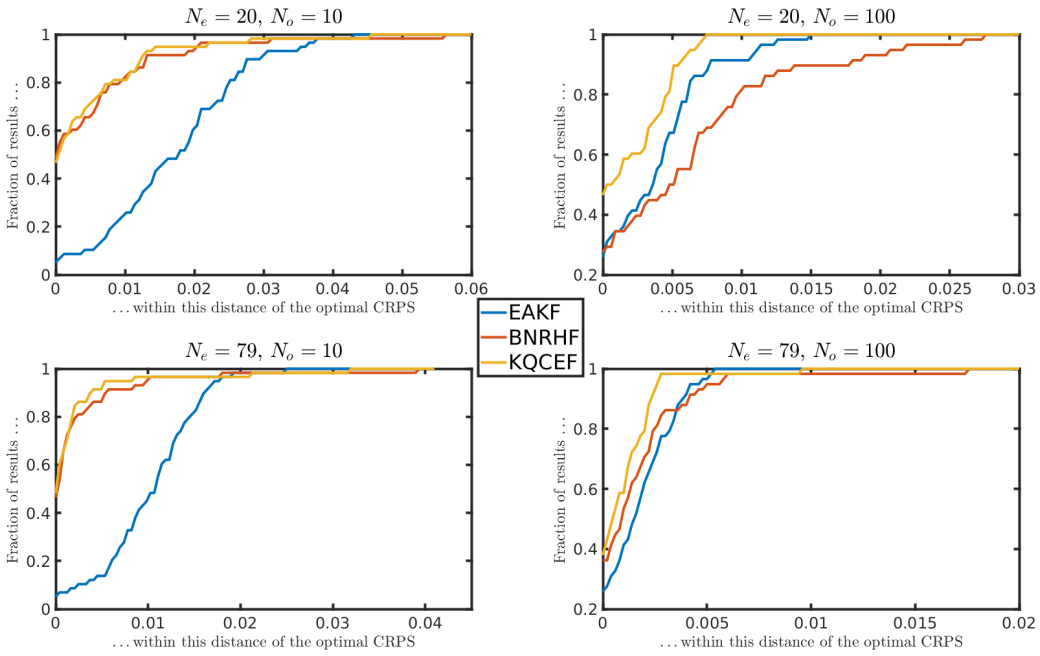

Figure 9. Forecast CRPS performance profiles, as described in the text.

QCEFs include the Rank Histogram Filter and Bounded Normal Rank Histogram Filter [6–8] and the improved Rank Histogram Filter [iRHF 4]. The iRHF also used kernel density estimation; the approach developed here improves on the iRHF by using a quadratic kernel rather than a piecewise-constant kernel, by enabling treatment of bounded distributions, by enabling treatment of mixed distributions with nonzero total probability on the boundaries, and by using quadrature to approximate the cdfs (whereas the iRHF interpolated the likelihood function and then integrated analytically).

The results of Section 3.1 show that non-parametric QCEFs like the BNRHF and KQCEF do not improve on EnKFs when the underlying distributions are Gaussian. Not only do they not improve, they can in fact perform worse than the EAKF: Both BNRHF and KQCEF perform worse than EAKF when the observation is in the tail of the prior, and BNRHF also performs worse than EAKF when the observation error variance is much smaller than the prior variance. This is an example of a broader principle that filters based on parametric distributions – e.g. ensemble Kalman filters and filters based on Gamma and Inverse Gamma distributions [19] – should be expected to outperform their nonparametric counterparts whenever the forecast and likelihood approximately match the assumptions of the parametric filters.

On the other hand, the results of Sections 3.2 and 3.3 show that BNRHF and KCQEF perform significantly better than EAKF in non-Gaussian problems, as expected. In these problems the KQCEF performs significantly better than BNRHF, especially when the observation error variance is small compared to the prior variance. The treatment of mixed distributions by KQCEF was shown to be better than BNRHF in Section 3.3, but this could be easily amended within the implementation of BNRHF. One potential advantage of BNRHF over KQCEF that was not explored here is that KQCEF can only handle mixed distributions with delta distributions on the boundaries, whereas BNRHF can accommodate delta distributions anywhere within the distribution [8]. A significant caveat on the results of Sections 3.1–3.3 is that the filter performance is quantified using the Kolmogorov-Smirnov test with the null hypothesis that the analysis ensemble is drawn from the known true posterior distribution. This is a stringent test; failing this test does not necessarily imply that the root mean squared error would be high, for example.

