A non-parametric way to estimate observation errors based on ensemble innovations

Received: 14 April 2023 | Revised: 16 February 2024 | Accepted: 20 February 2024

DOI: 10.1002/qj.4710

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

A non-parametric way to estimate observation errors based on ensemble innovations

Chih-Chi Hu[1] | **Peter Jan van Leeuwen**[1,2] | **Alan J. Geer**[3]

[1]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA

[2]Department of Meteorology, University of Reading, Reading, United Kingdom

[3]European Centre for Medium-Range Weather Forecasts, Reading, UK

Correspondence
Chih-Chi Hu, Department of Atmospheric Science, Colorado State University, Fort Collins, CO, 80526, USA.
Emails: chihchi@colostate.edu, chih.chi.hu23@alumni.colostate.edu

Funding information
Cooperative Institute for Research in the Atmosphere, Grant/Award Number: NA19OAR4320073

Abstract

Previous studies that inferred the observation error statistics from the innovation statistics can only provide the second moment of the error probability density function (pdf). However, the observation errors are sometimes non-Gaussian, for example, for observation operators with unknown representation errors, or for bounded observations. In this study, we propose a new method, the Deconvolution-based Observation Error Estimation (DOEE), to infer the full observation error pdf. DOEE does not rely on linear assumptions on the observation operator, the optimality of the data assimilation algorithm, or implicit Gaussian assumptions on the error pdf. The main assumption of DOEE is the availability of an ensemble of background forecasts following the independent and identically distributed (i.i.d.) assumption. We conduct idealized experiments to demonstrate the ability of the DOEE to accurately retrieve a non-Gaussian (bimodal, skewed, or bounded) observation error pdf. We then apply the DOEE to construct a state-dependent observation error model for satellite radiances by stratifying the observation errors based on cloud amount. In general, we find that the observation error pdfs in many categories are skewed. By adding a new predictor, total column water vapor (TCWV), into the state-dependent model, we find that for cloudy pixels, when TCWV is small, the observation error pdfs are quite similar and Gaussian-like, whereas when TCWV is large, the observation error pdfs differ for different cloud amount, while all of them are positively biased. This result suggests that exploring other predictors, like cloud type, might improve the stratification of the observation error model. We also discuss ways to include a non-parametric observation error pdf into modern data assimilation schemes, including a consideration of the strong-constraint 4D-Var perspective, as well as the implications for other observation types including observations with bounded range.

KEYWORDS

Bounded observation, data assimilation, non-Gaussian distribution, nonlinear observation, observation error, satellite radiance, state-dependent observation error

1 | BACKGROUND

Data assimilation (DA) is a method to sequentially estimate the probability density function (pdf) or the first and second moment of the model state, given an imperfect model forecast and noisy observations. DA is based on Bayes' theorem, which states how we can update the pdf from before knowing the observations (prior pdf) to after knowing the observations (posterior pdf). This update depends not only on the value of the observations, but also on the uncertainty of the observations via the likelihood function. Therefore, an accurate specification of the uncertainties of the observations is essential to any DA framework.

We note that the uncertainty of the model forecast, or the background error, is equally important in DA. Roughly speaking, the relative magnitude of the background and the observation error determines how much information from the observation is assimilated into the model. However, it is difficult to accurately estimate the background error for several reasons. For example, when the background error is quantified by the spread of an ensemble, an insufficiently large ensemble size can lead to sampling errors in the background error. In addition, failure to sufficiently address the model errors within the ensemble can result in an underestimation of the background error. There are some methods to tackle this issue by adaptively inflating the background error (e.g., Anderson, 2007, 2009; Minamide & Zhang, 2019), but how to accurately specify a flow-dependent background error is still an active research area. While careful specification of the background error is important, the main focus of this work is on the observation error.

The observation error not only includes the uncertainty due to the measurement error, but also the uncertainty due to the representation error. The measurement error is related to the measurement instrument design, which is often thought well known. The representation error is defined as the misfit between the observation and the simulated observation resulting from different representations of reality (see e.g., Hodyss & Nichols, 2015; Janjić et al., 2017; van Leeuwen, 2015) and hence is associated with the observation operator that we have access to, and the variability of the truth that is not represented by our model grid. However, the representation error is typically not well understood. We often do not know the true mapping between the model space and the observation space. For example, for radar simulators, we typically assume an empirical drop size distribution (e.g., Bringi & Chandrasekar, 2001), which is not always realistic. Furthermore, observations and models can represent different spatial scales, which is especially problematic if the observations represent smaller scales than the model can simulate. Nevertheless, there are some strategies to estimate the representation errors. For example, one is by conducting field campaigns and collecting extensive observations to estimate the relationship between observations at different scales, or perform high-resolution model runs to infer the observation operator (e.g., Hodyss & Nichols, 2015; Janjić et al., 2017; van Leeuwen, 2015). However, these methods are often less practical to estimate the representation error for every kind of observation operator.

Instead of having to estimate the measurement error and the representation error separately, there are some statistical approaches to quantify the total observation error by comparing the observations with several other collocated reference datasets. The reference could most typically be another observation type, or fields from a DA system (background or analysis). It is, however, still difficult to estimate the observation error this way since the reference against which we compare the observation will also have errors. The problem is ill-posed if we do not know the errors in the reference.

One solution is to make the problem well posed is to first estimate the errors in the reference, and then to subtract them. For example, if the reference is the model background, then its errors are already estimated as part of the DA system. Note that this is based on a standard relation (see Section 2.2 for detail) that has been used to diagnose the consistency of a DA system (e.g., Rutherford, 1972; Hollingsworth & Lönnberg, 1989; Daley, 1993; Andersson, 2003, etc.), while it has been named differently in the previous literature, for example, moment-based method, innovation covariance consistency, and so forth. Nevertheless, we will refer to this method as the "background subtraction method" in this paper, as we consider this nomenclature to be more intuitive.

In addition, Desroziers et al. (2005) showed how the product of the analysis and background departures can form an alternative approach to diagnose the observation error (their equation 3; this will be referred to as "DBCP method" hereafter). The DBCP method is an iterative approach that can jointly estimate the observation error and the background error covariance (in the observation space) at the same time. Note that although the DBCP method does not rely on the prior knowledge of the background error, it relies on the optimality in the DA algorithm. The DBCP method has become a popular method for estimating observation errors in the DA community (e.g. Bormann et al., 2016; Bormann & Bauer, 2010).

The other approach to make the problem well posed is the three-cornered hat (3CH) method (Anthes & Rieckh, 2018; Gray & Allan, 1974; Grubbs, 1948; Rieckh & Anthes, 2018; Sjoberg et al., 2021), which compares

the observation to two independent references. The 3CH method can estimate the error variances of the three datasets simultaneously, without any prior knowledge of the errors in the reference. Although the 3CH method seems unrelated to the DBCP method, Semane et al. (2022) and Todling et al. (2022) have recently shown the equivalence between the two methods, when the analysis and the background (along with the observations) are used as the reference datasets.

Unfortunately, all of the methods mentioned previously can only estimate the second moment of the observation error pdf. This is fine when the true observation error pdf is Gaussian. However, the second moment is not enough to represent the full non-Gaussian observation error pdf. There are many observations that have non-Gaussian observation errors. For example, for observations with complicated representation errors, there is no reason to assume their observation errors are still Gaussian. Furthermore, when the observed value is close to the boundary for bounded observations, for example, precipitation or the concentration of some variables, the observation error can be skewed, and hence clearly non-Gaussian (e.g., Bishop, 2016; Lien et al., 2013).

Note that in general, if the background error is independent of the observation error, the innovation pdf is the convolution of the observation error pdf and the background error pdf, following a standard statistic relationship. If we have access to the possibly non-Gaussian background error and innovation pdf, then in principle it is possible to solve the non-Gaussian observation error pdf by deconvolution. Note that deconvolution is a standard technique in signal processing, but in the DA context here it could be applied to recover the observation error pdf from the innovation pdf.

Therefore, the goal of this study is to propose a new method, the Deconvolution-based Observation Error Estimation (DOEE) method, that can estimate the full observation error pdf by solving the deconvolution equation. Note that DOEE does not rely on the linear approximation for the observation operator, the optimality of the DA algorithm, or the implicit Gaussian assumption for the error pdf (see Section 2.2 for details). In addition, we will also discuss ways to construct a state-dependent observation error pdf model using DOEE. Note that the focus of this study is on the theoretical aspect and the demonstration of the new method. We will briefly discuss some ways to apply this method for different DA schemes in Section 5, while the actual implementation of this new method in a real DA system is left for future work.

The remainder of this paper is organized as follows. In Section 2, we provide the mathematical background for estimating the observation error, and introduce the new framework DOEE. In Section 3, we conduct a series of idealized experiments to justify the new method. In Section 4, we apply DOEE to construct a state-dependent observation error model for satellite radiances. Finally, in Section 5 we conclude the findings in this paper and propose the use of DOEE in several applications.

2 | METHODOLOGY

2.1 | Definitions and notations

For clarity, we use capital letters and ε (with different superscripts) to denote random variables (vectors) and lower-case letters to denote non-random variables (vectors) in the following. Since the true atmosphere is a continuum, the true atmospheric state lives in a space of infinite dimension. Let's define $\tilde{x}^{\mathrm{t}}$ as the truth in this infinite dimension space (hereafter, "continuous space"), and define x^{t} as a vector representing the discretized truth in the "model space." Let r be the transformation from the continuous space to the model space:

$$x^{\mathrm{t}} = r\left(\tilde{x}^{\mathrm{t}}\right). \tag{1}$$

Note that we consider the model space variable given as soon as the truth is given. One could argue that x^{t} is a random variable as the transformation is not unique, in the sense that the averaging operation can be defined in many ways, and none of them satisfactorily, given notions as 'effective model resolution'. Our choice does not change any of the results discussed below. The "true observation" y^{t} can be related to $\tilde{x}^{\mathrm{t}}$ as follows:

$$y^{\mathrm{t}} = \tilde{H}^{\mathrm{t}}\left(\tilde{x}^{\mathrm{t}}\right), \tag{2}$$

where $\tilde{H}^{\mathrm{t}}$ is the true observation operator that maps from the continuous space to the observation space. Since the measurement noise is unavoidable in every measurement process, the "actual observed value" Y can be written as

$$Y = y^{\mathrm{t}} + \varepsilon^{\mathrm{m}} = \tilde{H}^{\mathrm{t}}\left(\tilde{x}^{\mathrm{t}}\right) + \varepsilon^{\mathrm{m}}, \tag{3}$$

where ε^{m} is the measurement error. Note that ε^{m} is treated as a random variable here, and so is Y. Denote the observation operator that we have access to as H. We can rewrite Equation (3) as

$$Y = H\left(x^{\mathrm{t}}\right) + \left[\tilde{H}^{\mathrm{t}}\left(\tilde{x}^{\mathrm{t}}\right) - H\left(x^{\mathrm{t}}\right)\right] + \varepsilon^{\mathrm{m}}, \tag{4}$$

and note the difference between $\tilde{H}^{\mathrm{t}}$ and H in Equation (4): $\tilde{H}^{\mathrm{t}}$ is the true mapping (which we do not know) from the continuous state to the observation space, while H is

the imperfect observation operator (which we have access to) from the model space to the observation space, which can include some unrealistic assumptions. We define this representation error ε^{r} as:

$$\varepsilon^{\mathrm{r}} = \left[\tilde{H}^{\mathrm{t}}(\tilde{x}^{\mathrm{t}}) - H(x^{\mathrm{t}})\right]. \tag{5}$$

Based on the likelihood function in Bayes' theorem, the observation error is defined as

$$\varepsilon^{\mathrm{o}} = Y - H(x^{\mathrm{t}}), \tag{6}$$

such that the observation error can also be written as

$$\varepsilon^{\mathrm{o}} = \varepsilon^{\mathrm{r}} + \varepsilon^{\mathrm{m}}. \tag{7}$$

2.2 | The background subtraction method

Based on the discussions in the introduction, estimating the representation error directly can be challenging. Nevertheless, the background subtraction method is one of the methods that can estimate the total observation error based on the innovation statistics, which will be described below. First consider the following random variable, the innovation D^{b} as:

$$D^{\mathrm{b}} = Y - H(X^{\mathrm{b}}), \tag{8}$$

where X^{b} represents the background model state in the model space. As in the Kalman filter we assume that X^{b} is unbiased, and has error covariance $\mathbf{B}$. We can write Equation (8) as:

$$\begin{aligned} D^{\mathrm{b}} &= \left[Y - H(x^{\mathrm{t}})\right] + \left[H(x^{\mathrm{t}}) - H(X^{\mathrm{b}})\right] \\ &= \varepsilon^{\mathrm{o}} + \left[H(x^{\mathrm{t}}) - H(X^{\mathrm{b}})\right]. \end{aligned} \tag{9}$$

Here we have used the definition of the observation error in Equation (6). We can further linearize the observation operator as follows:

$$H(x^{\mathrm{t}}) - H(X^{\mathrm{b}}) \approx \mathbf{H}(x^{\mathrm{t}} - X^{\mathrm{b}}), \tag{10}$$

where $\mathbf{H}$ denotes the linearized H. Therefore, we can rewrite Equation (9) as:

$$\begin{aligned} D^{\mathrm{b}} &= \varepsilon^{\mathrm{o}} + \left[H(x^{\mathrm{t}}) - H(X^{\mathrm{b}})\right] \\ &\approx \varepsilon^{\mathrm{o}} - \mathbf{H}(X^{\mathrm{b}} - x^{\mathrm{t}}) = \varepsilon^{\mathrm{o}} - \mathbf{H}\varepsilon^{\mathrm{b}}, \end{aligned} \tag{11}$$

where $\varepsilon^{\mathrm{b}} \equiv X^{\mathrm{b}} - x^{\mathrm{t}}$ is the background error. Assume ε^{o} and ε^{b} are independent, then we can estimate the second moment of the pdf of ε^{o} based on

$$E\left[\left(Y - H(X^{\mathrm{b}})\right)\left(Y - H(X^{\mathrm{b}})\right)^{T}\right] = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{T}, \tag{12}$$

where $E[\dots]$ is the expectation operator, T is the tranpose of a matrix, $\mathbf{R}$ is the observation error covariance, and $\mathbf{B}$ is the background error covariance.

