

© Copyright 2014 American Meteorological Society (AMS). Permission to use figures, tables, and brief excerpts from this work in scientific and educational works is hereby granted provided that the source is acknowledged. Any use of material in this work that is determined to be “fair use” under Section 107 of the U.S. Copyright Act or that satisfies the conditions specified in Section 108 of the U.S. Copyright Act (17 USC §108, as revised by P.L. 94-553) does not require the AMS’s permission. Republication, systematic reproduction, posting in electronic form, such as on a web site or in a searchable database, or other uses of this material, except as exempted by the above statement, requires written permission or a license from the AMS. Additional details are provided in the AMS Copyright Policy, available on the AMS Web site located at (<http://www.ametsoc.org/>) or from the AMS at 617-227-2425 or copyright@ametsoc.org.

Errors in Ensemble Kalman Smoother Estimates of Cloud Microphysical Parameters

DEREK J. POSSELT

University of Michigan, Ann Arbor, Ann Arbor, Michigan

DANIEL HODYSS AND CRAIG H. BISHOP

Naval Research Laboratory, Monterey, California

(Manuscript received 13 September 2013, in final form 30 November 2013)

ABSTRACT

If forecast or observation error distributions are non-Gaussian, the true posterior mean and covariance depends on the distribution of observation errors *and* the observed values. The posterior distribution of analysis errors obtained from ensemble Kalman filters and smoothers is independent of observed values. Hence, the error in ensemble Kalman smoother (EnKS) state estimates is closely linked to the sensitivity of the true posterior to observed values. Here a Markov chain Monte Carlo (MCMC) algorithm is used to document the dependence of the errors in EnKS-based estimates of cloud microphysical parameters on observed values. It is shown that EnKS analysis distributions are grossly inaccurate for nonnegative microphysical parameters when parameter values are close to zero. Furthermore, numerical analysis is presented that shows that, by design, the posterior distributions given by EnKS and even nonlinear extensions of these smoothers approximate the average of all possible posterior analysis distributions associated with all possible observations given the prior. Multiple runs of the MCMC are made to approximate this distribution. This empirically derived average of Bayesian posterior analysis errors is shown to be qualitatively similar to the EnKS posterior. In this way, it is demonstrated that, in the presence of nonlinearity, EnKS algorithms do not estimate the true posterior error distribution given the specific values of the observations. Instead, they produce an error distribution that is consistent with an average of the true posterior variance, weighted by the probability of obtaining each possible observation. This seemingly subtle distinction gives rise to fundamental differences between the approximate EnKS posterior and the true Bayesian posterior distribution.

1. Introduction

Data assimilation (DA) algorithms are used in a diverse set of applications to produce an estimate of the state of a system (the analysis) as well as an estimate of the analysis uncertainty (or error). The functionality of a DA system can be assessed by examining the degree to which it properly combines information from disparate sources; including prior knowledge of the system, observations, and the functional relationship between observations and the variables to be estimated. In nearly all modern DA algorithms, each piece of information entering the system is quantified by assigning it a probability distribution, and the analysis is computed as the solution to Bayes's relationship for conditional

probabilities. The analysis consists of an estimate of the center of mass (mean, median, mode, etc.) of the posterior probability distribution while the uncertainty (error) is a measure of its variability. The (large) dimensionality of geophysical systems often makes a direct computation of the Bayesian posterior distribution intractable, and over the past half century a number of approximations to Bayes's relationship have been developed. Among these, data assimilation systems based on the Kalman filter (Kalman 1960; Jazwinski 1970) have proven to be particularly useful, as they can be shown to return the minimum error variance estimate of the mean, as long as the model is linear and probability distributions additive and Gaussian (Cohn 1997).

Ensemble Kalman filter (ENKF)-type methods (Evensen 1994, 2009) are low-rank approximations to the Kalman filter that are tractable for use in high-dimensional nonlinear models. The fact that ENKF algorithms are relatively straightforward to implement, employ the full nonlinear model, and produce an estimate

Corresponding author address: Derek J. Posselt, Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, 2455 Hayward St., Ann Arbor, MI 48109-2143.
E-mail: dposselt@umich.edu

of the time-evolving analysis mean and error covariance has led to a proliferation of research into their use. In particular, the ability of EnKF-type DA methods to produce an ensemble estimate of the error covariance that is consistent with the current state of the system (the “error of the day”) is one of its most attractive features. In the Bayesian framework, the true covariance of the analysis distribution for a given state results from a product between the prior and the current set of observations. In the linear, Gaussian case, this Bayesian estimate is what results from the Kalman update equations. It is well known that, if the model is nonlinear and/or the prior non-Gaussian, the true Bayesian solution will depart from Gaussian, and the analysis ensemble returned by Kalman will be erroneous (Vukicevic and Posselt 2008; Hodyss 2011, 2012; Posselt and Bishop 2012, hereafter PB12). What is not clear is the *degree of departure*, which can be measured by examining how the posterior distribution changes for different sets of observations for a given prior. In this paper we examine the characteristics of the posterior analysis in the nonlinear and non-Gaussian case, paying particular attention to how the posterior distribution changes for different sets of observations for a given prior. Specifically, we use a Markov chain Monte Carlo (MCMC) algorithm to explore the characteristics of the posterior distribution for a set of cloud microphysical parameters that have a nonlinear nonmonotonic relationship with the model output state.

In previous work, Posselt and Vukicevic (2010, hereafter PV10) assessed parameter uncertainty, characterized posterior probability density functions (PDFs), and explored which combination of observations produce a unique solution. PB12 examined the evolution of the posterior distribution upon the addition of information from increasing numbers of observations distributed in time. PB12 also evaluated two forms of an ensemble Kalman filter and found that ensemble methods are overvariant and unable to track changes in the posterior associated with nonlinearity in the functional relationship between control parameters and model output. The results reported in PV10 and PB12 were produced for a single specific set of observations; however, Hodyss (2011) showed that ensemble filters approximate an integral over all possible observation sets given a prior and a model. Comparison of the solution returned by EnKF with a sample of the posterior PDF for a given set of observations illustrates the potential pitfalls in the Kalman-based approximation. However, analysis of the solution for a single set of true parameters does not reveal the effect of changes in true state on the posterior distribution, nor does it provide an assessment as to whether the integral approximation produced by ensemble

filters is correct. A more comprehensive assessment evaluates the Kalman analysis versus the ensemble produced by integrating over multiple Bayesian posterior distributions given multiple draws of observations from multiple true states. Such a comparison is more consistent with the formulation of the EnKF and its computation of an analysis that is the integral over all possible innovations for a given prior.

We approach this comparison systematically, first demonstrating the changes in the exact posterior solution that one obtains via direct computation of the Bayesian posterior, as well as the invariance of ensemble filter analysis ensembles. A response function analysis is provided to aid in interpretation of the parameter dependence of the posterior PDF. We then provide a tutorial example that demonstrates how the ensemble smoother analysis can be considered to be an integral of the analysis ensemble over all possible observations (or equivalently innovations). This is the expected posterior analysis covariance and, when the posterior PDF is non-Gaussian, is incorrect. We then compare the ensemble estimates produced by an ensemble Kalman smoother (EnKS; an ensemble whose error variance is consistent with the expected error made by the DA system given the specific observations obtained in the analysis cycle) with an analysis ensemble that accounts for skewness in the prior and an integral over an ensemble of Bayesian estimates of the posterior for multiple draws of observations from different true states.

The goals of this study are the following:

- 1) Explore the differences in posterior PDFs associated with changes in true state and provide physically based explanations for these differences.
- 2) Highlight the differences between the Bayesian posterior distribution and what is produced by ensemble filters, and examine whether the assumptions that undergird ensemble filters are appropriate when applied to a problem characterized by nonmonotonic nonlinearity.
- 3) Perform a comparison between a solution that consists of an integral over multiple Bayesian posterior distributions for random draws of observations and two forms of an ensemble filter.
- 4) Draw a distinction between the integral over all possible observations given a prior estimate and the true error of the day as provided by the Bayesian posterior PDF

The rest of this paper is structured as follows. Section 2 contains a description of the numerical model, and of the various data assimilation algorithms used in this work. Results from MCMC and from ensemble filters are presented in section 3 and a discussion and tutorial

example are contained in section 4. A summary of the major conclusions is presented in section 5.

2. Description of model, MCMC, and ensemble data assimilation algorithms

a. Column model

Intrinsic uncertainty in cloud microphysical parameters can be assessed by isolating the effect of changes in model physics parameters from any feedback to the cloud-scale dynamics. To this end, PV10 introduced a single-column version of the National Aeronautics and Space Administration (NASA) Goddard Cumulus Ensemble (GCE) model (Tao and Simpson 1993; Tao et al. 2003, 2014) that emulates the changes experienced by an atmospheric column during the evolution of a cloud system. The GCE model employs a single-moment bulk cloud microphysical scheme, in which two classes of liquid (rain and nonprecipitating cloud) and three classes of ice (nonprecipitating pristine crystals, precipitating unrimed pristine crystals and aggregates, and graupel) are assumed to have an exponential particle size distribution (Lin et al. 1983; Rutledge and Hobbs 1983, 1984; Tao et al. 2003; Lang et al. 2007). Clouds are generated in the column model by specifying time-varying vertical profiles of vertical motion and water vapor tendency. Advection is only allowed to operate on cloud liquid and ice condensate, and only in the vertical direction, and clouds are allowed to interact fully with both longwave and shortwave radiative transfer. By specifying appropriate base-state potential temperature and water vapor, along with time-varying vertical profiles of vertical motion and water vapor tendency, the model can be used to simulate the flow through a range of different cloud systems. Squall-line-type deep convection was used as a test case in PV10 and PB12 because it exhibits two distinct cloud morphologies: convective, in which precipitation is primarily generated by the collision-coalescence (warm rain) process, and stratiform, in which precipitation results from the melting of falling ice particles (e.g., snow and graupel). The details of the model configuration can be found in PV10, and simulation output is broadly consistent with idealized (Moncrieff 1992) and observed (Houze 2004) 3D squall lines.

As in PV10 and PB12, control variables consist of 10 parameters that govern the ice particle fall speed, ice and liquid particle size distribution, and ice density (Table 1). The ranges of variability for each parameter are taken from observations of ice crystal shape and particle size distribution (Locatelli and Hobbs 1974; Mitchell 1996; Heymsfield et al. 2002). As in PV10,

observations are column integral quantities: simulated precipitation rate (PCP), liquid water path (LWP), ice water path (IWP), and top-of-the-atmosphere outgoing longwave (OLR) and shortwave (OSR) radiative fluxes (Table 2). Motivation for the choice of observations is discussed in PV10, but we note that the use of radar reflectivity observations to characterize cloud microphysical parameter PDFs in the column framework is explored in van Lier-Walqui et al. (2012, 2014). In each of the MCMC and ensemble smoother experiments the true parameter values are specified, and observations are generated by running a 3-h forward integration of the model. PV10 and PB12 used specified parameter values consistent with those used in current cloud-resolving and mesoscale models (Table 1). Observations of precipitation rate, liquid and ice water path, and outgoing long- and shortwave radiative fluxes are drawn from the true state every 30 min starting 30 min into the simulation and ending at 180 min. Observation errors are assumed to be distributed Gaussian and uncorrelated, with variance consistent with the rigorous quantification of uncertainty performed on the analogous retrieved quantities from the Tropical Rainfall Measuring Mission (TRMM; L'Ecuyer and Stephens 2002, 2003; Table 2).

b. MCMC algorithm

Bayes's theorem provides a compact statement of the relationship between conditional probabilities:

$$P(\mathbf{x} | \mathbf{y}) = \frac{P(\mathbf{y} | \mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}, \quad (1)$$

where \mathbf{x} is the set of microphysical parameters and \mathbf{y} are the specific values of the observations. In (1) $P(\mathbf{x})$ represents prior knowledge of the control parameters \mathbf{x} , while $P(\mathbf{y})$ is a normalizing factor that consists of the integral of $P(\mathbf{x}, \mathbf{y})$ over all possible \mathbf{x} , holding \mathbf{y} constant. The optimal parameter set can be defined as either the point in the posterior conditional PDF $P(\mathbf{x} | \mathbf{y})$ with the maximum likelihood [maximum likelihood estimate (MLE)], or the mode of the posterior distribution [maximum a posteriori estimate (MAP)]. Parameter sensitivity can be quantified via calculation of the width of the posterior PDF [e.g., posterior (co)variance, interquartile range, etc.], while relationships between parameters and observations can be determined via examination of the likelihood $P(\mathbf{y} | \mathbf{x})$ or by examining the forward modeled response function.

Most modern data assimilation techniques and retrieval algorithms are based on the assumption that the probabilities represented in (1) are Gaussian (Rodgers 2000; Kalnay 2003; Lewis et al. 2006; Vukicevic and Posselt 2008; Evensen 2009; among many others). In

TABLE 1. Cloud microphysical parameters used in the MCMC algorithm, along with truth values for the simulated observation experiment and parameter ranges. Note that we have reported all values in CGS units to be consistent with what is used in the model formulation and inverse method.