We applied EAKF, BNRHF, and KQCEF in an idealized model for data assimilation of satellite observations of sea ice concentration. Broadly summarizing the results:

- Both QCEF methods (BNRHF and KQCEF) outperform EAKF when the problem is non-Gaussian,
- KQCEF and BNRHF produce similar results when the ensemble size is high, but KQCEF tends to produce better results than BNRHF when the ensemble size is small, and
- The gap between KQCEF and BNRHF is largest when the ensemble size is small and the observation error variance is also small compared to the prior variance.

5.1. Small ensembles and small observation errors

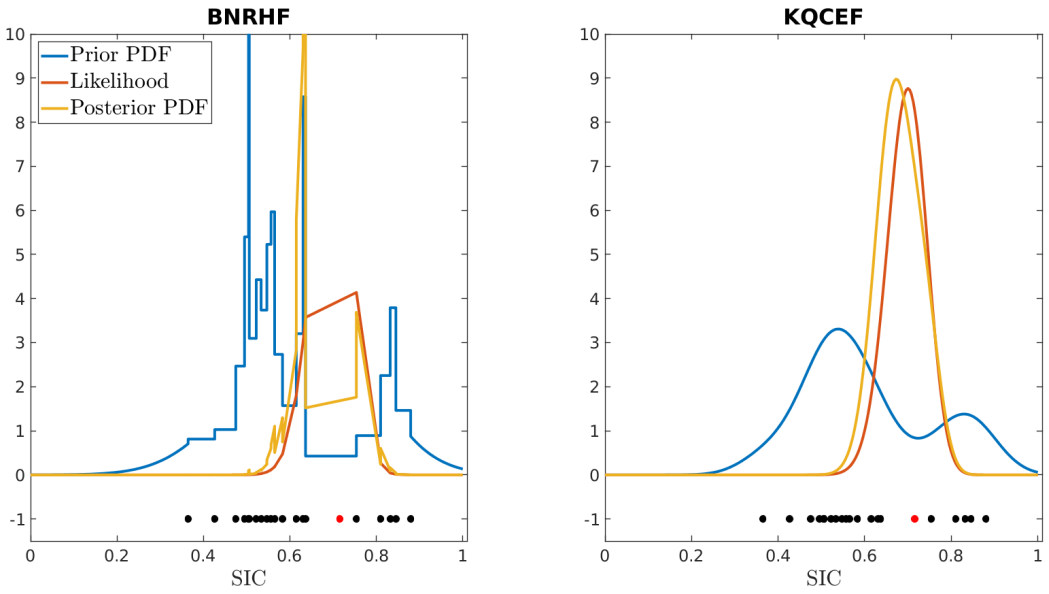

Figure 10. Forecast (blue) and analysis (yellow) pdfs of SIC for the BNRHF (left) and KQCEF (right) filters. The 20 ensemble members in both plots are drawn from a BNRHF forecast for September 27, 2011; they are shown as black dots beneath the pdfs. The reference value used to generate the observation is shown as a red dot beneath the pdfs. The true likelihood is shown in red in the right panel; it corresponds to an observation with $N_o = 100$ and $k = 30$. The piecewise-linear approximation to the likelihood used by the BNRHF is shown in red in the left panel.

The difference in accuracy between KQCEF and BNRHF when the observation error variance is small compared to the prior variance can be seen in both the idealized tests of Section 3 and Section 4. It can be traced back to the difference in how the likelihood function is treated in the two filters. The BNRHF only evaluates the likelihood function at ensemble members, so if the structure of the likelihood falls mainly between ensemble members then the BNRHF will not resolve it. In contrast the KQCEF evaluates the likelihood with infinite resolution, i.e. during the root finding phase of the algorithm it evaluates the likelihood wherever needed to obtain an accurate result.

