

© Copyright 2012 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.

Nonlinear Parameter Estimation: Comparison of an Ensemble Kalman Smoother with a Markov Chain Monte Carlo Algorithm

DEREK J. POSSELT

University of Michigan, Ann Arbor, Ann Arbor, Michigan

CRAIG H. BISHOP

Naval Research Laboratory, Monterey, California

(Manuscript received 15 September 2011, in final form 12 January 2012)

ABSTRACT

This paper explores the temporal evolution of cloud microphysical parameter uncertainty using an idealized 1D model of deep convection. Model parameter uncertainty is quantified using a Markov chain Monte Carlo (MCMC) algorithm. A new form of the ensemble transform Kalman smoother (ETKS) appropriate for the case where the number of ensemble members exceeds the number of observations is then used to obtain estimates of model uncertainty associated with variability in model physics parameters. Robustness of the parameter estimates and ensemble parameter distributions derived from ETKS is assessed via comparison with MCMC. Nonlinearity in the relationship between parameters and model output gives rise to a non-Gaussian posterior probability distribution for the parameters that exhibits skewness early and multimodality late in the simulation. The transition from unimodal to multimodal posterior probability density function (PDF) reflects the transition from convective to stratiform rainfall. ETKS-based estimates of the posterior mean are shown to be robust, as long as the posterior PDF has a single mode. Once multimodality manifests in the solution, the MCMC posterior parameter means and variances differ markedly from those from the ETKS. However, it is also shown that if the ETKS is given a multimode prior ensemble, multimodality is preserved in the ETKS posterior analysis. These results suggest that the primary limitation of the ETKS is *not* the inability to deal with multimodal, non-Gaussian priors. Rather it is the inability of the ETKS to represent posterior perturbations as nonlinear functions of prior perturbations that causes the most profound difference between MCMC posterior PDFs and ETKS posterior PDFs.

1. Introduction

It is increasingly recognized that errors and/or uncertainty in model physics parameterizations are a primary source of forecast error in weather and climate prediction (Stensrud et al. 2000; Murphy et al. 2004; Palmer et al. 2005; Stainforth et al. 2005; Järvinen et al. 2010; Neelin et al. 2010; Berner et al. 2011). In particular, simplifying assumptions about the form of the particle size distribution of ice and liquid condensate have an important effect on the details of cloud and precipitation development and feedback on the radiative fluxes, heating rates, and thermodynamic environment. It has been shown that, regardless of the details of

the microphysical formulation, model output is nearly uniformly highly sensitive to small changes in these parameters (Tao et al. 1995; Grabowski et al. 1999; Wu et al. 1999; Petch and Gray 2001; Grabowski 2003; Gilmore et al. 2004; van den Heever and Cotton 2004). In principle, the relative quality of a set of parameters can be judged by its ability to produce trajectories that stay close to time sequences of observations whose error distribution is known. However, in many circumstances, observations do not unequivocally identify one true set of parameters—at most they indicate some sort of probability distribution of parameters. The issue of how to quantitatively represent such uncertainties is one of the most challenging issues in probabilistic weather and climate forecasting, stochastic physics parameterization, and data assimilation.

Bayes' theorem is generally recognized as providing a rigorous framework within which to accurately characterize the probabilities of interest. However, for very

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

high-dimensional systems like the atmosphere and ocean, accurate solutions to Bayes' equation in the presence of non-Gaussian forecast and observation errors have, so far, proved to be computationally intractable. A number of recent studies have used data assimilation techniques based on Gaussian approximations to Bayes' theorem to study the effects of variation in model physics parameters (Aksoy et al. 2006; Tong and Xue 2008a,b) or parameterization schemes (Meng and Zhang 2007). These studies have met with some success, but have been hampered by the difficulty inherent in representing the functional relationship between parameters and model state (termed *parameter identifiability*; Anderson 2001) in a linear Gaussian framework. A diverse set of techniques that are not restricted to Gaussian statistics are currently being explored (e.g., Bocquet et al. 2010).

Posselt and Vukicevic (2010, hereafter PV10) examined the properties of a set of cloud microphysical parameters known to 1) control output liquid and ice content and precipitation (and, by extension, top-of-the-atmosphere radiative fluxes), 2) vary for differing cloud types and environment, and 3) be nonlinearly related to simulated cloud physical and radiative properties. They used a simplified model of deep convection along with a nonlinear non-Gaussian probability density function (PDF) sampling algorithm to characterize the parameter joint PDF conditioned on a set of observations of cloud liquid and ice content, precipitation rate, and radiative fluxes. They found that 1) nonlinearity in the model led to PDFs that deviated significantly from the Gaussian form, exhibiting strong skewness and in some cases multiple modes; and 2) the relationship between parameters and model output changed for convective versus stratiform environments.

In this paper, we extend the work of PV10 to 1) examine the time evolution of parameter sensitivity and 2) examine the extent to which an ensemble Kalman smoother (EnKS) type algorithm is able to represent the salient characteristics of the non-Gaussian PDF. We are especially interested in how the ensemble transform Kalman smoother (ETKS) handles skewness in the posterior PDF, as well as the transition between a single- and multiple-mode solution. The goals of this study are the following:

- 1) To examine how the PDFs of parameters evolve in time and identify whether (and when) the potential for nonmonotonic nonlinearity (e.g., a multimode posterior PDF) appears in the system.
- 2) To construct an EnKS algorithm for the same system to examine how well EnKS is able to characterize non-Gaussian posterior PDFs.

The remainder of this paper is organized as follows. Section 2 contains a description of the column model

used in our experiments, while sections 3 and 4 introduce the Markov chain Monte Carlo (MCMC) and ETKS algorithms. Results of the MCMC- and ETKS-based parameter sensitivity experiments are presented in sections 5 and 6, respectively. A discussion of the results is presented in section 7 and a summary and conclusions are offered in section 8.

2. Column model and observations

Uncertainty in cloud microphysical parameters affects simulated cloud and precipitation fields, which naturally feed back to the dynamics on multiple scales. Because of the complexity in the microphysics–dynamics relationship, a necessary first step in understanding the details of model error is to consider the effect of changes in model physics parameters on the model output in isolation from any feedback to the cloud-scale dynamics. Such an exercise requires a framework that accommodates realistic changes to the environment so that the cloud microphysical routine can respond as it would if it were embedded in a fully interactive model. To this end, PV10 constructed what they termed a Lagrangian column model designed to emulate the changes in environment experienced by an atmospheric column as it moves through a cloud system following the mean flow. The model consists of 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). The GCE model employs a single-moment bulk 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 adhere to an exponential particle size distribution (Lin et al. 1983; Rutledge and Hobbs 1983, 1984; Tao et al. 2003; Lang et al. 2007).

In the Lagrangian column framework, clouds are generated by forcing the model with 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. By specifying appropriate base state potential temperature and water vapor and time-varying vertical profiles of vertical motion and water vapor forcing, 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 is of interest in that it contains two discrete cloud morphologies; convective, in which precipitation is primarily generated by the collision–coalescence (warm rain) process, and stratiform, in which the melting of falling ice particles (e.g., snow and graupel) play a key role. The simulation is constructed so

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
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
Threshold cloud mass mixing ratio for autoconversion to rain	q_{c0}	g kg^{-1}	1	0.1	3

that the output is consistent with idealized (Moncrieff 1992) and observed (Houze 2004) squall lines. The details of the simulation and model setup are identical to what is used in PV10 and in the interest of brevity we will not repeat them here. As in PV10, we examine sensitivity of the model to changes in 10 parameters that govern the ice particle fall speed, ice and liquid particle size distribution, and ice density (Table 1). The range variability for each parameter is taken from observations of ice crystal shape and particle size distribution (Locatelli and Hobbs 1974; Mitchell 1996; Heymsfield et al. 2002).

As in PV10, we examine the influence of changes to cloud microphysical parameters on simulated precipitation rate (PCP), liquid water path (LWP), ice water path (IWP), top-of-the-atmosphere outgoing longwave (OLR) and shortwave (OSR) radiative flux, as they are directly related to cloud microphysical processes, relevant to the study of cloud-radiative feedbacks in the climate system, and available as retrievals from current satellite remote sensing platforms. These variables also serve as observations in our inverse modeling algorithm (see Table 2), and we note that because they are output directly from the model, we require no formal observation operator. Avoiding the use of a complicated (and likely nonlinear) observation operator (as would be necessary for radar reflectivity observations, for example), allows for a cleaner evaluation of the relationship between changes to parameters and the effect on model output. The pros and cons of our choice of observations are discussed in PV10, and the use of radar reflectivity observations to characterize cloud microphysical parameter PDFs is explored in van Lier-Walqui et al. (2011, manuscript submitted to *Mon. Wea. Rev.*).

3. Markov chain Monte Carlo algorithm

Most modern data assimilation techniques and retrieval algorithms are based on the assumption that the

probabilities represented in Bayes' theorem are Gaussian. The derivation of the solution to Bayes' theorem for the case of Gaussian PDFs has been well documented (Rodgers 2000; Kalnay 2003; Evensen 2006; Lewis et al. 2006; Vukicevic and Posselt 2008, among many others) and will not be repeated here. Instead, we address the question of how to compute each PDF in the case where PDFs are not assumed to be Gaussian, and may, in fact, not have a unique mode. MCMC algorithms sample the posterior probability distribution in such a way as to revisit those regions with high probability while avoiding regions with low probability (Tamminen and Kyrola 2001; Tamminen 2004; Tarantola 2005; Posselt et al. 2008; PV10). Specifically, the MCMC algorithm is made up of a random walk that consists of multiple successive iterations (see the flowchart in Fig. 4 of PV10). In each iteration, candidate values of all parameters are randomly drawn from a "proposal" distribution that depends on the current 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 that quantifies the goodness of fit between model and observations. Proposed parameter values are immediately accepted as the next sample in the posterior distribution if they generate a better fit to observations. If not, then a test value is drawn from a uniform $[0, 1]$ distribution. If this value is less than the ratio of the new

TABLE 2. Observations used in MCMC and ETKS parameter sensitivity experiments, along with their units and error estimates. TOA: top-of-the-atmosphere, LW: longwave, SW: shortwave.

