

This work was written as part of one of the author's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.

Public Domain Mark 1.0

<https://creativecommons.org/publicdomain/mark/1.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

**Please provide feedback**

Please support the ScholarWorks@UMBC repository by emailing [scholarworks-group@umbc.edu](mailto:scholarworks-group@umbc.edu) and telling us what having access to this work means to you and why it's important to you. Thank you.

# Determining Key Model Parameters of Rapidly Intensifying Hurricane Guillermo (1997) Using the Ensemble Kalman Filter

HUMBERTO C. GODINEZ AND JON M. REISNER

*Los Alamos National Laboratory, Los Alamos, New Mexico*

ALEXANDRE O. FIERRO

*Space and Remote Sensing Group, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico, and NOAA/Cooperative Institute for Mesoscale Meteorological Studies, Norman, Oklahoma*

STEPHEN R. GUIMOND

*Center for Ocean–Atmospheric Prediction Studies, The Florida State University, Tallahassee, Florida*

JIM KAO

*Los Alamos National Laboratory, Los Alamos, New Mexico*

(Manuscript received 17 January 2012, in final form 22 May 2012)

## ABSTRACT

In this work the authors determine key model parameters for rapidly intensifying Hurricane Guillermo (1997) using the ensemble Kalman filter (EnKF). The approach is to utilize the EnKF as a tool only to estimate the parameter values of the model for a particular dataset. The assimilation is performed using dual-Doppler radar observations obtained during the period of rapid intensification of Hurricane Guillermo. A unique aspect of Guillermo was that during the period of radar observations strong convective bursts, attributable to wind shear, formed primarily within the eastern semicircle of the eyewall. To reproduce this observed structure within a hurricane model, background wind shear of some magnitude must be specified and turbulence and surface parameters appropriately specified so that the impact of the shear on the simulated hurricane vortex can be realized. To identify the complex nonlinear interactions induced by changes in these parameters, an ensemble of model simulations have been conducted in which individual members were formulated by sampling the parameters within a certain range via a Latin hypercube approach. The ensemble and the data, derived latent heat and horizontal winds from the dual-Doppler radar observations, are utilized in the EnKF to obtain varying estimates of the model parameters. The parameters are estimated at each time instance, and a final parameter value is obtained by computing the average over time. Individual simulations were conducted using the estimates, with the simulation using latent heat parameter estimates producing the lowest overall model forecast error.

## 1. Introduction

Hurricanes are among the most destructive and costliest natural forces on Earth, and hence it is important to improve the ability of numerical models to forecast changes in their track, intensity, and structure. But, accurate prediction depends on minimizing errors associated with initial, environmental, and boundary conditions; numerical formulations; and physical parameterizations.

Though significant progress has been made over the past decade with regard to using data assimilation primarily to improve the initial state of a hurricane (Zhang et al. 2009; Torn and Hakim 2009; Zou et al. 2010), uncertainties still remain in other aspects of hurricane prediction.

One source of uncertainty within hurricane models comes from parameters that dominate the long-term behavior of the model. To explore whether these long-term uncertainties can be reduced, model parameters associated with both environmental and physical forcings are estimated for the sheared Hurricane Guillermo (1997; see Fig. 1) through the use of the ensemble Kalman filter (EnKF). Specifically, four parameters describing

---

*Corresponding author address:* Humberto C. Godinez, Los Alamos National Laboratory, MS B284, Los Alamos, NM 87545.  
E-mail: hgodinez@lanl.gov

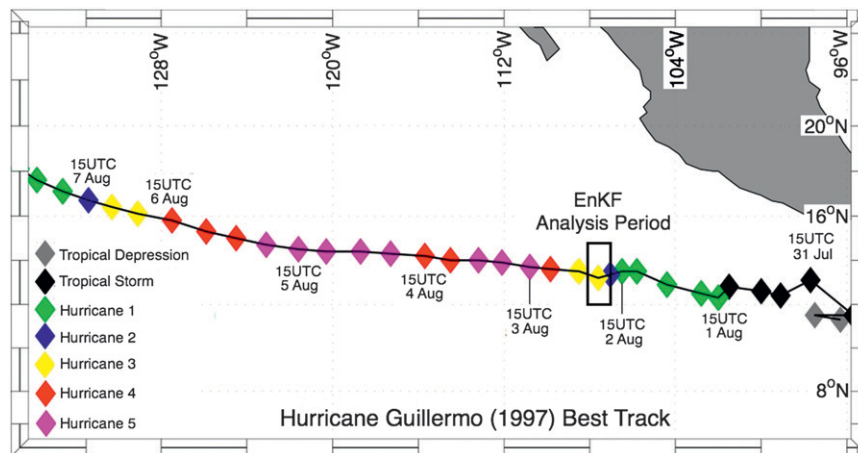


FIG. 1. Best track for Hurricane Guillermo (1997). Hurricane intensity is color coded based on the Saffir–Simpson scale with legend shown at the bottom left of the figure. Data courtesy of the NHC, NOAA. The EnKF analysis period is denoted by the small black rectangle.

momentum sinks and moisture sources in the planetary boundary layer, the unresolved transport of these quantities away from the boundary layer, and a parameter associated with describing the wind shear impacting Guillermo will be estimated. One of the key points of this paper is to illustrate that parameter uncertainty contributes significantly to the overall long-term uncertainty in a hurricane simulation.

Various authors (Anderson 2001; Annan et al. 2005; Hacker and Snyder 2005; Aksoy et al. 2006b,a; Tong and Xue 2008; Hu et al. 2010; Nielsen-Gammon et al. 2010) have documented the ability of the EnKF procedure to simultaneously evaluate model state and parameters. In these papers, the parameters are included as part of the model state in the assimilation. This combination of evolving elements (model variables) and nonevolving elements (model parameters) within the analysis introduces some difficulties for parameter estimation, such as parameter variance collapse and assimilation divergence. To mitigate these difficulties, the parameters are inflated to a prespecified variance, so as to avoid the collapse and keep a reasonable spread in the parameters. These techniques have proven to be effective for parameter estimation, but they require adjustments and tuning of the inflation to obtain good estimates. The approach in our current paper differs from previous work, in the sense that the parameters and model state are not combined in the assimilation procedure in order to estimate the parameters. In this work we use the EnKF as a tool to estimate only key model parameters for a given time-distributed observational dataset. Furthermore, since the parameters are assumed to be nonevolving, they are estimated independently from each observational dataset in time. Once the parameters are estimated at each time instance

where observations are available, a final estimate is obtained by computing the time-average value. A key aspect is to explore if estimating parameters through EnKF data assimilation can improve a model simulation, especially for such a highly nonlinear problem as a hurricane.

To test the applicability and viability of this approach, a twin experiment is performed for a hurricane model, where results indicate that the correct parameter values are recovered when considering sufficient observational data. It must be noted that the parameter estimates presented in the current work depends on two important factors: the numerical model being utilized and the dataset being assimilated. Nevertheless, this technique is applicable to estimating model parameters for any model and dataset, as long as the parameters have a strong connection to the type of observational data being used to estimate them.

One of the best dual-Doppler radar datasets (Reasor et al. 2009; Sitkowski and Barnes 2009) obtained within a hurricane will be used to estimate the parameters through assimilation. Some unique aspects of Guillermo's dual-Doppler radar data were that data from 10 flight legs over a 6-h period were individually processed to address temporal variability and that both derived fields of horizontal winds and latent heating (Guimond et al. 2011) were constructed from the flight legs. Taken together, this observational data will be used to quantify temporal variability in the four parameter estimates along with how these estimates change depending on which observational field is utilized. Note that assessing the temporal variability of the four parameters is important with regard to addressing the effectiveness of parameter estimation within this highly nonlinear system.

An important test of the viability of the particular approach and data used to estimate the parameters is their

ability to improve the solution of a hurricane model. To investigate this issue, three simulations were run using temporally averaged values of the parameters obtained from individual derived observational fields (horizontal winds or latent heat) or both fields. Hence, the final point of this paper will be to illustrate whether the parameter estimates improved overall model predictability with regard to the type of observations being assimilated.

The paper is organized as follows: Section 2 first describes the predictive component of the parameter estimation model including the analytic equation set of the hurricane model, the four model parameters present within this predictive model, the discretization, and a brief overview of the EnKF data assimilation method. In section 3 we describe various aspects of the parameter estimation model setup, including the ensemble setup and the processing of the derived observational data fields. Section 4 is broken up into three subsections with each section showing results relating to the three major points of this paper. A summary and final remarks are presented in section 5.

## 2. Parameter estimation model

The parameter estimation model is composed of a predictive model and a parameter estimation model. The chosen predictive model is the Navier–Stokes equation set coupled to a bulk cloud model with the four parameters utilized within this model, whereas the parameter estimation model employs the EnKF data assimilation. In addition to describing their representative analytical representations, discussions regarding their respective discrete formulations will also be presented in this section.

### a. Predictive model

The predictive model for hurricane simulations is based on the Los Alamos National Laboratory High Gradient (HIGRAD) large-eddy smooth cloud model (Reisner and Jeffery 2009).

#### 1) NAVIER–STOKES EQUATION SET

Since the analytical equation set representing the momentum, energy, and mass of the gas phase is identical to that utilized in Reisner and Jeffery (2009), interested readers can examine this manuscript for details regarding its formulation. Likewise the primary difference between the equation set found in that paper and the current equation set are additional terms associated with buoyancy in the vertical equation of motion related to various hydrometers, that is,  $g(\rho' + \rho_c + \rho_r + \rho_i + \rho_s + \rho_g)$ , where  $g$  is gravity,  $\rho'$  is a density perturbation,  $\rho_c$  is the cloud water density,  $\rho_r$  is the rainwater density,  $\rho_i$  is the ice water density,  $\rho_s$  is the density of snow, and  $\rho_g$  is the graupel density, with the next section briefly describing

the bulk microphysical model used to predict the evolution of the hydrometers.