Figure 10 illustrates this using a 20-member ensemble forecast of SIC from the BNRHF filter on September 27, 2011. The BNRHF prior pdf for this ensemble is shown in blue in the left panel, while the KQCEF prior pdf for this same ensemble is shown in blue in the right panel. In both panels the 20 ensemble members are shown as black dots below the pdfs, and the reference value of SIC is shown as a red dot. The forecast distribution is bimodal and the reference value lies between the two modes, where there are few ensemble members. The red line in the right panel corresponds to the likelihood for $N_o = 100$ and $k = 30$. The KQCEF obtains its posterior pdf by multiplying its prior pdf by the likelihood function (yellow line, right panel). The red line in the left panel evaluates the likelihood at the ensemble members and linearly interpolates between these values. The BNRHF obtains its posterior pdf by multiplying its prior pdf by this linearly-interpolated approximation of the likelihood (yellow line, left panel). In this example the reference value (and the

peak of the likelihood) is in a trough between two modes of the prior, where there are few ensemble members. Because the BNRHF only evaluates the likelihood at the ensemble members the peak of the likelihood is not correctly represented, and the BNRHF posterior is incorrect: in this example the BNRHF posterior pdf is bimodal, with peaks on either side of the reference value. With a larger ensemble there are fewer gaps, the BNRHF's piecewise-linear approximation of the likelihood is more accurate, and the result is more accurate.

5.2. Forecast vs Analysis performance with Icepack

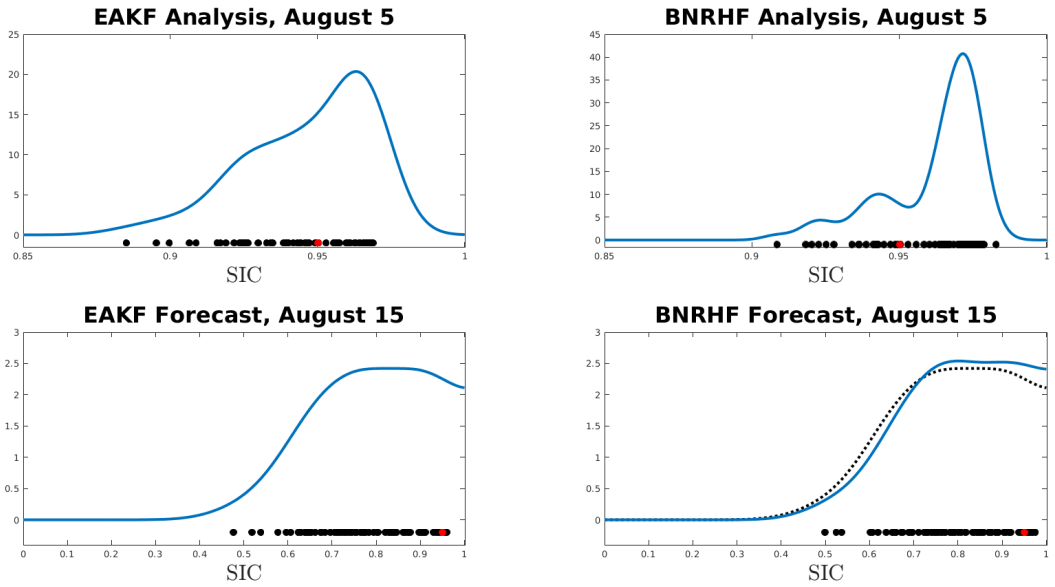

Figure 11. Top row: Ensemble members (black dots) for the analysis produced by the EAKF (left) and BNRHF (right) on August 5, 2011. Bottom row: Ensemble members (black dots) for the forecast produced by the EAKF (left) and BNRHF (right) on August 15, 2011, initialized from the analysis ensembles shown in the top row. In all panels the red dot indicates the reference value. The black dotted line in the lower-right panel is the same as the blue line in the lower left panel, reproduced for ease of comparison.

Figure 8 demonstrates that the EAKF outperforms the non-parametric methods when the observation error is small ($N_o = 100$), yet Figure 9 shows that KQCEF (and sometimes BNRHF) yield better forecasts in these cases. We attempt to explain this mismatch by examining a single analysis/forecast pair.

The top row of Figure 11 shows the analysis ensemble (black dots) produced by EAKF (left) and by BNRHF (right) on August 5, 2011; the blue line in each panel is a kernel-density estimate of the analysis pdf, and the red dot shows the reference SIC value. The BNRHF distribution is more tightly centered than the EAKF distribution, and is biased high compared to the true SIC value; as a result the CRPS associated with the EAKF is better (0.0058) than the CRPS associated with the BNRHF (0.0099). Notice that the EAKF produces an analysis ensemble with a longer tail towards low SIC values than the BNRHF. In particular, the smallest ensemble member in the EAKF analysis is noticeably smaller than the smallest ensemble member in the BNRHF analysis.