Parameter description	Abbreviation	Units	Truth	Min	Max
Snowfall speed coefficient	a_s	cm^{1-b_s}	200	50	1000
Snowfall speed exponent	b_s	none	0.3	0.1	1
Graupel fall speed coefficient	a_g	cm^{1-b_g}	400	50	1200
Graupel fall speed exponent	b_g	none	0.4	0.1	0.9
Threshold cloud mass mixing ratio for autoconversion to rain	q_{c0}	g kg^{-1}	1	0.1	3
Slope intercept of the rain particle size distribution	N_{0r}	cm^{-4}	0.5	0	5
Slope intercept of the snow particle size distribution	N_{0s}	cm^{-4}	0.5	0	5
Slope intercept of the graupel particle size distribution	N_{0g}	cm^{-4}	0.5	0	5
Snow particle density	ρ_s	g cm^{-3}	0.2	0.1	1
Graupel particle density	ρ_g	g cm^{-3}	0.4	0.1	1

contrast, MCMC algorithms make no prior assumption as to the form of any of the distributions in (1). Rather, MCMC methods produce a sample of the Bayesian posterior probability distribution via a random walk (a Markov chain) that revisits regions in the posterior parameter space with high probability, while avoiding regions with low probability (Tamminen and Kyrölä 2001; Tarantola 2005; Posselt et al. 2008; PV10; PB12; Posselt 2013). In each step of the chain, candidate values of all parameters are drawn from a “proposal” distribution that depends on the current parameter estimate. Proposed parameter values are used in the forward model, and the resulting simulated measurements and prior are compared with observations via a cost function consistent with the assumed observation uncertainty that quantifies the goodness of fit between model and observations. Proposed parameter values are accepted as a sample of the posterior probability distribution with probability:

$$Q(\mathbf{x}_i, \hat{\mathbf{x}}) = \min[1, \rho(\mathbf{x}_i, \hat{\mathbf{x}})], \quad (2)$$

where \mathbf{x}_i is the current parameter vector estimate and $\hat{\mathbf{x}}$ is the vector of proposed parameters. The acceptance ratio $\rho(\mathbf{x}_i, \hat{\mathbf{x}})$ is defined as

$$\rho(\mathbf{x}_i, \hat{\mathbf{x}}) = \frac{P(\hat{\mathbf{y}} | \hat{\mathbf{x}})P(\hat{\mathbf{x}})q(\hat{\mathbf{x}}, \mathbf{x}_i)}{P(\mathbf{y}_i | \mathbf{x}_i)P(\mathbf{x}_i)q(\mathbf{x}_i, \hat{\mathbf{x}})}, \quad (3)$$

where $q(\mathbf{x}_i, \hat{\mathbf{x}})$ is the aforementioned proposal distribution that quantifies the probability of randomly moving from \mathbf{x}_i to $\hat{\mathbf{x}}$ for each step in the Markov chain. Here $q(\cdot)$ is commonly defined to be reversible so that $q(\mathbf{x}_i, \hat{\mathbf{x}}) = q(\hat{\mathbf{x}}, \mathbf{x}_i)$, and in the experiments reported here the prior $P(\mathbf{x})$ is defined to be uniform. This implies $P(\hat{\mathbf{x}}) = P(\mathbf{x}_i)$ so that the acceptance ratio in (3) reduces to the ratio of likelihoods.

Note that the form of likelihood and prior are completely general. The value of the acceptance ratio determines

whether the proposed set of parameters is stored as a sample of the posterior distribution. If the new set of parameters produces an improved fit to the observations, then this set is saved as the next sample in the distribution. If not, then a test value is drawn from a uniform distribution. If this value is less than the acceptance ratio, then the proposed parameter values are saved; if not then the proposed set of parameters is rejected, the current set is stored as another sample, and new proposed parameter values are drawn. The accept/reject procedure is central to the operation of an MCMC algorithm and ensures that random moves in the direction of higher probability states are automatically accepted, while parameter values that provide a similar fit to observations are considered. A more complete description of the theoretical underpinnings of the MCMC algorithm can be found in Mosegaard and Tarantola (2002), Tarantola (2005), and Brooks et al. (2011); while tools for assessing convergence of MCMC to sampling a stationary distribution are discussed in Gelman et al. (1996, 2004).

Our implementation of the MCMC algorithm is documented in Posselt et al. (2008) and PV10, and uses an uncorrelated Gaussian proposal distribution $q(\cdot)$ with variance that is adaptively changed during a burn-in period (Posselt 2013) to strike a balance between efficient and thorough sampling. Samples of the posterior PDF obtained during burn-in are not included in the final analysis. The prior parameter distribution is bounded

TABLE 2. Observations used in MCMC and EnKS parameter sensitivity experiments, along with their units and error estimates.

Observation	Units	Error
Precipitation rate	mm h^{-1}	2.0 mm h^{-1}
Liquid water path	mm	0.5 mm
Ice water path	mm	1.0 mm
TOA LW radiative flux	W m^{-2}	5 W m^{-2}
TOA SW radiative flux	W m^{-2}	5 W m^{-2}

uniform, and bounds are enforced in the MCMC algorithm by rejecting any proposed parameter sets that lie outside the preset ranges. End points are set consistent with observations of ice particle fall speeds (Locatelli and Hobbs 1974; Mitchell 1996) and liquid and ice particle size distributions (Tokay and Short 1996; Heymsfield et al. 2002; Roy et al. 2005). Use of a bounded uniform as the parameter prior makes the prior distribution nearly noninformative, though there is some potential dependence on the specified parameter range. In previous work, we have tested a variety of ranges for parameter values, and found that the PDF returned by MCMC is insensitive to the exact range of parameter values used in the algorithm, as long as it spans the range variability observed in nature.

If the true posterior distribution is Gaussian, between 20 000 and 40 000 iterations should be sufficient to sample a 10-dimensional parameter space (Haario et al. 1999). In reality, the underlying PDF is non-Gaussian, and we find that approximately 100 000 iterations are necessary to sample the underlying PDF. To ensure a robust sample, each MCMC experiment is run for 5×10^5 iterations, and convergence to sampling a stationary distribution is ensured by drawing random samples of 100 000 parameter sets from the full MCMC sample and computing the integrated absolute difference between histograms. In each experiment, we found an average of less than 1% difference between any two given random samples, indicating robust sampling of the underlying (true) PDF.

c. Ensemble Kalman smoother

PB12 presented a variant of the ensemble transform Kalman smoother (ETKS) suitable for cases in which the number of ensemble members K exceeds the number of observations p . In our experiments, the ensemble size K is 10 000 while the number of total observations is 30 (5 observation types \times 6 observation times). The square root (deterministic) form was derived in PB12, as well as a perturbed-observations (stochastic) version that uses precisely the same gain matrix as that used in the square root smoother. PB12 showed that the posterior (analysis) ensemble produced by both perturbed-observations and deterministic forms had identical mean and variance in the limit of large ensembles. Hodyss and Campbell (2013) demonstrated that the primary difference between the deterministic and stochastic filter is a change in the higher-order (third and larger) moments of the distribution. For consistency with the perturbed-observations version of the quadratic ensemble smoother (see section 2d below) we employ the perturbed observations version of the ensemble Kalman smoother.

In brief, let the forecast error covariance matrix be defined as a product of its square roots

$\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{fT}$, where $\mathbf{Z}^f = \mathbf{X}^f / \sqrt{K-1}$, and \mathbf{X}^f is the matrix constructed from K perturbations around the forecast mean. Similarly, let the analysis error covariance matrix be $\mathbf{P}^a = \mathbf{Z}^a \mathbf{Z}^{aT}$, where the analysis error square root is (cf. PB12)

$$\begin{aligned} \mathbf{Z}^a &= \mathbf{Z}^f \{ \mathbf{I} - \mathbf{L} [\mathbf{I} - (\mathbf{\Gamma} + \mathbf{I})^{-1/2}] \mathbf{L}^T \} \\ &= \{ \mathbf{I} - \mathbf{Z}^f \mathbf{L} [\mathbf{I} - (\mathbf{\Gamma} + \mathbf{I})^{-1/2}] \mathbf{\Gamma}^{-1/2} \mathbf{E}^T \tilde{\mathbf{H}} \} \mathbf{Z}^f = \tilde{\mathbf{K}} \mathbf{Z}^f, \end{aligned} \quad (4)$$

where $\mathbf{L} = \mathbf{\Gamma}^{-1/2} \mathbf{E}^T \tilde{\mathbf{H}} \mathbf{Z}^f$; \mathbf{E} is a $p \times p$ orthonormal matrix obtained from the eigenvector decomposition $\mathbf{E} \mathbf{\Gamma} \mathbf{E}^T = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{P}^f \mathbf{H}^T \mathbf{R}^{-1/2}$, where $\mathbf{\Gamma}$ is the diagonal matrix of eigenvalues. As shown in PB12, the form of the square root given by (4) requires the eigenvector decomposition of a $p \times p$ matrix whereas Bishop et al.'s (2001) ensemble transform Kalman filter (ETKF) requires the eigenvector decomposition of a $K \times K$ matrix. Because of the high cost of eigenvector decompositions, (4) is more efficient than the form given in Bishop et al. (2001) when $K \gg p$, as is the case in this paper.

In the EnKS framework, the i th set of perturbed observations is generated by adding a random p vector with mean $\mathbf{0}$ and covariance \mathbf{R} to the original observation vector:

$$\mathbf{y}_i = \mathbf{y} + \boldsymbol{\varepsilon}_i^o, \quad \text{where } \boldsymbol{\varepsilon}_i^o \sim N(\mathbf{0}, \mathbf{R}). \quad (5)$$

The corresponding i th member of the analysis ensemble produced by EnKS is produced via

$$\begin{aligned} (\mathbf{x}_{\text{EnKS}}^a)_i &= \mathbf{x}_i^f + \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} [\mathbf{y}_i - \mathbf{H} \mathbf{x}_i^f] \\ &= \mathbf{x}_i^f + (\mathbf{Z}^a \mathbf{Z}^{aT}) \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y}_i - \mathbf{H} \mathbf{x}_i^f], \end{aligned} \quad (6)$$

where \mathbf{Z}^a is obtained from (4). Note that the analysis error covariance matrix \mathbf{P}^a formed from the square root \mathbf{Z}^a in (4) is only a function of the prior ensemble \mathbf{X}^f and the observation error covariance matrix \mathbf{R} , and can be generated independent of the actual observations of the day.

This can be shown another way by noting that we may use (5) in (6) and decompose (6) into a mean and perturbation to obtain the i th member of the analysis ensemble:

$$(\mathbf{x}_{\text{EnKS}}^a)_i = \bar{\mathbf{x}}_a + \boldsymbol{\varepsilon}_i^a, \quad (7)$$

where

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}^f + \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} [\mathbf{y} - \mathbf{H} \bar{\mathbf{x}}^f], \quad (8)$$

$$\boldsymbol{\varepsilon}_i^a = \boldsymbol{\varepsilon}_i^f + \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} [\boldsymbol{\varepsilon}_i^o - \mathbf{H} \boldsymbol{\varepsilon}_i^f]. \quad (9)$$

Note that operating on both sides of (9) with its transpose and taking an expectation over all random variables leads to the standard analysis error covariance matrix for Kalman filters [cf. (17) below]. In addition, it is important to realize that the ensemble perturbations around the mean in (9) are independent of the observations and therefore the error variance from this ensemble is also independent of observations. This we will show is crucial to understanding the behavior of typical ensemble generation algorithms in nonlinear systems. The parameter sets used to produce the 10 000-member EnKS forecast ensemble are drawn from a bounded uniform prior with ranges identical to those used in MCMC (Table 1).

d. Quadratic ensemble smoother

The quadratic ensemble filter (QEF; Hodyss 2011) extends the linear regression capability of the Kalman filter resulting in an algorithm that performs nonlinear polynomial regression via addition of a new term that is quadratically nonlinear in the innovation. One of the unique features of this technique is its remarkable mathematical similarity to the Kalman filter. This allows an existing Kalman filter algorithm to perform quadratic nonlinear regression for the update of the mean with only minor modifications. The update of the mean consistent with quadratic nonlinear regression allows for a significantly more accurate estimate of the posterior mean when the posterior distribution is skewed because a skewed posterior distribution is associated with a curved (nonlinear function of the innovation) posterior mean (Hodyss 2011).

In the current application, parameters are specified at the beginning of the forward integration, and as such the algorithm is technically a smoother. The quadratic ensemble smoother (QES) update of the posterior mean may be written as a weighted linear combination of the prior ensemble members:

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_f + \mathbf{Z}\mathbf{w}, \tag{10}$$

where the weights are generated according to

$$\mathbf{w} = \hat{\mathbf{Z}}^T \hat{\mathbf{H}}^T [\hat{\mathbf{H}}\hat{\mathbf{P}}_f\hat{\mathbf{H}}^T + \hat{\mathbf{R}}]^{-1} \hat{\mathbf{v}}'. \tag{11}$$

Here, $\hat{\mathbf{v}}'$ is the predictor vector [defined below in (17)], while $\hat{\mathbf{H}}$ and $\hat{\mathbf{R}}$ are the observation operator and observation error covariance matrix, respectively, in the extended state space that includes the higher moments. Clearly, the weights are isomorphic to the traditional Kalman update, but in this case the prior error covariance matrix is extended to include the higher moments:

$$\hat{\mathbf{P}}_f = \frac{\hat{\mathbf{z}}\hat{\mathbf{z}}^T}{K-1} = \begin{bmatrix} \mathbf{P}_f & \mathbf{T}_f \\ \mathbf{T}_f^T & \mathbf{F}_f - \mathbf{p}_f\mathbf{p}_f^T \end{bmatrix}, \tag{12}$$

whose square root form may be approximated with an ensemble as

$$\hat{\mathbf{Z}} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 & \boldsymbol{\varepsilon}_2 & \cdots & \boldsymbol{\varepsilon}_K \\ \boldsymbol{\varepsilon}_1^2 - \mathbf{p}_f & \boldsymbol{\varepsilon}_2^2 - \mathbf{p}_f & \cdots & \boldsymbol{\varepsilon}_K^2 - \mathbf{p}_f \end{bmatrix}. \tag{13}$$

Here K is the ensemble size and the vectorized covariance matrix, $\mathbf{p}_f = \text{vec}(\mathbf{P}_f)$, is an N^2 vector constructed from the concatenation of the N columns of \mathbf{P}_f , and whose organization follows that of the Kronecker product “ \otimes .” The unusual notation of “squaring” a vector in (13) refers to using the Kronecker product: $\boldsymbol{\varepsilon}^2 = \boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon}$. In Hodyss (2012) an approximate form was defined that produces a square root that has $2N$ rows rather than the N^2 rows defined in (13). We use the more accurate method of Hodyss (2011) here because our state dimension in parameter space is rather small.