#### 2) BULK MICROPHYSICAL MODEL

The mass conservation equation for a given particle type,  $\rho_{\text{part}} = \rho_c, \rho_r, \rho_i, \rho_s, \rho_g$ , within the bulk microphysical model can be written as follows:

$$\begin{aligned} \frac{\partial(\rho_{\text{part}})}{\partial t} + \frac{\partial[(u^{i'} - w_{\text{fall}_{\text{part}}}\delta_{i'3})\rho_{\text{part}}]}{\partial x^{i'}} \\ = f_{\text{density}_{\text{part}}} + \frac{\partial F'_{\rho_{\text{part}}}}{\partial x^{i'}}, \end{aligned} \quad (1)$$

where  $u^{i'}$  are fluid velocities in each spatial direction and the last term represents the turbulent diffusion of a particle type using a diffusion coefficient diagnosed from a turbulent kinetic energy (TKE) equation, see Eq. (3) in Reisner and Jeffery (2009).

The conservation equation for either cloud droplet number  $N_c$  or ice particle number  $N_i$ ,  $N_{\text{part}} = N_c, N_i$ , can be written as follows:

$$\begin{aligned} \frac{\partial(N_{\text{part}})}{\partial t} + \frac{\partial[(u^{i'} - w_{\text{fall}_{\text{part}}}\delta_{i'3})N_{\text{part}}]}{\partial x^{i'}} \\ = f_{\text{number}_{\text{part}}} + \frac{\partial F'_{N_{\text{part}}}}{\partial x^{i'}}, \end{aligned} \quad (2)$$

where  $w_{\text{fall}_{\text{part}}}$ ,  $f_{\text{density}_{\text{part}}}$ , and  $f_{\text{number}_{\text{part}}}$  represent the fall speed, density, and number sources or sinks from the bulk microphysical model for a given particle type, a hybrid of the activation and condensation model found in Reisner and Jeffery (2009) together with all of the other relevant bulk parameterizations found in Thompson et al. (2008). Note, because of significant differences in the particle distributions between winter storms and hurricanes, the slope–intercept formulas were modified following McFarquhar and Black (2004).

#### 3) PARAMETERS OF INTEREST

The initialization of the environmental or background horizontal homogeneous potential temperature, water vapor, and total gas density fields for all Guillermo simulations was achieved by examining vertical profiles from ECMWF analyses obtained near the time period of the dual-Doppler radar data (1830 UTC 2 August to 0030 UTC 3 August 1997) and using a representative composite. Though some uncertainty exists within the thermodynamic fields with regard to the actual environment versus the perturbed environment obtained from the ECMWF soundings, the impact of this uncertainty was deemed to be smaller than that associated with the

momentum fields, i.e., the simulated vortex is sensitive to small changes in wind shear. So to quantify this sensitivity, the horizontal velocity fields,  $u^{1'}$  and  $u^{2'}$ , were initialized as follows

$$u^{1'}(x^{3'}) = \phi_{\text{shear}}[\text{ecmwf}_u(x^{3'}) + 1.5], \quad (3)$$

$$u^{2'}(x^{3'}) = \phi_{\text{shear}}[\text{ecmwf}_v(x^{3'}) - 1.5], \quad (4)$$

where  $\phi_{\text{shear}}$  is a tuning coefficient that determines the shear impacting Hurricane Guillermo within a range of 0 and 1,  $x^{3'}$  is the height, and  $\text{ecmwf}_u$  and  $\text{ecmwf}_v$  represent mean soundings calculated from the ECMWF analyses.

Given the delicate balance in nature that is needed for a sheared hurricane to intensify, it is not entirely obvious whether numerical models, which are necessarily limited in resolution, can accurately represent boundary layer processes that are responsible for supplying water vapor to eyewall convection. The accurate representation of boundary layer processes implies the model has been somewhat tuned to represent the impacts of waves, sea spray, and air bubbles within the water; likewise, the accurate treatment of energy release in eyewall convection implies that the upward movement of, for example, moisture is being reasonably simulated by the hurricane model.

To examine this uncertainty, the diffusion coefficient for surface momentum calculations was specified as follows:

$$\kappa = \kappa_{\text{surface friction}} \tanh\left(\frac{\mathbf{V}_h}{80}\right), \quad (5)$$

where  $\kappa_{\text{surface friction}}$  is a tuning coefficient that ranges from 0.1 to 10  $\text{m}^2 \text{s}^{-1}$  and  $\mathbf{V}_h$  is the near-surface horizontal wind speed. A no-slip boundary condition was utilized in the horizontal momentum equations ( $u^{1'} = u^{2'} = 0$ ) with the magnitude of  $\kappa_{\text{surface friction}}$  the determining factor with regard to the impact of this boundary condition on the intensity and structure of Guillermo. Note, unlike for the horizontal momentum equations, all scalar equations use a diffusion coefficient estimated from the TKE equation within calculations of surface fluxes.

Another uncertain boundary layer process that has a significant impact on intensification rate is surface moisture availability and the unresolved vertical transport of this water vapor with the first term  $q_v^s$  being formulated as follows:

$$q_v^s = q_{vs}(0.75 + q_{v_{\text{surface}}}) \tanh\left(\frac{\mathbf{V}_h}{30}\right), \quad (6)$$

where  $q_{vs}$  is the saturated vapor pressure over water and  $q_{v_{\text{surface}}}$  is a tuning coefficient that ranges in value from 0.0 to 0.2. This term enters into surface diffusional flux calculations of water vapor  $F_{q_v}^{3'}$  in discrete form as follows:

$$F_{q_v}^{3'} = \kappa \frac{q_v^1 - q_v^s}{0.5\Delta x^{3'}}, \quad (7)$$

where  $q_v^1$  is the specific humidity of the first grid cell in the vertical direction. To address the uncertainty associated with the turbulent transport of water vapor (and all other fields) from the surface to the free atmosphere, the turbulent length scale was modified as follows:

$$L_s^m = \phi_{\text{turb}} L_s, \quad (8)$$

where the tuning coefficient  $\phi_{\text{turb}}$  ranged from 0.1 to 10, and the eddy diffusivity now being  $\kappa = 0.09L_s^m\sqrt{\text{TKE}}$ .

#### 4) DISCRETE MODEL

The discrete model for the Navier–Stokes equation set and the bulk microphysical model closely follow what was described in section 2c of Reisner and Jeffery (2009). This discrete equation set formulated on an Arakawa A grid (Arakawa and Lamb 1977) can utilize a variety of time-stepping procedures with the current simulations using a semi-implicit procedure (Reisner et al. 2005). The advection scheme used to advect gas and various cloud quantities was the quadratic upstream interpolation for convective kinematics advection scheme, including estimated streaming terms [Quadratic Upstream Interpolation for Convective Kinematics with Estimated Upstream Terms (QUICKEST); Leonard and Drummond 1995] with these quantities having the possibility of being limited by a flux-corrected transport procedure (Zalesak 1979).

The domain spans 1200 km in either horizontal direction and 21 km in the vertical direction. The stretched horizontal mesh employing 300 grid points has the highest resolution of 1 km at the center of the mesh and lowest resolution of 7 km at the model edges. Because of the addition of a mean wind intended to keep the vortex centered in the middle of the domain, the coarsest resolution resolving the highest wind field of Guillermo is approximately 2 km. The stretched vertical mesh is resolved by 86 grid points with the highest resolution of 50 m at the surface and the lowest of 500 m at the model top.

#### b. Parameter estimation model

In this section the EnKF method is briefly described. In the current work, the EnKF is mainly utilized to estimate the model parameters.

### 1) PARAMETER ESTIMATION WITH ENKF

The EnKF is a Monte Carlo approach of the Kalman filter that estimates the covariances between observed variables and the model state variables through an ensemble of predictive model forecasts. The EnKF for the geosciences was first introduced by Evensen (1994) and is discussed in detail in Evensen and van Leeuwen (1996) and in Houtekamer and Mitchell (1998). For the current study, only the parameters will be estimated, not the state vector of the model. The EnKF procedure is directly applied to the parameters, that is, the state vector contains only the parameter values. Nevertheless, the model covariance matrix is still required for the innovation with observations. The following EnKF description will be concerned with model parameters.

Let  $\mathbf{p} \in \mathbb{R}^\ell$  be a vector holding the different model parameters and  $\mathbf{x}^f \in \mathbb{R}^n$  be the model state forecast. Let  $(\mathbf{p}_i, \mathbf{x}_i^f)$  for  $i = 1 \dots N$  be an ensemble of model parameters and state forecasts, and  $\mathbf{y}^o \in \mathbb{R}^m$  a vector of  $m$  observations, and then the estimated parameter values  $\mathbf{p}_i^a$  given by the EnKF equations are

$$\mathbf{p}_i^a = \mathbf{p}_i + \tilde{\mathbf{K}}(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f), \quad i = 1, \dots, N \quad (9)$$

$$\tilde{\mathbf{K}} = \mathbf{C}^T \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}, \quad (10)$$

where the matrix  $\tilde{\mathbf{K}} \in \mathbb{R}^{\ell \times m}$  is a modified Kalman gain matrix (see appendix B),  $\mathbf{P}^f \in \mathbb{R}^{n \times n}$  is the model forecast covariance matrix,  $\mathbf{C} \in \mathbb{R}^{n \times \ell}$  is the cross-correlation matrix between the model forecast and parameters,  $\mathbf{R} \in \mathbb{R}^{m \times m}$  is the observations covariance matrix, and  $\mathbf{H} \in \mathbb{R}^{m \times n}$  is an observation operator matrix that maps state variables onto observations. In the EnKF, the vector  $\mathbf{y}_i^o$  is a perturbed observation vector, defined as

$$\mathbf{y}_i^o = \mathbf{y}^o + \boldsymbol{\varepsilon}_i, \quad (11)$$

where  $\boldsymbol{\varepsilon}_i \in \mathbb{R}^m$  is a random vector sampled from a normal distribution with zero mean and a specified standard deviation  $\sigma$ . Usually  $\sigma$  is taken as the variance or error in the observations.