Without relying on the optimality assumption (i.e., DBCP), there are two possibilities for inferring $\mathbf{R}$. If $\mathbf{B}$ is known, then Equation (12) can be used to estimate $\mathbf{R}$ by subtracting $\mathbf{H}\mathbf{B}\mathbf{H}^{T}$ from $E\left[\left(Y - H(X^{\mathrm{b}})\right)\left(Y - H(X^{\mathrm{b}})\right)^{T}\right]$. When $\mathbf{B}$ is not available, one can use an ensemble to provide a flow-dependent estimate of the background covariance $\mathbf{B}_{\mathrm{e}}$. In this case the ensemble mean takes the role of X^{b}, and we find, see e.g. Bonavita et al. (2020):

$$E\left[(Y - \overline{H(x)})(Y - \overline{H(x)})^{T}\right] \approx \mathbf{R} + \mathbf{H}\mathbf{B}_{\mathrm{e}}\mathbf{H}^{T}. \tag{13}$$

There are two major limitations of the background subtraction method: the linear approximation in Equation (11) and the implicit Gaussian assumption for the observation error pdf. First, for nonlinear observations, the linear assumption can only be used when either the background error is small or the observation operator is weakly nonlinear. This assumption might not be true for some observations, for example, satellite radiances (e.g., Bauer et al., 2010; Bonavita et al., 2018). Second, the second moment is not enough to provide the full observation error when the observation error is non-Gaussian (e.g., Pires et al., 2010).

2.3 | A new non-parametric method for estimating the observation error — The DOEE method

In the following, we propose a new method that can estimate the full non-parametric observation error pdf. It is based on the commonly used assumption that each ensemble member can be seen as an independent draw from the same distribution from which the true state x^{t} is drawn. This is referred to as the "i.i.d. assumption" in statistics, and is the basics behind the use of rank histograms (e.g., Anderson, 1996; Hamill, 2001).

We explain the algorithm in the following three steps:

1. We consider a random variable D related to a random forecast vector X, defined via

$$D = Y - H(X) = \varepsilon^{\mathrm{o}} + \left[H(x^{\mathrm{t}}) - H(X)\right]. \tag{14}$$

Instead of linearizing Equation (14), we define $\varepsilon = H(x^t) - H(X)$ and rewrite Equation (14) as

$$D = \varepsilon^o + \varepsilon. \tag{15}$$

2. Analogous to the assumption that ε^o and ε^b are independent, we assume ε^o and ε are independent. We now use a general result from statistics that the pdf of a sum of two independent variables is the convolution of the pdfs of those two variables. Therefore, based on Equation (15), we have:

$$f_D(d) = \int_{-\infty}^{\infty} f_{\varepsilon^o}(\varepsilon^o) f_\varepsilon(d - \varepsilon^o) d\varepsilon^o, \tag{16}$$

where f_D, f_ε, and f_{ε^o} are the pdfs of the random variables D, ε, and ε^o, respectively. If the pdfs f_ε and f_D are known, we can estimate f_{ε^o} by deconvolution. In our method we derive f_ε and f_D from an ensemble of model forecasts.

The ensemble members can be used directly to generate samples from f_D as

$$d^i = y - H(x^i) \tag{17}$$

in which y is the actual observation (obtained from the true state), and the ensemble members are x^i, with $i \in \{1, 2 \ldots, N_e\}$.

3. To find samples from f_ε we use the i.i.d. assumption. Specifically, we assume that the pdf from which the true state x^t is drawn (by nature), is the same as the pdf from which the ensemble members are drawn. Consequently, $f_{H(X^t)} = f_{H(X)}$. In other words, the sample distribution of $H(x^i) - H(x^t)$ should be the same as the sample distribution of $H(x^i) - H(x^j)$ for arbitrary $i \neq j$. Therefore, to estimate the distribution of ε, we can randomly pick one ensemble member as "the truth" and project it into the observation space. The differences between this reference member and the other ensemble members in the observation space then provide us independent samples of f_ε. The above estimation of f_ε could be sensitive to which member we pick as the truth member. However, we find that when the sample size is large enough, the result does not change much when we pick a different member (as will be verified in the sensitivity experiments in Section 3.2). When the sample size is small, we can mitigate the sampling error by using the pairwise differences of the ensemble members as samples drawn from f_ε. Therefore, we will also call f_ε the ensemble difference pdf in the following.

The method described above is called the "Deconvolution-based Observation Error Estimation (DOEE)" method. Note that Equation (16) is completely general: we neither require a linear (or weakly nonlinear) observation operator, nor a Gaussian (or any prescribed distribution) distributed model or truth state. Furthermore, DOEE can provide us the full observation error pdf, instead of just the second moment of the observation error pdf. We will discuss how to solve Equation (16) in Section 2.5 and in the Appendix A.

It is interesting to understand the meaning behind Equation (16). The convolution relationship states that f_D is a smoothed version of f_{ε^o}, in which the smoothing kernel is pdf f_ε. The smoothness depends on the width and the shape of f_ε. In general, the wider f_ε, the stronger the smoothing effect is on f_D. Similarly, for the same innovation pdf f_D, if the pdf f_ε (i.e., the smoothing kernel) is narrower, we can expect a wider observation pdf f_{ε^o}.

2.4 | State-dependent observation errors

Deconvolution-based observation error estimation can also be used to estimate state-dependent observation errors. Suppose the observation error pdf depends on a predictor c, which is a function of the true state, $c(x^t)$. We can rewrite Equation (16) when $c(x^t) = \tilde{c}$ as

$$\begin{aligned} D_{|c(x^t)=\tilde{c}} = Y_{|c(x^t)=\tilde{c}} - H(X) = \varepsilon^o{}_{|c(x^t)=\tilde{c}} \\ + \left[H(x^t{}_{|c(x^t)=\tilde{c}}) - H(X)\right]. \end{aligned} \tag{18}$$

To estimate the pdf of $D_{|c(x^t)=\tilde{c}}$, we only use the observations with $c_{\text{obs}} = \tilde{c}$, where c_{obs} is the predictor associated with the observation. Note that in real cases, since we do not know the true predictor, we assume that the observation predictor is a good approximation to the true predictor $c_{\text{obs}} \approx c(x^t)$. On the other hand, we can also leverage the i.i.d. assumption to estimate $H(x^t{}_{|c(x^t)=\tilde{c}}) - H(X)$. Specifically, since the pdf of $H(X^t{}_{|c(X^t)=\tilde{c}})$ is assumed to be equal to the pdf of $H(X_{|c(X)=\tilde{c}})$, we draw from a subset of ensemble whose predictor satisfies $c(x^i) = \tilde{c}$ to sample $H(X^t{}_{|c(X^t)=\tilde{c}})$, instead of arbitrarily picking one from the whole ensemble. We call this subset of ensemble the "indistinguishable subset." Similarly, when the sample size is small, we can average the histograms of the differences between each member in the indistinguishable subset and the other ensemble members to approximate the pdf.

For example, for all-sky satellite radiances we could assume the predictor c to be the cloud amount. Then, we can use DOEE to estimate the observation error pdf for each cloud amount. Note that it can be very difficult to identify the exact predictor. In practice, one could propose a predictor based on the knowledge of possible error sources associated with the observation operator. The observation error model can also depend on multiple predictors. See Section 4.2 for examples.

2.5 | The way to solve the deconvolution equation in the DOEE method

In this section, we present some of the mathematical details of how we solve the deconvolution equation in Equation (16) in DOEE. Note that we consider a univariate pdf in the following. Although the convolution relation in Equation (16) is valid for multivariate pdfs as well, solving the multivariate deconvolution can be more complicated. We leave the extension of DOEE to multivariate pdfs as a future work.

We express the pdfs f_D and f_{ε^o} using histograms with bins centered at $[-n, -n+1, \dots, n]\Delta x$, where Δx is the bin width. We choose Δx based on the Freedman–Diaconis rule (Freedman & Diaconis, 1981), which is designed to select the bin width that minimizes the squared difference between the histogram generated by the samples and the underlying true pdf. Specifically, the bin width Δx is chosen based on

$$\Delta x = 2\frac{\mathrm{IQR}}{\sqrt[3]{N}}, \tag{19}$$

where IQR is the interquartile range of the data, and N is the sample size. Then, we choose n so that $n\Delta x$ is large enough and we can make sure the values of pdf close to the boundaries are all zeros. If we discretize Equation (16) using histograms defined by the bins above, the value of pdf f_D at $m\Delta x$ can be written as:

$$f_D(m\Delta x) = \sum_{j=-n}^{n} f_\varepsilon((m-j)\Delta x) f_{\varepsilon^o}(j\Delta x). \tag{20}$$

Note that we have totally $2n+1$ equations (with $m = -n, \dots, n$ in Equation (20)) and $2n+1$ unknowns $(f_{\varepsilon^o}(j\Delta x), j = -n, \dots, n)$, and therefore we can form a linear system of equations. For convenience, we can write the linear system of equations into a matrix form:

$$\mathbf{f}_D = \mathbf{A}\mathbf{f}_{\varepsilon^o}, \tag{21}$$

where $\mathbf{f}_D$ and $\mathbf{f}_{\varepsilon^o}$ are column vectors with the values from $f_D(j\Delta x)$ and $f_{\varepsilon^o}(j\Delta x), j = -n, \dots, n$, respectively, and $\mathbf{A}$ is a matrix containing the values of $f_\varepsilon(j\Delta x)$ that fulfills Equation (20).

Although the number of equations and the unknowns is the same, it turns out that directly solving Equation (21) by $\mathbf{A}^{-1}\mathbf{f}_D$ does not give us a desirable solution. Specifically, the pdf can have negative values, and the solution tends to be very noisy. Therefore, we try to seek the solution from a constrained minimization problem, in which we formulate a cost function with a penalty term to make the solution smooth, and also constrain the solution to be non-negative. There is a "smoothness parameter" in the cost function to control the smoothness of the solution. Please see the Appendix A for details of the cost function and how we determine the smoothness parameter.

Finally, we summarize how the DOEE method works in the following:

1. Collect samples of observations and the collocated ensemble of model equivalences in observation space. Calculate the histograms of $\mathbf{f}_D$ and $\mathbf{f}_\varepsilon$. Define the indistinguishable subset if constructing a state-dependent error model (we will give an example in Section 4.1).
2. Form the cost function with the optimal smoothness parameter (see the Appendix A).
3. The observation error pdf $\mathbf{f}_{\varepsilon^o}$ is the solution that minimizes the cost function with the optimal smoothness parameter.

3 | IDEALIZED EXPERIMENTS

In this section, we conduct a series of idealized experiments to demonstrate and justify the DOEE method. In Section 3.1, we will first show that DOEE not only can retrieve a Gaussian observation error pdf, but also a non-Gaussian one (bimodal, skewed), which is the strength of DOEE. In Section 3.2, we will examine the sensitivity of DOEE to the sample size, the quality of the ensemble, and the member we pick as the reference member.

3.1 | Estimate of Gaussian and non-Gaussian observation error distribution

The setup for the idealized experiments is as follows. First, we generate the truth from a given distribution, and then add noise drawn from an (assumed unknown) observation error pdf, which we would like to estimate, to the truth to generate the observation. We generate the ensemble by drawing from the same distribution as we draw the truth from. Then, we use the observation and the ensemble as input to the methodology, and try to estimate the unknown observation error pdf. We then compare the estimated observation error pdf with the true observation error pdf to evaluate the performance of DOEE. If not mentioned differently, the number of observations (grid points) is 10,000, and the ensemble size is 100 in the following experiments.