Obs	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}

to the old likelihood, then the proposed parameter set is also accepted as a sample in the distribution. This is referred to as “probabilistic acceptance.” Otherwise, the proposed set of parameters is rejected, the current set is stored as another sample of the PDF, and new proposed parameter values are drawn based on the original set. The accept/reject procedure ensures that parameter sets that provide a better fit to the observations are immediately accepted, those that provide a similar fit are considered, and those that lead to simulated observations that are very different from the measurements are rejected. In the process, regions of the parameter space with relatively high probability are preferentially sampled, while regions with low probability are avoided, and a sample of the posterior distribution is produced using far fewer iterations than direct computation of the PDF. A more complete description of the theoretical underpinnings of the MCMC algorithm can be found in Mosegaard and Tarantola (2002) and in Tarantola (2005), 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 with variance that is adaptively changed during a burn-in period 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. For a discussion on optimal sampling rates, the reader is referred to Gelman et al. (1996) and Haario et al. (1999). We assign a bounded uniform PDF as the prior parameter distribution, and set the parameter ranges to correspond to 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). In specifying a bounded uniform as the parameter prior, we make the prior distribution nearly noninformative, though there is some potential dependence on the specified parameter range. We 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.

In each experiment, a truth state is generated by running a 3-h integration of the model with parameter values consistent with what is currently used in cloud-resolving and mesoscale models (Table 1). Observations of precipitation rate, liquid and ice water path, and outgoing longwave and shortwave radiative fluxes are drawn from this truth state every 30 min starting 30 min into the simulation and ending at 180 min. Observation errors

are assumed to have a Gaussian form, 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). The reduced uncertainty estimates reported in PV10 are used for the radiative fluxes, as these values more effectively illustrate the inherent nonlinearity in the parameter–state relationship. It should be noted that the form of the observation error is not constrained to be Gaussian (Posselt et al. 2008); we use this form in our experiments to be consistent with the error estimates associated with TRMM retrievals. Observation uncertainty standard deviations are listed in Table 2.

According to the results presented by Haario et al. (1999), between 20 000 and 40 000 iterations should be sufficient to sample a 10-dimensional parameter space if the target (posterior) distribution is multivariate Gaussian. In reality, as will be shown later in section 5, 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, we ran MCMC out to 4×10^6 iterations. In our experiments, convergence to sampling a stationary distribution was ensured by drawing random samples of 100 000 parameter sets from the full MCMC sample and computing the integrated absolute difference between histograms. We found an average of less than 1% difference between any two given random samples, indicating robust sampling of the underlying (true) PDF.

4. Ensemble transform Kalman smoother and ensemble Kalman smoother

Bishop et al.'s (2001) ensemble transform Kalman filter (ETKF) provides a rapid method for solving the Kalman filter equations in the case that (i) the forecast error covariance matrix is given by the sample covariance of an ensemble forecast, and (ii) the number of observations p exceeds the number of ensemble members K . Here, we present a variant of this algorithm suitable for the problem of parameter estimation and also for cases in which $K \gg p$.

A model trajectory initialized from the i th plausible initial state using the i th plausible set of model parameters is uniquely defined by the vector \mathbf{x}_i^f listing all of the discrete 3D states defining the 4D trajectory from the i th initial condition using the i th set of model parameters. Suppose that one can generate a large set $\mathbf{X}^f = [\mathbf{x}_1^f, \mathbf{x}_2^f, \dots, \mathbf{x}_K^f]$ of such vectors that randomly sample a presumed prior distribution of plausible model trajectories and model parameters. Further suppose that one had access to a vector of imperfect observations

\mathbf{y} ($\mathbf{y} = \mathbf{y}^t + \varepsilon^0$) of some linear function $\mathbf{y}^t = \mathbf{H}\mathbf{x}^t$ of the true trajectory \mathbf{x}^t . It can be shown that if both forecast and observation errors had Gaussian distributions with covariances \mathbf{P}^f and \mathbf{R} , respectively, then Bayes' theorem implies a posterior mean $\bar{\mathbf{x}}^a$ given by

$$\begin{aligned} \bar{\mathbf{x}}^a &= \bar{\mathbf{x}}^f + \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} [\mathbf{y} - \bar{\mathbf{H}} \bar{\mathbf{x}}^f] \\ &= \bar{\mathbf{x}}^f + \mathbf{P}^a \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y} - \bar{\mathbf{H}} \bar{\mathbf{x}}^f], \end{aligned} \quad (1)$$

where \mathbf{P}^a is the posterior covariance of the errors of the analysis mean $\bar{\mathbf{x}}^a$. If the Gaussian error PDF assumptions of the Kalman filter are satisfied then

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{P}^f = \mathbf{Z}^a \mathbf{Z}^{aT}, \quad (2)$$

where $\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{fT}$. In the appendix, it is shown that the square root \mathbf{Z}^a of \mathbf{P}^a is given by

$$\begin{aligned} \mathbf{Z}^a &= \mathbf{Z}^f \{ \mathbf{I} - \mathbf{L} [\mathbf{I} - (\mathbf{\Gamma} + \mathbf{I})^{-1/2}] \mathbf{L}^T \} \\ &= \{ \mathbf{I} - \mathbf{L} [\mathbf{I} - (\mathbf{\Gamma} + \mathbf{I})^{-1/2}] \mathbf{\Gamma}^{-1/2} \mathbf{E}^T \tilde{\mathbf{H}} \} \mathbf{Z}^f = \tilde{\mathbf{K}} \mathbf{Z}^f, \end{aligned} \quad (3)$$

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 and $\mathbf{\Gamma}$ is a diagonal eigenvalue matrix both of which are defined by the eigenvector decomposition $\mathbf{E} \mathbf{\Gamma} \mathbf{E}^T = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{P}^f \mathbf{H}^T \mathbf{R}^{-1/2}$. As discussed in the appendix, the form of the square root given by (3) requires the eigenvector decomposition of a $p \times p$ matrix, whereas Bishop et al's (2001) ETKF requires the eigenvector decomposition of a $K \times K$ matrix. Because of the high cost of eigenvector decompositions, (3) is more efficient than the form given in Bishop et al. (2001) when $K \gg p$, as is the case in this paper.

Having obtained the square root \mathbf{Z}^a from (3), we use $\mathbf{P}^a = \mathbf{Z}^a \mathbf{Z}^{aT}$ in (1) to obtain $\bar{\mathbf{x}}^a$. The posterior ensemble members defining the ETKS posterior distribution of ensemble members are then given by

$$\mathbf{x}_i^a = \bar{\mathbf{x}}^a + (\sqrt{K} - 1) \mathbf{z}_i^a, \quad (4)$$

where \mathbf{z}_i^a is the i th column of \mathbf{Z}^a . Given sufficient ensemble members, the ensemble members given by (4) can be used to estimate the ETKS posterior PDF, which, in turn, can be compared to the MCMC posterior PDF.

For comparisons sake, we will also show the posterior PDF produced by a perturbed observations or stochastic EnKS (e.g., Houtekamer and Mitchell 1998; Burgers et al. 1998; Tippett et al. 2003). In most ensemble Kalman filter literature, the perturbed observations form is referred to as simply, the ensemble Kalman filter (EnKF)

or EnKS when the desired state estimate is valid toward the beginning of the data assimilation window. To emphasize the fact that our implementation of the EnKS uses precisely the same gain matrix as that used in (1) to update the mean of the posterior PDF of the ETKS, we shall refer to it as the ETKS_PO method. To create the i th perturbed observation, we use the following equation:

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

In other words, the i th observation vector is created by adding a random p vector with mean $\mathbf{0}$ and covariance \mathbf{R} to the original observation vector. The formula for the i th member of the posterior ETKS_PO is then given by

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

where \mathbf{Z}^a is obtained from (3).

Although the Kalman gain $\mathbf{P}^a \mathbf{H}^T \mathbf{R}^{-1}$ is a *nonlinear* function of the forecast error variance, (1) represents the mean of the ETKS posterior PDF as a linear function of the forecast error of the prior mean (because the innovation is a linear function of the actual forecast error). Similarly, though the modified gain $\tilde{\mathbf{K}}$ is a nonlinear function of forecast error variance, the form $\mathbf{Z}^f \{ \mathbf{I} - \mathbf{L} [\mathbf{I} - (\mathbf{\Gamma} + \mathbf{I})^{-1/2}] \mathbf{L}^T \}$ in (3) and (4) give the ETKS posterior perturbations as strictly linear functions of the prior perturbations. In contrast, the true posterior mean and perturbations obtained from the MCMC are arbitrarily high-order nonlinear functions of the error of the prior mean and the prior perturbations. Equation (6) shows that the ETKS_PO posterior perturbations are given by linear combinations of the forecast perturbations, while ETKS_PO is again a linear function of the actual forecast error.

In our particular application, we simplify the estimation problem by assuming that we precisely know the initial atmospheric state while being uncertain about the cloud model parameters defining the evolution of the atmosphere. Hence, at the initial time, the only uncertainties in the system are the cloud model parameters and it is these parameters that are the focus of this study. Hence, we only need to evaluate the elements pertaining to cloud parameters of the ETKS posterior vector \mathbf{x}_i^a and the ETKS_PO posterior vector $(\mathbf{x}_{\text{ETKS_PO}}^a)_i$ to define the ETKS and EnKS posterior cloud parameter PDFs.

In our experiments, we use an ensemble of 10 000 members drawn from a multivariate bounded uniform distribution, with bounds set equal to the MCMC parameter ranges (Table 1). Use of an ensemble that is large compared to the dimension of the state space avoids

sampling issues reported in earlier parameter estimation studies (e.g., Tong and Xue 2008a,b). We examine the effect on the ETKS and ETKS_PO posterior PDFs of increasing the number of observations constraining the parameter estimation problem by simply lengthening the data assimilation time window. For short time window experiments, the sizes of the vectors and matrices describing the 4D-state estimate are correspondingly smaller than when long time window experiments are performed. The form of the above equations is preserved regardless of whether the time window is short or long.