One of the main advantages of the EnKF is that the model forecast covariance matrix is approximated using the ensemble of model forecasts as shown:

$$\mathbf{P}^f \approx \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i^f - \bar{\mathbf{x}}^f)(\mathbf{x}_i^f - \bar{\mathbf{x}}^f)^T, \quad (12)$$

where  $\bar{\mathbf{x}}^f \in \mathbb{R}^n$  is the model forecast ensemble average. The use of an ensemble of model forecast to approximate  $\mathbf{P}^f$  enables the evolution of this matrix for large nonlinear models at a reasonable computational cost. Additionally, the cross-correlation matrix is defined as

$$\mathbf{C} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i^f - \bar{\mathbf{x}}^f)(\mathbf{p}_i - \bar{\mathbf{p}})^T, \quad (13)$$

where  $\bar{\mathbf{p}} \in \mathbb{R}^\ell$  is the parameter ensemble average.

For our particular implementation, the system of Eqs. (9) and (10) is rewritten as

$$(\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R}) \mathbf{z}_i = \mathbf{y}_i^o - \mathbf{H} \mathbf{x}_i^f \quad (14)$$

$$\mathbf{p}_i^a = \mathbf{p}_i + \mathbf{C}^T \mathbf{H}^T \mathbf{z}_i, \quad (15)$$

for  $i = 1, \dots, N$ , where  $\mathbf{z}_i \in \mathbb{R}^m$  is the solution of the linear system (14) for ensemble  $i$ . For our implementation,  $\mathbf{R}$  is taken as a diagonal matrix, with  $\sigma$  in its main diagonal.

### 2) CONSIDERATIONS FOR PARAMETER ESTIMATION WITH ENKF

Several studies have utilized the EnKF data assimilation to simultaneously estimate the model state and parameter. Among them are the studies by Aksoy et al. (2006a,b), Tong and Xue (2008), Hu et al. (2010), and Nielsen-Gammon et al. (2010). Typically, the parameters are included as part of the model state in the assimilation. This combination of evolving (dynamic) and nonevolving elements within the analysis introduces some difficulties for parameter estimation, such as parameter variance collapse and assimilation divergence. To mitigate these difficulties, the parameters are inflated to a prespecified variance, so as to avoid the collapse and keep a reasonable spread in the parameters. These techniques have proven to be effective to estimate parameters, but they require adjustments and tuning of the inflation to obtain good estimates.

The particular approach taken in this work is to use the EnKF as a tool only to estimate the model parameters using the available data. This approach is significantly different from the studies mentioned above, in the sense that only the parameters are estimated with the EnKF; that is, the model state is not being estimated. The motivation behind this approach is that model parameters are assumed to be constant; they do not evolve through the model, although they affect the dynamics of the solution. For this reason, determining parameters can be viewed as a stationary or static optimization. Our objective is to estimate a constant parameter value for the given dataset, over the given time window. To achieve this, the assimilation of the data is performed on each time instance, where observations are available, independently. The reasoning behind this technique is to treat the parameters as constants and nonevolving elements in the model; hence, for each time period we compute an estimated parameter value.

The procedure used to estimate the parameters is the following: Let  $t_1, \dots, t_k$  be the time instances where observations are available. For each time instance  $t_j$ ,  $j = 1, \dots, k$ ; the EnKF data assimilation [Eqs. (14) and (15)] provides parameter estimates for the ensemble,  $\mathbf{p}_i^a(t_j), i = 1, \dots, N$ . A final parameter estimate is then computed by first taking the ensemble average and then the time average of the parameters, that is,

$$\mathbf{p}^a = \frac{1}{k} \sum_{j=1}^k \left[ \frac{1}{N} \sum_{i=1}^N \mathbf{p}_i^a(t_j) \right]. \quad (16)$$

One advantage is that this approach avoids the problem of parameter collapse and filter divergence, since the data assimilation is used to estimate the parameters at each time instance independently. Additionally, since the state is not being updated in the assimilation, and only the parameter is being estimated, localization is not required for the EnKF.

The number of available data points for Hurricane Guillermo is about 200 000 at any given time. This rich dataset can be used to investigate a number of aspects of hurricane intensification. To exploit this dataset for assimilation, we used an efficient matrix-free EnKF algorithm developed by Godinez and Moulton (2012). The algorithm works by efficiently solving the linear system in Eq. (14) using a solver based on the Sherman–Morrison formulas. In their paper, Godinez and Moulton (2012) show that this algorithm is more efficient than traditional implementations of the EnKF, by several orders of magnitude, and enables the assimilation of vast amounts of data. Additionally, the algorithm provides an analysis that is qualitatively and quantitatively the same as more traditional implementations. Readers are referred to the work of Godinez and Moulton (2012) for more details.

### 3. Parameter estimation model setup

This section describes the three necessary steps to conduct the parameter estimates: the processing of derived observational data fields, the setting up of the ensemble, and the setup of the EnKF system.

#### *a. Derived Doppler data fields*

The primary driver of a hurricane is the release of latent heat in clouds, which arises mainly from condensation. Latent heat cannot be observed directly and instruments, such as Doppler radars, only measure the reflectivity and radial velocity of precipitation particles averaged over the pulse volume. As a result, retrievals of dynamically relevant quantities (e.g., the Cartesian wind components and latent heat) are required. Guillermo's 3D wind field was retrieved using a variational approach

on a system of equations that includes the radar projection equations, the anelastic mass continuity equation, and a Laplacian filter (Gao et al. 1999; Reasor et al. 2009). This wind field and estimates of the precipitation water content (derived from the reflectivity measurements) are used to retrieve the latent heat of condensation/evaporation following Guimond et al. (2011). There are two main steps in the latent heat retrieval algorithm: 1) determine the saturation state at each grid point in the radar analysis using the precipitation continuity equation and 2) compute the magnitude of heat released using the first law of thermodynamics and the vertical velocity estimates described in Reasor et al. (2009). There are several potential sources of error in the latent heat retrievals, and a detailed treatment of these errors can be found in Guimond et al. (2011). Here, we summarize the most relevant information.

The uncertainty in the latent heat retrievals reduces to uncertainties in two main fields: reflectivity and vertical velocity. Guimond et al. (2011) were able to reduce and/or document the uncertainty in these fields to a level where the latent heat retrievals have a reasonably acceptable accuracy. For example, Guimond et al. (2011) focus on the inner portion of the eyewall and often use two aircraft to construct the radar analyses (Reasor et al. 2009), which reduce the effects of attenuation. In an attempt to correct the known calibration bias in the National Oceanic and Atmospheric Administration (NOAA) P-3 tail radar reflectivity, 7 dB was added to the fields (J. Gamache and P. Reasor 2010, personal communication). More importantly, however, the reflectivity is only used to determine the condition of saturation in the latent heat retrieval. Thus, the algorithm is not dependent on the precise value of the reflectivity, rendering the retrievals somewhat insensitive to errors (Guimond et al. 2011).

The uncertainties in the magnitude of the retrieved heating are dominated by errors in the vertical velocity. Using a combination of error propagation and Monte Carlo uncertainty techniques, biases are found to be small, and randomly distributed errors in the heating magnitudes are 16% for updrafts greater than  $5 \text{ m s}^{-1}$  and 156% for updrafts of  $1 \text{ m s}^{-1}$  (Guimond et al. 2011). Even though errors in the vertical velocity can lead to large uncertainties in the latent heating field for small updrafts/downdrafts, in an integrated sense, the errors are not as drastic. Figure 2 shows example horizontal views (averaged over all heights) of the latent heating rate of condensation/evaporation for 4 of the 10 aircraft sampling periods of Guillermo.