For all the experiments, we generate the truth and the ensemble by drawing from a Gamma distribution $\mathrm{Gamma}(2+S_\mathrm{h}, 2+S_\mathrm{c})$, where S_h and S_c are independent draws from $\mathrm{Uniform}(-0.5, 0.5)$. The reason we choose a Gamma distribution is that it is often used to simulate

the distribution of variables in the geophysical models, for example, the precipitation-related variables (e.g., Bishop, 2016). In addition, it is skewed, which can mimic the situation when the observation operator is non-linear, which could transform a symmetric background error in state space (e.g., Gaussian) into a skewed distribution in the observation space. Note that we choose the distribution of the truth differently at different grid points, and this is designed to mimic the realistic situation as well. In addition, the method should be based on the collection of observations whose observation errors are sampled from the same pdf at different grid points.

We first examine cases where the observation error is Gaussian but with non-zero mean, specifically, $\text{Normal}(2, 2^2)$ or $\text{Normal}(-2, 2^2)$. These cases are conducted to mimic situations where there is a bias in the prediction of the observation operator. The results show that we can well reproduce the true observation error (Figure 1a,b), suggesting this method can also be applied for bias correction, if the i.i.d. assumption for the ensemble is valid.

Next, we turn to some more difficult cases where the observation error is non-Gaussian. Specifically, Figure 1c shows the result when the true observation error is drawn from a bi-Gaussian distribution $\frac{1}{2}\left[\text{Normal}(-4, 1^2) + \text{Normal}(4, 1^2)\right]$, and Figure 1d is when the observation error is from Gamma(2, 2). We can see the estimated error pdfs are very close to the true error pdfs. These results are encouraging, since they suggest that DOEE has the ability to well capture non-Gaussianities in the observation error. Especially, we can accurately retrieve the error pdf when the true error is bounded and skewed (Figure 1d). For example, for semi-positive definite observations, when the observed value is zero, the realization of the observation error can only be either negative or zero. In this case, the observation error pdf has to be skewed if the error variance does not vanish. Previous methods are not able to provide any information of the skewness for this kind of observation error, while our method can quite accurately capture the full pdf.

We examine all the pdfs from Equation (16) in Figure 2. As discussed before, the innovation pdf (red curve) is the smoothed observation error pdf (blue curve), where the smoothing kernel is the ensemble difference pdf (orange curve). We also calculate the convoluted pdf (brown dashed curve) from the estimated observation error pdf from DOEE (blue curve) and the ensemble difference pdf (orange curve). The convoluted pdf almost overlaps with the innovation pdf, which suggests that the estimated observation error pdf from DOEE well satisfies Equation (16).

3.2 | Sensitivity tests

Recall that one of the assumptions in the method is that the ensemble is indistinguishable from the truth (the i.i.d. assumption). However, in practice, it is difficult to know

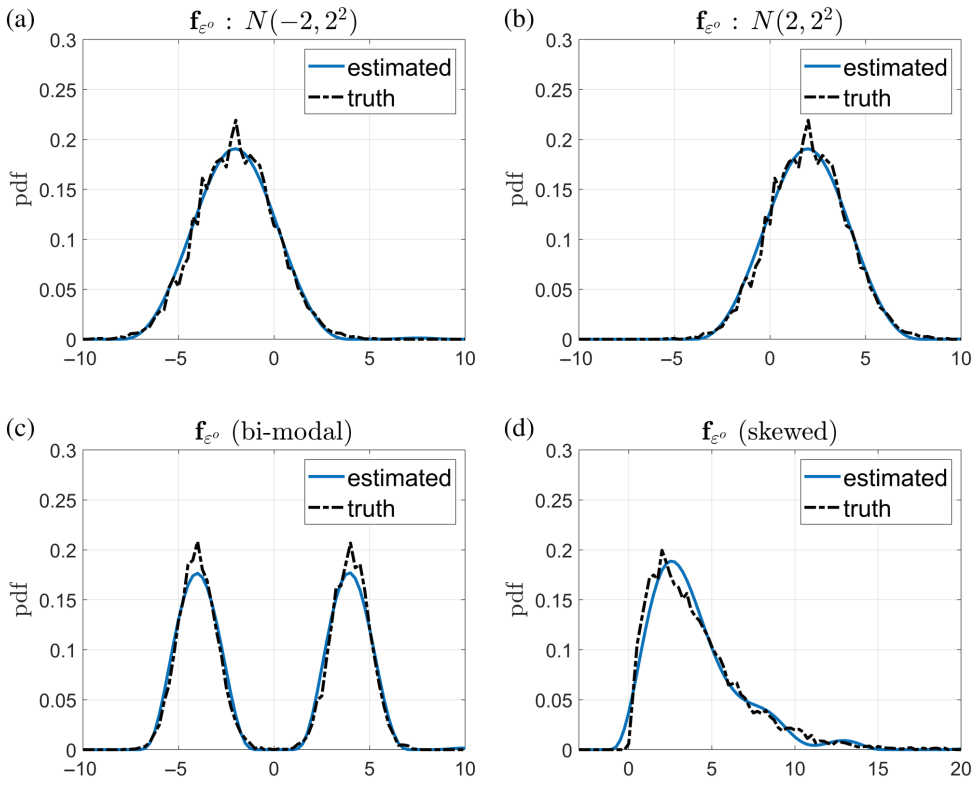

FIGURE 1 The true observation error pdf (black dashed line), and the estimated (blue solid line) observation error pdf from DOEE method. These are the cases where the observation error pdf is (a) Gaussian with a negative mean, (b) Gaussian with a positive mean, (c) bimodal Gaussian and (d) Gamma. [Colour figure can be viewed at wileyonlinelibrary.com]

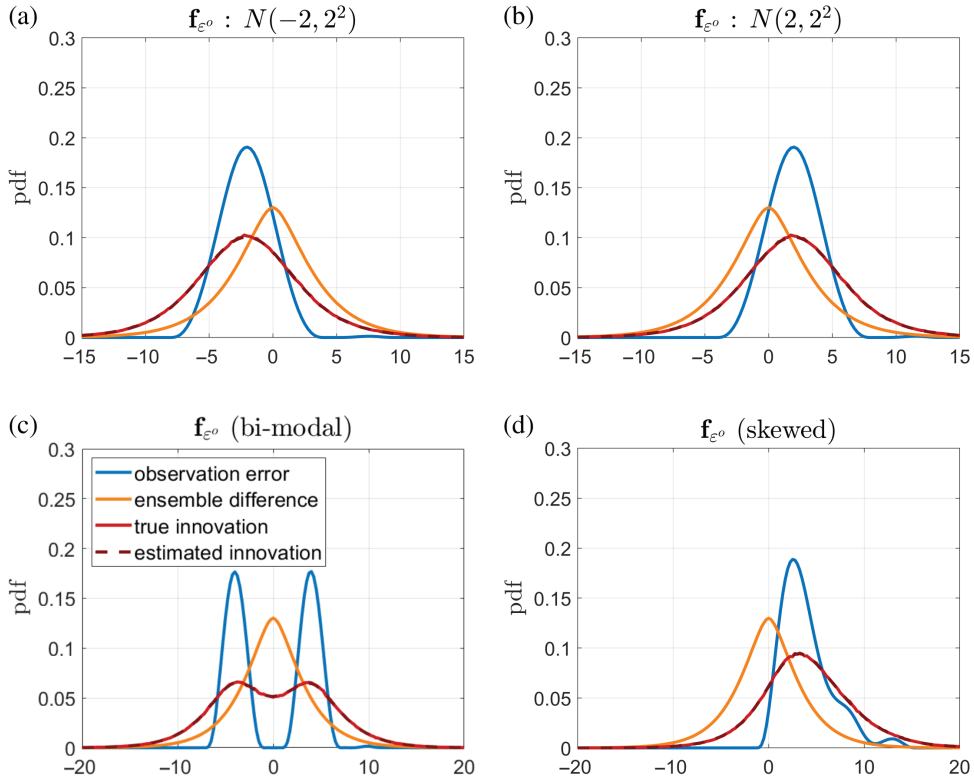

FIGURE 2 The decomposition of the pdfs in Equation (16). The blue line is the estimated observation error pdf f_{ε^o} (the same as the blue lines in Figure 1), the orange line is the ensemble difference pdf f_ε, the solid red line is the innovation pdf f_D, and the dashed brown line is the estimated innovation pdf given by the convolution between the estimated observation error pdf f_{ε^o} and the ensemble difference pdf f_ε. These are the cases where the observation error pdf is (a) Gaussian with a negative mean, (b) Gaussian with a positive mean, (c) bimodal Gaussian and (d) Gamma. [Colour figure can be viewed at wileyonlinelibrary.com]

if this assumption is correct. Therefore, here we examine how the estimates change when the i.i.d. assumption is not true. First, we examine the case that the ensemble is underdispersive. Specifically, we generate the truth and the ensemble based on the following distributions:

$$H\left(X^{\mathrm{t}}\right) \sim \mathrm{Gamma}(2+S_{\mathrm{h}}, 2+S_{\mathrm{c}}) + \mathrm{Normal}\left(0, 1.5^2\right),$$

$$H(X) \sim \mathrm{Gamma}(2+S_{\mathrm{h}}, 2+S_{\mathrm{c}}).$$

The ensemble is generated from the same distribution as in the previous sections, while the truth is generated by a wider distribution. We take all the observation errors to be Gaussian in this section. We conduct a similar experiment as described in Section 3.1, but in this case the estimation of f_ε is based on the underdispersive ensemble. Figure 3a shows that the estimated observation error becomes overdispersive. We conduct another experiment when the ensemble is overdispersive (by swapping the distribution of the truth and the ensemble from the previous experiment), and the estimated observation error becomes underdispersive (Figure 3b).

Figure 3c,d show the results when the ensemble is biased from the truth state. When the ensemble has a positive bias, we can expect there is a negative bias in the estimated observation error (Figure 3c), and vice versa (Figure 3d). Recall that in Figure 1a,b we have also seen an observation error pdf that is not centered at zero. However, note that in Figure 1a,b, the ensemble is drawn from the same distribution as the truth, while that is not the case for Figure 3c,d. The biases in Figure 1a,b are totally attributed to the errors in the observation operator (or the measurement noise, although less likely), while the biases in Figure 3c,d are totally attributed to the poor quality of the ensemble, and hence the errors in the model. In theory, we should deal with these two kinds of biases differently: when the error comes from the observation operator (like in Figure 1a,b), we should include them into the observation error. In contrast, if the error comes from the ensemble, we should include them into the background error (like in Figure 3c,d), that is, remove the bias in the prior ensemble model state. In practice, however, since we do not know if the truth observation error is biased or not, we cannot distinguish which is the case when we obtain a biased observation error pdf like in Figure 1a,b or Figure 3c,d.