The extended “observation” error covariance matrix in (11) is defined as

$$\hat{\mathbf{R}} = \begin{bmatrix} \mathbf{R} & \mathbf{0}_{p \times p^2} \\ \mathbf{0}_{p^2 \times p} & \mathbf{R}_4 + \mathbf{A} + \mathbf{B} + \mathbf{C} - \langle \mathbf{v}^2 \rangle \langle \mathbf{v}^{2T} \rangle + \mathbf{p}_o\mathbf{p}_o^T \end{bmatrix}, \tag{14}$$

$$\langle \mathbf{v}^2 \rangle \langle \mathbf{v}^{2T} \rangle = \mathbf{r}\mathbf{r}^T + \mathbf{r}\mathbf{p}_o^T + \mathbf{p}_o\mathbf{r}^T + \mathbf{p}_o\mathbf{p}_o^T, \tag{15}$$

$$\hat{\mathbf{H}} = \begin{bmatrix} \mathbf{H} & \mathbf{0}_{p \times N^2} \\ \mathbf{0}_{p^2 \times N} & \mathbf{H} \otimes \mathbf{H} \end{bmatrix}, \tag{16}$$

where $\mathbf{r} = \text{vec}(\mathbf{R})$ with length p^2 , $\mathbf{p}_o = \text{vec}(\mathbf{H}\mathbf{P}_f\mathbf{H}^T)$, and $\langle \rangle$ denotes an operation that finds the expected value over all random variables. The predictor vector can be written as

$$\hat{\mathbf{v}}' = [\mathbf{v}^T \quad \mathbf{v}^{2T}]^T - [\langle \mathbf{v} \rangle^T \quad \langle \mathbf{v}^{2T} \rangle^T]^T, \tag{17}$$

with $\mathbf{v} = \mathbf{y} - \mathbf{H}\bar{\mathbf{x}}_f$ being the innovation vector and \mathbf{H} is the $p \times N$ observation operator. Here \mathbf{R} is the observation error covariance matrix and \mathbf{R}_4 is the fourth moment matrix of the observation likelihood. The $\hat{\mathbf{H}}$ defined in (16) is the observation operator in the extended state space. Note that the extended observation covariance ($\hat{\mathbf{R}}$) matrix includes information from the prior forecast error statistics; this is a direct result of including information from moments higher than the second. The final matrices in (14) are

$$\mathbf{A} = \langle \boldsymbol{\varepsilon}_o^2 (\mathbf{H} \boldsymbol{\varepsilon}_f)^{2\text{T}} \rangle + \langle (\mathbf{H} \boldsymbol{\varepsilon}_f)^2 \boldsymbol{\varepsilon}_o^{2\text{T}} \rangle, \quad (18a)$$

$$\mathbf{B} = \mathbf{R} \otimes \mathbf{H} \mathbf{P}_f \mathbf{H}^{\text{T}} + \mathbf{H} \mathbf{P}_f \mathbf{H}^{\text{T}} \otimes \mathbf{R}, \quad (18b)$$

$$\mathbf{C} = \langle (\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_f^{\text{T}} \mathbf{H}^{\text{T}}) \otimes (\mathbf{H} \boldsymbol{\varepsilon}_f \boldsymbol{\varepsilon}_o^{\text{T}}) \rangle + \langle (\mathbf{H} \boldsymbol{\varepsilon}_f \boldsymbol{\varepsilon}_o^{\text{T}}) \otimes (\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_f^{\text{T}} \mathbf{H}^{\text{T}}) \rangle, \quad (18c)$$

where $\boldsymbol{\varepsilon}_o$ is drawn from the observation statistics, which in this case is $N(\mathbf{0}, \mathbf{R})$, and $\mathbf{x} = \bar{\mathbf{x}}_f + \boldsymbol{\varepsilon}_f$ is a member of the prior.

Finally, we also make use of the ensemble generation scheme of Hodyss (2011), which used a new version of perturbed observations consistent with quadratic nonlinear regression. We emphasize that, as in Hodyss (2011), the perturbed observation algorithm presented here, while appearing different, is in fact absolutely consistent with traditional perturbed observation algorithms as long as the observation likelihood is symmetric. In the case where the observation likelihood is non-Gaussian, however, the typical form for perturbed observations [(5) and (6)] will return the wrong expected third moment for both the EnKS and the QES, while the form of perturbed observations described below will return the correct value. Rather than perturb the observations, we resample new observations from members of the prior:

$$\mathbf{y}_i = \mathbf{H} \mathbf{x}_i^f + \boldsymbol{\varepsilon}_i^o. \quad (19)$$

This definition allows one to write the i th innovation as

$$\mathbf{v}_i = \mathbf{y}_i - \mathbf{H} \bar{\mathbf{x}}^f = \boldsymbol{\varepsilon}_i^o + \mathbf{H} \boldsymbol{\varepsilon}_i^f, \quad (20)$$

which effectively resamples the innovation with the correct frequency as long as the prior and observation statistics are correct. This resampled innovation is used as the Monte Carlo equivalent of the indicated integration specified in (25). In this case, the i th analysis ensemble perturbation is

$$\boldsymbol{\varepsilon}_i^a = \boldsymbol{\varepsilon}_i^f - \mathbf{Z} \mathbf{w}, \quad (21a)$$

where the weights are generated according to

$$\mathbf{w} = \hat{\mathbf{Z}}^{\text{T}} \hat{\mathbf{H}}^{\text{T}} [\hat{\mathbf{H}} \hat{\mathbf{P}}_f \hat{\mathbf{H}}^{\text{T}} + \hat{\mathbf{R}}]^{-1} \hat{\mathbf{v}}'. \quad (21b)$$

The resulting ensemble perturbation is then added to the analysis to form a member. As in the EnKS experiments, each 10 000 member QES forecast (prior) ensemble consists of random draws of parameter sets from the multivariate bounded uniform prior (Table 1).

3. Comparison between exact and ensemble filter solutions

The first set of experiments addresses the question of how much the posterior PDFs change if the true parameter values are allowed to vary. In so doing, we run a small ensemble of perturbed-parameter experiments, varying parameter values around the control set used in PV10 and PB12 (see Table 1 for the control values and parameter ranges). Perturbation guidelines are as follows: 1) vary parameters at least 10% of the prior range, 2) modify only those parameters that exhibited sensitivity in the original experiments, and 3) if two parameters are closely related, vary them together. The parameters found to have the largest sensitivity in PV10 and PB12 were a_s , b_s , a_g , b_g , N_{0r} , q_{c0} , ρ_g , ρ_s , and N_{0s} . Parameters with significant interrelationships were a_s – b_s and a_g – b_g . Sets of individual and paired parameters chosen for perturbation are 1) a_s – b_s , 2) N_{0r} – q_{c0} , 3) a_g – b_g , 4) ρ_s , 5) ρ_g , and 6) N_{0s} – N_{0g} . It is sensible to vary a_s – b_s and a_g – b_g together, as these were shown to covary in PV10 and PB12, and are closely related at a process level in the microphysics code. Both N_{0s} and N_{0g} are perturbed together to examine the effect of changing the mean precipitating ice particle sizes represented in a given grid volume. Covarying N_{0r} with q_{c0} crudely emulates the first aerosol indirect effect. If autoconversion is specified to happen at a larger value of cloud mass mixing ratio, as would happen if a larger concentration of cloud condensation nuclei (CCN) were present, the result would be a larger number of small raindrops in a given volume. In the absence of an explicit aerosol nucleation scheme in this microphysics code, increases in autoconversion with an increase in rain PSD slope intercept can be used to emulate the effect of increased CCN on the rain PSD. Snow and graupel densities were largely independent of one another in PV10, and as such are varied independently here.

Twelve sets of parameters are produced by varying these six sets around the values used in PV10 and PB12 (Fig. 1), and the results are designed to 1) assess the degree to which posterior PDFs change with changes in true parameter value, 2) compare the posterior PDFs with those obtained from the ensemble smoothers, and 3) diagnose the differences between posterior PDFs using a response function analysis. Note that all parameter values are near the low end of the range of realistic values. There are several reasons for this. First, the default values are consistent with what is used in the operational version of the scheme. While larger parameter values are consistent with the limited sets of direct observations reported in the literature, values near the lower end of the range tend to produce a superior fit

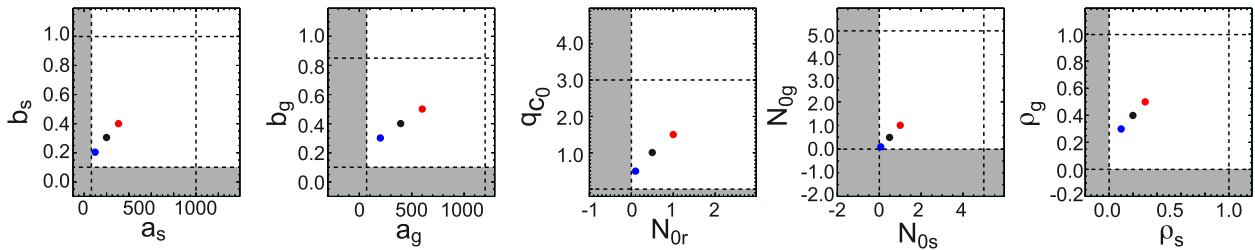


FIG. 1. Parameter values used in the 12 perturbed-truth experiments. The default parameter values are depicted in black, while values perturbed low and high are depicted in blue and red, respectively.

to observed cloud content, radar reflectivity, and top-of-the-atmosphere radiative flux (S. Lang 2012, personal communication). Perturbation around the default values also allows a more direct comparison with the PDFs examined in PV10 and PB12, each of which exhibited non-Gaussianity; both via nonzero skewness as well as, in certain cases, multimodality. Finally, use of parameter values that are consistently near a hard bound, in this case zero, illustrates a number of the persistent challenges for ensemble smoothers.

Observations in each of the 12 experiments are generated by running a single forward integration of the model for each parameter set, extracting observations from each simulation, and adding Gaussian noise (Table 2). A sample of the posterior PDF is obtained by running an MCMC algorithm on each of the 12 sets of observations. As in PV10 and PB12, the prior parameter PDF is set to uniform with bounds obtained from the literature (Table 1). Convergence of the MCMC algorithm is assessed in each case via visual inspection, as well as via computation of the integrated absolute difference between PDFs of subsamples drawn from the full sample.

a. Variability in posterior solutions

Comparison between posterior PDFs reveals several of the 12 parameter sets lead to posterior PDFs with very similar structure. In the interest of brevity, and for the purpose of focusing the analysis, 6 of the 12 cases are selected for closer inspection (Table 3; Fig. 2). PV10 and PB12 found the parameters that exert the greatest control over the simulation are the warm rain (slope intercept of the rain particle size distribution N_{or} and cloud to rain autoconversion threshold q_{c0}) and the coefficient and exponent in the ice (snow and graupel) fall speed power laws. For these parameters (rows 1–3, Fig. 2), the general structure of the posterior distribution does not change with true parameter value; sensitive parameters remain sensitive parameters, regardless of their true value, and correlated parameters remain correlated. While the interparameter structural relationship

is consistent, the posterior variance scales with true value for several of the parameters (N_{or} , a_s , and a_g). The closer the true PSD parameter values are to zero, the greater the sensitivity (smaller variance), while posterior PDFs in cases for which true parameters are closer to the center of the range exhibit reduced sensitivity (larger variance). In essence, this means that when the true value is relatively large, nearly any parameter value will produce forward observations consistent with the “true” observations. In contrast, when the true value is relatively small, only small parameter values produce realistic forward solutions. Certainly one factor is the proximity to the hard bound at zero. The MCMC algorithm prevents the selection of parameter values smaller than this bound. This does not, however, explain the fact that, for parameter values near zero, relatively large parameter values (greater than the midpoint of the prior range) produce highly unlikely (near-zero probability) solutions.

A consideration of the physical nature of the parameters is useful here. Large values of the coefficient and exponent of the snow and graupel fall speed power laws correspond to more rapid particle fall speeds. If the true parameter values are set so that ice particles precipitate more rapidly, a larger range of fall speeds produces solutions consistent with the observations. Consider the response of model output to changes in only the fall

TABLE 3. Values of each of the 10 variable parameters in the six selected parameter perturbation experiments.