For assimilation, only latent heat data where reflectivity observations are available is included into the dataset. The errors for the retrieved wind fields are set to

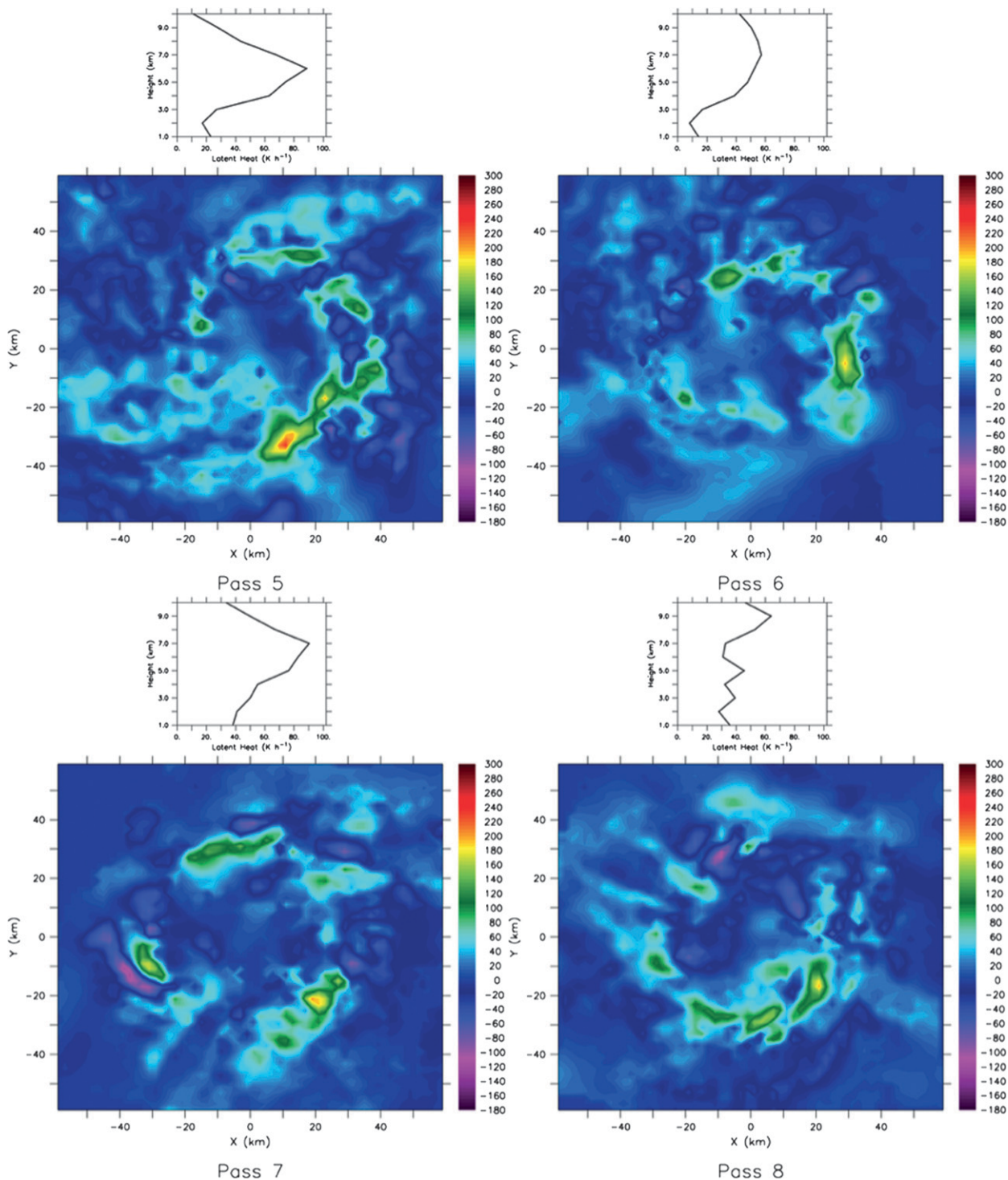


FIG. 2. Horizontal views (averaged over all heights) of the latent heating rate ( $\text{K h}^{-1}$ ) of condensation/evaporation retrieved from airborne Doppler radar observations in Hurricane Guillermo (1997) at four select times of the dual-Doppler radar data. Pass 5 corresponds to 2117 UTC from Fig. 6 of Reasor et al. (2009). Note that grid points without latent heating were assigned zero values after the vertical averaging. Vertical profile of the azimuthal mean latent heating rate at the RMW (30 km) is shown above each contour plot. The first level of data is at 1 km due to ocean surface contamination.

5.0%–6.0%, and the errors in the retrieved latent heating, as a percentage, are specified following Guimond et al. (2011):

$$\delta \mathbf{y}_{\text{lh}}^o = \left| \frac{\delta \mathbf{w}}{\mathbf{w}} \right| \times 100, \quad (17)$$

where  $\delta \mathbf{w} = 1.56 \text{ m s}^{-1}$  represents the overall uncertainty in the vertical wind velocity field  $\mathbf{w}$  Reasor et al. (2009) and  $\delta \mathbf{y}_{\text{lh}}^o$  is the observation error for the latent heat fields. It is worth noticing that observational latent heat errors are sometimes overestimated or underestimated. Thus, in an integral sense, the errors are not so drastic; the bias is only of  $+0.16 \text{ m s}^{-1}$ .

### b. Guillermo ensemble setup

Since the primary goal is to examine the impact of various model parameters containing high uncertainty on the intensity and structure of Guillermo, but not the track, all simulations comprising the ensemble have been undertaken in which a mean wind of  $1.5 \text{ m s}^{-1}$  was added or subtracted to the respective environmental wind components to prevent the movement of Guillermo from a region containing high spatial resolution found in the domain center. Specifically, this high-resolution patch in Cartesian space,  $\Delta x_c^{1'}$  and  $\Delta x_c^{2'}$ , is defined as follows:

$$\Delta x_c^{1'} = 6000 \sin^2(\phi_g x^*) + 1000, \quad (18)$$

$$\Delta x_c^{2'} = 6000 \sin^2(\phi_g y^*) + 1000, \quad (19)$$

where  $\phi_g = \pi/\text{Ngp}_i$  determines how quickly the grid spacing changes from 7 km near the model edges to 1 km near the center with  $\text{Ngp}_i$  the number of grid points in either direction and  $x^*$ ,  $y^*$  representing grid values for a normalized grid with a domain employing  $0.5\text{Ngp}_i$  away from a center location in which  $x^* = y^* = 0$ . Like the horizontal direction, the vertical direction also employs stretching with the highest resolution near the ocean boundary, approximately 50 m, and the coarsest near the model top, 500 m, with 86 vertical grid points being utilized to resolve a domain extending upward to 21 km. Note, because of the relatively high vertical spatial resolution, time step size was limited to 1 s to avoid any instabilities associated with exceeding the advective Courant number limit.

The ensemble is generated by perturbing only the four parameters discussed in section 2a(3), which are  $\phi_{\text{shear}}$ ,  $\kappa_{\text{surface friction}}$ ,  $q_{\text{surface}}$ , and  $\phi_{\text{turb}}$ . The parameter values are generated by the Latin hypercube sampling technique with a uniform distribution, where each parameter is sampled over a specified interval. All ensembles have the same initial, where the background fields are initialized as

TABLE 1. Initial parameter intervals for sampling with Latin hypercube strategy using a uniform distribution.

Parameter	Interval
Surface moisture ( $\text{kg kg}^{-1}$ )	[0.05, 0.2]
Wind shear ( $\text{m s}^{-1}$ )	[0.1, 1.0]
Turbulent length scale (m)	[0.1, 10.0]
Surface friction ( $\text{m}^2 \text{s}^{-1}$ )	[0.1, 10.0]

described in section 2a(3), where the background winds that are independent of the wind field associated with Guillermo are initialized using ECMWF data. Whereas to initialize the wind field associated with Guillermo, a composite of the radar winds and a bogus vortex is employed within a nudging procedure over a 1-h period. Afterward, the hurricane model is simulated in free mode for 5 h to become balanced with the parameters, and develops a hurricane vortex. It is important to mention that at this stage, the behavior of the ensemble simulations is mainly dominated not by the initial conditions but by the parameter values. Hence, utilizing derived fields from the same hurricane for a vortex initialization does not bias the results of the parameter estimation experiments, as long as the subsequent spinup is sufficiently long to allow the parameters to dominate the long-term behavior of the simulation.

### c. EnKF and observation selection

The EnKF data assimilation is used to estimate only the parameters of interest. The time distribution of the parameters is obtained by assimilating each time period independently, as stated in section 2b(2). A final parameter value is obtained by averaging in time the results of the assimilation. The particular algorithm used for the parameter EnKF is a matrix-free algorithm presented in Godinez and Moulton (2012).

The initial set of parameter values is generated using a Latin hypercube sampling strategy with a uniform distribution. The interval for each parameter, shown in Table 1, is chosen according to the values mentioned in section 2a, which are based on prior sensitivity simulations and from a physical intuition of their interaction with the model. Figure 3 shows the initial distribution for each parameter of interest.

Although the optimal ensemble size for estimating reliable model uncertainty is still under active research, an ensemble of 120 members was deemed appropriate to capture essential model statistics for parameter estimation. Upon completion of the 120-member ensemble and due to the small movement of simulated hurricanes within the ensemble, information from each ensemble member was independently interpolated to the center location of the grid so as to be coincident with the

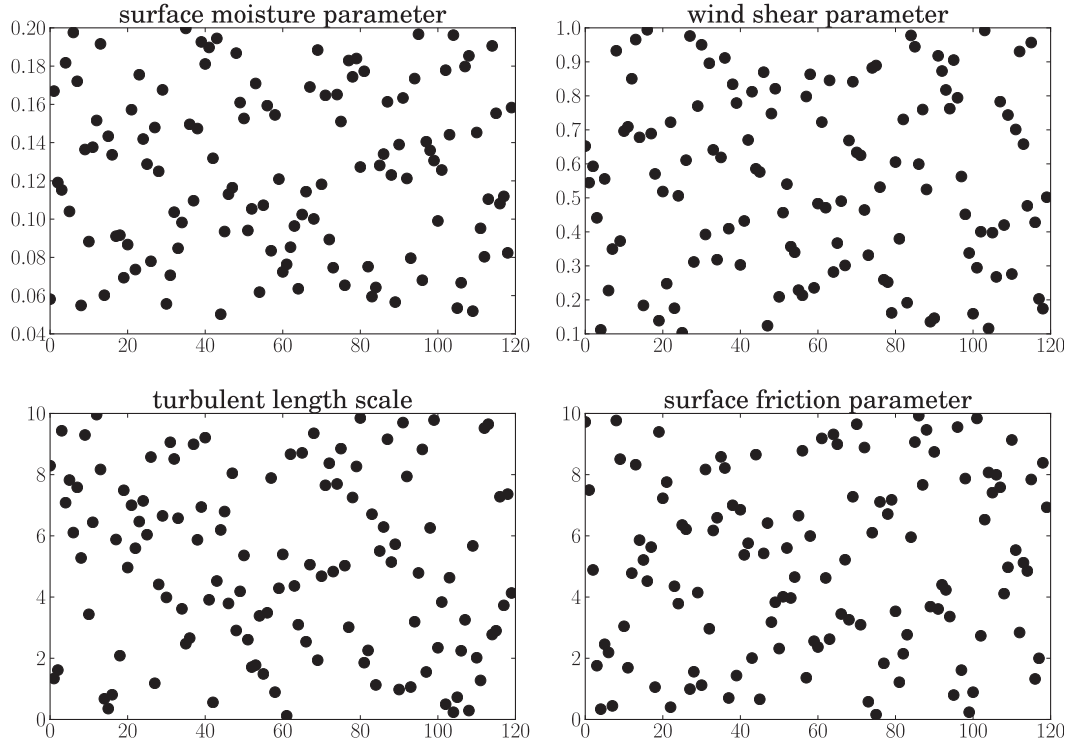


FIG. 3. Parameter spread of the 120 ensemble members obtained by utilization of the Latin hypercube sampling strategy within the limits shown in Table 1.

observations. Note, only a portion of the computational domain was utilized within the EnKF corresponding to the portion of the domain that contains the derived fields of winds and latent heat. Finally, after this interpolation step, all interpolated model and derived data fields were read into the EnKF to conduct the parameter estimation for a given time period.