Finally, we examine the sensitivity of DOEE to the sample size. Specifically, we are trying to answer the question: how does the choice of the reference member affect the estimation of the ensemble difference pdf f_ε and the observation error pdf f_{ε^o}. Here we examine three ways to estimate the ensemble difference pdf: (1) We always choose the first member as the reference. (2) Every time we randomly pick a different ensemble member as the reference. (3) We calculate the pdf of the pairwise ensemble differences (there is in total $\frac{N_{\mathrm{e}}(N_{\mathrm{e}}-1)}{2}$ number of pairs where N_{e} is the ensemble size). We compare the true ensemble difference pdf and the estimated one from the

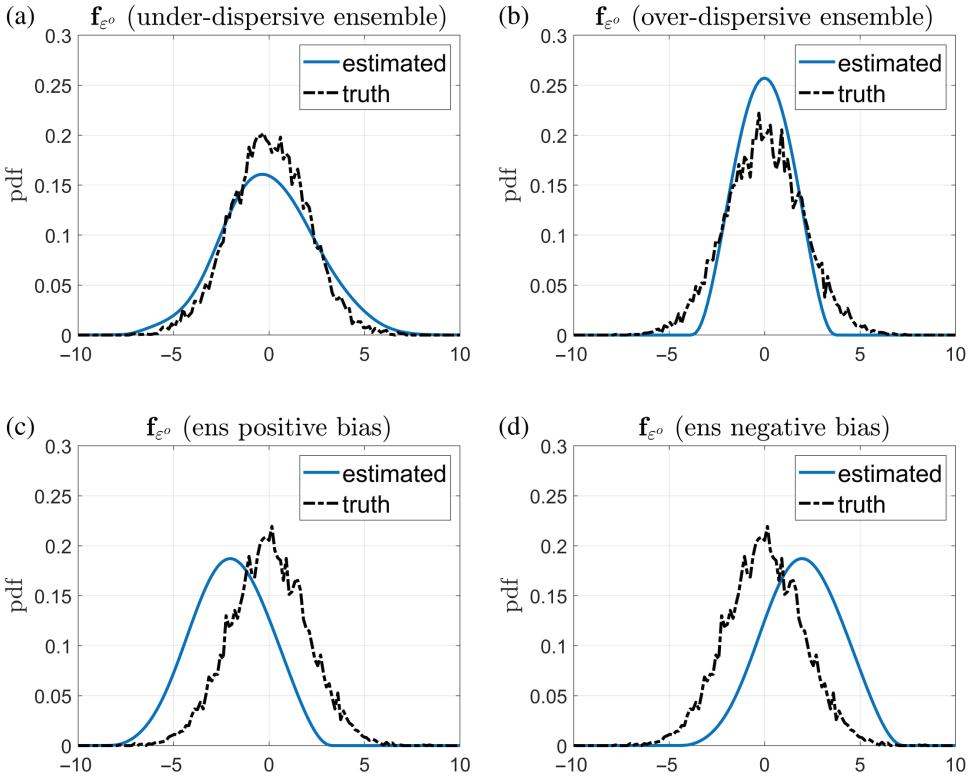

FIGURE 3 Similar to Figure 1, but when the i.i.d. assumption for the ensemble is invalid. (a) When the ensemble is too narrow, (b) when the ensemble is too wide, (c) when the ensemble has a positive bias, and (d) when the ensemble has a negative bias. [Colour figure can be viewed at wileyonlinelibrary.com]

three methods with 100 ensemble members and 10,000 observations in Figure 4a,c. It is hard to tell the differences between the true pdf and the three estimated pdfs, suggesting that all of the methods provide an accurate estimate of f_ε. Since the ensemble difference pdfs are very similar, it is not surprising that the estimated observation error pdfs are also very similar for the three methods (not shown). We redo the experiment, but now reduce the ensemble size from 100 to 10, and the number of observations from 10,000 to 100. The results are shown in Figure 4d–f. We do see some minor differences in the fine structure of the estimated ensemble difference pdfs, while the overall shape of the pdfs are still very similar, as are the estimated observation error pdfs (not shown).

We conclude that the estimated observation error pdf is not sensitive to the choice of the reference member, and DOEE works well even for small ensemble sizes and number of observations.

4 | AN APPLICATION OF DOEE TO REAL DATA

4.1 | Background and data descriptions

In this section, we apply DOEE to a real dataset. We will look at the observation at 37 GHz vertical polarization measured by the Special Sensor Microwave Imager/Sounder (SSMIS) (Kunkee et al., 2008). SSMIS is onboard the DMSP (Defense Meteorological Satellite Program) satellite, which is polar-orbiting and provides global coverage about twice per day. We collect the observations during September 2015. The observations are paired with an ensemble, which is from the local ensemble transform Kalman Filter (LETKF) from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Bonavita et al., 2020), and the ensemble size is 100. Due to the limitation of the radiative transfer model, observations over land and over the regions where the latitude is over 60 degrees are excluded from the analyses. After data screening, there are in total around 900,000 pixels, which should provide enough samples for the analysis.

Previous studies have suggested that the observation error for the all-sky radiance observations can depend on the cloud amount (e.g., Geer & Bauer, 2011). Specifically, we assume that f_{ε^o} depends on the cloud amount c, which in this study is defined as,

$$c = 1 - \frac{T_{\mathrm{B;19V}} - T_{\mathrm{B;19H}}}{T_{\mathrm{Bclear;19V}} - T_{\mathrm{Bclear;19H}}} \in [0, 1], \tag{22}$$

where $T_{\mathrm{B;19V}}$, $T_{\mathrm{B;19H}}$ are the vertical and horizontal polarized brightness temperatures at 19 GHz, and $T_{\mathrm{Bclear;19V}}$, $T_{\mathrm{Bclear;19H}}$ are the vertical and horizontal polarized brightness temperatures at 19 GHz assuming the cloud is absent. The advantage of using Equation (22) is that we can have a consistent definition for the cloud amount among the ensemble members and the observations. $T_{\mathrm{Bclear;19V}}$,

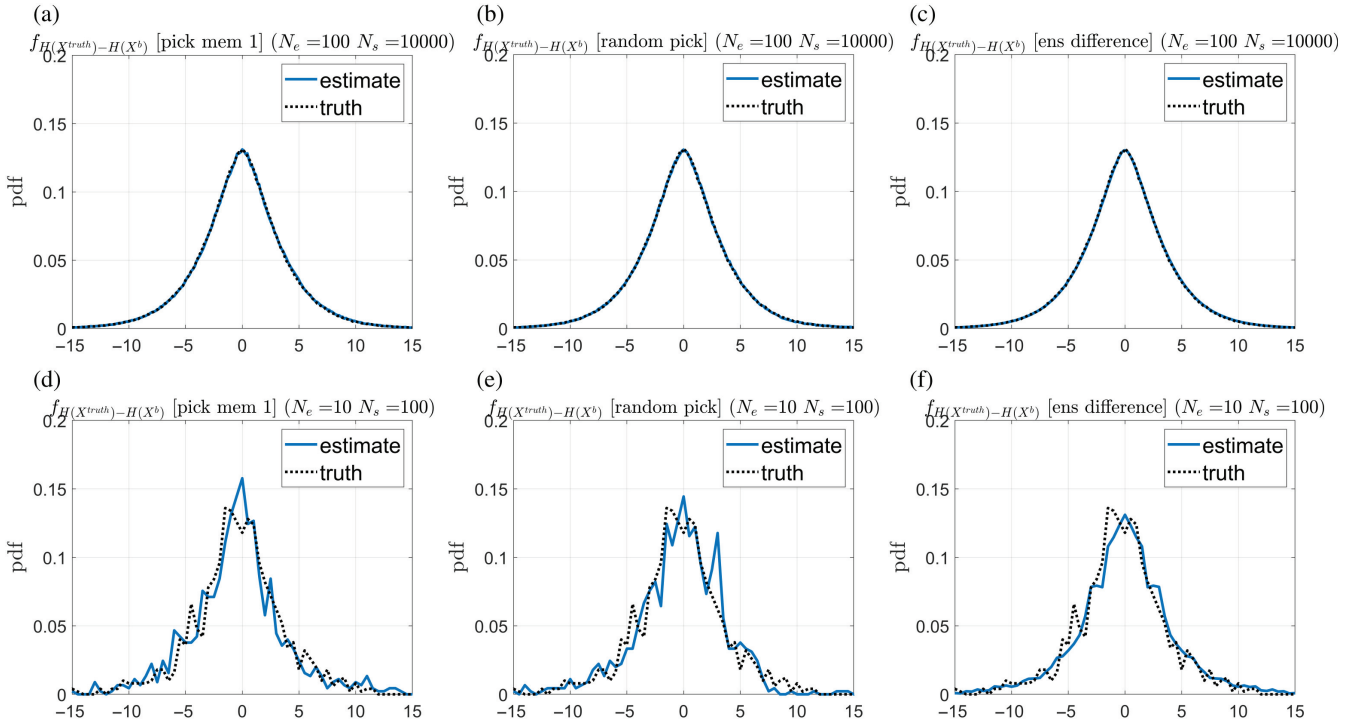

FIGURE 4 The true (black dashed line) and the estimated ensemble difference pdf f_ε (blue solid line) for the sensitivity experiments in Section 3.2. To estimate f_ε, (a, d) the first member is always chosen as the truth (b, e) the truth is randomly chosen from the ensemble members (c, f) the ensemble differences are used for the estimation. The ensemble size N_e and the number of observation N_s are (a–c) $N_e = 100$, $N_s = 10000$ and (d–f) $N_e = 10$, $N_s = 100$. [Colour figure can be viewed at wileyonlinelibrary.com]

$T_{\mathrm{Bclear;19H}}$ are estimated using a high-resolution control run and they are assumed to be the same among the observation and all the ensemble members at the same grid point. Although $T_{\mathrm{Bclear;19V}}$ and $T_{\mathrm{Bclear;19H}}$ depend on variations in temperature, specific humidity and skin temperature, which can be different for each ensemble member and the control run, the ensemble spread in these variables only cause minor variations in simulated brightness temperature across the ensemble, when compared to the variations caused by the cloud amount. This supports our approximations for $T_{\mathrm{Bclear;19V}}$, $T_{\mathrm{Bclear;19H}}$ using the control run. However, since the approximation is still not perfect, especially for the clear-sky end of the cloud amount predictor (e.g., see Lonitz & Geer, 2020), sometimes it could lead to a negative cloud amount. Therefore, here we impose a lower bound of zero on the calculated cloud amount. Note that the 19 GHz cloud predictor used here differs from the 37 GHz predictor used in e.g. Geer and Bauer (2011). This is because the 37 GHz h-polarized channel is not actively assimilated and prior brightness temperatures for this channel are not available in the LETKF.

Since the cloud amount is a continuous variable, we discretize the cloud amount when building the cloud-amount-dependent observation error pdf. In practice, we assume that the observation error pdf f_{ε^o} is the same pdf within a finite range of cloud amount. Recall Section 2.4 summarizes the way to estimate this cloud-amount-dependent error model. For example, assume we would like to estimate the observation error when cloud amount is within [0.2, 0.3]. We collect the observations with observed cloud amount $c_{\mathrm{obs}} \in$[0.2, 0.3] and calculate the difference between the observations and all of the ensemble members (regardless of their cloud amount), to estimate the pdf of $Y_{|c(x^t)\in[0.2,0.3]} - H(X)$. To estimate the pdf of $H\left(x^{\mathrm{t}}{}_{|c(x^t)\in[0.2,0.3]}\right) - H(X)$, we pick an ensemble member from the indistinguishable subset, which in this case is the subset of the ensemble with cloud amount within [0.2, 0.3], and calculate the differences between the member in the indistinguishable subset and the other ensemble members. Note that a partly similar idea was used by Chambon et al. (2014) to estimate the biases of all-sky microwave radiances using a sample in which both observations and background had similar cloud amounts.

To have an idea of the relation between the observed cloud amount and the ensemble cloud amount, we first examine the logarithm of their joint distribution in Figure 5. It Is interesting to see that the joint histogram is not symmetric. Although it seems that the ensemble has too large cloud amount in general (e.g., compare the values along the 1–4 line and the 4–1 line in Figure 5), the ensemble has too small cloud amount in the bin with cloud

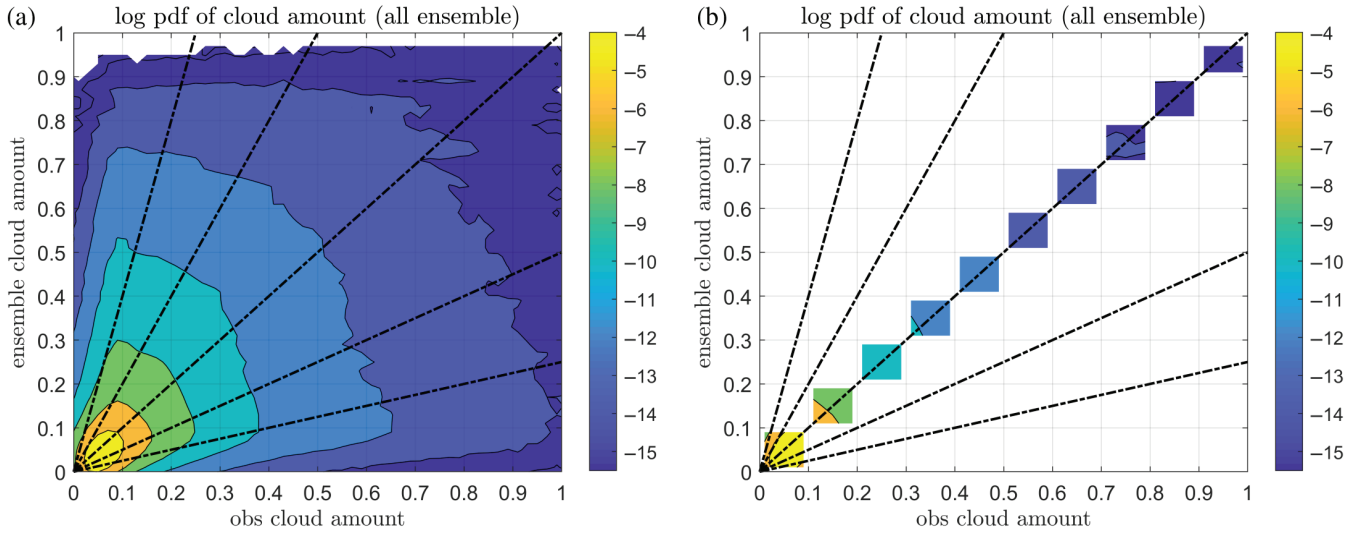

FIGURE 5 The logarithm of the joint distribution of the observed cloud amount and the ensemble cloud amount. The bin size of the joint histogram is 0.02. The black dashed lines are the 1–4, 1–2, 1–1, 2–1, and 4–1 line for reference. (a) for all ensemble members (b) for the ensemble members in the indistinguishable subset (see text) [Colour figure can be viewed at wileyonlinelibrary.com]

amount less than 0.1. In summary, the observed cloud amount can sometimes be very different from the ensemble cloud amount, and the spread of the ensemble cloud amount can be very large (we can see, for example, by looking at the marginal pdf of the ensemble cloud amount given a value of the observed cloud amount in Figure 5a).