5. MCMC-based temporal evolution of parameter PDFs and manifestation of (nonmonotonic) nonlinearity in the parameter–state relationship

a. Posterior analysis

The joint posterior parameter PDF obtained from MCMC is 10-dimensional, and examination of 1-dimensional marginals (not shown) indicates that the integration over the other 9 dimensions of the space obscures key features of the probability space that are more effectively revealed when higher-dimensional marginals are plotted. Since presentation of two-dimensional PDFs for every parameter pair at every time would be tedious, we have chosen to plot joint PDFs of those parameters to which the model is most sensitive (Fig. 1). There is a high degree of complexity in the parameter–state relationship revealed in these plots and it is worth spending some time examining them in detail. Note that each column in Fig. 1 corresponds to a different number of accumulated 30-min observation times used in the inversion, from a single time (at 30 min) at left with successively greater numbers of observations used up to a maximum of six times (30, 60, 90, 120, 150, and 180 min) at the far right. As such, these plots depict the degree of accumulated influence of parameters on the model solution.

Early on, the properties of simulated clouds are controlled by warm rain processes. Consequently, parameters that control the evolution of ice species in the model and conversions between liquid and ice have little effect on the model output and exhibit a posterior PDF with a nearly uniform distribution. Parameters that directly influence the properties of cloud and rain (first row, Fig. 1), by contrast, have a significant influence on the output from the model, as evidenced by the fact that their PDFs have a well-defined mode. Note that N_{0r} and q_{c0} are positively correlated—this makes physical sense: increasing N_{0r} translates to an increasing proportion of large rain drops in a given volume particles relative to

small. Increasing numbers of large droplets correspond to a higher fall speed and collision efficiency and a larger precipitation rate for a given autoconversion threshold. As the number of large droplets increases, the threshold cloud mixing ratio for autoconversion of cloud to rain must increase so that the conversion of cloud to rain produces the same precipitation rate and liquid water path.

Between 30 and 60 min of simulated time, the system transitions from shallow to troposphere-deep convection (see PV10, their Fig. 2). Simulated cloud extends through the depth of the model and all five cloud species are present in the domain. Ice phase processes begin to have an effect on simulated output, as evidenced by the fact that there are now large regions of the PDFs of snow and graupel fall speed parameters characterized by low probability mass. The snow and graupel density affect the radiative fluxes through their influence on the particle size distribution, and affect conversion between species via the influence they have on the collision–collection process. However, these effects are secondary to the influence of fall speed parameters on the snow and graupel.

By 90 min into the simulation, the system is in transition between convective and stratiform and the precipitation process is consequently in transition between warm rain and melting of ice species. As such, parameters that control the fall speed of ice particles begin to have a greater influence on the simulation output, and this is reflected in the shapes of the PDFs of snow and graupel fall speed parameters. Nonuniqueness in the relationship between parameters and model output is evident in the fact that the probability mass is nearly identical over a large range of combinations of a_g/b_g —an indication that any of these combinations of parameters would produce very similar model output.

Note that the particle velocity is proportional to particle diameter [$V(D) \propto aD^b$], and changes to the fall speed coefficient and exponent represent changes to particle shape. One can readily see that, if particle diameter <1 , then as fall speed coefficient increases, the fall speed exponent must also increase to keep the particle fall velocity constant. For this reason, the fall speed coefficient and exponent are positively correlated for both snow and graupel.

At 120 min into the simulation and following, clouds and precipitation in the system are controlled largely by stratiform processes. Precipitation results primarily from melting ice particles, and as the updraft velocity slows, ice particle settling begins to have a greater influence on the cloud-top particle size distribution and in turn on the radiative fluxes. As such, at times after 120 min into the simulation, parameters that control the distribution of snow and graupel begin to have a much greater

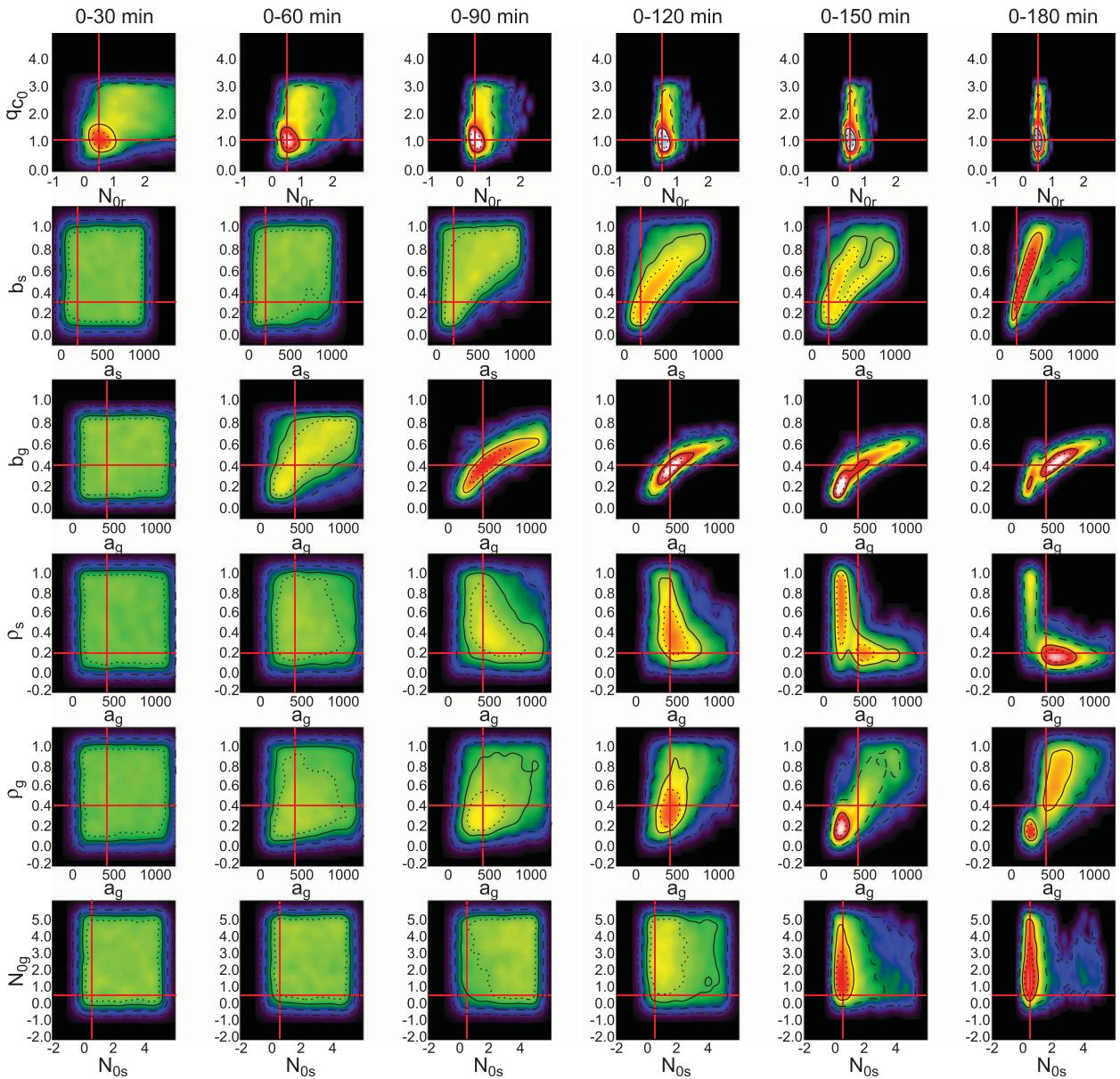


FIG. 1. Posterior two-dimensional marginal parameter PDFs from MCMC. Each row corresponds to a different pair of parameters, while each column represents a different set of assimilated observation times. True parameter values are indicated in the red crosshairs. The thin dash-dot, dashed, solid, and dotted black contours enclose the 99.7%, 95%, 68.3%, and 38.3% probability contours, respectively.

influence on the model output. It follows that, when observations at 120 min and later in the simulation are assimilated, the posterior PDFs of snow and graupel microphysical parameters are more well defined. It is interesting that the slope intercept of the snow particle size distribution appears to have more influence on the solution at later times than does the graupel. This is likely because snow is more likely to be concentrated near the cloud top where it can affect the outgoing short-wave radiation.

Interestingly, the appearance of discrete multimodality appears in the parameter PDFs when observations at times following 120 min are assimilated. There is a pronounced second mode evident in marginal PDFs of a_s/b_s and a_g/b_g . As was noted in PV10, this second mode arises from the fact that, when parameters defining the fall speed and density of snow and graupel are interchanged, the model produces identical output. PV10 found that application of suitable physically based constraints in the MCMC inversion led to a 1:1 relationship between

parameters and model output. *We now see that this discrete multimodality only arises when observations from stratiform times are assimilated.* At earlier times, there is nonuniqueness in the parameter–state relationship, but not discrete multimodality. In practice, this means that, when fewer observations are assimilated, the posterior PDF has a unique mode and algorithms that iterate to find the maximum likelihood solution are likely to converge. In contrast, addition of information from observations late in the simulation has the effect of enhancing the multimodality in the solution and potentially leading to nonconvergence (or convergence to the wrong solution) in the case of maximum likelihood estimators. It is interesting to note that the addition of observations at 180 min into the simulation appears to lead to the concentration of probability mass in the 2D PDFs of a_g/b_g and a_s/b_s around the correct values.

b. Characteristics of the forecast

Since nonmonotonic nonlinearity, and the consequent appearance of discrete modes in select posterior parameter PDFs, arises when times after 90 min of integration time are assimilated, it is natural to wonder whether multimodality in the parameter space translates to multimodality in the observation space as well. In effect, we would like to determine the characteristics of the forecast produced from parameter estimates that utilize only observations from 30–90 min. In this section, we briefly compare the model output at 180-min forecast from the parameter ensemble associated with assimilation between 30–90 min with the observations taken at 180 min.