The data selection procedure followed in this paper is similar to the rank parameter-observation correlation procedure presented by Tong and Xue (2008). The parameter cross-correlation matrix in Eq. (13) provides information about the correlation between the model state variables and the parameters. By applying the observation operator to the model state, we have

$$\tilde{\mathbf{C}} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{H}\mathbf{x}_i^f - \mathbf{H}\bar{\mathbf{x}}^f)(\mathbf{p}_i - \bar{\mathbf{p}})^T, \quad (20)$$

which is a cross-correlation matrix that can be used to identify regions of high correlation between parameters and model state variables in observation space. To identify the location of the observations, we add the correlations in  $\tilde{\mathbf{C}}$  and sort them from largest to smallest. To clarify, let  $\tilde{\mathbf{C}} = [\mathbf{c}_1 \cdots \mathbf{c}_l]$ . Each of the columns  $\mathbf{c}_i$  contains the correlation of parameter  $\mathbf{p}_i$  to the state

variable in observation space. We add all of these columns to obtain  $\mathbf{c} = \sum_{i=1}^l \mathbf{c}_i$ , which is then sorted from smallest to largest. This will provide the location of the largest correlation of the model in observation space to assimilate. It is important to note that the regions of high correlation can be identified from different model variables fields, such as latent heat, horizontal winds, or reflectivity. This enables the use of one of these fields as a proxy for observation localization in order to better capture the physical processes of interest.

## 4. Results

This section will highlight the three main points of this paper: 1) the highly nonlinear interactions between the various parameters, leading to large variations in the simulated intensity and structure of Guillermo among the 120 ensemble members; 2) the large amounts of derived data, horizontal winds or latent heat, required to reasonably estimate the four model parameters; and 3) the newly estimated parameters lead to an overall reduction in the model forecast error.

### a. Ensemble spread and structure

Since the EnKF method depends on proper model statistics to determine optimally the parameter values

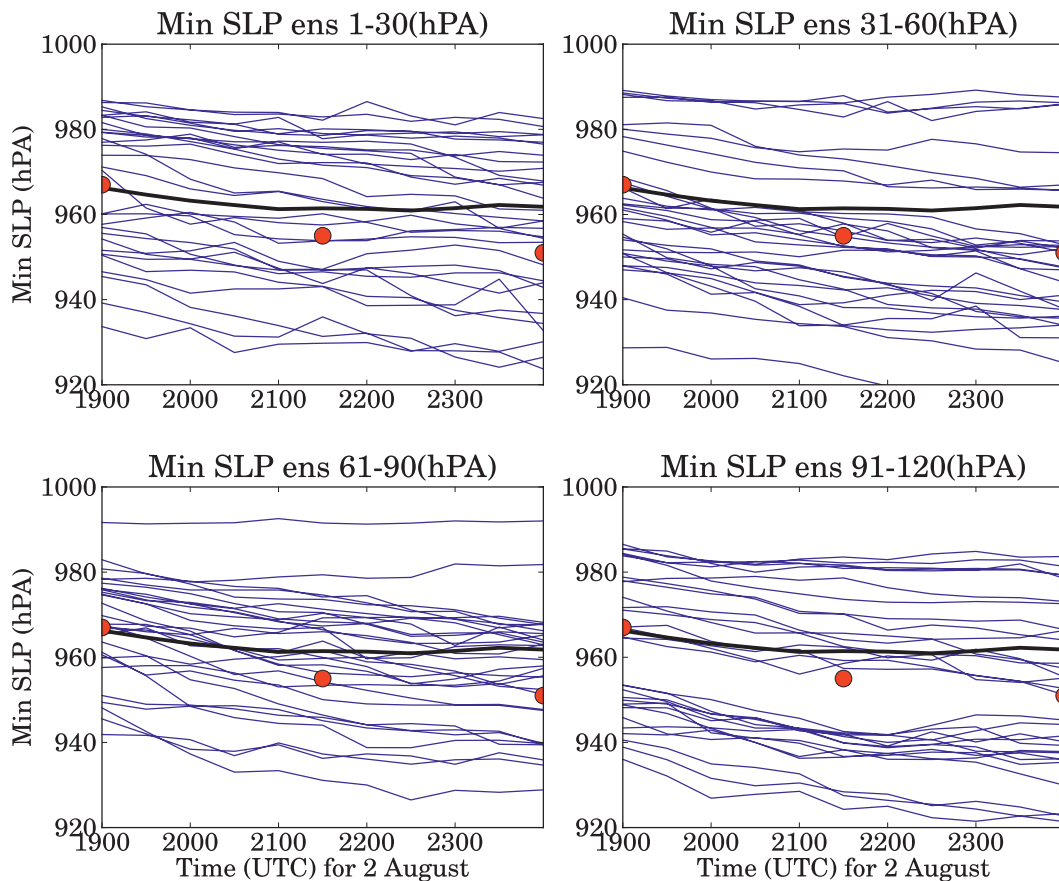


FIG. 4. Minimum sea level pressure vs simulation time for each ensemble member (blue lines), ensemble average (black line), and observations (red dots), for (top left) ensemble 1–30, (top right) ensemble 31–60, (bottom left) ensemble 61–90, and (bottom right) ensemble 91–120.

(Anderson 2001), this section will illustrate this necessary variability across the various members of the ensemble. Likewise, for reasonable parameter estimation, it is important that the ensemble produces statistics that are within the range of the observations and this is another aspect of the ensemble that will be presented. For example, Fig. 4 shows the simulated pressure traces for the 120 ensemble members (blue lines), the ensemble average (black line), and the 3-hourly observations from the National Hurricane Center (NHC) advisories (red dots), with a large spread in hurricane intensity being denoted within the ensemble. Further, even though this result may be somewhat fortuitous, the ensemble-averaged pressure trace is in remarkably good agreement with observations with a difference of less than 5 hPa. To highlight differences between the ensemble average and a given member, in addition to displaying ensemble-average statistics, results from a control simulation that also reasonably reproduced the observed pressure trace will be shown (parameter values shown in Table 4). The parameter values for the control run were obtained through

brute-force trial-and-error experiments until a hurricane simulation that matches the observed minimum sea level pressure was obtained.

Because Hurricane Guillermo was embedded in an environment characterized with vertical wind speed shear, its observed eyewall horizontal structure was asymmetric with a dominant wavenumber-1 mode, for example, Reasor et al. (2009) and Sitkowski and Barnes (2009). To illustrate the models' ability to reproduce this asymmetry, Fig. 5 shows both the ensemble-averaged layer-averaged vertical velocities and the corresponding layer-averaged fields from the control at two different layers. As evident in this figure, the vortex in both the average sense and for the control is asymmetric with a dominant wavenumber-1 mode being readily apparent. Likewise, the impact of averaging across all members is clearly evident in Fig. 5, with a significant smoothing and reduction in vertical motions being noted with regard to the vertical motion fields produced by the control simulation. Furthermore, using a similar Fourier spectral decomposition procedure as in Reasor et al. (2009), Fig. 6