4.2 | A non-parametric state-dependent observation error model

We examine the cloud-amount-dependent observation error model constructed based on Equation (18) in this section. We divide the cloud amount into 10 categories, from 0 to 1 and with bin size 0.1. Figures 6 and 7 show the pdf f_{ε^o} for each cloud amount category, and all the pdfs used in DOEE in Equation (16). Note that for the mostly-clear-sky case (i.e., the cloud amount being [0.0, 0.1]), the pdf is right-skewed and the tail is long at the positive values (Figure 6a). We attribute this right-skewness to the cloud amount asymmetry between the observation and the ensemble in this category. (Note that this asymmetry only occurs in the mostly-clear-sky category.) In this category, the joint histogram has larger mass in the region where the observed cloud amount is larger than the ensemble cloud amount (Figure 5b). In addition, in this channel, the brightness temperature in general increases with the increasing cloud amount (Figure 8). Therefore, if the observation has larger cloud amount compared to the ensemble, the brightness temperature of the observation will be larger than the ensemble, leading to positive observation errors.

In general, the standard deviation of the observation error pdf increases with cloud amount when the cloud amount is between 0.0 and 0.5 (Figure 6a–e), and it remains constant or slightly decreases with cloud amount when the cloud amount is between 0.5 and 1.0 (Figure 6f–j). This result is qualitatively consistent with Geer and Bauer (2011), but quantitatively the standard deviations are smaller here (around 5 K here compared to 15–20 K there). The reason is that Geer and Bauer (2011) estimate the errors from the strong-constraint 4D-Var perspective, where the observation error includes the errors in the forecast model during the assimilation window, while here our estimations do not account for the errors in the forecast model. We will elaborate on the differences between these two errors in Section 5.

The mean of the observation error pdf, in general, increases with cloud amount as well. Most of the pdfs are not Gaussian-like, and especially some are skewed, for example, [0.0, 0.1], [0.2, 0.3], and [0.4, 0.5]. This suggests that Gaussian assumption for the observation errors used in many current DA schemes can be problematic when assimilating the satellite radiances.

We note that the cloud amount is just one choice for the predictor for the state-dependent observation error model for satellite radiances. However, we realize that the true observation error model can be more complicated. Therefore, we also examine a slightly more complicated state-dependent observation error model by adding the total column water vapor (TCWV) as a new predictor. In other words, we have two predictors for the observation error model: the cloud amount defined in Equation (22) and the TCWV. Specifically, we further divide the pixels in each category of cloud amount into two: the TCWV being larger or smaller than 40 $kg \cdot m^{-2}$ in the high-resolution

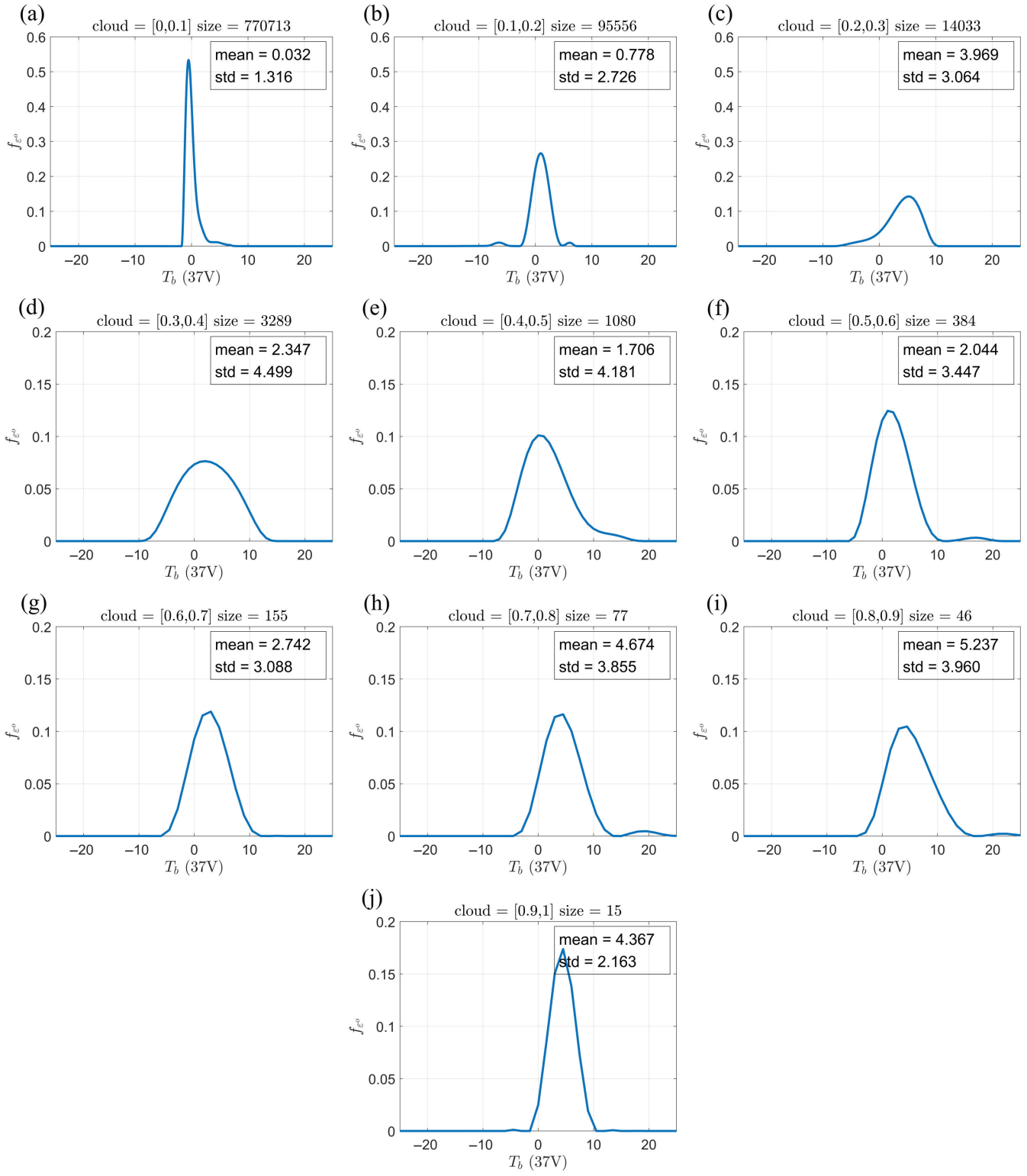

FIGURE 6 The cloud-amount-dependent observation error model for satellite radiance at 37 GHz vertical polarization, for cloud amount (a) [0.0, 0.1], (b) [0.1, 0.2], (c) [0.2, 0.3], (d) [0.3, 0.4], (e) [0.4, 0.5], (f) [0.5, 0.6], (g) [0.6, 0.7], (h) [0.7, 0.8], (i) [0.8, 0.9], and (j) [0.9, 1.0]. [Colour figure can be viewed at wileyonlinelibrary.com]

control run. (Since TCWV for each ensemble member and the observation are not available in the current dataset, we only use TCWV in the high-resolution control run. However, the variations of TCWV are believed to be small, so the TCWV in the control run should be representative enough.) The reason for choosing 40 $kg \cdot m^{-2}$ as the criterion is based on the previous studies showing that 40–60 $kg \cdot m^{-2}$ is the threshold for tropical convection (e.g.,

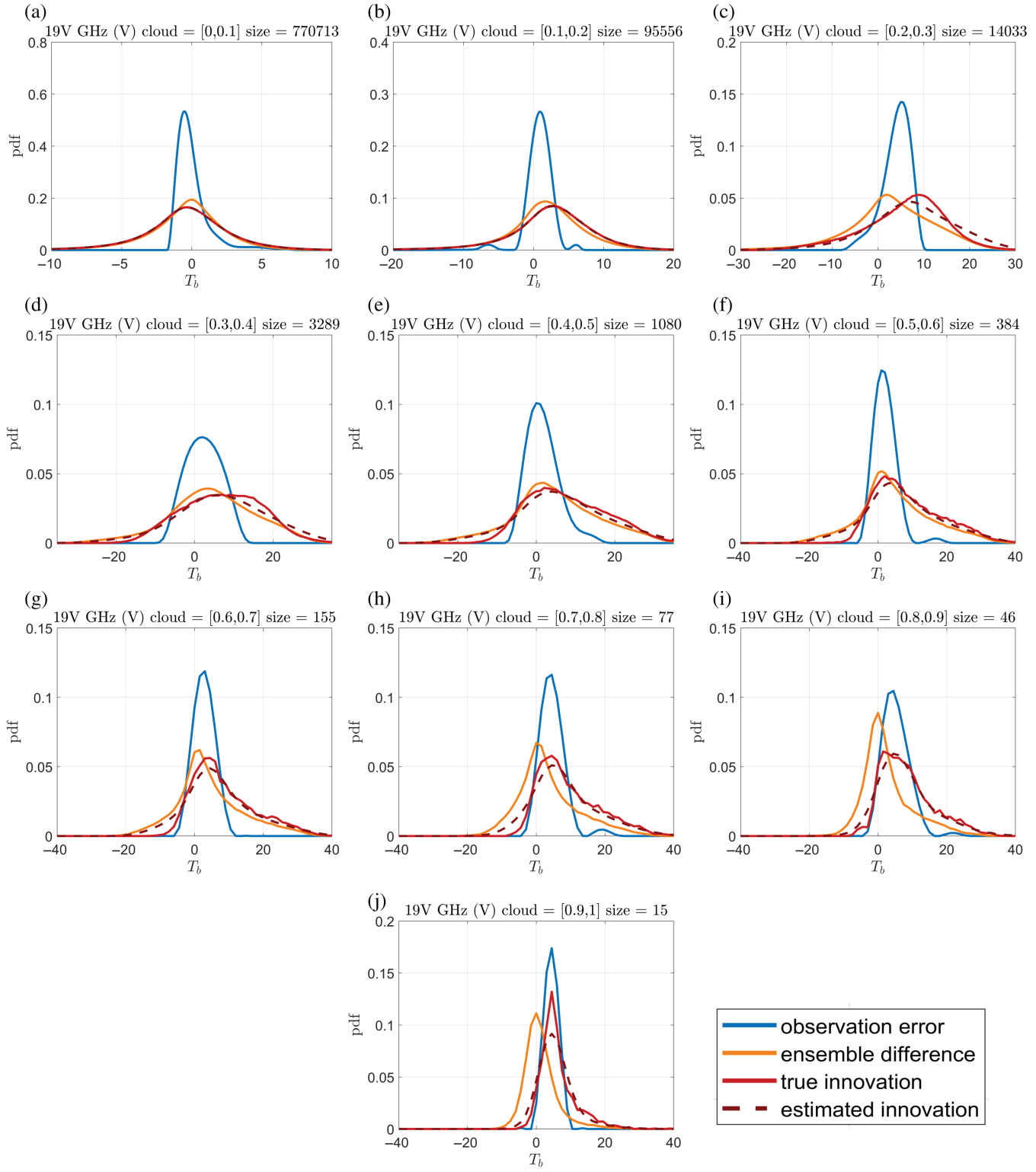

FIGURE 7 Similar to Figure 2, but for the cloud-amount-dependent observation error model for satellite radiance at 37 GHz vertical polarization. The blue line in each subplot is exactly the same as in Figure 6. [Colour figure can be viewed at wileyonlinelibrary.com]

Bretherton et al., 2004; Schiro et al., 2016), which we believe could have very different observation error properties compared to other cloud types.