The bottom row of Figure 11 shows the ensembles (black dots) that result from forecasting the analysis ensembles in the top row 10 days forwards to August 15, 2011; the blue line in each panel is a kernel-density estimate of the forecast pdf, and the red dot shows the reference SIC value. The black dotted line in the lower right panel shows the forecast pdf associated with the EAKF, for easier comparison with the forecast pdf associated with the BNRHF (blue). The atmospheric forcing and perturbed parameters of the reference case conspire to keep its SIC value high throughout the 10-day forecast, but the atmospheric

forcing and perturbed parameters of the ensemble cause many of the ensemble members to melt out to values of SIC much lower than the reference case. The main difference between the EAKF and BNRHF forecasts is that the EAKF forecast has a longer tail towards low SIC values and has fewer ensemble members close to the large reference SIC value. This results in a larger CRPS for the EAKF forecast (0.0443) compared to the BNRHF forecast (0.0391).

This behavior is common across the time window of interest. The smallest analysis ensemble member produced by the EAKF is almost always significantly smaller than the smallest analysis ensemble members produced by BNRHF and KQCEF across the time window of interest, which is indicative of a longer tail towards low SIC values in the EAKF analysis. This can be traced back to the way in which the filters update the ensemble. The EAKF shifts and scales the ensemble, but does not otherwise change its shape. Thus, if the atmospheric forcing and perturbed parameters conspire to create a forecast with a long tail towards small SIC values, the EAKF analysis will also have a tail towards small SIC values. In contrast both BNRHF and KQCEF change the shape of the ensemble, e.g. moving from a bimodal forecast to a unimodal analysis as shown in the right panel of Figure 10. In this specific context of the Icepack model with perturbed parameters, the result is poorer performance of EAKF.

5.3. Computational Cost

The computational cost of these filters is negligible compared to the forecast step, even when the forecast is an extremely low-cost single-column ice model. On a simple workstation the forecast step with 79 ensemble members took approximately 8 seconds. The typical time required to run a single assimilation cycle was computed by using the free-running forecast from Figure 7 on September 27, 2011 as the forecast ensemble. The ensemble size was varied from 10 to 79 by selecting the first N_e members of the free-running ensemble. The likelihood used $N_o = 100$ and varied k from 20 through 80. The median time to perform the assimilation was computed across these values of k, and the results are shown as a function of ensemble size in Figure 12. Both EAKF and BNRHF have total cost approximately independent of ensemble size at a value near 0.007 seconds per assimilation cycle. The code is not parallelized, so the independence with respect to ensemble size means that the total cost is dominated by overhead costs. KQCEF has total cost that scales approximately linearly with N_e from being comparable to the cost of EAKF and BNRHF at $N_e = 10$ to being a little more than three times more expensive at $N_e = 79$. Even at $N_e = 79$ though, the total cost of the filter is negligible compared to the cost of the ensemble forecast.

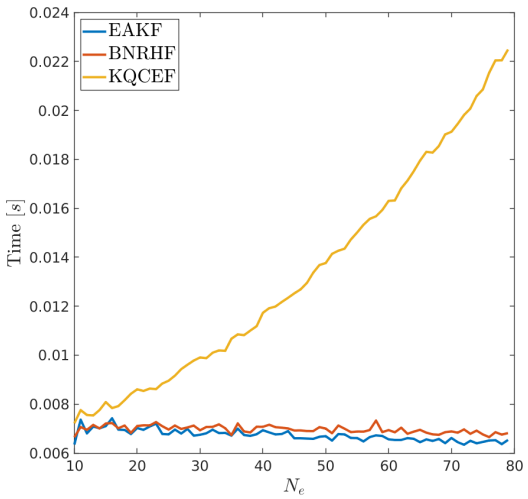

Figure 12. Median time for a single assimilation cycle for each filter as a function of ensemble size.

The focus here has been on the first step of the two-step framework, i.e. the analysis in observation space. Anderson [7] recently advocated using the bounded normal rank histogram approximation to the prior cdf as part of a generalized linear regression approach

to the second step (regression from observation space back to state variables). The KQCEF approximation of the prior cdf could similarly be used in such a context.

The KQCEF has been implemented in DART [41,42] so that it can be available to the data assimilation research community.

Author Contributions: I.G.: Conceptualization, methodology, software, writing—original draft preparation; C.R: software, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the US National Science Foundation grant number 2152814.