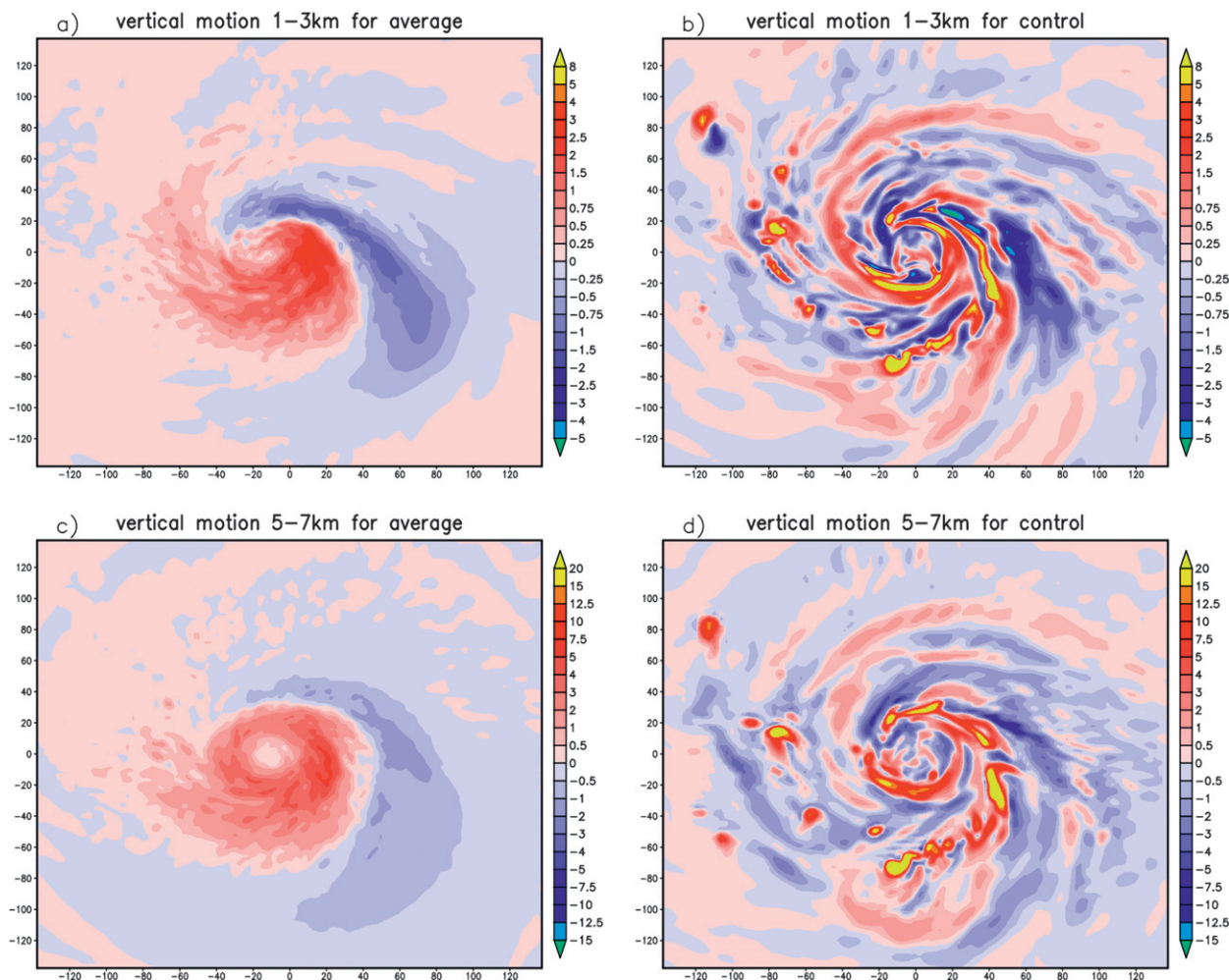


FIG. 5. Ensemble average vertical motion fields at 2300 UTC (11 h into the simulations) for ensemble averaged (a) between 1 and 3 km and (c) between 5 and 7 km. Corresponding layer-averaged vertical motions fields from a control simulation (b) between 1 and 3 km and (d) between 5 and 7 km.

reveals significant amounts of the vertical motion fields being in wavenumber-0 and -1 components, with the magnitude of the wavenumber-1 vertical motion field from the control reasonably agreeing with the observations [see Fig. 15a of Reasor et al. (2009)].

To provide another view of the simulated storm structure, both the ensemble-averaged and control simulation azimuthal structures were compared to observations and are presented for two flight legs in Figs. 7 and 8. Common disagreement with observations can be seen in both figures: First the simulated radius of maximum wind (RMW), and hence eyewall size, is about 5–10 km smaller than in the observations. While the simulated ensemble-averaged azimuthal tangential component of the wind lies within about  $10 \text{ m s}^{-1}$  of the observations, more notable differences are seen in the radial component of the wind with magnitudes rarely exceeding  $10 \text{ m s}^{-1}$  in the observations and the simulation

consistently producing magnitudes exceeding  $15 \text{ m s}^{-1}$ . Similar overestimation is produced in the latent heat fields, with values exceeding 25 000 or even 30 000  $\text{K h}^{-1}$  in the simulation and with the observations showing values marginally reaching 20 000  $\text{K h}^{-1}$ .

Despite these noteworthy differences, the HIGRAD model (section 2a) is able to reproduce the slope/tilt of radial, tangential, and latent heat fields with a reasonable degree of realism. Moreover, the heights above sea level of the contours encompassing the largest simulated values of those three fields are in overall good agreement with observations. Note that the larger values of latent heat are required to compensate for the impact of spurious evaporation (Reisner and Jeffery 2009) common to most cloud models, with the consequences of this evaporation being discussed later in this manuscript. Additional plots were made for the remaining eight flight legs (not shown) and displayed similar attributes.

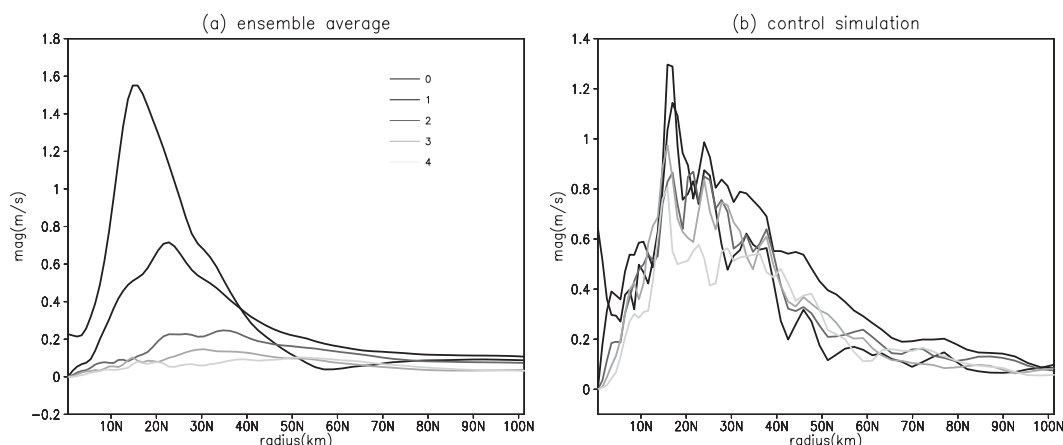


FIG. 6. Time-averaged axisymmetric (black solid line) and azimuthal wavenumber-14 amplitudes of vertical velocity in the 13-km layer within a 200-km radial distance from the storm center from the (a) ensemble average and (b) control simulation. Averaging times are between 2230 and 2300 UTC.

### b. Twin experiments for parameter estimation

To assess the reliability of parameter estimation within the current context, and the amount of observational

data needed for the estimation, a series of twin experiments was performed. A synthetic observational dataset is produced from a reference model run with the specific parameter values given in Table 2, and initialized

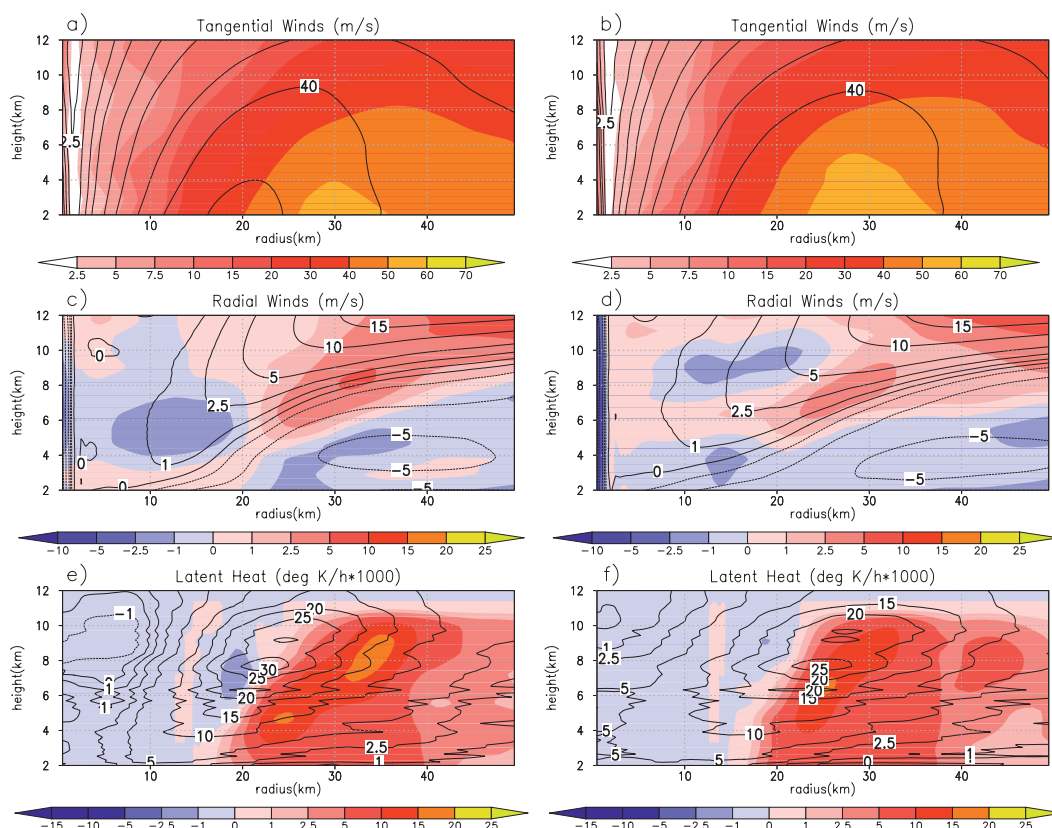


FIG. 7. Comparisons between azimuthally averaged profiles for the ensemble average (contours) and observations (shaded) for (a), (b) tangential winds, (c), (d) radial winds, and (e), (f) latent heat. Time periods for comparisons are at flight legs 5 (2117 UTC) and 9 (2333 UTC).