Figure 9 shows the observation error pdfs as a function of cloud amount and TCWV. We first note that if the observation error pdf is only a function of cloud amount, then we can expect that the pdfs for TCWV < 40 (blue curves in Figure 9) be similar to the pdfs for TCWV > 40 (red curves in Figure 9) in each category, which is however not the case here. On the other hand, the observation error

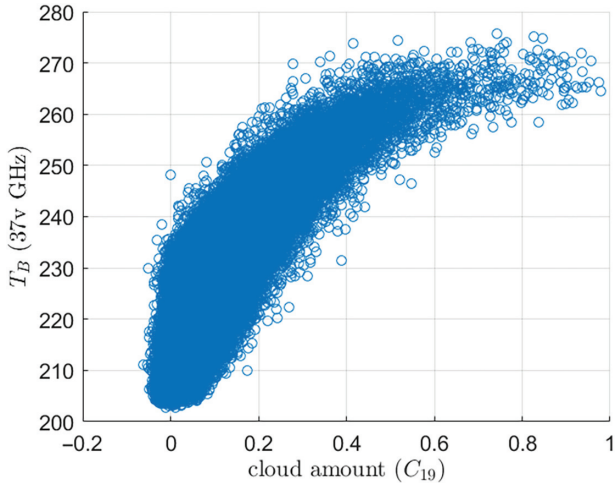

FIGURE 8 The relation between the observed cloud amount defined by the 19 GHz normalized polarization difference in Equation (22) and the observed radiance at 37 GHz vertical polarization brightness temperature. [Colour figure can be viewed at wileyonlinelibrary.com]

pdf is not only a function of TCWV either. If it is only a function of TCWV, then all the red curves in Figure 9 should be similar, and the same for all the blue curves. This suggests that the true state-dependent observation error model for radiances at this channel is indeed more complicated. Constructing a more accurate state-dependent observation error model is important as we can better utilize the information from the observations.

It is interesting to find that the positive tail of the observation errors in the mostly-clear-sky category (cloud amount [0.0, 0.1]) is largely associated with pixels with smaller TCWV. Specifically, Figure 9a shows that when TCWV < 40, there is a similar right-skewness in the observation error pdf (Figure 6a), while when TCWV > 40, the observation error pdf is symmetric and close to Gaussian (Figure 9a). For cloud amount [0.1, 0.2], the observation error pdf for TCWV<40 is wider than TCWV > 40, and both of their means are slightly larger than 0 (Figure 9b). This is consistent with Figure 6b.

The difference between the pdfs for TCWV < 40 and TCWV > 40 becomes clear when the cloud amount is larger than 0.2 (Figure 9c–f). It is interesting to see that, when TCWV < 40, the pdfs for cloud amount larger than 0.2 (blue curves in Figure 9c–f) are very similar: they are symmetric, Gaussian-like, and centered close to 0, whereas when TCWV > 40, the pdfs (red curves in Figure 9c–f) are very different but all of them have mean larger than 0. Note that the pixels with cloud and TCWV > 40 are likely

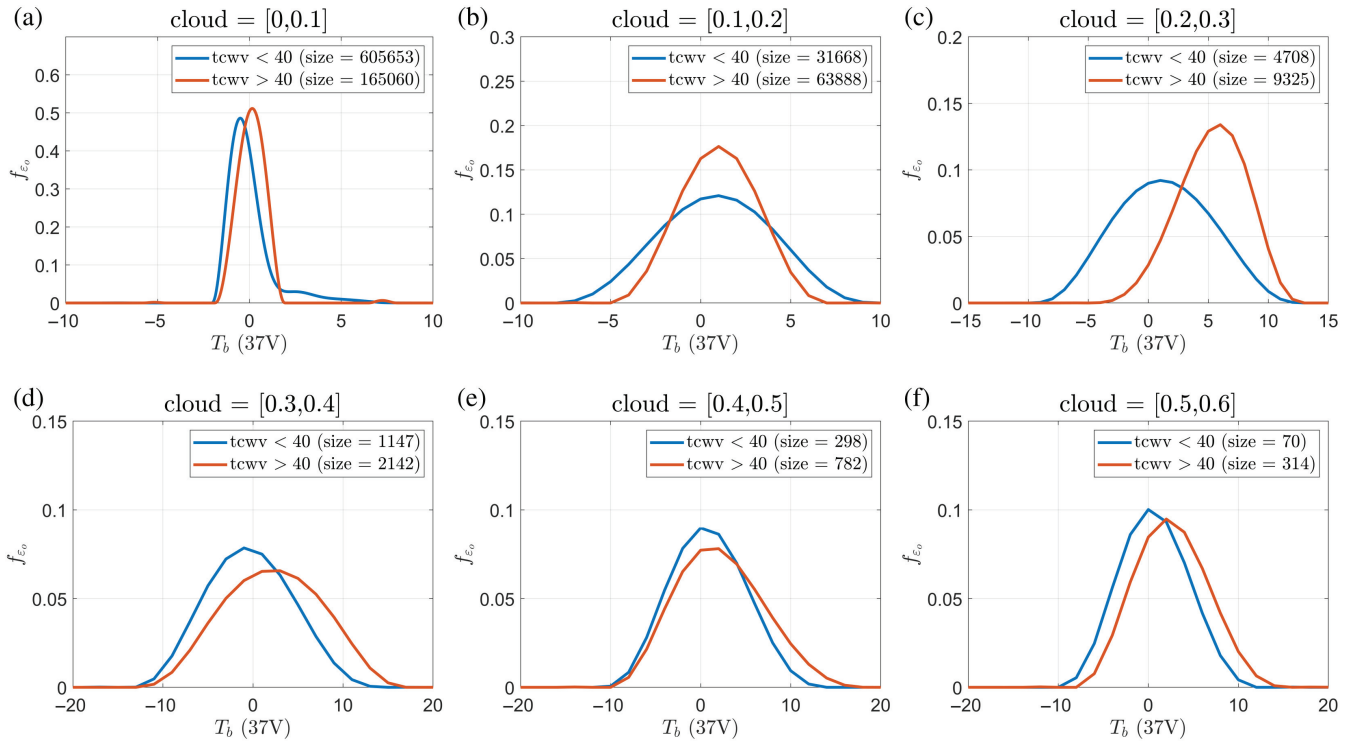

FIGURE 9 Similar to Figure 6, but the samples in each cloud amount category are further divided into two based on the total column water vapor (TCWV). The blue curve is the observation error pdf when TCWV < 40 kg·m^{-2}, and the red curve is when TCWV > 40 kg·m^{-2}. Note that only the pdfs for cloud amount smaller than 0.6 are shown, since there are less than 10 samples for TCWV < 40 kg·m^{-2} in the cloud amount category when the cloud amount is larger than 0.6. The error model for cloud amount being (a) [0.0,0.1], (b) [0.1,0.2], (c) [0.2,0.3], (d) [0.3,0.4], (e) [0.4,0.5], (f) [0.5,0.6]. [Colour figure can be viewed at wileyonlinelibrary.com]

to be associated with convective clouds. These results at least suggest two things: first, the non-Gaussianity and skewness in the observation error for cloud amount larger than 0.2 (Figure 6c–f) can be mostly attributed to the pixels with larger TCWV. Second, there are biases in the pixels with TCWV > 40 when cloud amount is larger than 0.2. As discussed in Section 3.2, we cannot distinguish where the biases originate by simply looking at the results: they might either come from the model biases, or from the errors in the observation operators. Note that we only examine the pdfs up to cloud amount being 0.6 in Figure 9 since the sample size for the sub-category TCWV<40 becomes too small when cloud amount is larger than 0.6.

5 | DISCUSSIONS AND CONCLUSIONS

In this study, we propose a new methodology, the Deconvolution-based Observation Error Estimation (DOEE), to estimate the full probability density function (pdf) of the observation error. DOEE does not rely on the linear approximation for the observation operator, the optimality of the DA algorithm, or the Gaussian assumptions for the error pdf. DOEE can theoretically retrieve any form of the observation error pdf based on the statistics of the observations and an ensemble of model equivalences. We demonstrate a series of idealized experiments in Section 3, and find that DOEE can successfully retrieve the observation error pdf when it is either a Gaussian or a non-Gaussian (bimodal, bounded or skewed) distribution.

The key equations we use in the DOEE are Equation (16). To have an estimate of the pdf of the difference between the true state and a random forecast state in observation space, we propose to randomly pick an ensemble member as a sample for $H(x^t)$. This is based on the assumption that the ensemble is indistinguishable from the truth (the i.i.d. assumption). In addition, we can construct a state-dependent observation error model based on Equation (18). In this case, instead, we propose to select the ensemble member from the "indistinguishable subset", which is a subset of the ensemble with a certain value or a certain range of the predictor $c(x) \in C$, to sample $H(X^t{}_{|c(X^t)\in C})$.

We applied the DOEE method to construct state-dependent observation error models for the satellite radiances at 37 GHz vertical polarization. We first construct a cloud-amount-dependent observation error model (see Figure 6). The results suggest that: (1) the standard deviation of the pdf generally increases with cloud amount. (2) Many of the pdfs are skewed and non-Gaussian. Furthermore, we examine a slightly more complicated state-dependent observation error model by adding the TCWV as a second predictor (see Figure 9). What we can learn from this slightly more complicated model is: (1) the cloud amount alone may not be the optimal predictor for the observation error model, since the pdfs for TCWV < 40 and TCWV > 40 at the same cloud amount category can be very different. (2) For the mostly-clear-sky category (cloud amount [0.0, 0.1]), the non-Gaussianity in the observation error pdf is associated with the pixels with TCWV < 40, whereas for the cloudy categories (cloud amount 0.2–0.6), the non-Gaussianities are associated with the pixels with TCWV > 40. (3) For cloudy categories (cloud amount 0.2–0.6), the pdfs for TCWV < 40 are similar, which might suggest the observation errors in these pixels come from the same distribution. In contrast, the pdfs for TCWV > 40 differ for different cloud amount categories, while all of them have positive mean. This suggests there are some biases in the simulated radiances from these cloudy pixels with TCWV >40. Exploring other predictors for the observation error models, like cloud type, would be an interesting topic for the future work. Note that with a better set of predictors for the state-dependent observation error model, we can better utilize the information from the observations.

We should emphasize that, as in Section 4, we concentrate on the observation operator from a filter perspective. When DOEE is applied for a strong-constraint 4D-Var, the representation errors will also include errors in the model equations. Specifically, we denote the radiative transfer model (RTM) as H and the model forecast from the beginning of the window to the observation time as M. For the observation error of RTM alone, we are looking at the errors in H, while for the observation error from the strong-constraint 4D-Var perspective, we are looking at the errors in the composite function $H \circ M$. There is no fundamental difference in estimating the observation error for either H or $H \circ M$ using DOEE. However, when constructing a state-dependent model in Equation (18), we need to be cautious about the assumption $c(x^t) \approx c_{\text{obs}}$. For example, when it applies to constructing a similar cloud-amount- (but at the observation time) dependent error model for $H \circ M$, the assumption that $c(M(x^t)) \approx c_{\text{obs}}$ becomes less appropriate because we do recognize that the model M typically has large errors in predicting clouds. Therefore, it becomes difficult to construct a state-dependent observation error model for $H \circ M$ when the predictor is the cloud amount at the observation time. Instead, we may choose the TCWV at the observation time as a predictor and assume that $\text{TCWV}(M(x^t)) \approx \text{TCWV}_{\text{obs}}$ since the model will typically predict TCWV more accurately than the cloud amount. Exploring a proper predictor for the state-dependent observation error models from the strong-constraint 4D-Var perspective is left for future work.

In this work, DOEE is applied to estimate the univariate pdf for each scalar observation, which provides enough information if the observation errors are independent. When the observation errors are not independent, e.g., due to interchannel or spatial correlations for satellite radiances, we can still apply DOEE to estimate the full joint pdf of the observation error pdf. This can be achieved by generalizing Equations (20) and (21) for the multivariate pdf. We also leave this exploration as future work.

It is also important to note the limitations of the assumptions in DOEE. First, although the i.i.d. assumption is commonly used in ensemble methods, it may not always be appropriate, for example, when the model error is significant but not (sufficiently) accounted for in the ensemble. DOEE becomes suboptimal when the i.i.d. assumption is compromised (e.g., see Figure 3). To address this issue, we could discard the samples in which the ensemble is considered to violate the i.i.d. assumption in the DOEE estimate. We could also modify the ensemble before the ensemble data are used in DOEE. For example, if the ensemble is considered underdispersive in some cases, model error can be added to the ensemble before DOEE. We may explore using anchor observations, or cross-validation methods (e.g., Marseille et al., 2016; Ménard & Deshaies-Jacques, 2018; Tandeo et al., 2020) to justify the output from DOEE, when there are doubts about the validity of the i.i.d. assumption. Second, the assumption that the background error and the observation error are independent can be problematic for some cases, especially for the observation error model of the all-sky radiances from the strong-constraint 4D-Var perspective. While it is possible in principle to include dependent background and observation errors, the innovation pdf will not be a simple convolution anymore, and more work is needed to extend DOEE to this case.