Figure 2 shows the 1D marginal PDFs of model output variables after 180 min of simulation time. Each column in the figure corresponds to a different set of ensembles: those in the center column were generated via assimilation of observations at 30, 60, and 90 min, while those in the right-hand column were produced by assimilation of observations at 30, 60, 90, 120, 150, and 180 min. When these are compared with the ensemble of output states derived from the initial (random) distribution of parameters (Fig. 2, left-hand column), it is clear that assimilation of observations in the first 90 min (Fig. 2, middle column) results in a reduction in the range of likely solutions. However, the structure of the PDFs of the model output state variables are very similar. Indeed, it appears that the primary effect of assimilation of observations at convective times (Fig. 2, middle column) is to eliminate the thin cloud solutions; those characterized by a combination of low precipitation rates, small LWP and IWP, and relatively large OLR. The bottom line is that, with a perfectly constrained environment, observations taken at convective times (through 90 min)

do not necessarily lead to an accurate forecast at stratiform times.

6. EnKS PDFs compared with MCMC PDF

The manifestation of a profoundly non-Gaussian posterior PDF with assimilation of observations from the stratiform region of the cloud has a number of implications, both for the use of parameters as control variables in data assimilation, as well as for the implementation of stochastic parameterization schemes. While it is known that the presence of such non-Gaussian PDFs will cause affordable ensemble-based minimum-variance-estimating assimilation schemes (e.g., ETKS) to produce erroneous posterior PDFs, little work has been done to describe, analyze, and understand posterior PDF errors given by these methods. To better understand these PDF errors, in this section, we apply both the deterministic (ETKS) and perturbed observations ETKS_PO form of an ensemble Kalman smoother to the parameter estimation problem outlined above.

a. Assimilation of observations at convective times

The 2D marginal posterior PDFs for the same parameter pairs shown in the MCMC results (Fig. 1) and for successively greater numbers of observation times are plotted in Figs. 3 and 4 for ETKS and ETKS_PO, respectively. We focus first on the evolution of posterior PDFs when increasing numbers of observations are assimilated in the first 90 min of the simulation. Comparison between the posterior PDFs obtained from MCMC and those in ETKS reveals the following. Parameter marginal PDFs are similar between the three algorithms through approximately 90 min into the simulation, though there is much greater skewness in the posterior PDF of N_{or} for MCMC compared with ETKS and ETKS_PO (Table 3). As increasing numbers of observation times are assimilated, there is a successive and incremental reduction in probability mass at large values of a_s and small values of b_s in the ETKS (Fig. 3) and ETKS_PO (Fig. 4). By contrast, the change in probability mass is much sharper between 60 and 90 min in MCMC (Fig. 1)—the ETKS and ETKS_PO fail to keep pace with this rapid change. It is possible that changes in the posterior PDF during this 60–90-min period correspond to an increase in the importance of nonlinear terms in the equations governing the evolution of perturbations about the true state. Such nonlinearity causes true posterior errors to be *nonlinear* functions of forecast errors (Hodyss 2011). Presumably, the reason for the failure of the ETKS and ETKS_PO PDFs to match the MCMC PDF during this period is a consequence of the fact that both the ETKS and ETKS_PO falsely

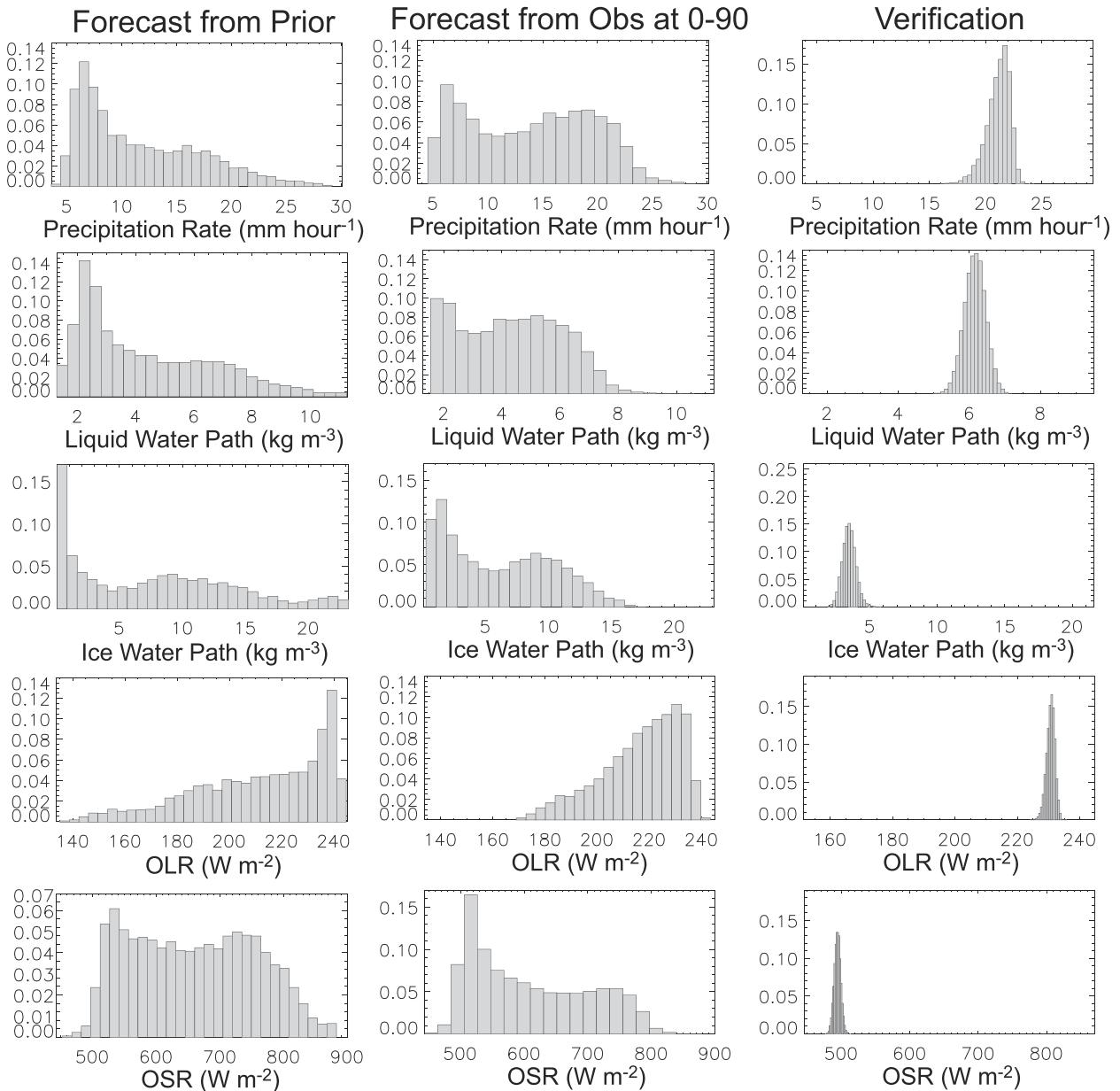


FIG. 2. Marginal PDFs of model output state variables at 180 min of simulation time for forecasts generated from (left) a uniformly distributed prior ensemble, (middle) the posterior parameter ensemble produced by assimilating observations at 30, 60, and 90 min, and (right) the posterior parameter ensemble produced by assimilating observations at 30, 60, 90, 120, 150, and 180 min.

assume that posterior errors are a *linear* function of prior errors. This restrictive property may also be responsible for the presence of negative (unphysical) values of the parameters shown in Figs. 3 and 4. The MCMC accept-reject procedure automatically rejects any values outside of the allowable parameter range and thus avoids such unphysical values.

Note that the observation uncertainty is the same at each time and the background error is neglected (a uniform distribution is assumed for each of the parameters

initially). As such, observations have the same influence at each time in the ETKS analysis. If observation uncertainty were reduced at later times, the analysis would draw more to these observations. The issue of whether this would cause the analysis ensemble to look more like that from MCMC will be explored in greater detail momentarily.

After assimilating observations from convective times, the estimates of posterior mean and variance estimate are robust in both versions of the ETKS, with the exception

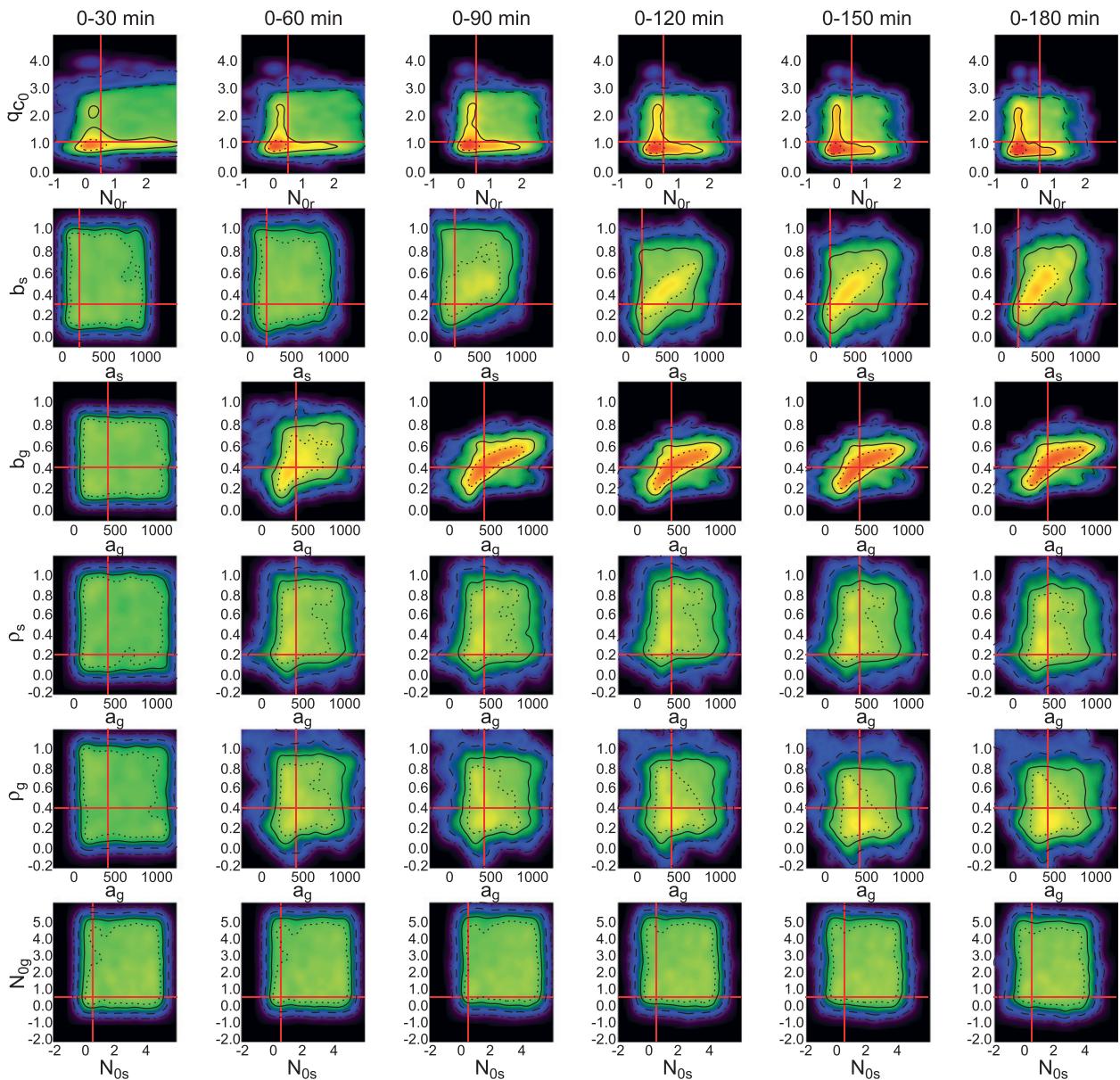


FIG. 3. As in Fig. 1, but for posterior two-dimensional marginal parameter PDFs from the deterministic ETKS.