Data Availability Statement: The code and data are available at https://doi.org/10.6084/m9.figshare.25802203.v1 [43].

Acknowledgments: The authors are grateful to JL Anderson for guidance in implementing KQCEF into DART, and to two anonymous reviewers whose comments led to an improved presentation and discussion of the results.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BRNHF	Bounded Normal Rank Histogram Filter
EAKF	Ensemble Adjustment Kalman Filter
EnKF	Ensemble Kalman Filter
KDE	Kernel Density Estimation
KS	Kolmogorov-Smirnov
KL	Kullback-Leibler
KQCEF	Kernel-based Quantile Conserving Ensemble Filter
QCEF	Quantile Conserving Ensemble Filter

Appendix A

Suppose that

$$F_X(x) = P[X < x] = \int_{-\infty}^{x} p_X(s)\mathrm{d}s \tag{A1}$$

is the cdf of a scalar random variable X. Then $U = F_X(X)$ is a random variable with uniform pdf on $[0,1]$:

$$P[U = F_X(X) < u] = P[X < F_X^{-1}(u)] = F_X(F_X^{-1}(u)) = u \tag{A2}$$

for $0 \leq u \leq 1$. Since the cdf is linear on $[0,1]$, the pdf is constant on $[0,1]$, which shows that U has a uniform distribution. The derivation above assumes that F_X is invertible in the sense that $F_X^{-1}(F_X(x)) = x$ for every x in the range of X (the first equality), and also that $F_X\left(F_X^{-1}(u)\right) = u$ for every $u \in [0,1]$ (the last equality).

Suppose that U is uniform on $[0,1]$ and F_X is the cdf of a random variable X. Then the distribution of $\hat{X} = F_X^{-1}(U)$ is the same as the distribution of X:

$$P\left[\hat{X} = F_X^{-1}(U) < x\right] = P\left[F_X\left(F_X^{-1}(U)\right) = \tilde{U} < F_X(x)\right] = F_X(x). \tag{A3}$$

The last step in this derivation assumes that $F_X(F_X^{-1}(U)) = \tilde{U}$ is a uniformly-distributed random variable on $[0,1]$. Note that it does not require $F_X\left(F_X^{-1}(u)\right) = u$ for every $u \in [0,1]$.

The probability integral transform Equation (1) acts in two steps: First transform from the prior distribution to a uniform distribution using the prior cdf, then transform to the

posterior distrbution using the posterior quantile function. What happens when the prior cdf has a jump discontinuity at x_*? Let the limit from below be

$$\lim_{x \to x_*^-} F_X(x) = u_- \tag{A4}$$

and from above be

$$\lim_{x \to x_*^+} F_X(x) = u_+. \tag{A5}$$

If we define $F_X^{-1}(u) = x_*$ for every $u \in [u_-, u_+]$, then we will be able to satisfy $F_X^{-1}(F_X(x)) = x$ for every x, as long as $F_X(x_*) \in [u_-, u_+]$. But if we define $F_X(x_*)$ to be any single value in $[u_-, u_+]$, then we will not be able to satisfy $F_X(F_X^{-1}(u)) = u$ for every $u \in [0, 1]$. In this case we cannot directly use Equation (1) to transform from the prior to a uniform distribution.

But in order to sample from the posterior distribution, we only need to sample from a uniform distribution and then apply the posterior quantile function. In fact, we could completely skip the first step (transforming from the prior to a uniform) and instead just sample from the posterior directly using the posterior quantile function. The problem with such an approach only appears in the second step of two-step ensemble filters, i.e. the step where observation space increments are regressed back to state variables. A completely random sampling in observation space scrambles the ensemble increments of the state variables, which can lead to dynamical imbalances.

To avoid this while still making use of the probability integral transform, we need to define a map from the prior (in observation space) to a uniformly-distributed random variable. We can achieve this by defining F_X so that $F_X(x_*)$ is a random variable uniformly distributed on $[u_-, u_+]$. With this definition F_X is no longer a deterministic function, but $F_X(X)$ is uniformly distributed on $[0, 1]$.