Finally, we briefly comment on ways to include the non-parametric observation errors in some current DA methods. DA methods that do not require a parametric form of the observation error, for example, particle filters and particle flow filters (e.g., Ades & van Leeuwen, 2015; Daum & Huang, 2011; Hu & Van Leeuwen, 2021; Poterjoy, 2016; Pulido & van Leeuwen, 2019; van Leeuwen et al., 2019) or the quantile-conserving filter (Anderson, 2022), would be natural choices to assimilate observations with non-parametric observation errors. For methods that require a parametric form for the observation error, we may, for example, locally approximate the non-parametric pdf by the required parametric pdf. If the observation error is state-dependent and we are using a variational method, we may need to re-approximate the non-parametric pdf by a different Gaussian in each outer loop. In ensemble Kalman filters, the observation error variance should be the expected value of $(\varepsilon^o)^2$ over the prior distribution (Bishop, 2019). In this case, we have to estimate $(\varepsilon^o)^2$ for the pdf in each category, and average these values based on the prior distribution of the predictors.

In summary, DOEE is an important methodological step forward in diagnosing the non-Gaussianity of the observation error. In addition to the non-Gaussianity, the state-dependent error, spatial and temporal correlations of the error, and the separability of the background and the observation error are also important aspects of the observation error. Although it is not clear that whether the non-Gaussianity is the main challenge that needs priority to be addressed in a numerical weather prediction (NWP) system, the diagnosed non-Gaussianity in real observations is clear. We are currently implementing DOEE and testing the inclusion of the non-Gaussian observation error in a full-scale NWP system. The impact of the non-Gaussianity on the DA and NWP will be explored in future studies.

ACKNOWLEDGEMENTS

The authors thank the Cooperative Institute for Research of the Atmosphere (CIRA) for the funding support under the project number NA19OAR4320073 from NOAA, and support via the Assisting Students, Cultivating Excellence, Nurturing Talent (ASCENT) award from the Department of Atmospheric Science at Colorado State University (CSU). Valuable comments from Neils Bormann, Tony McNally, Ricardo Todling and an anonymous reviewer that helped improve the quality of the manuscript are highly appreciated.

DATA AVAILABILITY STATEMENT

The data and code in the idealized experiment are available at: https://github.com/ChihChiHu1125/DOEE.

ORCID

Chih-Chi Hu https://orcid.org/0000-0002-3020-8975
Peter Jan van Leeuwen https://orcid.org/0000-0003-2325-5340
Alan J. Geer https://orcid.org/0000-0002-9476-5519

REFERENCES

Ades, M. & van Leeuwen, P.J. (2015) The equivalent-weights particle filter in a high-dimensional system. *Quarterly Journal of the Royal Meteorological Society*, 141, 484–503. Available from: https://doi.org/10.1002/qj.2370

Anderson, J.L. (1996) A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9(7), 1518–1530.

Anderson, J.L. (2007) An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, 59, 210–224. Available from: https://doi.org/10.1111/j.1600-0870.2006.00216.x

Anderson, J.L. (2009) Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A*, 61, 72–83. Available from: https://doi.org/10.1111/j.1600-0870.2008.00361.x

Anderson, J.L. (2022) A quantile-conserving ensemble filter framework. *Part I: Updating an Observed Variable, Monthly Weather Review*, 150(5), 1061–1074. Available from: https://doi.org/10.1175/MWR-D-21-0229.1

Andersson, E. (2003) Modelling of innovation statistics. In: *Of proceedings of workshop on recent developments in data assimilation for atmosphere and ocean*. Reading: ECMWF, pp. 153–164.

Anthes, R. & Rieckh, T. (2018) Estimating observation and model error variances using multiple data sets. *Atmospheric Measurement Techniques*, 11(7), 4239–4260. Available from: https://doi.org/10.5194/amt-11-4239-2018

Bauer, P., Geer, A.J., Lopez, P. & Salmond, D. (2010) Direct 4D-Var assimilation of all-sky radiances. Part I: implementation. *Quarterly Journal of the Royal Meteorological Society*, 136, 1868–1885. Available from: https://doi.org/10.1002/qj.659

Bishop, C.H. (2016) The GIGG-EnKF: ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. *Quarterly Journal of the Royal Meteorological Society*, 142, 1395–1412. Available from: https://doi.org/10.1002/qj.2742

Bishop, C.H. (2019) Data assimilation strategies for state-dependent observation error variances. *Quarterly Journal of the Royal Meteorological Society*, 145(718), 217–227. Available from: https://doi.org/10.1002/qj.3424

Bonavita, M., Geer, A.J. & Hamrud, M. (2020) All-sky microwave radiances assimilated with an ensemble Kalman filter. *Monthly Weather Review*, 148(7), 2737–2760. Available from: https://doi.org/10.1175/MWR-D-19-0413.1

Bonavita, M., Lean, P. & Holm, E. (2018) Nonlinear effects in 4D-Var. *Nonlinear Processes in Geophysics*, 25, 713–729. Available from: https://doi.org/10.5194/npg-25-713-2018

Bormann, N. & Bauer, P. (2010) Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: methods and application to ATOVS data. *Quarterly Journal of the Royal Meteorological Society*, 136, 1036–1050. Available from: https://doi.org/10.1002/qj.616

Bormann, N., Bonavita, M., Dragani, R., Eresmaa, R., Matricardi, M. & McNally, A. (2016) Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, 142, 1767–1780. Available from: https://doi.org/10.1002/qj.2774

Bretherton, C.S., Peters, M.E. & Back, L.E. (2004) Relationships between water vapor path and precipitation over the tropical oceans. *Journal of Climate*, 17(7), 1517–1528. Available from: https://doi.org/10.1175/1520-0442(2004)017<1517:RBWVPA>2.0.CO;2

Bringi, V. & Chandrasekar, V. (2001) *Polarimetric Doppler weather radar: principles and applications*. Cambridge: Cambridge University Press. Available from: https://doi.org/10.1017/CBO9780511541094

Chambon, P., Zhang, S.Q., Hou, A.Y., Zupanski, M. & Cheung, S. (2014) Assessing the impact of pre-GPM microwave precipitation observations in the Goddard WRF ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 140, 1219–1235. Available from: https://doi.org/10.1002/qj.2215

Daley, R. (1993) *Atmospheric data analysis (No. 2)*. Cambridge: Cambridge University Press.

Daum, F. & Huang, J. (2011) Particle flow for nonlinear filters, In *Proceedings under International Conference on Acoustics, Speech, and Signal Processing*, ICASSP 2011, Prague. 22–27 May 2011, pp. 5920–5923. https://doi.org/10.1109/ICASSP.2011.5947709

Desroziers, G., Berre, L., Chapnik, B. & Poli, P. (2005) Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131, 3385–3396. Available from: https://doi.org/10.1256/qj.05.108

Freedman, D. & Diaconis, P. (1981) On the histogram as a density estimator: L2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57, 453–476.

Geer, A.J. & Bauer, P. (2011) Observation errors in all-sky data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 137, 2024–2037. Available from: https://doi.org/10.1002/qj.830

Gray, J.E. & Allan, D.W. (1974, May) A method for estimating the frequency stability of an individual oscillator. In: *28th annual symposium on frequency control*. Atlantic City, NJ, IEEE, pp. 243–246.

Grubbs, F.E. (1948) On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association*, 43, 243–264. Available from: https://doi.org/10.1080/01621459.1948.10483261

Hamill, T.M. (2001) Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129, 550–560. Available from: https://doi.org/10.1175/15200493(2001)129<0550:IORHFV>2.0.CO;2

Hodyss, D. & Nichols, N. (2015) The error of representation: basic understanding. *Tellus*, A67, 24822. Available from: https://doi.org/10.3402/tellusa.v67.24822

Hollingsworth, A. & Lönnberg, P. (1989) The verification of objective analyses: diagnostics of analysis system performance. *Meteorology and Atmospheric Physics*, 40, 3–27.

Hu, C.C. & Van Leeuwen, P.J. (2021) A particle flow filter for high-dimensional system applications. *Quarterly Journal of the Royal Meteorological Society*, 147(737), 2352–2374. Available from: https://doi.org/10.1002/qj.4028

Janjić, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L. et al. (2017) On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society 2018*, 144, 1257–1278. Available from: https://doi.org/10.1002/qj.3130

Kunkee, D.B., Poe, G., Boucher, D., Swadley, S., Hong, Y., Wessl, J. et al. (2008) Design and evaluation of the first special sensor microwave imager/sounder. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 863–883. Available from: https://doi.org/10.1109/TGRS.2008.917980

Lien, G.Y., Kalnay, E. & Miyoshi, T. (2013) Effective assimilation of global precipitation: simulation experiments. *Tellus A: Dynamic Meteorology and Oceanography*, 65, 19915. Available from: https://doi.org/10.3402/tellusa.v65i0.19915

Lonitz, K. & Geer, A.J. (2020) Reducing the drying effect through a water vapour correction to the all-sky error model. *ECMWF Tech. Memo.*, 53, 27. Available from: https://doi.org/10.21957/qmy8utbgb

Marseille, G.-J., Barkmeijer, J., de Haan, S. & Verkley, W. (2016) Assessment and tuning of data assimilation systems using passive observations. *Quarterly Journal of the Royal Meteorological*

Society, 142, 3001–3014. Available from: https://doi.org/10.1002/qj.2882

Ménard, R. & Deshaies-Jacques, M. (2018) Evaluation of analysis by cross-validation, part II: diagnostic and optimization of analysis error covariance. *Atmosphere*, 9(2), 70. Available from: https://doi.org/10.3390/atmos9020070

Minamide, M. & Zhang, F. (2019) An adaptive background error inflation method for assimilating all-sky radiances. *Quarterly Journal of the Royal Meteorological Society*, 145, 805–823. Available from: https://doi.org/10.1002/qj.3466

Pires, C.A., Talagrand, O. & Bocquet, M. (2010) Diagnosis and impacts of non-Gaussianity of innovations in data assimilation. *Physica D*, 239, 1701–1717. Available from: https://doi.org/10.1016/j.physd.2010.05.006

Poterjoy, J. (2016) A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, 144(1), 59–76. Available from: https://doi.org/10.1175/MWR-D-15-0163.1

Pulido, M. & van Leeuwen, P.J. (2019) Sequential Monte Carlo with kernel embedded mappings: the mapping particle filter. *Journal of Computational Physics*, 396, 400–415. Available from: https://doi.org/10.1016/j.jcp.2019.06.060

Rieckh, T. & Anthes, R. (2018) Evaluating two methods of estimating error variances using simulated data sets with known errors. *Atmospheric Measurement Techniques*, 11(7), 4309–4325. Available from: https://doi.org/10.5194/amt-11-4309-2018

Rutherford, I.D. (1972) Data assimilation by statistical interpolation of forecast error fields. *Journal of the Atmospheric Sciences*, 29, 809–815. Available from: https://doi.org/10.1175/1520-0469(1972)029<0809:DABSIO>2.0.CO;2

Schiro, K.A., Neelin, J.D., Adams, D.K. & Lintner, B.R. (2016) Deep convection and column water vapor over tropical land versus Tropical Ocean: a comparison between the Amazon and the tropical Western Pacific. *Journal of the Atmospheric Sciences*, 73(10), 4043–4063. Available from: https://doi.org/10.1175/JAS-D-16-0119.1

Semane, N., Anthes, R., Sjoberg, J., Healy, S. & Ruston, B. (2022) Comparison of Desroziers and three-cornered hat methods for estimating COSMIC-2 bending angle uncertainties. *Journal of Atmospheric and Oceanic Technology*, 39(7), 929–939. Available from: https://doi.org/10.1175/JTECH-D-21-0175.1

Sjoberg, J.P., Anthes, R.A. & Rieckh, T. (2021) The three-cornered hat method for estimating error variances of three or more atmospheric datasets. Part I: overview and evaluation. *Journal of Atmospheric and Oceanic Technology*, 38(3), 555–572. Available from: https://doi.org/10.1175/JTECH-D-19-0217.1

Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M. et al. (2020) A review of innovation-based methods to jointly estimate model and observation error covariance matrices in ensemble data assimilation. *Monthly Weather Review*, 148(10), 3973–3994. Available from: https://doi.org/10.1175/MWR-D-19-0240.1

Todling, R., Semane, N., Anthes, R. & Healy, S. (2022) The relationship between two methods for estimating uncertainties in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 148(747), 2942–2954.

van Leeuwen, P.J. (2015) Representation errors and retrievals in linear and nonlinear data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 141, 1612–1623. Available from: https://doi.org/10.1002/qj.2464

van Leeuwen, P.J., Künsch, H.R., Nerger, L., Potthast, R. & Reich, S. (2019) Particle filters for high-dimensional geoscience applications: a review. *Quarterly Journal of the Royal Meteorological Society*, 2019(145), 2335–2365. Available from: https://doi.org/10.1002/qj.3551