of the slope intercept of the rain size distribution, which is closer to its true value in MCMC and with significantly lower variance (Table 3). The strong similarity of the posterior mean and variance from the ETKS and ETKS_PO is consistent with the large ensemble size causing errors in sample variance estimates from the ETKS_PO to be small. The posterior PDF sampled by MCMC has in general larger skewness than does the PDF in ETKS or ETKS_PO. This reflects the fact that, while the full nonlinear model is used to propagate perturbations to parameters in the ETKS, these perturbations are restricted by the linear assumption and information on

the higher-order moments of the prior PDF are not explicitly included in the solution (Hodyss 2011). Results for N_{or} and q_{c0} are consistent with the findings of VP08, who noted that, in cases for which the posterior PDF was skewed, an EnKF-type estimate was biased in the direction of the tail and with overestimated variance.

Larger differences between the three estimates appear when the posterior mode [maximum a posteriori (MAP) estimate] is examined. Note that, in the case of a discrete sample and for a method in which iterative convergence to a solution is not used, the MAP estimate is a more robust estimate of the mode than is the

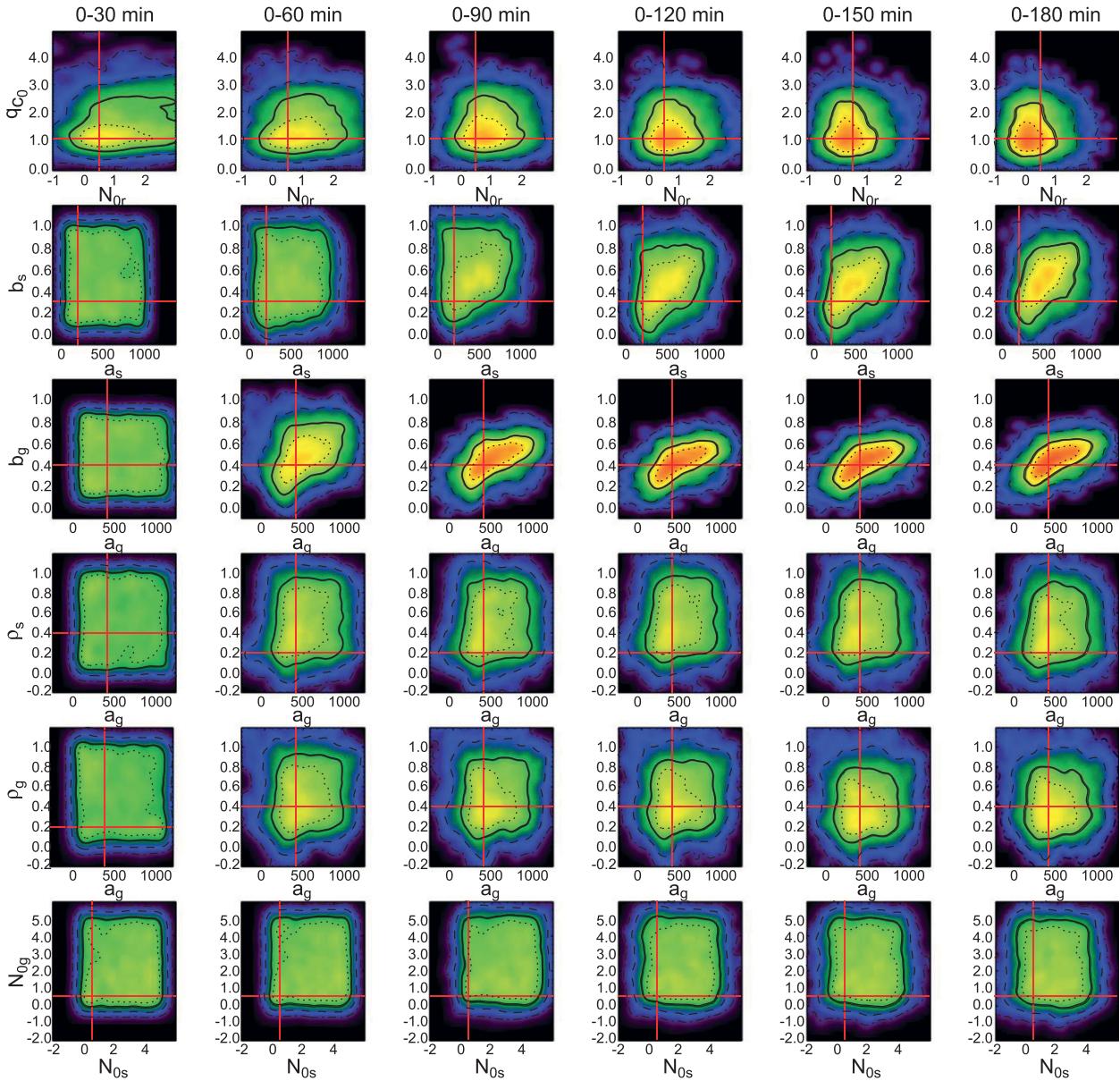


FIG. 4. As in Fig. 1, but for posterior two-dimensional marginal parameter PDFs from the stochastic (perturbed observations) ETKS.

maximum likelihood point. Whereas the mean and variance are nearly identical for the ETKS and ETKS_PO, they have differing modes for all parameters, but especially for N_{0r} . This is associated with the fact that the probability mass is more concentrated around the mode in the deterministic filter as compared with the stochastic filter. The fact that observations taken at convective times contain insufficient information to constrain the parameter values to a unique solution is evidenced by the fact that the posterior mode differs from the true value in all of the analyses. The mode of the PDF produced by MCMC is near the true parameter values for a_s , a_g , N_{0r} , and q_{c0} , but not for the other parameters,

while the mode in the PDF produced by ETKS and ETKS_PO is relatively far from the truth for all parameters but q_{c0} .

b. Characteristics of the ETKS-based ensemble when stratiform times are assimilated

Between 90 and 120 min of simulated time, decreases in the prescribed vertical motion and water vapor tendency cause the model to transition from a convective state to stratiform. As such, the radiation, cloud liquid and ice content, and precipitation are governed to far greater extent by ice processes as well as interactions between liquid and ice. This transition from lower to

TABLE 3. Statistics from the posterior mean for MCMC and deterministic and perturbed obs forms of the ETKS using observations at 30, 60, and 90 min of simulated time.

	a_s	b_s	a_g	b_g	N_{0r}	N_{0s}	N_{0g}	ρ_s	ρ_g	q_{c0}
Mean										
MCMC	472.27	0.64	591.475	0.455	0.652	2.803	2.67	0.479	0.513	1.33×10^{-3}
ETKS	437.54	0.59	535.944	0.454	0.955	2.716	2.533	0.503	0.508	1.47×10^{-3}
ETKS PO	437.54	0.59	535.944	0.454	0.955	2.716	2.533	0.503	0.508	1.47×10^{-3}
Variance										
MCMC	67 169.02	0.054	76 105.29	0.021	0.061	1.905	1.994	0.06	0.057	3.88×10^{-7}
ETKS	60 408.19	0.055	71 819.6	0.019	0.483	2.054	1.973	0.063	0.058	4.20×10^{-7}
ETKS PO	60 234.56	0.055	71 668.06	0.019	0.479	2.054	1.976	0.063	0.058	4.20×10^{-7}
Skewness										
MCMC	0.296	-0.408	0.387	-0.179	1.279	-0.214	-0.113	0.352	0.256	0.992
ETKS	0.161	0.059	0.188	-0.257	0.482	-0.087	0.122	0.089	0.149	0.512
ETKS PO	0.198	0.009	0.166	-0.308	0.385	-0.082	0.12	0.101	0.211	0.479
Mode										
MCMC	227.53	0.74	431.45	0.51	0.54	4.1	4.06	0.32	0.33	9.70×10^{-4}
ETKS	310.45	0.5	353.62	0.51	0.27	3.28	1.39	0.28	0.28	8.30×10^{-4}
ETKS PO	353.57	0.56	357.16	0.5	0.76	3.56	1.38	0.31	0.36	1.06×10^{-3}

higher ice-parameter sensitivity is reflected in the evolution of the parameter PDFs sampled by MCMC in Fig. 1 and discussed above; there are rapid changes to the structure of the posterior PDFs with increasing numbers of assimilated observations after 90 min into the simulation. In particular, modes in the posterior solution become rapidly more well defined.

Comparisons of the posterior PDFs produced by ETKS (Fig. 3) with those output from ETKS_PO (Fig. 4) reveal that, aside from N_{0r} and q_{c0} , analysis PDFs from ETKS and ETKS_PO are very similar through all observation times, though it is evident that the mode consistently shifts in the direction of the tail of the distribution for the perturbed observation case as compared with the deterministic filter. Stated another way, when the skewness is positive (negative), the mode shifts toward larger (smaller) values in ETKS_PO relative to ETKS.