Parameter	Expt					
	1	2	3	4	5	6
a_s	300	100	100	300	100	300
b_s	0.4	0.2	0.2	0.4	0.2	0.4
a_g	600	200	200	600	200	600
b_g	0.5	0.3	0.3	0.5	0.3	0.5
q_{c0}	0.5	1.5	0.5	1.5	0.5	1.5
N_{or}	0.1	1.0	0.1	1.0	0.1	1.0
N_{os}	0.1	1.0	1.0	0.1	0.1	1.0
N_{og}	0.1	1.0	1.0	0.1	0.1	1.0
ρ_s	0.1	0.3	0.3	0.1	0.1	0.3
ρ_g	0.5	0.3	0.5	0.3	0.3	0.5

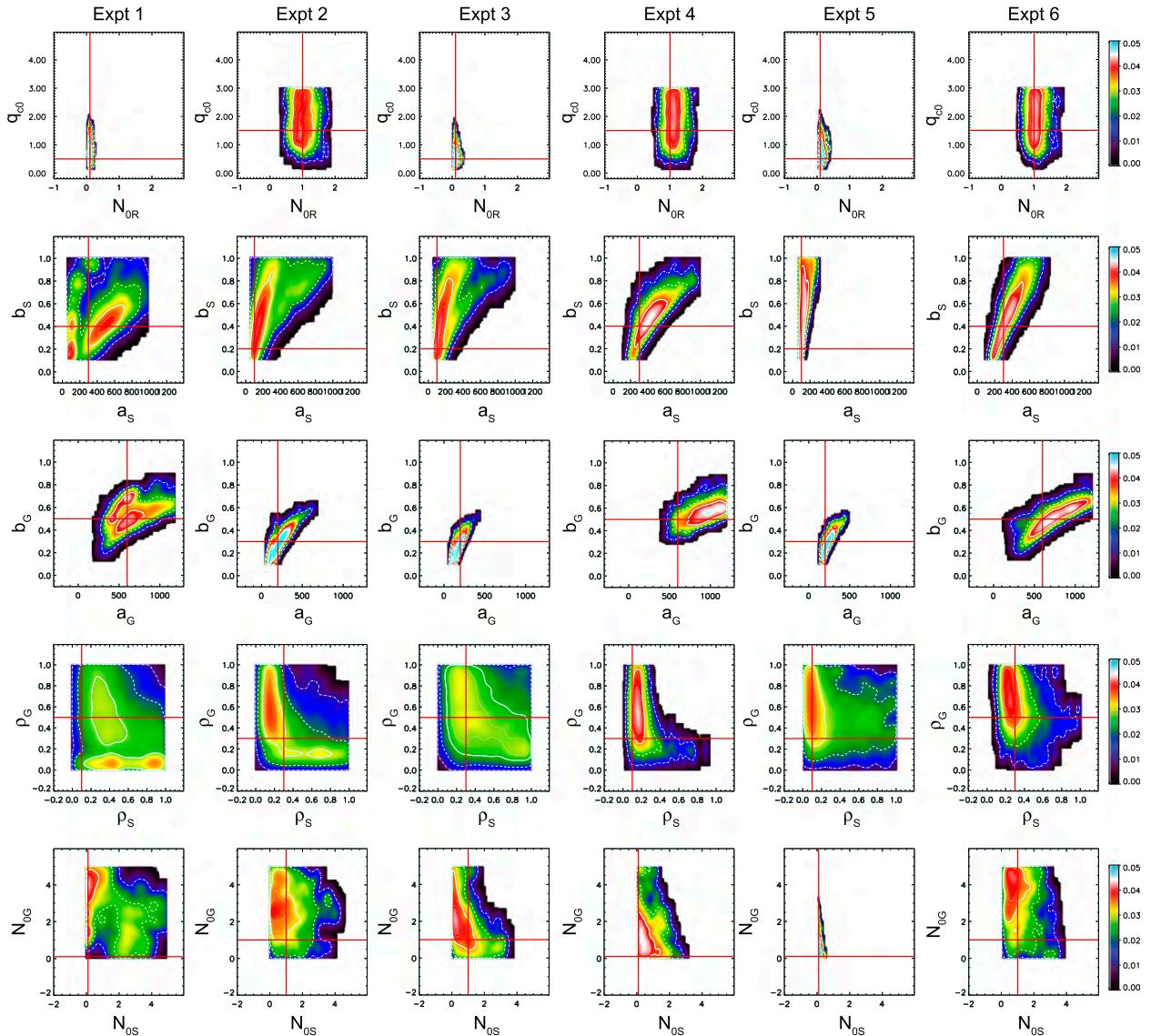


FIG. 2. Posterior PDFs (color contours) for pairs of parameters from 6 of the 12 perturbed-truth experiments. Each column corresponds to a different experiment, while each row corresponds to a different pair of parameters. True parameter values are indicated by the red cross hairs, while the thin dash-dot, dashed, solid, and dotted white contours enclose the 99.7%, 95%, 68.3%, and 38.3% probability contours, respectively.

speed of rimed ice (graupel) particles (Fig. 3), holding all other parameters constant at their default values (Table 1). Parameter sensitivity, as represented by the slopes of the blue lines in Fig. 3, is generally large when graupel fall speeds are small, decreasing at larger values. The physical reasoning behind the change in response function derivative with parameter value follows from a consideration of the influence of the graupel fall speed on simulated output. As fall speeds increase, precipitation rate and LWP first increase then decrease, and IWP, OLR, and OSR first decrease, then increase. As fall speeds increase from the low end of the range, particles

settle faster, accretion and melting increase, and LWP and precipitation rate increase. As particle fall speeds continue to increase, graupel particles have less time to accrete liquid droplets as they fall. Growth is inhibited and the cold rain process decreases. At the same time, because ice is falling fast enough that supercooled liquid does not have time to freeze on it, greater amounts are lofted vertically where they then freeze homogeneously, leading to greater amounts of ice aloft and increased longwave (LW) and shortwave (SW) fluxes.

In addition to changes in posterior variance, Fig. 2 demonstrates the potential to realize a multimode posterior

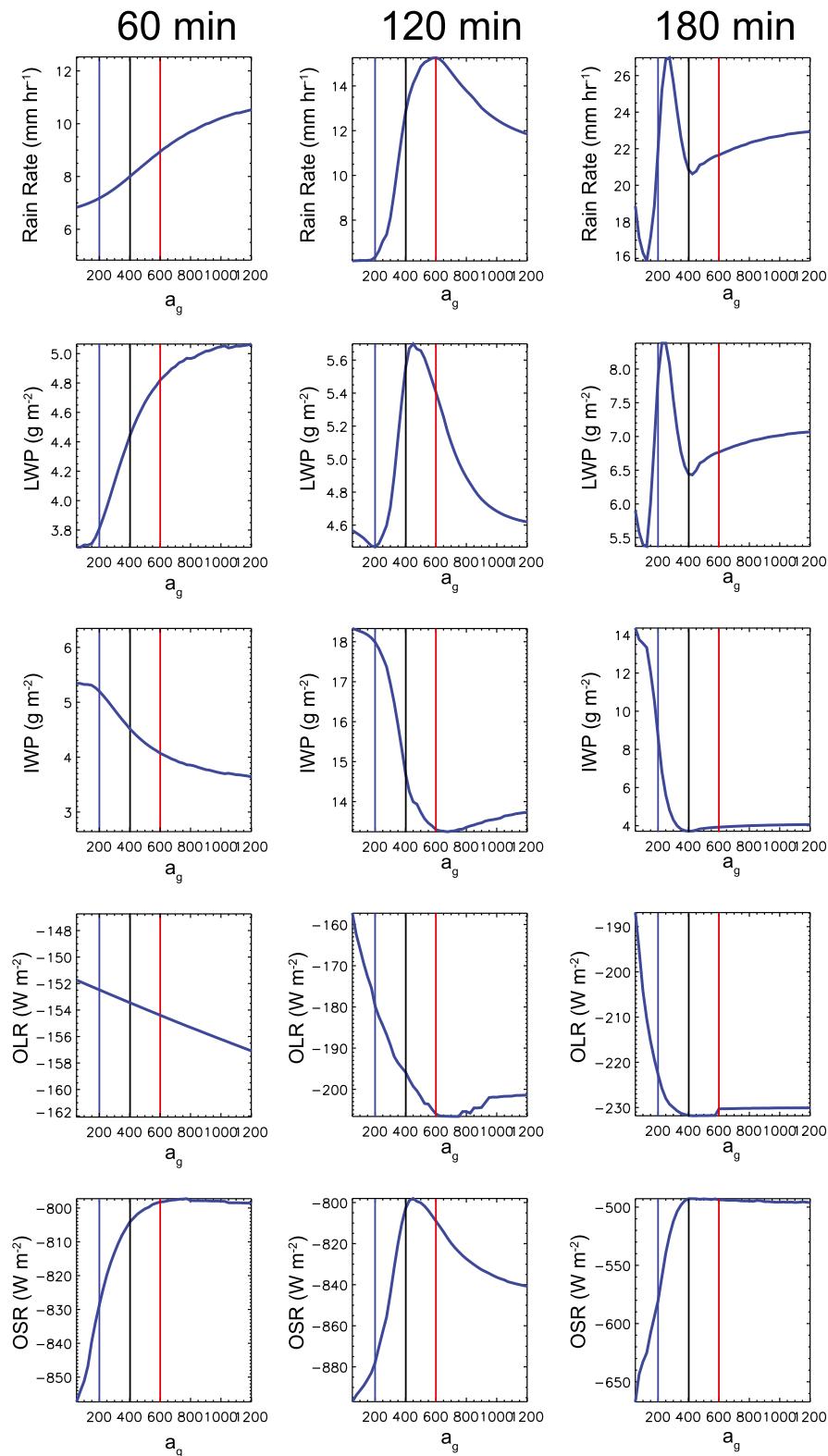


FIG. 3. Response of each model output variable at (left) 60, (middle) 120, and (right) 180 min to changes in graupel fall speed coefficient a_g . Vertical blue, black, and red lines indicate the position of the true parameter values used in the perturbation experiments.

PDF also depends critically on the true value for certain parameters. The warm rain (N_{or} , q_{c0}) parameter PDFs have a single mode, independent of parameter value, while snow and graupel fall speed parameters (a_s – b_s and a_g – b_g) exhibit parameter-dependent manifestation of multimodality. Again the response functions provide insight as to the cause. For example, a_g is clearly bimodal when the true value is set to 600 cm^{-1} (experiment 1, Fig. 2) and also, upon careful inspection, when set equal to 200 cm^{-1} in experiments 2 and 3. Examination of the a_g response functions (Fig. 3) indicates that the response to changing a_g is nonmonotonic over nearly the entire range, but in particular for values between about 50 and 800 over which the response function changes sign as many as 3 times. A nonmonotonic response means that there are multiple values of the parameter that produce identical model response, and gives rise to multimodality in the posterior PDF (Vukicevic and Posselt 2008).

The posterior variance of q_{c0} , b_s , b_g , and ρ_g does not change very much with the true value. For these parameters, the magnitude of the response function derivative is constant with parameter value over the range of true values considered (cf. graupel density in Fig. 4). Note that, while the response function for ρ_g is monotonic over the full range of values, the posterior joint PDF of ρ_g exhibits multiple modes in experiments 1 and 2 (Fig. 2, row 4). These are cases for which the posterior PDF of a_g has multiple modes, and examination of the joint posterior PDF of a_g and ρ_g indicates only certain combinations of a_g and ρ_g will satisfy the observations (cf. Fig. 1 in PB12).

The posterior variance of ρ_s , N_{0s} , and N_{0g} exhibits significant variability, with little relationship to the parameter value. For example, the true ice particle size distribution (PSD) intercept (N_{0s} and N_{0g}) is set small in experiments 1, 4, and 5; however, the posterior variance is much larger in experiments 1 and 4 than experiment 5. The explanation follows from a consideration of the cloud microphysical environment. First, ice number concentration (of, in particular, suspended ice) has a large effect on the top-of-the-atmosphere radiative fluxes, but minimal effect on the cloud content and precipitation rate (Fig. 5). In experiments 1 and 4, ice fall speeds are set large and warm rain (liquid precipitation) parameters (N_{or} and q_{c0}) are set small. This means that precipitating liquid and ice particles both fall rapidly (the raindrop fall speed depends only on the liquid content and the particle size distribution). Ice is therefore more concentrated at lower levels than it would be if it were constrained to fall more slowly, and as such has a smaller radiative impact on the simulation. Consistent with a small autoconversion parameter, cloud transfers to (rapidly falling) rain at smaller values resulting in

relatively less cloud particles available to freeze higher in the cloud. In experiment 5, ice fall speeds are set low and as such, ice is suspended at higher levels, resulting in a larger radiative effect. The ice slope intercept is thus constrained to values near the truth. In addition, when the slope-intercept values are small, there are relatively greater numbers of large particles, and in essence smaller effect on the radiative transfer due to the shortwave dependence on effective radius. In the radiative transfer code, the effective radius for snow and graupel are computed as

$$R_{e_x} = \frac{3}{\left(\frac{\pi N_{0x} \rho_x}{\rho_a q_x}\right)^{1/4}}, \quad (22)$$

which is consistent with the assumption that the particle size distribution is inverse-exponential (Marshall–Palmer). It is clear that the effective radius scales as $1/\rho_x$ and $1/N_{0x}$, and that the fourth root enhances sensitivity to changes in both density and slope intercept at small values.

b. Comparison of the exact solution with EnKS and QES

Two ensemble smoothers (EnKS and QES) are used to generate 10 000 member posterior ensembles for each of the 12 perturbed parameter sets; the same subset of the resulting distributions as were examined in the MCMC experiments are plotted in Figs. 6 and 7 for EnKS and QES, respectively. Examination of the results indicates the following. All distributions produced by the ensemble smoothers for a given parameter pair appear to be identical in shape, but the center of mass shifts according to the location of the true parameters. It should be emphasized that MCMC, EnKS, and QES all have the same prior distribution; a bounded uniform with mean centered between the upper and lower bounds. Both QES and EnKS produce a reasonable estimate of the distribution mean when parameter values are relatively far from zero, but exhibit significant bias for experiments in which parameter values are set close to their specified lower bound. In fact, both EnKS and QES routinely produce unphysical (in this case, less than zero in value) ensemble members with nonzero probability, while MCMC restricts the posterior parameter distributions to the predefined realistic range. This is notable for q_{c0} – N_{or} , which is consistently a high-sensitivity combination of parameters. In cases for which the true values of these two parameters lie close to zero (experiments 1, 3, and 5), not only is there a significant portion of the posterior probability mass located in a nonphysical region of the space, but the *mode* of the posterior distribution for EnKS is less than zero in