How to cite this article: Hu, C.-C., van Leeuwen, P.J. & Geer, A.J. (2024) A non-parametric way to estimate observation errors based on ensemble innovations. *Quarterly Journal of the Royal Meteorological Society*, 1–20. Available from: https://doi.org/10.1002/qj.4710

APPENDIX A. THE COST FUNCTION AND THE SMOOTHNESS PARAMETER IN DOEE

In order to find a smooth pdf that can best satisfy Equation (21), we introduce prior information in the problem formulation as follows:

$$q(\mathbf{f}_{\varepsilon^o}) = \|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|^2 + (\mathbf{F}\mathbf{f}_{\varepsilon^o})^T\mathbf{S}^{-1}(\mathbf{F}\mathbf{f}_{\varepsilon^o}), \tag{A1}$$

$$\min q(\mathbf{f}_{\varepsilon^o}) \; with \; \mathbf{f}_{\varepsilon^o} \geq 0, \tag{A2}$$

where $\mathbf{F}$ is a finite difference operator which calculates the derivative of $\mathbf{f}_{\varepsilon^o}$. Condition (A2) ensures that all pdf estimates are non-negative, while the constraint with the derivatives (second term on the RHS of Equation A2) enforces a smooth solution. The derivative matrix $\mathbf{F}$ can be written as:

$$\mathbf{F} = \begin{bmatrix} -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & -1 & 1 \end{bmatrix}, \tag{A3}$$

which is a one-sided derivative. $\mathbf{S}$ is a covariance matrix that controls the smoothness of $\mathbf{F}\mathbf{f}_{\varepsilon^o}$ and is constructed as:

$$\mathbf{S} = \alpha \begin{bmatrix} 1 & r_1 & r_2 & \cdots & \cdots & \cdots & r_{2n} \\ r_1 & 1 & r_1 & r_2 & \cdots & \cdots & r_{2n-1} \\ r_2 & r_1 & 1 & r_1 & r_2 & \cdots & r_{2n-2} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ r_{2n-1} & \cdots & \cdots & r_2 & r_1 & 1 & r_1 \\ r_{2n} & \cdots & \cdots & \cdots & r_2 & r_1 & 1 \end{bmatrix}, \tag{A4}$$

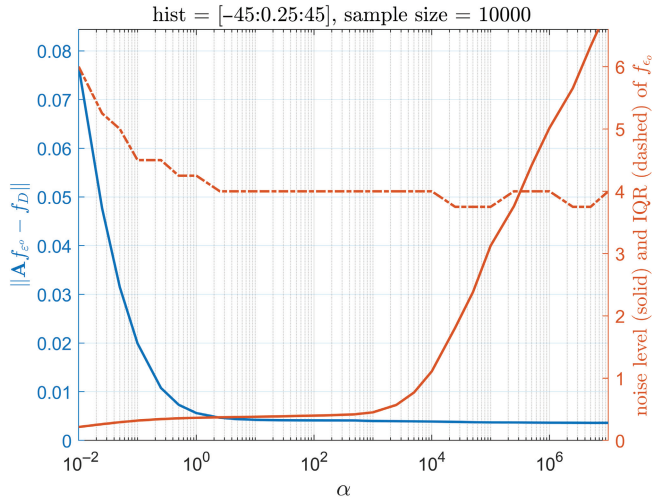

FIGURE A1 The statistical properties of the observation error pdf $\mathbf{f}_{\varepsilon^o}$ that minimizes the cost-function $q(\mathbf{f}_{\varepsilon^o})$ in Equation (A1) when $q(\mathbf{f}_{\varepsilon^o})$ is constructed by different values of the smoothness parameter α. The blue line is the value of $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|$, the brown solid line is the noise level of $\mathbf{f}_{\varepsilon^o}$, which is defined as 50 times of the root-mean-square of the derivative of $\mathbf{f}_{\varepsilon^o}$, and the brown dashed line is the interquartile range (IQR) of $\mathbf{f}_{\varepsilon^o}$. [Colour figure can be viewed at wileyonlinelibrary.com]

where α is a scalar controlling the magnitude of $\mathbf{S}$, which is called the "smoothness parameter", and $r_i = \exp\left(-\left(\frac{i\Delta x}{r_{corr}}\right)^2\right)$ in which r_{corr} is chosen as Δx here and it is the correlation length scale for the matrix $\mathbf{S}$. Note that the DOEE result is not sensitive to the choice of r_{corr} as long as it is of the order of Δx.

The first term $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|^2$ in the cost function in Equation (A1) is a measure of the difference between the estimated innovation pdf based on a solution of $\mathbf{f}_{\varepsilon^o}$ and the true innovation pdf, which we would like to minimize. The second term $(\mathbf{F}\mathbf{f}_{\varepsilon^o})^T\mathbf{S}^{-1}(\mathbf{F}\mathbf{f}_{\varepsilon^o})$ is a penalty term, which is analogous to a prior constraint in the cost function of the variational methods. Therefore, the inverse of $\mathbf{F}^T\mathbf{S}^{-1}\mathbf{F}$ is analogous to the prior covariance matrix, where α determines the prior variance. When α is larger, the constraint from the penalty term is weaker, so we can expect a less smooth $\mathbf{f}_{\varepsilon^o}$, which can, however, fit $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|^2$ better.

Unfortunately, there is no physically based way to determine the size of the smoothness parameter α. In order to determine an optimal α, we seek the following criteria for a good solution for $\mathbf{f}_{\varepsilon^o}$. First, the solution $\mathbf{f}_{\varepsilon^o}$ should make $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|$ as small as possible. Second, $\mathbf{f}_{\varepsilon^o}$ is expected to be smooth. Of course, the second criterion is a little ad-hoc and can be problematic when the true underlying distribution is rugged. However, we believe that most of the error distributions that we are looking at are smooth. Based on the two criteria, we examine the relation between the minimum $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|$ that we can obtain for each value of α, which is shown in Figure A1. We can see that $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|$ decreases with increasing α. This is because as α increases, the penalty term becomes smaller (the penalty term vanishes as $\alpha \to \infty$), and minimizing the cost function becomes almost equivalent to minimizing $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|$ as $\alpha \to \infty$. Note that $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|$ never really goes to zero because we also have the constraint $\mathbf{f}_{\varepsilon^o} \geq 0$ in Equation (A2). Most importantly, Figure A1 shows that when α is larger than a certain threshold, $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|$ almost "saturates" and does not decrease much if we further increase α. At the same time, the noise level of $\mathbf{f}_{\varepsilon^o}$ increases if we further increase α. This motivates us to heuristically choose the optimal smoothness parameter α_{optimal} to be around this threshold, since this should be the smoothest $\mathbf{f}_{\varepsilon^o}$ that best fits the equation $\mathbf{f}_D = \mathbf{A}\mathbf{f}_{\varepsilon^o}$, which satisfies both criteria. For example, this threshold should be around 1–100 in Figure A1.

We demonstrate how the smoothness parameter can affect the solution in an idealized experiment where the true observation error is Gaussian. See Section 3 for the details of the experiment setup. Figure A2 shows the pdf of

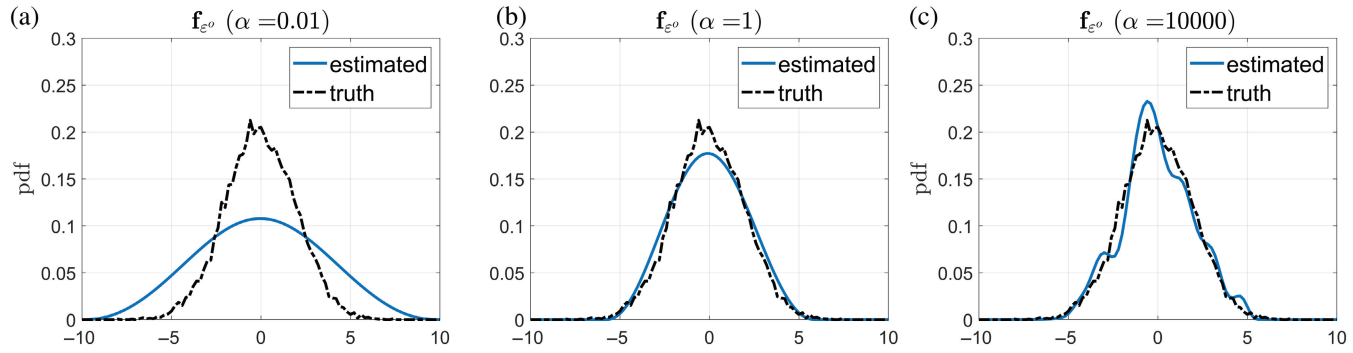

FIGURE A2 The true observation error pdf (black dashed line), which is a Gaussian Normal$(0, 2^2)$, and the estimated (blue solid line) observation error pdf when the estimate is based on (a) a too small smoothness parameter α (b) an appropriate smoothness parameter and (c) a too large α. [Colour figure can be viewed at wileyonlinelibrary.com]

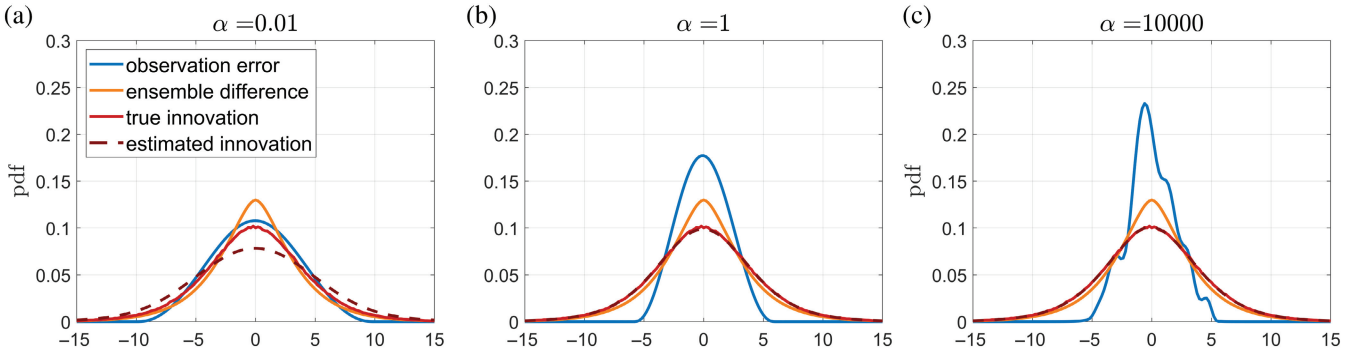

FIGURE A3 The decomposition of the pdfs in Equation (16) when the true observation error is drawn from Normal$(0, 2^2)$. The blue line is the estimated observation error pdf f_{ε^o} (the same as the blue lines in Figure A2), the orange line is the ensemble difference pdf f_ε, the solid red line is the innovation pdf f_D, and the dashed brown line is the estimated innovation pdf given by the convolution between the estimated observation error pdf f_{ε^o} and the ensemble difference pdf f_ε. The estimate is obtained based on (a) a too small smoothness parameter α (b) an appropriate smoothness parameter and (c) a too large α. [Colour figure can be viewed at wileyonlinelibrary.com]

the actual draws for the observation errors (black dashed line) and the estimated observation errors (blue solid line) using different smoothness parameters α. We can see that when we choose an α that is too small (Figure A2a), the estimated distribution is too smooth, whereas a too large α leads to a noisy distribution (Figure A2c). When we pick an appropriate smoothness parameter (which is chosen based on the example in Figure A1), the estimated pdf is a smooth bell-shaped curve that is similar to the true distribution (Figure A2b). To understand how α affects the estimation in detail, we plot each of the pdfs in the convolution equation from Equation (16) when using a different α in Figure A3. Note that the innovation pdf (solid red line) and the ensemble difference pdf (orange line) are the input to this methodology and therefore are the same for all the subplots in Figure A3. We also plot the reconstructed innovation pdf (dashed brown line) obtained from the convolution of the estimated observation error pdf (blue line) and the ensemble difference pdf (orange line). With a small α (Figure A3a), the difference between the reconstructed innovation pdf and the actual innovation pdf and is large, which corresponds to the large value of $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|$ in Figure A1. In contrast, when α is larger than the α_{optimal}, the reconstructed innovation pdf becomes very similar to the actual innovation pdf (Figure A3b,c), which corresponds to the small values of $\|\mathbf{A}\mathbf{f}_{\varepsilon^o} - \mathbf{f}_D\|$ in Figure A1. Note that although both of the reconstructed innovation pdfs in Figure A3b,c fit the actual innovation pdf very well, the observation error pdf in Figure A3c is rugged, which seemingly manages to fit the noise in the actual observation error (black dashed line in Figure A2). The results here justify that the way we choose α_{optimal} tends to give us good estimates, as long as the true observation error pdf is smooth.