Interestingly, the mode of N_{0r} in both ETKS and ETKS_PO shifts toward progressively smaller values with addition of information from successive observation times, to such an extent that eventually the mode of the deterministic ETKS ensemble becomes negative (Table 4). As was the case when observations from convective times were assimilated, the posterior mean and variance of the ensembles produced by ETKS and ETKS_PO remain virtually identical; as mentioned above, this can be attributed to the large sample size and the consequent lack of sampling error. However, both the deterministic and stochastic versions of the ETKS produce significant deviations from the MCMC posterior mean and variance for a_s , a_g , N_{0r} , N_{0s} , ρ_s , and ρ_g . These are all parameters for which the posterior PDF changes shape significantly as increasing numbers of stratiform times are assimilated. In general, the variance is overestimated for each of these parameters.

TABLE 4. As in Table 3, but for observations assimilated at 30, 60, 90, 120, 150, and 180 min of simulated time.

	a_s	b_s	a_g	b_g	N_{0r}	N_{0s}	N_{0g}	ρ_s	ρ_g	q_{c0}
Mean										
MCMC	376.816	0.583	518.546	0.418	0.504	0.668	2.268	0.262	0.578	1.16×10^{-3}
ETKS	526.555	0.501	482.387	0.451	0.233	1.746	2.346	0.466	0.445	1.29×10^{-3}
ETKS PO	526.555	0.501	482.387	0.451	0.233	1.746	2.346	0.466	0.445	1.29×10^{-3}
Variance										
MCMC	28 688.88	0.047	43 957.61	0.012	0.008	0.448	1.576	0.047	0.072	2.47×10^{-7}
ETKS	48 140.65	0.043	68 392.81	0.016	0.304	1.816	1.947	0.06	0.052	4.11×10^{-7}
ETKS PO	48 306.11	0.043	68 401.56	0.016	0.303	1.828	1.946	0.06	0.051	4.07×10^{-7}
Skewness										
MCMC	1.4788	-0.088 46	0.268 36	-0.450 48	0.448 13	3.720 35	0.378 64	2.058 17	-0.317 96	1.486
ETKS:	0.2896	-0.075 37	0.192 98	-0.208 66	0.554 55	-0.031 17	0.124 79	0.152 99	0.275 51	0.518 62
ETKS PO	0.295 71	-0.117 29	0.165 19	-0.2102	0.447 92	-0.014 81	0.123 15	0.181 32	0.373 37	0.4588
Mode										
MCMC	243.284	0.398	488.06	0.398	0.512	0.468	1.224	0.155	0.726	0.001
ETKS	416.5838	0.488 56	344.7761	0.495 74	-0.230 51	2.068 55	1.166 39	0.260 48	0.274 41	0.000 67
ETKS PO	429.8507	0.472 75	346.545	0.476 34	0.074 63	1.719 18	1.215 04	0.2953	0.305 36	0.000 91

As mentioned earlier, the ETKS and ETKS_PO assume that posterior errors are linear functions of forecast errors. In the 0–90-min (predominantly convective) time period, this results in an approximately correct characterization of the posterior conditional PDF. However, as stratiform cloud processes begin to become more important after 90 min of simulated time, the true posterior PDF changes rapidly. As such, the size of the error associated with the assumption in the ETKS and ETKS_PO that posterior errors are linear functions of prior errors becomes markedly greater during the 90–180-min time period. Neither the ETKS nor the ETKS_PO is able to reproduce the true multimode behavior in the posterior during this period, nor are they capable of tracking the rapid evolution of the posterior structure with increasing numbers of observing times. Analogous behavior was noted in the parameter estimation study conducted by Tong and Xue (2008a,b). While Tong and Xue (2008a,b) hypothesized that the loss of parameter identifiability might have been due to a combination of small ensemble size, nonunique solutions, and poor assumed observation quality, we can confirm in this case that the key issue is nonuniqueness in the parameter–state relationship. In addition, our results indicate that this nonuniqueness is not due to nonlinear interaction between model physics and dynamics as we have intentionally removed the feedback from radiation and cloud microphysics to the model dynamics and thermodynamic state.

7. Remarkable robustness of ETKS in the presence of multimodal priors

The fundamental limitation of the ETKS that has been emphasized in the preceding discussion is that it assumes that the posterior error representations are strictly linear functions of prior error representations. This limitation is separate from that associated with the fact that the Kalman filter is often derived using Gaussian prior and likelihood error statistics. Here, we hypothesize that it is the former limitation rather than the latter limitation that causes ETKS posterior PDFs to be qualitatively different from MCMC posterior PDFs. To test this hypothesis, we use draws from the MCMC-based posterior PDF at 90, 120, and 150 min as the forecast (prior) ensemble in the ETKS.

An ETKS experiment that uses the exact prior from assimilation of observations from 30–90 min then assimilates observations between 120–180 min (Fig. 5, second column) results in a structure that, while it differs from the sample produced by MCMC, captures the primary mode in the marginal PDFs for all parameter pairs except N_{0s} and N_{0g} . However, covariance between N_{0s} and N_{0g} in the forecast ensemble is of different sign than the

true posterior reflecting the fact that the algorithm is unable to concentrate probability mass around mode in N_{0s} . That said, we do note a reduction in probability mass at larger values of N_{0s} , reflecting the information contained in the observations. This reduction was not observed in the ETKS_PO analysis derived from an initially random parameter sample (Fig. 3).

Similar results are obtained when the 0–120-min MCMC is used as the prior for the ETKS (Fig. 5, third column). The posterior PDF trends in the direction of the true posterior, but multimodality is not observed. Model output and parameters covary in a consistent manner and this is reflected in the posterior PDFs; however, multimodality and changes in the third moment arise from nonlinear relationships between parameters and between parameters and model output. Note that, because the covariance between N_{0s} and N_{0g} is of the same sign at 120 min as it is at 180 min (Fig. 1), the posterior PDF of N_{0s} and N_{0g} looks much more like the MCMC solution.

The 0–150-min posterior PDF exhibits clear multimodality in the 2D marginal PDFs (Fig. 1). When this PDF is used as the ETKS prior (Fig. 5, fourth column), this multimodality is preserved. However, whereas there is a clear change in which mode is preferred in the MCMC solution between 150 and 180 min, no such transition takes place in the ETKS. These results are consistent with our hypothesis that multimodal, non-Gaussian priors are not the primary limitation of the ETKS, rather it is the inability of the ETKS to represent posterior perturbations as nonlinear functions of prior perturbations that causes the most profound difference between MCMC posterior PDFs and ETKS posterior PDFs.

8. Conclusions

Our major findings consist of the following:

- 1) The nonlinear dynamics associated with the *transition* from convective precipitation to stratiform precipitation is the primary cause of the multimodal posterior PDFs found by PV10. This finding is supported by the fact that multimodal posteriors only begin to appear when observations taken after the transition from convective to stratiform rainfall are included in the analysis. Early in the simulation (0–90 min), the model is strongly forced with large vertical motion and water vapor tendency, and cloud evolution is controlled by warm rain processes. Information from observations assimilated at these times strongly affects the posterior PDF of the warm rain parameters, while the PDF of the ice phase parameters is left relatively unaffected. Because the relationship between changes in parameters and changes to model output is monotonic, the posterior

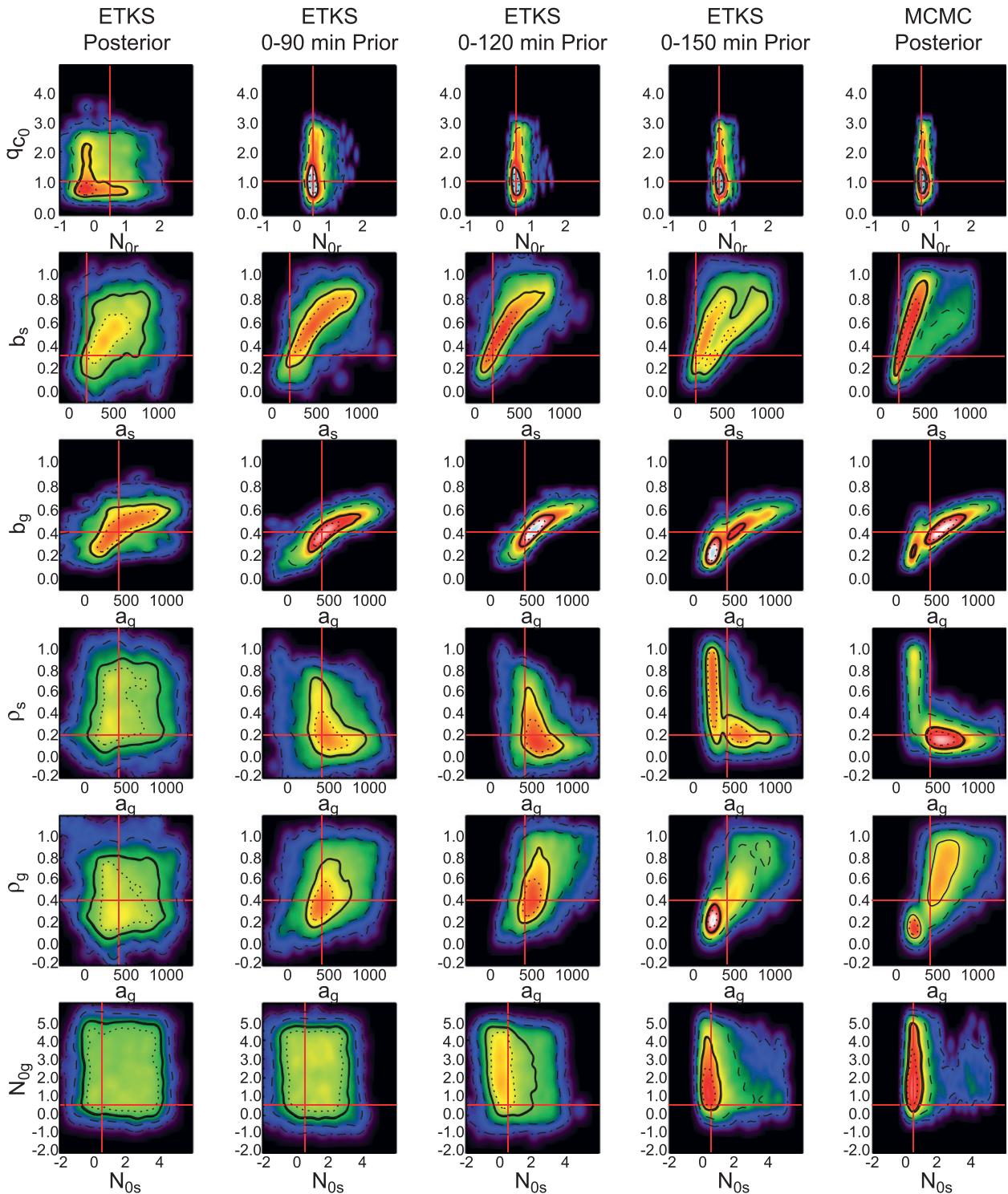


FIG. 5. Posterior two-dimensional marginal parameter PDFs. Each row is as in Figs. 1, 3, and 4, however, the first column shows the posterior PDF from the deterministic ETKS with observations assimilated at 30, 60, 90, 120, 150, and 180 min of simulated time. The second through fourth columns depict the posterior PDF from the deterministic ETKS using as the prior ensemble the posterior from MCMC assimilating observations from (second column) 30–90, (third column) 30–120, and (fourth) 30–150 min of simulated time. The fifth column (at far right) depicts the posterior solution from the MCMC with observations assimilated at 30–180 min.