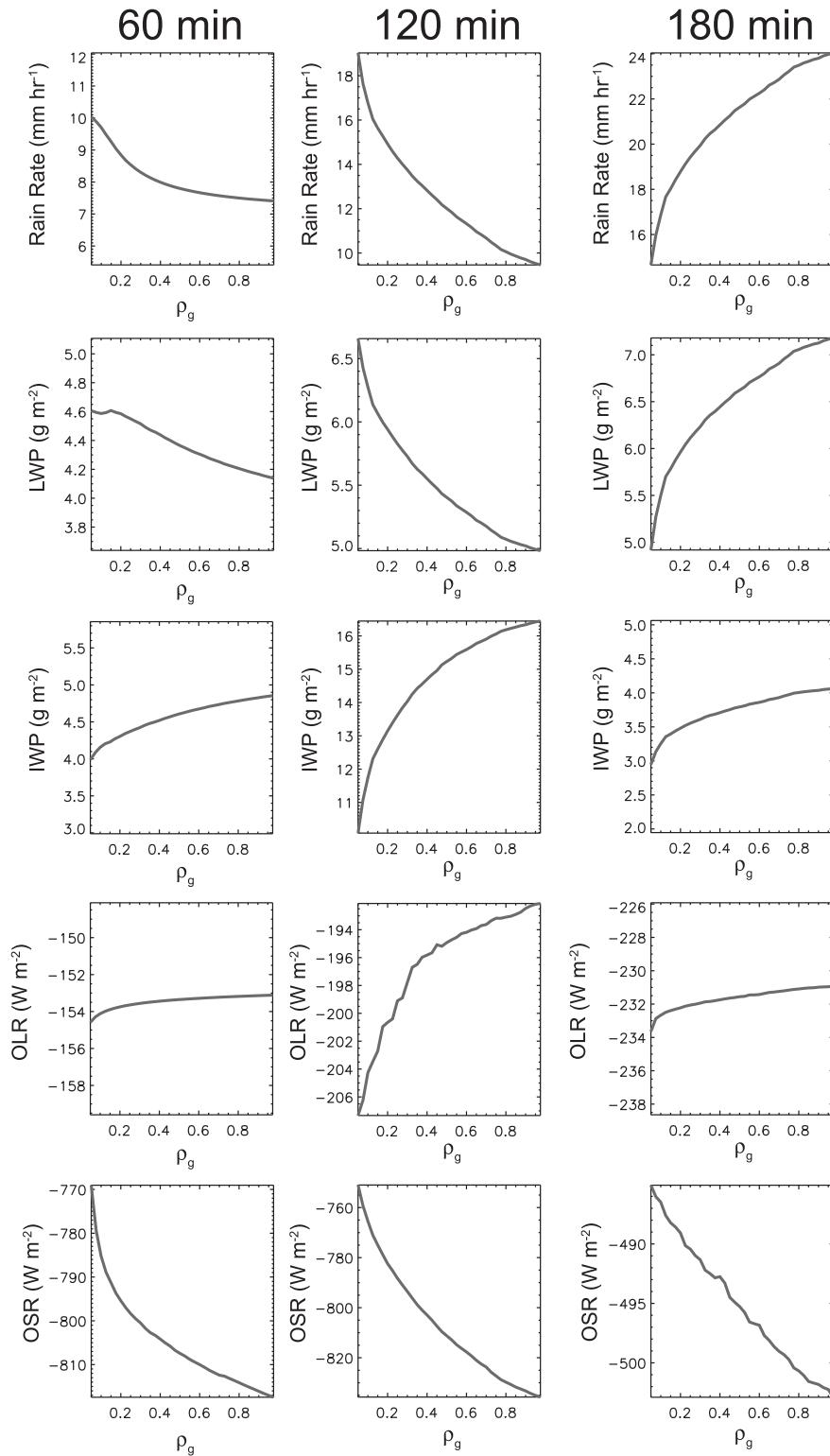


FIG. 4. As in Fig. 3, but for graupel density ρ_g .

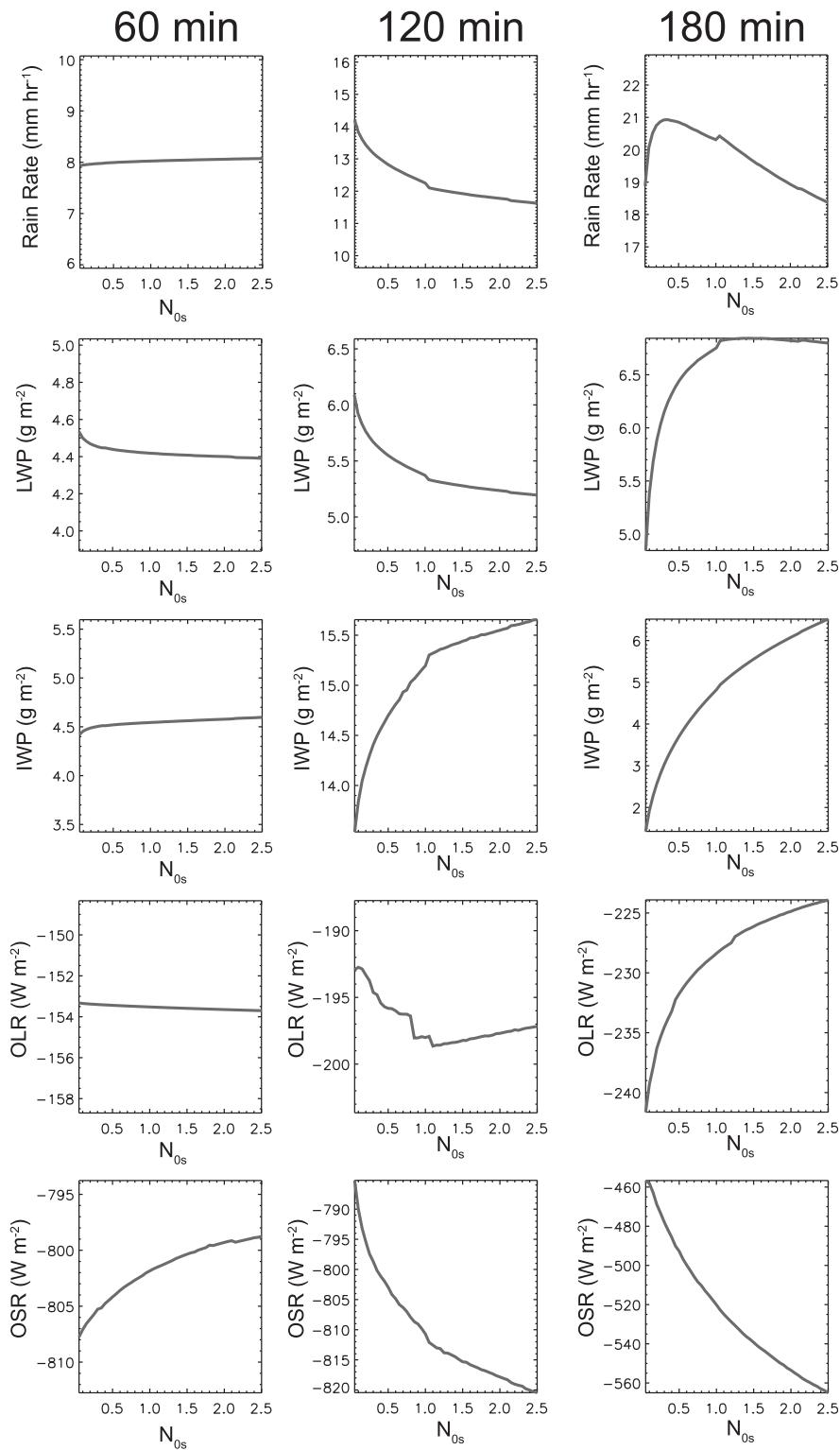


FIG. 5. As in Figs. 3 and 4, but for the slope intercept of the snow particle size distribution N_{0s} .

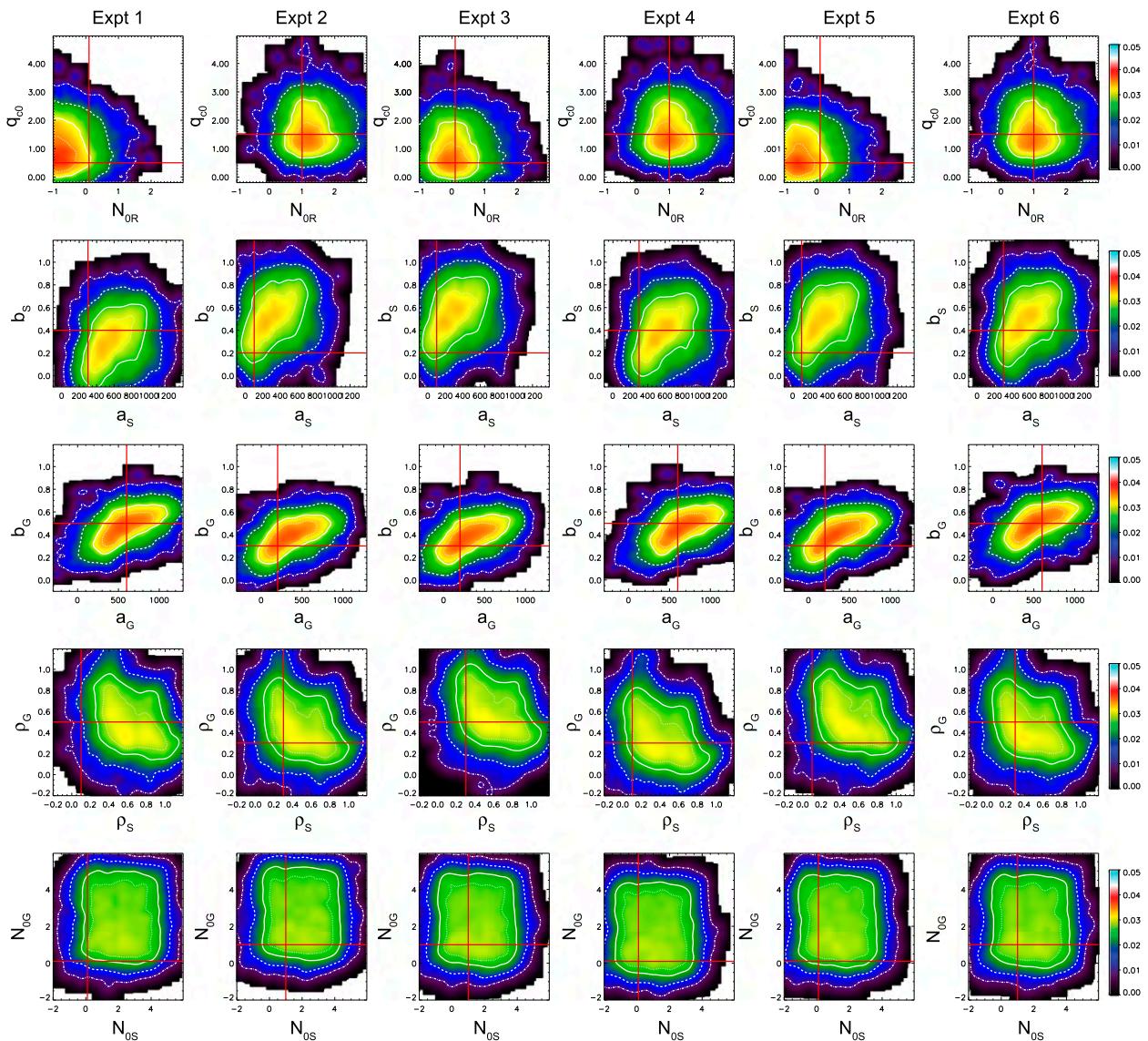


FIG. 6. As in Fig. 2, but for the perturbed observations EnKS.

experiment 1. In terms of reducing the number of non-physical values, the q_{c0} – N_{0r} plot for QES, experiment 1 is considerably better than the EnKS. However, in experiment 5, row 5 it is the QES posterior that has more unphysical values (including the mode) than the EnKS posterior.

Inspection of Figs. 6 and 7 also reveals the error variance for each distribution to be larger than the corresponding posterior PDFs returned by MCMC (Fig. 2). We shall explore the reason for this in more detail shortly, but note in brief that this must be the case when a linear or quadratic update is applied to an ensemble generated by a nonlinear model function. The error variance about a flawed estimate of the mean of a distribution

must necessarily be larger than the error variance about the true mean. The error variance about the QES estimate is smaller than or equal to the Kalman estimate because the QES estimate of the posterior mean is closer to the true mean than Kalman. As pointed out by Hodyss (2011), this is because the QES accounts for the third moment of the distribution in the update of the distribution mean. We elaborate on this point in section 4a. Consideration of the third moment is also the reason the QES distribution shifts a larger distance in parameter space in experiment 5 relative to EnKS and relative to QES in experiments 1–4 and 6. The true values are not only near zero in this case, but the posterior distributions in experiment 5 exhibit relatively large skewness.

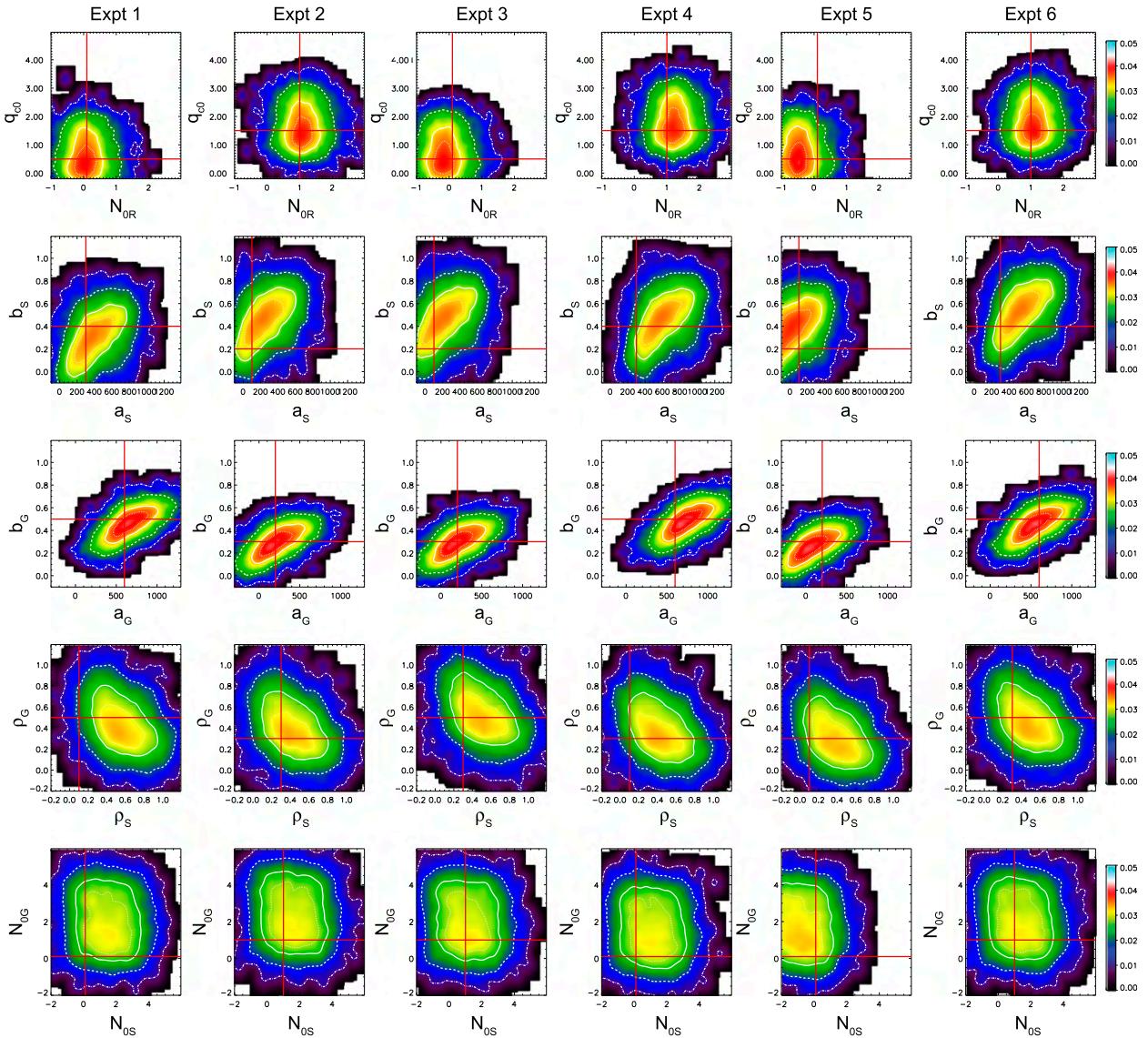


FIG. 7. As in Fig. 2, but for the quadratic ensemble smoother.