The foregoing analysis was focused on mapping from the prior to a uniform distribution when the prior cdf has a jump. What if the posterior cdf has a jump discontinuity at $x = x_*$? The second step in the transform, from uniform to posterior, does not require $F_X\left(F_X^{-1}(u)\right) = u$ for every $u \in [0, 1]$. It simply requires that $F_X(F_X^{-1}(U)) = \tilde{U}$ is a uniformly-distributed random variable on $[0, 1]$. We can define the cdf F_X to be a random map, as above, which satisfies this requirement. From a practical standpoint this is not necessary, since the second step does not actually use the cdf F_X; it uses the quantile function F_X^{-1}. It suffices to define $F_X^{-1}(u) = x_*$ for every $u \in [u_-, u_+]$.

References

1. Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.-Oceans* **1994**, *99*, 10143–10162.
2. Evensen, G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics* **2003**, *53*, 343–367.
3. Anderson, J.L. A local least squares framework for ensemble filtering. *Mon. Weather Rev.* **2003**, *131*, 634–642.
4. Grooms, I. A comparison of nonlinear extensions to the ensemble Kalman filter. *Computational Geosciences* **2022**, pp. 1–18.
5. Anderson, J.L. A Quantile-Conserving Ensemble Filter Framework. Part I: Updating an Observed Variable. *Mon. Weather Rev.* **2022**, *150*, 1061–1074.
6. Anderson, J.L. A non-Gaussian ensemble filter update for data assimilation. *Mon. Weather Rev.* **2010**, *138*, 4186–4198.
7. Anderson, J.L. A Quantile-Conserving Ensemble Filter Framework. Part II: Regression of Observation Increments in a Probit and Probability Integral Transformed Space. *Mon. Weather Rev.* **2023**, *151*, 2759–2777.
8. Anderson, J.; Riedel, C.; Wieringa, M.; Ishraque, F.; Smith, M.; Kershaw, H. A Quantile-Conserving Ensemble Filter Framework. Part III: Data Assimilation for Mixed Distributions with Application to a Low-Order Tracer Advection Model. *Mon. Weather Rev.* **2024**. https://doi.org/10.1175/MWR-D-23-0255.1.
9. Silverman, B. *Density estimation for statistics and data analysis*; CRC Press, 1998.
10. Epanechnikov, V.A. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications* **1969**, *14*, 153–158.
11. Jones, M.C. Simple boundary correction for kernel density estimation. *Statistics and computing* **1993**, *3*, 135–146.
12. Jones, M.; Foster, P. A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica* **1996**, pp. 1005–1013.