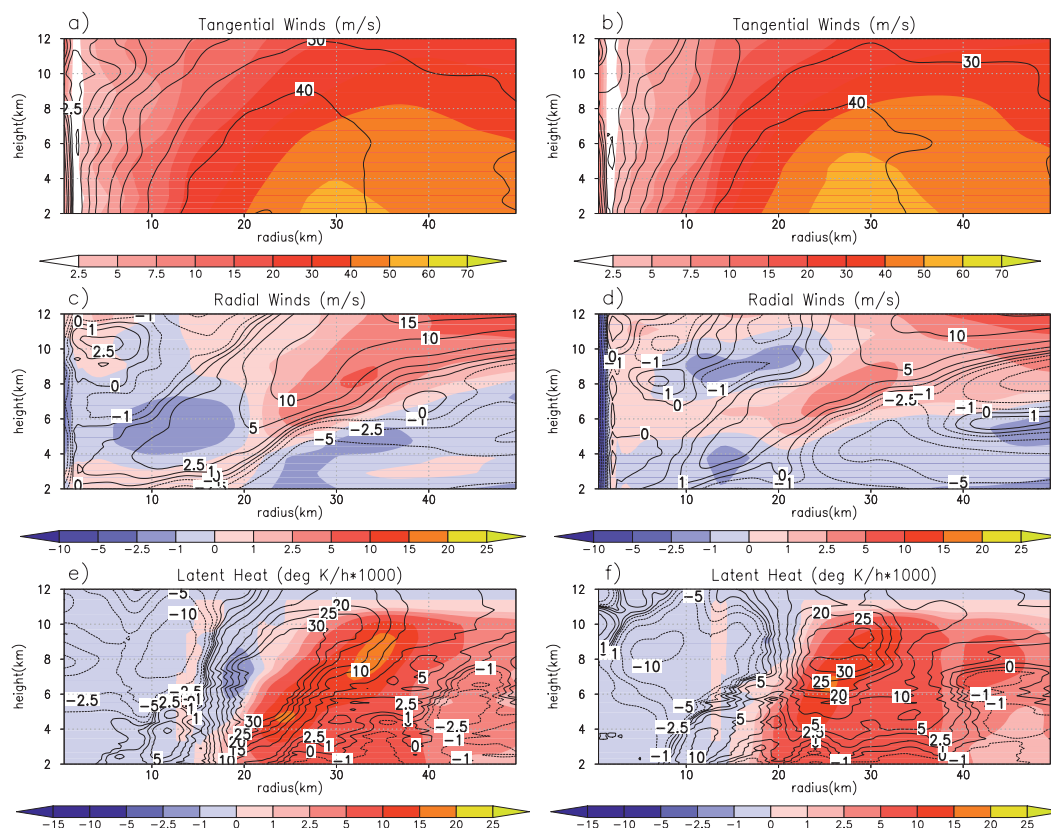


FIG. 8. As in Fig. 7, but for the control simulation (contours) and observations (shaded).

according to section 3b. These reference parameters were selected near the ensemble average with white noise added to them.

In their paper Tong and Xue (2008) found that to estimate simultaneously the model state and five parameters with the ensemble square root filter, using their rank parameter-observation procedure, they needed only 30 observational data points. Given that our parameter estimation setting is significantly different from their approach, it is not entirely obvious whether only small amounts of data are required to conduct the parameter estimation or whether, in the other limiting situation, more data than are currently available are needed for undertaking the estimates. To address this issue, nine parameter estimation experiments, named TE1–TE9, were performed for different amounts of data, given in Table 3. The data being used for parameter estimation are the latent heat field from the reference run, where the data are selected according to the procedure described in section 3c. The first observation set was taken at  $t = 6$  h of the simulation time, afterward 9 more observation sets were taken at 30-min intervals, which makes a total of 10 observational dataset over a 5-h window. For each experiment, the parameters are

estimated according to section 2b(2); that is, parameter estimates for the ensemble are computed with the EnKF at each observational time, and then a final parameter estimate is computed by averaging over ensemble and then time, as in Eq. (16). Figure 9 shows the parameter estimates for each experiment, where the vertical lines indicate the time variance of the parameter estimate. The figure clearly shows the impact of the additional data on the parameter estimates with a noticeable reduction in the error for all for parameters. Furthermore, it is only when approximately 200 000 observations are used that all four parameters converge to the correct values. This is highly relevant since it indicates that a significant amount of data is required, in this context, to estimate correctly the values of the parameters. A similar result was

TABLE 2. Parameter values used for the reference run, from where synthetic data are used for the twin experiments.

Parameter	Value
Surface moisture ( $\text{kg kg}^{-1}$ )	$9.325\,522 \times 10^{-2}$
Wind shear ( $\text{m s}^{-1}$ )	$4.968\,604 \times 10^{-1}$
Turbulent length scale (m)	3.753 693
Surface friction ( $\text{m}^2 \text{s}^{-1}$ )	1.443 062

TABLE 3. Number of observations (obs) used in the twin experiments for parameter estimation.

Experiment	No. of obs
TE1	20
TE2	63
TE3	200
TE4	632
TE5	2000
TE6	6325
TE7	20 000
TE8	63 246
TE9	200 000

obtained when the horizontal wind field or reflectivity was used to estimate the parameters. Additionally, this twin-experiment parameter estimation result is consistent with the results of Yussouf and Stensrud (2012), where the simultaneous estimation of all parameters leads to an improved analysis compared to single-parameter estimation.

### c. Parameter estimation experiments with Hurricane Guillermo observations

Given the large ensemble spread in various model fields, such as derived intensity, and the ability of the

model to reproduce observed data suggests that estimation of the four model parameters is not only possible with the EnKF but it should produce parameter estimates that hopefully reduce model forecast errors. The sensitivity of the hurricane simulation, with regard to structure and intensity, is shown in appendix A. In that appendix, it is observed that the intensity of the storm is highly sensitive to the turbulent length scale, surface friction, and surface moisture parameters, while the structure of the storm is highly sensitive to the turbulent length scale wind shear and surface friction parameters. The sensitivity results suggest that all four parameters must be estimated to obtain both structure and intensity correct. This is consistent with Yussouf and Stensrud (2012), who found that the simultaneous estimation of all parameters will provide a more accurate set of parameter values, in the presence of model error, compared to single-parameter estimation experiments with the EnKF.

The number of observations used for parameter estimation, using the EnKF in all subsequent experiments, were those identified with the highest ensemble sensitivity located at 200 000 model grid points (see selection of observations in section 3c). The assimilation is performed over latent heat (DA1), horizontal winds (DA2), or both fields (DA3) started 6 h into the ensemble simulation (corresponding to 1900 UTC). With the inclusion

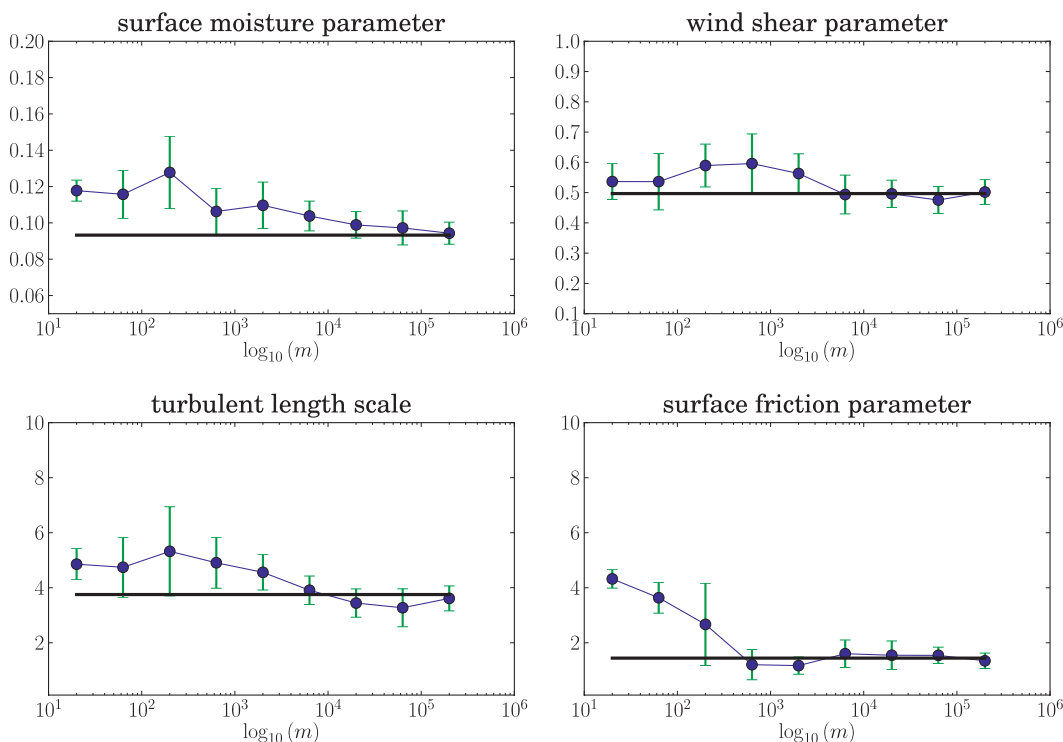


FIG. 9. EnKF parameter estimation as a function of number of latent heat observations assimilated. Latent heat observations were added in locations where the ensemble is most sensitive to changes in the parameters (section 3c).