PDF has a single mode, and nonlinearity manifests in the posterior skewness. Once the system transitions from convective to stratiform (120–180 min of simulated time), the prescribed vertical motion and moisture source are weaker, and ice phase processes control the precipitation rate, the partition between liquid and ice in the system, and the optical depth near the top of the cloud. As such, simulated rain rate, integrated liquid and ice, and radiative fluxes are far more sensitive to changes in ice particle fall speed, density, and size distribution. The processes controlled by these parameters interact in a complex manner, leading to the rapid manifestation of multiple modes in the posterior PDF.

- 2) Ensemble Kalman smoothers perform poorly when the posterior mean and perturbations are strongly nonlinear functions of the forecast error. This is evidenced by the failure of the ETKS and ETKS_PO to represent the transition of the posterior PDF from a unimodal to multimodal form when observations of the stratiform phase of cloud evolution were assimilated. Though the ETKS is unable to produce a multimode analysis from a unimode prior, the posterior PDF produced by both the deterministic and perturbed observations versions of the ETKS is clearly non-Gaussian. The effect of model nonlinearity on the posterior PDF is included in ETKS in that the nonlinear model is allowed to respond to changes in model parameters; the full nonlinear model propagates perturbations forward in time. However, these changes are constrained by the requirement that the ETKS posterior perturbations be strictly linear functions of the prior perturbations. In contrast, the accept–reject procedure of the MCMC finds posterior perturbations that can be *any* function of the prior perturbations.
- 3) Ensemble Kalman smoothers can preserve key aspects of non-Gaussian priors. Specifically, the ETKS was found to give a qualitatively accurate multimodal posterior PDF when given an accurate multimodal prior (see Fig. 5). The implication is that it is not the Gaussian assumption used in the derivation of the ETKS that causes misrepresentation of the posterior. Instead, it is the lack of information on higher moments and/or multiple modes in the prior ensemble.

Our results suggest that the uncertainty characteristics of model physics parameterizations depend critically on the characteristics of the environment. Abrupt changes in the physical environment can lead to similarly abrupt changes in parameter uncertainty. Ensemble Kalman filter-type algorithms, while not capable of capturing rapid (nonlinear) transitions in the nature of the posterior PDF, can be shown to perform well if provided with a robust prior ensemble. As such, ensemble Kalman filter-type algorithms have promise as cost-efficient methods for model parameter estimation.

Finally, though our column model simulations are highly idealized (there is no feedback between model physics and dynamics), the transition from convective to stratiform processes and the associated change in model parameter sensitivity conform to what we might expect of more realistic three-dimensional models. Convective time periods in both are driven by strong upward vertical motion and a large moisture source, leading to a relative lack of sensitivity to changes in ice microphysical parameters. Similarly, the late-stage development is weakly forced and depends to a greater extent on the interaction between, and melting of, ice particles. The extent to which parameter sensitivity in our experiments is representative of the uncertainty characteristics of free-running simulations of squall-line-type convection can be assessed by running MCMC experiments with unforced three-dimensional simulations. These experiments are already under way.

Acknowledgments. DJP’s contribution to this paper was funded by Office of Naval Research Grant N00173-10-1-G035, as well as by NASA Modeling, Analysis, and Prediction Grants NNX09AJ43G and NNX09AJ46G, while CHB was supported by the U.S. Office of Naval Research Grant 4596-0-1-5.

APPENDIX

Efficient ETKF for the Case when Ensemble Size K Exceeds the Number of Observations p

Let \mathbf{X}^f denote the $n \times K$ matrix whose K columns list the K ensemble perturbations around the ensemble mean so that $\mathbf{Z}^f = \mathbf{X}^f / \sqrt{K - 1}$ is the square root of the forecast error covariance matrix because $\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{fT}$. From (2)

$$\begin{aligned}
 \mathbf{P}^a &= \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{P}^f \\
 &= \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T [\mathbf{R}^{1/2} (\tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}^T + \mathbf{I}) \mathbf{R}^{1/2}]^{-1} \mathbf{H} \mathbf{P}^f, \quad \text{where } \tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \mathbf{H} \\
 &= \mathbf{P}^f - \mathbf{P}^f \tilde{\mathbf{H}}^T [\tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}^T + \mathbf{I}]^{-1} \tilde{\mathbf{H}} \mathbf{P}^f \\
 &= \mathbf{Z}^f [\mathbf{I} - \mathbf{Z}^{fT} \tilde{\mathbf{H}}^T (\tilde{\mathbf{H}} \mathbf{Z}^f \mathbf{Z}^{fT} \tilde{\mathbf{H}}^T + \mathbf{I})^{-1} \tilde{\mathbf{H}} \mathbf{Z}^f] \mathbf{Z}^{fT}.
 \end{aligned}
 \tag{A1}$$

In Bishop et al. (2001), the square root of this matrix is derived from the $K \times K$ eigendecomposition of the matrix $\mathbf{Z}^{fT} \tilde{H}^T \tilde{H} \mathbf{Z}^f$. This is an appropriate choice when the number of observations p exceeds the number of ensemble members K . However, in the converse situation when $K > p$, the eigendecomposition of the $p \times p$ matrix,

$$\tilde{H} \mathbf{Z}^f \mathbf{Z}^{fT} \tilde{H}^T = \mathbf{E} \mathbf{\Gamma} \mathbf{E}^T, \quad (\text{A2})$$

is less computationally expensive than that of $\mathbf{Z}^{fT} \tilde{H}^T \tilde{H} \mathbf{Z}^f$. Now, it may be shown that $\tilde{H} \mathbf{Z}^f$ has the concise singular value decomposition:

$$\tilde{H} \mathbf{Z}^f = \mathbf{E} \mathbf{\Gamma}^{1/2} \mathbf{L}^T. \quad (\text{A3})$$

By *concise* singular value decomposition, we mean the singular value decomposition that removes all of the left and right singular vectors corresponding to zero singular values. In general, there will be as many nonzero singular values as there are observations p . (While it is still possible to proceed when this is not true, for simplicity, we will hereafter assume that this is true.) Hence, in (A3) \mathbf{E} is a $p \times p$ orthonormal matrix containing left singular vectors of $\tilde{H} \mathbf{Z}^f$, $\mathbf{\Gamma}$ is a diagonal $p \times p$ matrix, and \mathbf{L} is a $K \times p$ matrix containing right singular vectors of $\tilde{H} \mathbf{Z}^f$. Note that \mathbf{L} is orthonormal in the sense that $\mathbf{L}^T \mathbf{L} = \mathbf{I}$. Note that in Bishop et al. (2001), the right singular vectors of $\tilde{H} \mathbf{Z}^f$ were denoted by the square matrix \mathbf{C} , while here we use the symbol \mathbf{L} in order to remind the reader that in this case, the concise right singular vectors do not form a square matrix. Because, by definition, only singular vectors corresponding to nonzero singular values are included in (A3), it follows that $\mathbf{\Gamma}^{-1/2}$ exists.

A key aspect of the following derivation is that once \mathbf{E} has been deduced from the eigendecomposition (A2), one can use (A3) to derive \mathbf{L} from $\tilde{H} \mathbf{Z}^f$ because from (A3)

$$\mathbf{\Gamma}^{-1/2} \mathbf{E}^T \tilde{H} \mathbf{Z}^f = \mathbf{L}. \quad (\text{A4})$$

Using (A3) in (A1) and noting that $[\mathbf{E} \mathbf{\Gamma} \mathbf{E}^T + \mathbf{I}]^{-1} = (\mathbf{E}[\mathbf{\Gamma} + \mathbf{I}]\mathbf{E}^T)^{-1} = \mathbf{E}[\mathbf{\Gamma} + \mathbf{I}]^{-1} \mathbf{E}^T$, (A1) can be rewritten in the following form:

$$\begin{aligned} \mathbf{P}^a &= \mathbf{Z}^f [\mathbf{I} - \mathbf{L} \mathbf{\Gamma}^{1/2} \mathbf{E}^T (\mathbf{E}[\mathbf{\Gamma} + \mathbf{I}]^{-1} \mathbf{E}^T) \mathbf{E} \mathbf{\Gamma}^{1/2} \mathbf{L}^T] \mathbf{Z}^{fT} \\ &= \mathbf{Z}^f (\mathbf{I} - \mathbf{L} \mathbf{\Gamma} [\mathbf{\Gamma} + \mathbf{I}]^{-1} \mathbf{L}^T) \mathbf{Z}^{fT}. \end{aligned} \quad (\text{A5})$$