With the exception of the snow and graupel particle size distribution parameters in experiment 5 (Fig. 7: row 5, column 5), the mode of the QES posterior tracks the true posterior mode more accurately than in the EnKS, likely because accounting for the third moment in the ensemble update produces on average a superior estimate of the distribution mean. This does not prevent the QES from producing a negative (unphysical) mode for the q_{c0} - N_{0r} parameter combination in experiments 3 and 5 and for the N_{0s} - N_{0g} combination in experiment 5. This reflects the fact that the value of observations are not accounted for in the QES ensemble generation, only in the update of the ensemble mean. Indeed, the fact that the QES draws more accurately to the mean and

mode is a problem in cases for which both mean and mode are near zero. The posterior mean estimate is on average an improvement over the EnKS, but the ensemble generated from the QES update extends in these cases even farther into unphysical parameter territory. In this case a more useful ensemble generation scheme would be informed and adjusted by an observation's value and the proximity of the observation to zero and would subsequently reduce the error variance significantly and/or develop a strong third moment to ensure placement of ensemble members in physically realistic parts of phase space. See, in particular, the posterior ensemble estimates for a_s - b_s (Fig. 7, row 2) in experiment 5. Interestingly, even though the true values of a_s ,

and b_s are identical in experiments 2, 3, and 5, the posterior distribution of a_s – b_s in experiment 3 contains fewer subzero ensemble members. The only difference between experiment 3 and experiments 2 and 5 is fact that the snow and graupel densities and PSD slope intercepts are set large in experiment 3.

4. Discussion

a. A tutorial example using a chi-squared distribution

It was noted above that both the EnKS and QES construct the posterior analysis error ensemble without use of the observations drawn from the true state. We now show explicitly how the posterior distributions associated with both EnKS and QES consist of integrals over all possible innovations for a given prior ensemble. We begin by noting that both square root and perturbed observation ensemble generation are constrained to satisfy

$$\bar{\mathbf{P}}_a = (\mathbf{I} - \mathbf{KH})\mathbf{P}_f, \quad (23)$$

where \mathbf{P}_f is the error covariance matrix derived from the prior distribution, \mathbf{H} is the observation operator, \mathbf{K} is the Kalman gain,

$$\mathbf{K} = \mathbf{P}_f \mathbf{H}^T (\mathbf{H} \mathbf{P}_f \mathbf{H}^T + \mathbf{R})^{-1}, \quad (24)$$

\mathbf{R} being the observation error covariance matrix, and $\bar{\mathbf{P}}_a$ is the analysis error covariance matrix meant to describe the covariance matrix of the posterior distribution. A systematic discussion of (23) may be found in chapter 7 of Jazwinski (1970) as well as in Ghil and Malanotte-Rizzoli (1991).

There are a number of different ways of deriving (23) and (24) (e.g., Ghil and Malanotte-Rizzoli 1991; Daley 1991; and more recently Hodyss 2011). The derivation of (24) given in Hodyss (2011) is identical to these earlier derivations but it explicitly highlights the fact that (23) can be written in the following form:

$$\bar{\mathbf{P}}_a = \int_{-\infty}^{\infty} \mathbf{P}_a(\mathbf{v}) \rho(\mathbf{v}) d\mathbf{v}, \quad (25a)$$

where \mathbf{v} is the innovation vector and $\mathbf{P}_a(\mathbf{v})$ is the posterior error covariance matrix,

$$\mathbf{P}_a(\mathbf{v}) = \int_{-\infty}^{\infty} (\mathbf{x} - \bar{\mathbf{x}}_a)(\mathbf{x} - \bar{\mathbf{x}}_a)^T \rho(\mathbf{x} | \mathbf{v}) d\mathbf{x}, \quad (25b)$$

and $\rho(\mathbf{v})$ is the pdf that describes the distribution of innovations, $\rho(\mathbf{x} | \mathbf{v})$ is the pdf that describes the posterior distribution, and $\bar{\mathbf{x}}_a$ is the state estimate (commonly

referred to as an “analysis”) whose error variance is being measured. The nontraditional overbar in (25a) is designed to make it clear that this calculation returns the mean (or *expected*) posterior error covariance matrix, which is only equal to the true posterior error covariance matrix (25b) when the posterior has no skewness (Hodyss 2011). The fact that the estimate (23) is entirely independent of the innovation is implicitly evident in Daley (1991) and Ghil and Malanotte-Rizzoli (1991)’s derivations because expectations are taken over all possible forecast errors and all possible observation errors.

Because we wish to generate an ensemble whose error variance is consistent with the expected squared error in a data assimilation system given the specific observations obtained from the present analysis cycle, the algorithm should satisfy the stronger constraint in (25b) and produce an analysis ensemble consistent with a particular innovation rather than the mean analysis error distribution in (25a). It is important to keep in mind that even if the model is perfect and the ensemble has an infinite number of members that (23) will still be independent of innovation and therefore incorrect when the posterior distribution has a nonzero third moment. Because the analysis covariance matrix in (23) is therefore obviously incorrect we then might ask the following question: *What ensemble distribution do we generate when we generate an ensemble using (23)?*

To answer this question we note that by employing a change of variable of $\boldsymbol{\varepsilon} = \mathbf{x} - \bar{\mathbf{x}}_a$, combining (25a) and (25b), and then simply switching the order of integration we obtain the following:

$$\bar{\mathbf{P}}_a = \int_{-\infty}^{\infty} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \bar{\rho}(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon}, \quad (26a)$$

where we refer to $\bar{\rho}(\boldsymbol{\varepsilon})$ as a *mean error distribution*, which we define as

$$\bar{\rho}(\boldsymbol{\varepsilon}) \equiv \int_{-\infty}^{\infty} \rho(\boldsymbol{\varepsilon} | \mathbf{v}) \rho(\mathbf{v}) d\mathbf{v}. \quad (26b)$$

Equation (26b) describes the average posterior distribution of errors around the posterior mean averaged over all possible innovation vectors. Hence, the ensemble distribution that we generate when we use (23), or equivalently (25a), in a perturbed observation framework is the mean error distribution. As shown by Hodyss and Campbell (2013), in a square root ensemble generation framework only the second moment is consistent with the mean error distribution; all higher moments depend on the arbitrary rotation that defines the square root form [i.e., the arbitrary rotation that defines the difference between, say, Anderson (2001) and Bishop

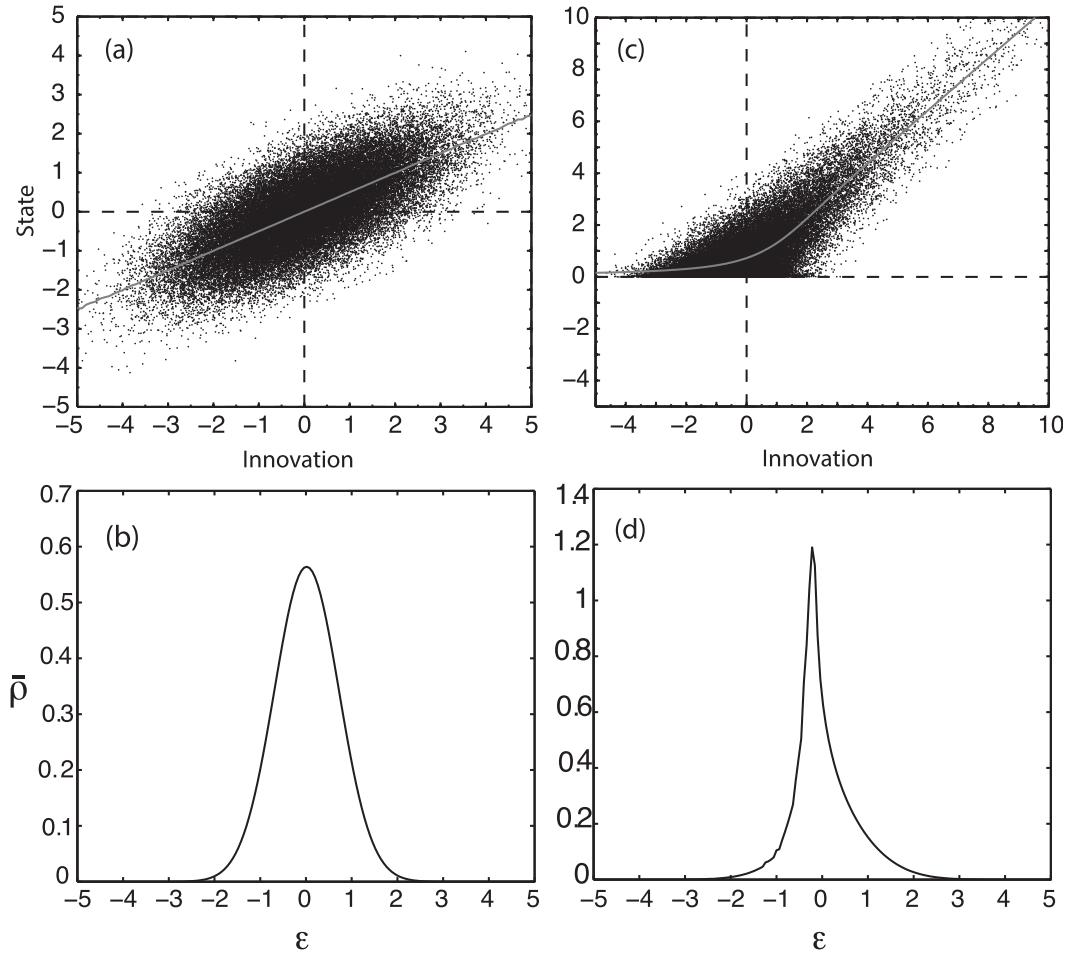


FIG. 8. Posterior and mean error distributions. (a) A Gaussian posterior and (b) its associated mean error distribution is shown. (c) A non-Gaussian posterior and (d) its associated mean error distribution is shown. The non-Gaussian posterior in (c) is constructed from a chi-square prior and a Gaussian observation likelihood. The gray line in (a) and (c) is the posterior mean.

et al. (2001)]. Next, we illustrate this result for two examples of mean error distributions next.

In the first example we construct a posterior distribution from a Gaussian prior distribution with mean 0 and a variance of 1 and an observation likelihood that has errors about the truth that is also Gaussian with mean 0 and a variance of 1 (Fig. 8a). We employ this posterior distribution in (26b) to obtain the mean error distribution shown in Fig. 8b. This mean error distribution is the Gaussian distribution that our ensemble generation schemes create, and is typically centered on the analysis obtained from our data assimilation system. Hence, we would center the distribution in Fig. 8a on our analysis to obtain our ensemble distribution. Because the error variance for Gaussian posteriors is independent of innovation we would, in this case, create the correct distribution for any value of the innovation. However,

a different situation arises when we employ this same technique for non-Gaussian posteriors.

If we alter the prior to be chi-square but keep the observation likelihood Gaussian as before, we obtain the mean error distribution shown in Fig. 8d. This mean error distribution is obviously non-Gaussian. Moreover, as shown by Hodyss (2011) the true error variance for this case should be a strong function of the innovation. This can be seen by the varying dispersion of states around the gray line as a function of innovation. Hence, it is not possible to obtain the correct posterior distribution by centering the fixed, mean error distribution in Fig. 8b on the analysis (gray line in Fig. 8c) for an arbitrary innovation. In fact, note that if the prior is chi-square this implies that we are examining a physical system in which there is zero probability of occurrence of negative values of the state. Unfortunately, an ensemble

generation scheme that produces the fixed, mean analysis distribution shown in Fig. 8d will produce negative values of the state for a wide range of innovations (i.e., for an innovation less than $v \approx 2$).