13. Atkinson, K. *An introduction to numerical analysis*; John wiley & sons, 1991.
14. Oliveira, I.F.; Takahashi, R.H. An enhancement of the bisection method average performance preserving minmax optimality. *ACM Transactions on Mathematical Software (TOMS)* **2020**, *47*, 1–24.
15. Anderson, J.L. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **2001**, *129*, 2884–2903.
16. Berger, V.W.; Zhou, Y. Kolmogorov–Smirnov test: Overview. *Wiley statsref: Statistics reference online* **2014**.
17. Pérez-Cruz, F. Kullback-Leibler divergence estimation of continuous distributions. In Proceedings of the 2008 IEEE international symposium on information theory. IEEE, 2008, pp. 1666–1670.
18. Bulinski, A.; Dimitrov, D. Statistical estimation of the Kullback–Leibler divergence. *Mathematics* **2021**, *9*, 544.
19. Bishop, C.H. The GIGG-EnKF: Ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. *Quarterly Journal of the Royal Meteorological Society* **2016**, *142*, 1395–1412.
20. Morzfeld, M.; Hodyss, D. Gaussian approximations in filters and smoothers for data assimilation. *Tellus A: Dynamic Meteorology and Oceanography* **2019**, *71*, 1600344.
21. Grooms, I.; Robinson, G. A hybrid particle-ensemble Kalman filter for problems with medium nonlinearity. *PLoS ONE* **2021**, *16*, e0248266.
22. Wieringa, M.; Riedel, C.; Anderson, J.A.; Bitz, C. Bounded and categorized: targeting data assimilation for sea ice fractional coverage and non-negative quantities in a single column multi-category sea ice model. *EGUsphere* **2023**, *2023*, 1–25.
23. Hunke, E.; Allard, R.; Bailey, D.A.; Blain, P.; Craig, A.; Dupont, F.; DuVivier, A.; Grumbine, R.; Hebert, D.; Holland, M.; et al. CICE-Consortium/Icepack: Icepack 1.4.0 **2023**. https://doi.org/10.5281/zenodo.10056496.
24. Delworth, T.L.; Broccoli, A.J.; Rosati, A.; Stouffer, R.J.; Balaji, V.; Beesley, J.A.; Cooke, W.F.; Dixon, K.W.; Dunne, J.; Dunne, K.; et al. GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate* **2006**, *19*, 643–674.
25. Rampal, P.; Bouillon, S.; Ólason, E.; Morlighem, M. neXtSIM: a new Lagrangian sea ice model. *The Cryosphere* **2016**, *10*, 1055–1073.
26. Lisæter, K.A.; Rosanova, J.; Evensen, G. Assimilation of ice concentration in a coupled ice–ocean model, using the Ensemble Kalman filter. *Ocean Dyn.* **2003**, *53*, 368–388.
27. Lindsay, R.W.; Zhang, J. Assimilation of ice concentration in an ice–ocean model. *J. Atmos. Ocean. Tech.* **2006**, *23*, 742–749.
28. Stark, J.D.; Ridley, J.; Martin, M.; Hines, A. Sea ice concentration and motion assimilation in a sea ice- ocean model. *J. Geophys. Res.-Oceans* **2008**, *113*.
29. Metref, S.; Cosme, E.; Snyder, C.; Brasseur, P. A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation. *Nonlinear Proc. Geoph.* **2014**, *21*, 869–885.
30. Barth, A.; Canter, M.; Van Schaeybroeck, B.; Vannitsem, S.; Massonnet, F.; Zunz, V.; Mathiot, P.; Alvera-Azcárate, A.; Beckers, J.M. Assimilation of sea surface temperature, sea ice concentration and sea ice drift in a model of the Southern Ocean. *Ocean Model.* **2015**, *93*, 22–39.
31. Massonnet, F.; Fichefet, T.; Goosse, H. Prospects for improved seasonal Arctic sea ice predictions from multivariate data assimilation. *Ocean Model.* **2015**, *88*, 16–25.
32. Chen, Z.; Liu, J.; Song, M.; Yang, Q.; Xu, S. Impacts of assimilating satellite sea ice concentration and thickness on Arctic sea ice prediction in the NCEP Climate Forecast System. *J. Climate* **2017**, *30*, 8429–8446.
33. Zhang, Y.F.; Bitz, C.M.; Anderson, J.L.; Collins, N.; Hendricks, J.; Hoar, T.; Raeder, K.; Massonnet, F. Insights on sea ice data assimilation from perfect model observing system simulation experiments. *J. Climate* **2018**, *31*, 5911–5926.
34. Riedel, C.; Anderson, J. Exploring Non-Gaussian Sea Ice Characteristics via Observing System Simulation Experiments. *EGUsphere* **2023**, *2023*, 1–30.
35. Abdalati, W.; Zwally, H.J.; Bindschadler, R.; Csatho, B.; Farrell, S.L.; Fricker, H.A.; Harding, D.; Kwok, R.; Lefsky, M.; Markus, T.; et al. The ICESat-2 laser altimetry mission. *Proceedings of the IEEE* **2010**, *98*, 735–751.
36. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous univariate distributions, volume 2*; Vol. 289, John wiley & sons, 1995.
37. Raeder, K.; Hoar, T.J.; El Gharamti, M.; Johnson, B.K.; Collins, N.; Anderson, J.L.; Steward, J.; Coady, M. A new CAM6+ DART reanalysis with surface forcing from CAM6 to other CESM models. *Scientific Reports* **2021**, *11*, 16384.
38. Danabasoglu, G.; Lamarque, J.F.; Bacmeister, J.; Bailey, D.; DuVivier, A.; Edwards, J.; Emmons, L.; Fasullo, J.; Garcia, R.; Gettelman, A.; et al. The community earth system model version 2 (CESM2). *J. Adv. Model. Earth Syst.* **2020**, *12*, e2019MS001916.
39. Hersbach, H. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecast* **2000**, *15*, 559–570.
40. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc* **2007**, *102*, 359–378.
41. Anderson, J.L.; Collins, N. Scalable implementations of ensemble filter algorithms for data assimilation. *J Atmos Ocean Tech* **2007**, *24*, 1452–1463.
42. Anderson, J.; Hoar, T.; Raeder, K.; Liu, H.; Collins, N.; Torn, R.; Avellano, A. The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society* **2009**, *90*, 1283–1296.
43. Grooms, I. Data for "A quantile-conserving ensemble filter based on kernel density estimation" **2024**. https://doi.org/10.6084/m9.figshare.25802203.v1.