To find a square root of the analysis error covariance matrix (A5), we seek a matrix \mathbf{B} that is a solution to the following equation:

$$\begin{aligned} \mathbf{I} - \mathbf{L} \mathbf{\Gamma} [\mathbf{\Gamma} + \mathbf{I}]^{-1} \mathbf{L}^T &= (\mathbf{I} - \mathbf{L} \mathbf{B} \mathbf{L}^T) (\mathbf{I} - \mathbf{L} \mathbf{B} \mathbf{L}^T) \\ &= \mathbf{I} - 2 \mathbf{L} \mathbf{B} \mathbf{L}^T + \mathbf{L} \mathbf{B}^2 \mathbf{L}^T. \end{aligned} \quad (\text{A6})$$

Equation (A6) is true, provided that

$$\begin{aligned} 0 &= \mathbf{L} \mathbf{\Gamma} [\mathbf{\Gamma} + \mathbf{I}]^{-1} \mathbf{L}^T - 2 \mathbf{L} \mathbf{B} \mathbf{L}^T + \mathbf{L} \mathbf{B}^2 \mathbf{L}^T \\ &= \mathbf{L} (\mathbf{\Gamma} [\mathbf{\Gamma} + \mathbf{I}]^{-1} - 2 \mathbf{B} + \mathbf{B}^2) \mathbf{L}^T. \end{aligned} \quad (\text{A7})$$

Equation (A7) is satisfied provided $\mathbf{\Gamma} [\mathbf{\Gamma} + \mathbf{I}]^{-1} - 2 \mathbf{B} + \mathbf{B}^2 = \mathbf{0}$. Since $\mathbf{\Gamma} [\mathbf{\Gamma} + \mathbf{I}]^{-1}$ is a diagonal matrix, it follows that we can use the standard formula for the solution of a quadratic equation to solve (A7). Applying this formula gives

$$\begin{aligned} \mathbf{B} &= \mathbf{I} \pm [\mathbf{I} - \mathbf{\Gamma} (\mathbf{\Gamma} + \mathbf{I})^{-1}]^{-1/2} \\ &= \mathbf{I} \pm [(\mathbf{\Gamma} + \mathbf{I}) (\mathbf{\Gamma} + \mathbf{I})^{-1} - \mathbf{\Gamma} (\mathbf{\Gamma} + \mathbf{I})^{-1}]^{-1/2} \\ &= \mathbf{I} \pm (\mathbf{\Gamma} + \mathbf{I})^{-1/2}. \end{aligned} \quad (\text{A8})$$

Hence, the square root \mathbf{Z}^a of \mathbf{P}^a is given by

$$\mathbf{Z}^a = \mathbf{Z}^f (\mathbf{I} - \mathbf{L} \mathbf{B} \mathbf{L}^T) = \mathbf{Z}^f \{ \mathbf{I} - \mathbf{L} [\mathbf{I} \pm (\mathbf{\Gamma} + \mathbf{I})^{-1/2}] \mathbf{L}^T \}. \quad (\text{A9})$$

The occurrence of the term \pm in (A9) means that there are two distinct valid square roots. We wish to choose the square root that enables posterior ensemble members to approximate solutions to the governing nonlinear equations. Consider the fact that in the limit of infinite observation error covariance, $\mathbf{Z}^a = \mathbf{Z}^f$ is a valid posterior square root that would produce posterior ensemble members that would precisely satisfy the governing nonlinear equations. Since in the limit of infinite observation error variance $\mathbf{\Gamma} = \mathbf{0}$ so that $\{ \mathbf{I} - \mathbf{L} [\mathbf{I} \pm (\mathbf{\Gamma} + \mathbf{I})^{-1/2}] \mathbf{L}^T \} = \{ \mathbf{I} - \mathbf{L} [\mathbf{I} \pm \mathbf{I}] \mathbf{L}^T \}$, it follows that replacing “ \pm ” by “ $-$ ” gives the square root that sets the analysis perturbations equal to the forecast perturbations in the special case of infinite observation error variance.

To connect the form of the square root given by (A9) with those given in Anderson (2001) and Whitaker and Hamill (2002), note that using (A4) in (A9) allows it to be rewritten in the following form:

$$\mathbf{Z}^a = \{ \mathbf{I} - \mathbf{L} [\mathbf{I} - (\mathbf{\Gamma} + \mathbf{I})^{-1/2}] \mathbf{\Gamma}^{-1/2} \mathbf{E}^T \tilde{H} \} \mathbf{Z}^f = \tilde{\mathbf{K}} \mathbf{Z}^f, \quad (\text{A10})$$

where $\tilde{\mathbf{K}} = \{ \mathbf{I} - \mathbf{L} [\mathbf{I} - (\mathbf{\Gamma} + \mathbf{I})^{-1/2}] \mathbf{\Gamma}^{-1/2} \mathbf{E}^T \tilde{H} \}$ is the ETKF counterpart of the modified Kalman gain appropriate for ensemble perturbations discussed in Whitaker and Hamill (2002). Since (A9) solves exactly the same equations as those solved by the forms of the square root filter

given in Anderson (2001) and Whitaker and Hamill (2002), we expect it to give essentially equivalent performance to the square roots given in those papers. The key difference between the form given here and those papers is that (A9) makes an explicit connection between the square root of the analysis error covariance matrix and eigenvector decomposition of the observation space forecast error covariance matrix $\tilde{H}\mathbf{Z}^f\mathbf{Z}^{fT}\tilde{H}^T$. In some circumstances, this may make the performance characteristics/behavior of the square root filter and the effect of approximations such as covariance localization easier to understand.

REFERENCES

- Aksoy, A., F. Zhang, and J. W. Nielsen-Gammon, 2006: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Wea. Rev.*, **134**, 2951–2970.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Berner, J., S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995.
- 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.
- Bocquet, M., C. A. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Wea. Rev.*, **138**, 2997–3023.
- Burgers, G., P. Jan van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Evensen, G., 2006: *Data Assimilation: The Ensemble Kalman Filter*. Springer, 279 pp.
- Gelman, A. G., 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.
- Gilmore, M. S., J. M. Straka, and E. N. Rasmussen, 2004: Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics scheme. *Mon. Wea. Rev.*, **132**, 2610–2627.
- Grabowski, W. W., 2003: Impact of cloud microphysics on convective-radiative quasi equilibrium revealed by cloud-resolving convective parameterization. *J. Climate*, **16**, 3463–3475.
- , X. Wu, and M. W. Moncrieff, 1999: Cloud-resolving modeling of tropical cloud systems during phase III of GATE. Part III: Effects of cloud microphysics. *J. Atmos. Sci.*, **56**, 2384–2402.
- Haario, H., E. Saksman, and J. Tamminen, 1999: Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Stat.*, **14**, 375–395.
- Heysmsfield, A. J., A. Bansemmer, 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.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Houze, R. A., Jr., 2004: Mesoscale convective systems. *Rev. Geophys.*, **42**, RG4003, doi:10.1029/2004RG000150.
- Järvinen, H., P. Räisänen, M. Laine, J. Tamminen, A. Ilin, E. Oja, A. Solonen, and H. Haario, 2010: Estimation of ECHAM5 climate model closure parameters with adaptive MCMC. *Atmos. Chem. Phys.*, **10**, 9993–10 002.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge University Press, 341 pp.
- 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.
- 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.
- Lewis, J., S. Lakshminarayanan, 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.
- Meng, Z., and F. Zhang, 2007: Tests of an ensemble Kalman filter for mesoscale and regional scale data assimilation. Part II: Imperfect model experiments. *Mon. Wea. Rev.*, **135**, 1403–1423.
- 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 A)*, W. H. K. Lee et al., Eds., International Geophysical Series, Vol. 81A, Academic Press, 237–265.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772.
- Neelin, J. D., A. Bracco, H. Luo, J. C. McWilliams, and J. E. Meyerson, 2010: Considerations for parameter optimization and sensitivity in climate models. *Proc. Natl. Acad. Sci. USA*, **107**, 21 349–21 354.
- Palmer, T. N., G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung, and M. Leutbecher, 2005: Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.*, **33**, 163–193.

- Petch, J. C., and M. E. B. Gray, 2001: Sensitivity studies using a cloud-resolving model simulation of the tropical west Pacific. *Quart. J. Roy. Meteor. Soc.*, **127**, 2287–2306.
- Posselt, D. J., and T. Vukicevic, 2010: Robust characterization of model physics uncertainty for simulations of deep moist convection. *Mon. Wea. Rev.*, **138**, 1513–1535.
- , 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.
- Stainforth, D. A., and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.
- Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107.
- Tamminen, J., 2004: Validation of nonlinear inverse algorithms with Markov chain Monte Carlo method. *J. Geophys. Res.*, **109**, D19303, doi:10.1029/2004JD004927.
- , and E. Kyrola, 2001: Bayesian solution for nonlinear and non-Gaussian inverse problems by Markov chain Monte Carlo method. *J. Geophys. Res.*, **106**, 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.
- , J. R. Scala, B. S. Ferrier, and J. Simpson, 1995: The effect of melting processes on the development of a tropical and a midlatitude squall line. *J. Atmos. Sci.*, **52**, 1934–1948.
- , and Coauthors, 2003: Microphysics, radiation, and surface processes in the Goddard Cumulus Ensemble (GCE) model. *Meteor. Atmos. Phys.*, **82**, 97–137.
- Tarantola, A., 2005: *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 342 pp.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- 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.
- Tong, M., and M. Xue, 2008a: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part I: Sensitivity analysis and parameter identifiability. *Mon. Wea. Rev.*, **136**, 1630–1648.
- , and —, 2008b: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part II: Parameter estimation experiments. *Mon. Wea. Rev.*, **136**, 1649–1688.
- van den Heever, S. C., and W. R. Cotton, 2004: The impact of hail size on simulated supercell storms. *J. Atmos. Sci.*, **61**, 1596–1609.
- Vukicevic, T., and D. J. Posselt, 2008: Analysis of the impact of model nonlinearities in inverse problem solving. *J. Atmos. Sci.*, **65**, 2803–2823.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- Wu, X., W. D. Hall, W. W. Grabowski, M. W. Moncrieff, W. D. Collins, and J. T. Kiehl, 1999: Long-term behavior of cloud systems in TOGA COARE and their interactions with radiative and surface processes. Part II: Effects of ice microphysics on cloud-radiation interaction. *J. Atmos. Sci.*, **56**, 3177–3195.