This simple example serves to illustrate two important points regarding the EnKS and QES. First, the shape of the posterior distribution does not change for different true parameter values because it is, by construction, formed via integration over all possible values of the innovation for a given prior. As such, the posterior analysis ensemble is essentially designed to be independent of the true state. The observations (or equivalently innovation) serve only to recenter the mean of the analysis error distribution. Second, when the prior is non-Gaussian, the posterior analysis error distribution should be a function of the innovation. Recentering a fixed distribution around an arbitrary innovation for the case of a non-Gaussian prior and nonlinear model is virtually guaranteed to produce an erroneous posterior analysis distribution. It is for this reason that the posterior distributions produced by EnKS and QES depicted in Figs. 6 and 7, respectively, 1) do not change shape when the true parameters are changed, and 2) contain non-physical (disallowed) solutions.

b. Comparing integrals over multiple draws of truth

A comparison is now conducted between the integral over all possible values of the innovation, as produced by the ensemble filters, and an integral over posterior PDFs produced by multiple draws of observations from a prior parameter distribution. As shown above, the analysis ensemble produced by EnKS and QES are independent of the specific observation, and can be obtained from a single analysis ensemble update. In contrast, the integral over accurate estimates of the Bayesian posterior distribution (as provided by MCMC) requires an ensemble of posterior PDFs consistent with random draws of true parameter sets (or equivalently, random draws of observations from the prior). This ensemble is produced by drawing 100 parameter sets at random from the bounded uniform prior, generating observations for each set by running the forward model with each parameter set specified to be the true values and adding Gaussian noise, and then using MCMC to produce a posterior ensemble from each set of observations. Note that, as before, the prior parameter distribution for each of the 100 cases is specified to be bounded uniform, and independent of true parameter value. A sample of 400 000 model integrations is generated for each of the 100 MCMC experiments, each posterior distribution is centered by removing its mean, and the resulting centered ensembles are combined in a unified posterior ensemble to produce an integral over

multiple parameter sets. As in the more limited ensemble described in section 3a, convergence of the MCMC algorithm is assessed via computation of differences between subsamples drawn from the full distribution for each of the 100 experiments.

A quantitative assessment of the differences between the integrated Bayesian ensemble, EnKS, and QES can be obtained by comparing the posterior error variance and covariance between distributions. The posterior error covariance is computed for the $N = 100$ experiments as

$$\bar{\mathbf{P}}_e = \frac{1}{(N-1)} \sum_{i=1}^N (\mathbf{x}_t - \bar{\mathbf{x}}_a)(\mathbf{x}_t - \bar{\mathbf{x}}_a)^T, \quad (27)$$

where \mathbf{x}_t is the true parameter vector. The error variance is

$$\langle \bar{\mathbf{E}}_e \rangle^2 = \frac{1}{(N-1)} \sum_{i=1}^N (\mathbf{x}_t - \bar{\mathbf{x}}_a)^2. \quad (28)$$

Comparison of the integrated MCMC distributions (row 1, Fig. 9) with those valid for a single realization of truth (Fig. 2) indicates that the structure of the PDFs produced by EnKS and QES (Fig. 9, rows 2 and 3, respectively) are far more representative of the integrated Bayesian ensemble than a single realization. Multimodality, present in the Bayesian posterior distributions for select true parameter sets, is not present in the integrated ensemble, and the sign and structure of the covariance is similar. However, posterior variance and error variance (Tables 4 and 5, respectively) are larger for both QES and EnKS than for MCMC. As noted by Hodyss and Campbell (2013), the posterior ensemble (co)variance is an estimate of the posterior error (co)variance; hence, the values in Tables 4 and 5 are consistent to within sampling error. As mentioned above in section 3b and in the tutorial example in section 4a, on average the posterior error variance is required to be larger than the true error variance if the ensemble update produces a flawed estimate of the mean; the mean returned by the EnKS and QES will not be the error minimizing estimate. Equation (23) is the true statement of the error variance around the Kalman estimate of the mean, but not around a given set of observations. The mean error variance for the QES will be on average less than or equal to that for EnKS because of the fact that the QES accounts for the third moment in its update of the ensemble mean. This is reflected in the plots of centered PDFs (Fig. 9): distributions are broader for EnKS and QES as compared with MCMC. In addition, the effects of the hard lower and upper parameter value bounds can be seen in the MCMC distributions: there

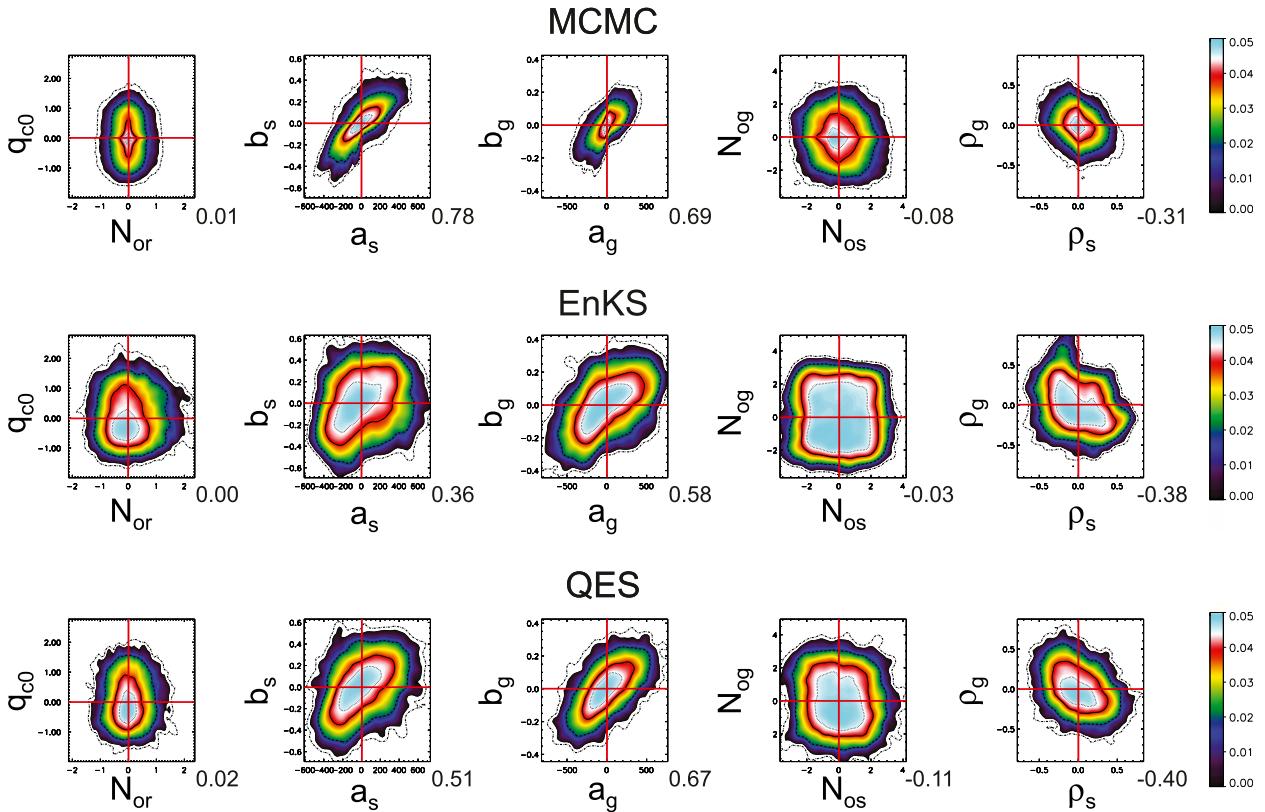


FIG. 9. Centered combined posterior PDFs for each pair of parameters represented in each column of Figs. 2, 6, and 7. Results from (top) MCMC, (middle) EnKS, and (bottom) QES are shown. Numerical values located to the bottom right of each subplot indicate the correlation of the analysis ensemble in units of the parameters (Table 1). The line and contour color scheme, as well as the specifics of the line and contour plots are as described in Fig. 2.

are “kinks” in the centered PDFs at zero; in contrast, EnKS and QES are smooth.

Examining individual parameter pairs, both ensemble smoothers produce a distribution for warm rain parameters N_{or} and q_{c0} (column 1, Fig. 9) that is qualitatively similar (in shape and correlation) to the ensemble produced by MCMC, but with a few important differences. The mode of the MCMC distribution is centered at zero; differencing the posterior distribution from the distribution mean results in an ensemble whose mean and maximum a posteriori estimate (mode) are identical. The posterior mean is equivalent to the mode only because each parameter range is bounded both above and below. In the tutorial example (Fig. 8d) the posterior

mean is not equal to the mode because there is a single lower bound. In contrast to the integrated MCMC distribution, when both EnKS and QES ensembles are standardized about their mean, the mode of the distribution is biased consistently toward low values of q_{c0} . Inspection of the posterior PDFs for individual parameter sets (Fig. 2) reveals that, when values of q_{c0} are relatively large, there is little skewness in the posterior distribution, and the mode extends from the true value toward larger values (experiments 2, 4, and 6). This means that, for relatively large true values of q_{c0} (greater than 1.0 g kg^{-1}), the center of the distribution mode is offset from the true value. In such cases, both EnKS and QES will produce a posterior analysis that is biased high

TABLE 4. Posterior ensemble standard deviations for the MCMC, QES, and EnKS posterior distributions. The units are consistent with those reported in Table 1.

	a_s	b_s	a_g	b_g	N_{or}	N_{os}	N_{og}	ρ_s	ρ_g	q_{c0}
MCMC	156.54	0.15	133.15	0.08	0.33	0.99	1.06	0.16	0.16	0.581
QES	201.63	0.19	210.82	0.11	0.41	1.20	1.29	0.22	0.20	0.607
EnKS	219.54	0.21	261.56	0.13	0.55	1.35	1.40	0.24	0.23	0.637

TABLE 5. Error standard deviations for the posterior parameter distributions from MCMC, QES, and EnKS. As in Table 4, the units are consistent with those reported in Table 1.

	a_s	b_s	a_g	b_g	N_{0r}	N_{0s}	N_{0g}	ρ_s	ρ_g	q_{c0}
MCMC	175.36	0.17	126.55	0.08	0.39	1.05	0.95	0.16	0.14	0.631
QES	238.98	0.19	225.20	0.10	0.45	1.20	1.33	0.21	0.21	0.683
EnKS	256.14	0.20	249.89	0.12	0.62	1.30	1.46	0.23	0.21	0.687

compared with the truth. There is nothing inherently wrong with the ensemble smoothers in this case; once q_{c0} exceeds a threshold, the forecast ensemble is relatively insensitive to changes in its value. As in Fig. 7, the QES produces a slightly improved estimate of the posterior distribution because it accounts for skewness in the update of the mean.

Snow and graupel fall speed parameter distributions (columns 2 and 3, respectively, Fig. 9) are similar across all three algorithms; the correlation between parameters that is present by construction in the cloud microphysics algorithm appears in each centered distribution. The same is true of the snow and graupel joint density. In contrast, N_{0s} and N_{0g} are centered, but tremendously overdispersive in the mean. This is largely because both of these parameters have a near-zero effect on the posterior until late in the simulation. PB12 demonstrated that the model forecast state at 120 min is somewhat sensitive to changes in these parameters, but these parameters do not begin to have any substantial effect on model output until 150 and 180 min into the simulation. PB12 hypothesized that EnKS is not able to track the rapid change in response function at these later times, and the QES, while it accounts more accurately for nonlinear updates, is similarly unable to keep pace with rapid changes in response function. Examination of the response function bears this out. The N_{0s} has little effect on the model output at 60 and 120 min other than on the shortwave radiation (Fig. 5), while the effect is stronger at 180 min. It is notable that the derivative of the response function changes sign between 120 and 180 min.

As with the PDFs of ice fall speed and intercept, the structure of the snow and graupel density PDF returned by the ensemble smoothers appears to be consistent with the MCMC ensemble. Parameters are weakly and negatively covariant, and the variance is inflated in the ensemble posteriors. Comparison of the response function of snow density (Fig. 10) versus that of graupel is quite informative here. At 60 min, the graupel and snow density response covaries strongly and positively; with increases in both snow and graupel density, rain rate, LWP, and OLR decrease, while IWP and OSR increase. At 120 min, sensitivity to graupel density is much larger than sensitivity to snow density, and snow density

response is of a different sign for rain rate, LWP, and IWP. At 180 min, sensitivity to snow outweighs graupel for radiative fluxes, but is similar for cloud content and rain rate. The sign of response in rain rate and liquid cloud is opposite that at 120 min and remains opposite each other (the sign reverses between ρ_s and ρ_g and between 120 and 180 min).

5. Summary and major conclusions

This paper explored how changes in observations map to changes in the posterior parameter probability distributions for a set of cloud microphysical parameters nonlinearly related to the output of a column cloud model. Samples of the exact Bayesian posterior distribution provided by a Markov chain Monte Carlo algorithm were contrasted with approximate solutions generated by an ensemble Kalman smoother and a quadratic ensemble smoother. The major conclusions of this study are the following:

- 1) The true analysis ensemble, as constructed from samples of the Bayesian posterior distribution, changes shape significantly with changes in the true parameter set for a model in which control parameters are nonlinearly related to the observations.
- 2) Multimodality is realized only in certain regions of the parameter space, and is associated with nonmonotonicity in the parameter-observation response function.
- 3) The slope of the parameter-observation response function determines parameter sensitivity, and, by extension, the posterior variance. A constant response function derivative leads to posterior variance that is independent of the true parameter value. This is consistent with results found by Hodyss (2011), who showed that the first derivative of the response function with respect to observations determines the posterior variance while the second derivative determines the posterior third moment.
- 4) The true error distribution for a given set of observations (the “error of the day”) consists of the Bayesian posterior PDF formed via the product of the prior forecast error distribution with the likelihood error distribution constructed from the observations. In the presence of nonlinearity, ensemble

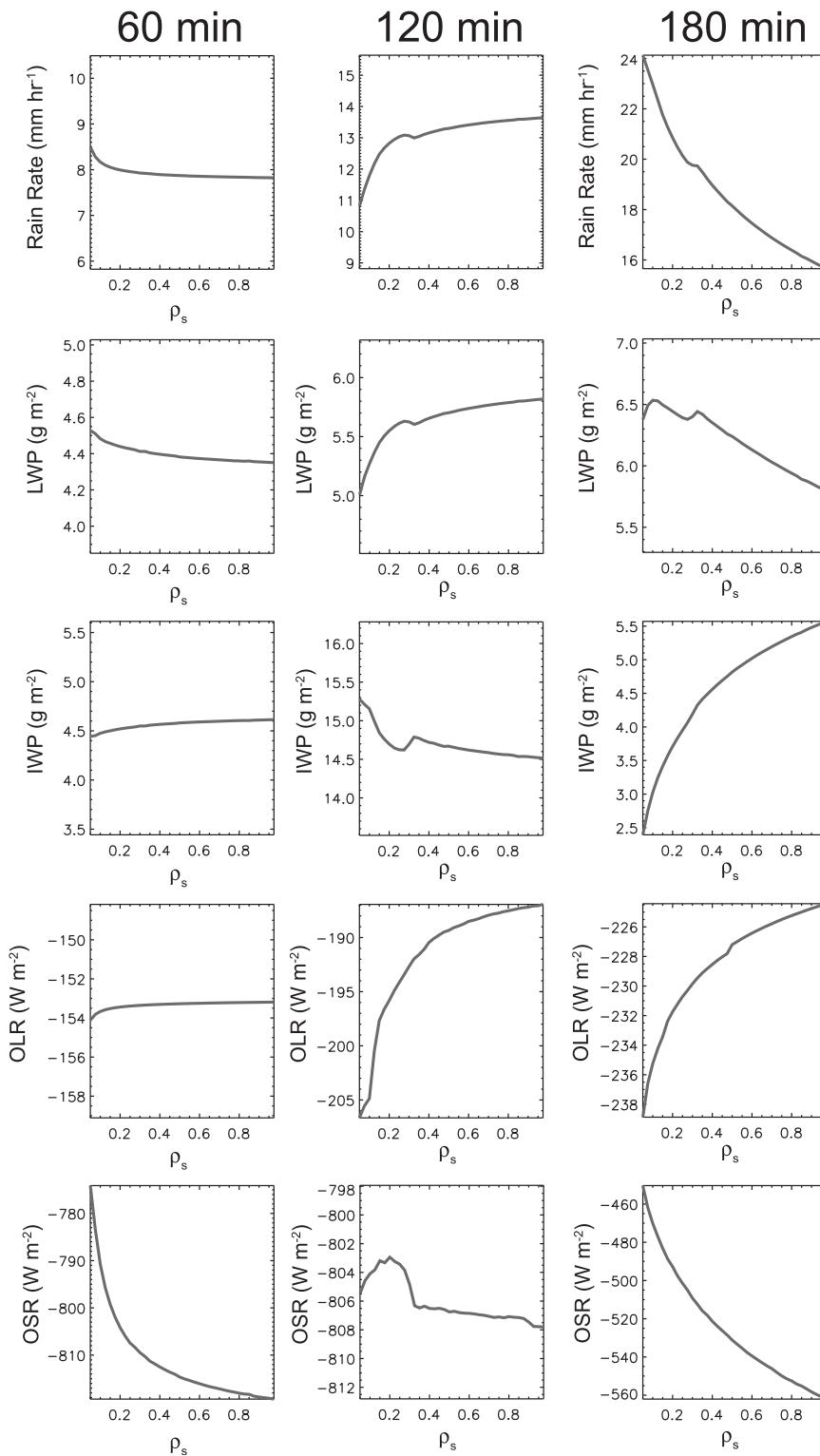


FIG. 10. As in Figs. 3, 4, and 5, but for snow density ρ_s .

smoothers do not reproduce this analysis distribution. Instead, the distribution produced by ensemble Kalman smoothers consists of the expected analysis error covariance matrix produced by integrating over all possible innovations for a given prior. As this ensemble is not consistent with the true Bayesian posterior distribution, ensemble smoothers or filters or any method based on a gain matrix (linear or nonlinear) that is computed without using information from the observations of the day does not, in general, reproduce the correct error distribution of the day.

- 5) It is illustrative, but not fully informative, to compare EnKS and QES analysis ensembles with the posterior PDF obtained from a single Bayesian posterior solution. When approximate ensemble solutions are compared with the integral over multiple Bayesian posteriors constructed from multiple draws of the true parameters from the prior, both EnKS and QES can be shown to provide a realistic estimate of the average posterior analysis distribution, but with larger ensemble variance than that of the average of the true posterior analysis distributions.
- 6) A filter constructed with a nonlinear update that accounts for the effects of skewness in the prior and posterior distribution produces on average an estimate that is more consistent with the true posterior ensemble mean, but which still fails for cases with nonmonotonic nonlinearity. For these cases, the mean is closer to the true mean than an EnKS algorithm, but, like the EnKS, is also not restricted to regions of phase space where known physically consistent solutions exist. In addition, for state estimates hard bounded at some value (e.g., zero for concentrations of scalar quantities), a significant portion of the posterior ensemble density may lie in an unphysical region of the parameter space. This becomes more marked when the observed concentrations and/or parameter values approach the specified limit. As noted in section 3, a more useful ensemble generation scheme would be informed and adjusted by an observation's value and the closeness of the observation to zero. As yet, an ensemble generation scheme consistent with the EnKS or QES that develops the correct observation dependent moments does not exist.

We leave the reader with three general comments on the topic of future research. First, there are two general aspects missing from current ensemble assimilation algorithms: 1) the ability to properly estimate the mean and error distribution for quantities that are hard-bounded (and whose error asymptotes to zero as the quantity itself asymptotes to zero), and 2) the ability to

account for a multimode posterior distribution. These are topics that continue to be actively researched, and it is possible that modifications to the EnKF framework, combined with selection of modified or transformed control parameters, will lead to significant progress in the near future.

Second, a primary application of ensemble filter-based data assimilation algorithms is production of an improved forecast, via construction of an analysis with error characteristics consistent with the current state of the system. While we have been primarily concerned with exploring differences between the Bayesian posterior analysis and the approximate analysis generated by ensemble methods, a discrepancy between the true and approximate analysis mean and covariance is bound to have an effect on the forecast as well. The specific effect will, of course, depend on the state of the system; for example, the true parameter values, their covariances, and the nature of the deviation from the exact Bayesian estimate. While it is tempting to draw general conclusions about the nature of the forecast errors, this is simply not possible without generating sample forecasts from a variety of analysis distributions, then comparing them with the forecast generated from the Bayesian posterior analysis. We reserve presentation of this work for a future manuscript.

Finally, it should be noted in closing that, while ensemble smoothers do not incorporate observations into the generation of the analysis ensemble, there are two aspects of the ensemble update that can result in incorporation of the "error of the day" into the ensemble distribution. First, the current set of observations is used to create the ensemble mean, and this step is done in a self-consistent manner in both the EnKS and QES algorithms. To the extent that cycling ensemble assimilation systems contain information from past observations that has been incorporated into multiple successive updates of the mean, they are consistent with recent error structures, but this information is not provided by the observations (or equivalently innovations) of the day. Second, the generation of ensemble perturbations around the analysis mean can be performed taking into account the properties of the physical system and the most recent observations to ensure ensemble members are restricted to the correct locations within phase space. This second step is not presently done correctly in any standard ensemble generation algorithm.

Acknowledgments. DJP's contribution to this paper was funded by the U.S. Office of Naval Research Grant N00173-10-1-G035 while CHB was supported by the U.S. Office of Naval Research Grant 4596-0-1-5. Support

for DH was provided by the Chief of Naval Research PE-0601153N.

REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Brooks, S., A. Gelman, G. Jones, and X. L. Meng, Eds., 2011: *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 619 pp.
- Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*, **75**, 257–288.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10 143–10 162.
- , 2009: *Data Assimilation: The Ensemble Kalman Filter*. 2nd ed. Springer, 307 pp.
- Gelman, A., G. O. Roberts, and W. R. Gilks, 1996: Efficient Metropolis jumping rules. *Bayesian Statistics V*, J. M. Bernardo et al., Eds., Oxford University Press, 599–608.
- , J. B. Carlin, H. S. Stern, and D. B. Rubin, 2004: *Bayesian Data Analysis*. 2nd ed. Chapman and Hall/CRC, 696 pp.
- Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Advances in Geophysics*, Vol. 33, Academic Press, 141–266.
- Haario, H., E. Saksman, and J. Tamminen, 1999: Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Stat.*, **14**, 375–395.
- Heymtsfield, A. J., A. Bansemer, P. R. Field, S. L. Durden, J. L. Smith, J. E. Dye, W. Hall, and C. A. Grainger, 2002: Observations and parameterizations of particle size distributions in deep tropical cirrus and stratiform precipitating clouds: Results from in situ observations in TRMM field campaigns. *J. Atmos. Sci.*, **59**, 3457–3491.
- Hodyss, D., 2011: Ensemble state estimation for nonlinear systems using polynomial expansions in the innovation. *Mon. Wea. Rev.*, **139**, 3571–3588.
- , 2012: Accounting for skewness in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 2346–2358.
- , and W. F. Campbell, 2013: Square root and perturbed observation ensemble generation techniques in Kalman and quadratic ensemble filtering algorithms. *Mon. Wea. Rev.*, **141**, 2561–2573.
- Houze, R. A., 2004: Mesoscale convective systems. *Rev. Geophys.*, **42**, RG4003, doi:10.1029/2004RG000150.
- Jazwinski, H., 1970: *Stochastic Processes and Filtering Theory*. Vol. 64, *Mathematics in Science and Engineering*, Academic Press, 376 pp.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems, transactions of the ASME. *J. Basic Eng.*, **82**, 35–45.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge University Press, 341 pp.
- L’Ecuyer, T. S., and G. L. Stephens, 2002: An estimation-based precipitation retrieval algorithm for attenuating radars. *J. Appl. Meteor.*, **41**, 272–285.
- , and —, 2003: The tropical oceanic energy budget from the TRMM perspective. Part I: Algorithm and uncertainties. *J. Climate*, **16**, 1967–1985.
- Lang, S., W.-K. Tao, R. Cifelli, W. Olson, J. Halverson, S. Rutledge, and J. Simpson, 2007: Improving simulations of convective systems from TRMM LBA: Easterly and westerly regimes. *J. Atmos. Sci.*, **64**, 1141–1164.
- Lewis, J., S. Lakshmiarahan, and S. Dhall, 2006: *Dynamic Data Assimilation: A Least Squares Problem*. Cambridge University Press, 680 pp.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092.
- Locatelli, J. D., and P. V. Hobbs, 1974: Fall speeds and masses of solid precipitation particles. *J. Geophys. Res.*, **79**, 2185–2197.
- Mitchell, D. L., 1996: Use of mass- and area-dimensional power laws for determining precipitation particle terminal velocities. *J. Atmos. Sci.*, **53**, 1710–1723.
- Moncrieff, M. W., 1992: Organized convective systems: Archetypal dynamical models, mass and momentum flux theory, and parameterization. *Quart. J. Roy. Meteor. Soc.*, **118**, 819–850.
- Mosegaard, K., and A. Tarantola, 2002: Probabilistic approach to inverse problems. *International Handbook of Earthquake & Engineering Seismology (Part 2)*, W. H. K. Lee et al., Eds., Academic Press, 237–265.
- PosseLT, D. J., 2013: Markov chain Monte Carlo methods: Theory and applications. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, 2nd ed. S. K. Park and L. Xu, Eds., Springer, 59–87.
- , and T. Vukicevic, 2010: Robust characterization of model physics uncertainty for simulations of deep moist convection. *Mon. Wea. Rev.*, **138**, 1513–1535.
- , and C. H. Bishop, 2012: Nonlinear parameter estimation: Comparison of an ensemble Kalman smoother with a Markov chain Monte Carlo algorithm. *Mon. Wea. Rev.*, **140**, 1957–1974.
- , T. S. L’Ecuyer, and G. L. Stephens, 2008: Exploring the error characteristics of thin ice cloud property retrievals using a Markov chain Monte Carlo algorithm. *J. Geophys. Res.*, **113**, D24206, doi:10.1029/2008JD010832.
- Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding, Theory and Practice*. World Scientific, 240 pp.
- Roy, S. S., R. K. Datta, R. C. Bhatia, and A. K. Sharma, 2005: Drop size distributions of tropical rain over south India. *Geofizika*, **22**, 105–130.
- Rutledge, S. A., and P. V. Hobbs, 1983: The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. VIII: A model for the “seeder-feeder” process in warm-frontal rainbands. *J. Atmos. Sci.*, **40**, 1185–1206.
- , and —, 1984: The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. Part XII: A diagnostic modeling study of precipitation development in narrow cold frontal rainbands. *J. Atmos. Sci.*, **41**, 2949–2972.
- Tamminen, J., and E. Kyrölä, 2001: Bayesian solution for nonlinear and non-Gaussian inverse problems by Markov chain Monte Carlo method. *J. Geophys. Res.*, **106** (D13), 14 377–14 390.
- Tao, W.-K., and J. Simpson, 1993: Goddard Cumulus Ensemble Model. Part I: Model description. *Terr. Atmos. Oceanic Sci.*, **4**, 35–72.

- , and Coauthors, 2003: Microphysics, radiation, and surface processes in the Goddard Cumulus Ensemble (GCE) model. *Meteor. Atmos. Phys.*, **82**, 97–137.
- , and Coauthors, 2014: The Goddard Cumulus Ensemble (GCE) model: Improvements and applications for studying precipitation processes. *Atmos. Res.*, in press.
- Tarantola, A., 2005: *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 342 pp.
- Tokay, A., and D. A. Short, 1996: Evidence from tropical raindrop spectra of the origin of rain from stratiform versus convective clouds. *J. Appl. Meteor.*, **35**, 355–371.
- van Lier-Walqui, M., T. Vukicevic, and D. J. Posselt, 2012: Quantification of cloud microphysical parameterization uncertainty using radar reflectivity. *Mon. Wea. Rev.*, **140**, 3442–3466.
- , —, and —, 2014: Linearization of microphysical parameterization uncertainty using multiplicative process perturbation parameters. *Mon. Wea. Rev.*, **142**, 401–413.
- Vukicevic, T., and D. J. Posselt, 2008: Analysis of the impact of model nonlinearities in inverse problem solving. *J. Atmos. Sci.*, **65**, 2803–2823.