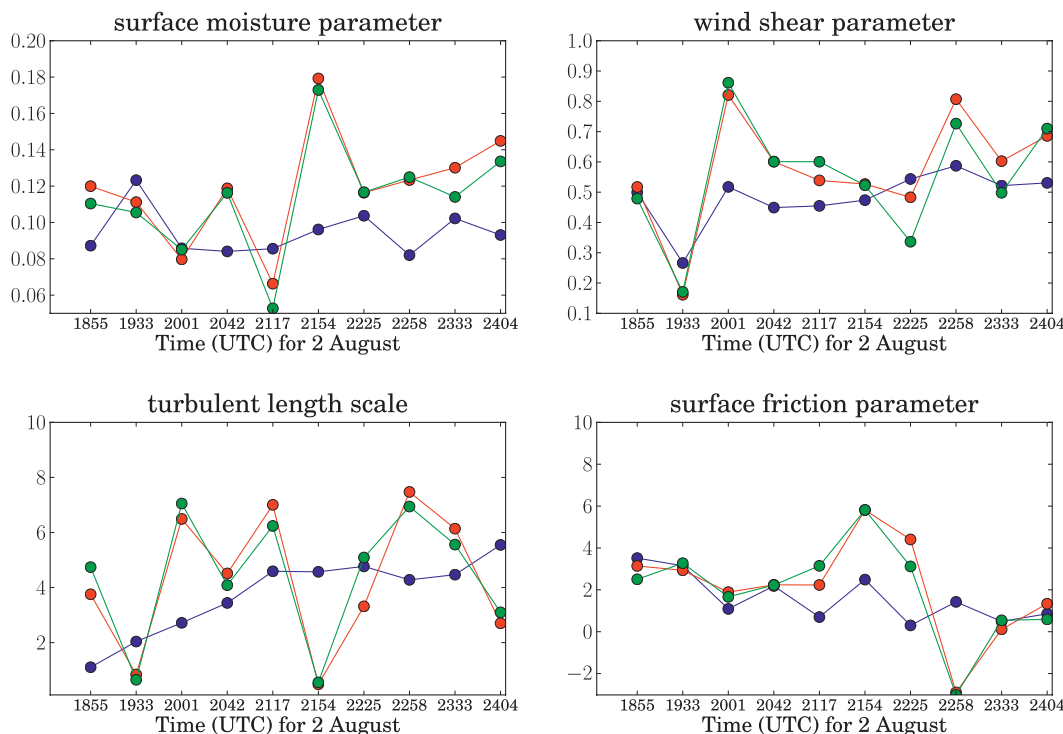


FIG. 10. Time distribution of the ensemble-average parameter estimates with EnKF from DA1 (blue line, latent heat), DA2 (red line, horizontal winds), and DA3 (green line, both latent heat and horizontal winds).

of DA3, an assessment regarding how the EnKF procedure weights two different observational datasets and model results can be made and analyzed. Hence, this section will not only highlight how the various parameter estimates change when using different observational fields but also how these estimates change over time.

Figure 10 shows the time distribution of the ensemble-average parameter estimates with EnKF using the 10 data time periods for DA1, DA2, and DA3. The largest temporal oscillations in the parameter estimates are associated with DA2 and suggest that the wind fields produced by the ensemble tend to oscillate more in time than the latent heat fields. Likewise, the surface moisture and the wind shear parameter estimates do not appear to change significantly in time, whereas the turbulent length scale and surface friction estimates either increase or decrease with time. Note that the temporal changes in these two parameters could be due to numerical errors and/or the impact of initial condition errors. For example, Fig. 7 shows that with time the areas of maximum latent heating and/or winds from the ensemble are expanding outward and hence this outward expansion, probably the result of numerical diffusion, could explain the temporal changes in these two parameter estimates.

Figure 11 shows the control parameter values and the time-averaged parameter values for each of the assimilation experiments (SDA1–SDA3). The vertical

lines indicate the 95% confidence interval for each parameter value. The parameter estimate for the surface moisture parameters shows the most difference between the assimilation experiments, mainly between the control, SDA1, and SDA2. These small differences in the parameters do lead to rather significant differences in both the structure and intensity of the simulated hurricanes (see appendix A for a brief sensitivity-to-parameters study).

Table 4 shows the time average parameter values from the assimilation experiments, as well as the parameter values used for control simulation. Figure 11 also reveals the various parameter estimates are different from the parameter values used in the control simulation, suggesting the importance of using a technique, such as the EnKF, to obtain the estimates, instead of simply using estimates obtained from a simulation that matches an observable, such as minimum sea level pressure. This ability of the EnKF to fit the parameter values reasonably to the chosen observational dataset is also illustrated by the time-averaged estimates from DA3 that, as expected, lie somewhere between the parameter estimates from DA1 and DA2, except for the turbulent length scale parameter.

#### d. Parameter error assessment

To illustrate the ability of the parameter estimates to reduce model forecast error, three simulations (SDA1,

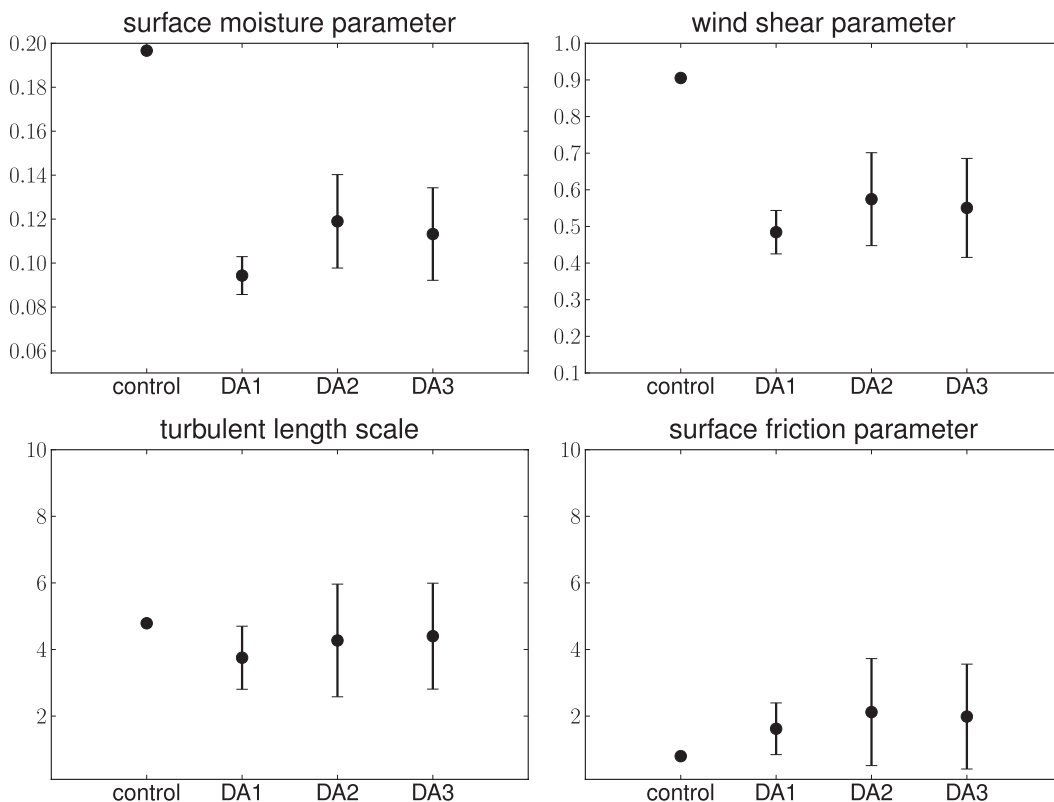


FIG. 11. Analysis parameters averaged over time for the control simulation, DA1, DA2, and DA3. Vertical lines from the dots indicate 95% confidence intervals of the parameter estimates for each experiment.

SDA2, and SDA3) were run using the time-average parameter estimates from DA1, DA2, and DA3 shown in Table 4. The setup for each simulation is the same as described in section 3b. Both qualitatively (see Figs. 12 and 13) and quantitatively (see Fig. 14) the parameter estimates produce model fields that are in better agreement with observed fields, especially the estimates associated with DA1 or those derived from using the latent heat fields. Although it is not surprising that SDA1 best matches the observed latent heat fields, what was surprising was the errors associated with the wind fields were lower in SDA1 than those associated with SDA2. This suggests the possible utility of using latent heat instead of more traditional observational fields, such as horizontal winds or radar reflectivity within data assimilation procedures to estimate model parameters and/or portions of

the model state vector. In fact, since for sheared hurricanes the majority of convection can occur on one side of the hurricane, the utilization of latent heat fields may be especially important for parameter estimation as compared to the incorporation of the much smoother and more symmetric wind fields that are a consequence of the asymmetric convective activity. Thus, for Hurricane Guillermo, it is important to estimate properly the parameters responsible for the observed asymmetric convective activity, and by doing so the overall model error is reduced not only in the latent heat fields but also in the wind fields.

When a combination of both observational fields are utilized for parameter estimation, error estimates from SDA3 reveal that the response is weighted toward producing results closer to SDA2, suggesting the

TABLE 4. Time-average parameter values for each of the experiments DA1–DA3, and the parameter values for the control simulation.

Parameter/simulation or DA	Control	DA1	DA2	DA3
Surface moisture ( $\text{kg kg}^{-1}$ )	$1.966\ 640 \times 10^{-1}$	$9.431\ 537 \times 10^{-2}$	$1.189\ 942 \times 10^{-1}$	$1.132\ 110 \times 10^{-1}$
Wind shear ( $\text{m s}^{-1}$ )	$9.052\ 798 \times 10^{-1}$	$4.843\ 292 \times 10^{-1}$	$5.744\ 430 \times 10^{-1}$	$5.506\ 012 \times 10^{-1}$
Turbulent length scale (m)	4.787 677	3.753 072	4.271 998	4.401 512
Surface friction ( $\text{m}^2 \text{s}^{-1}$ )	$7.975\ 932 \times 10^{-1}$	1.619 931	2.120 377	1.986 076

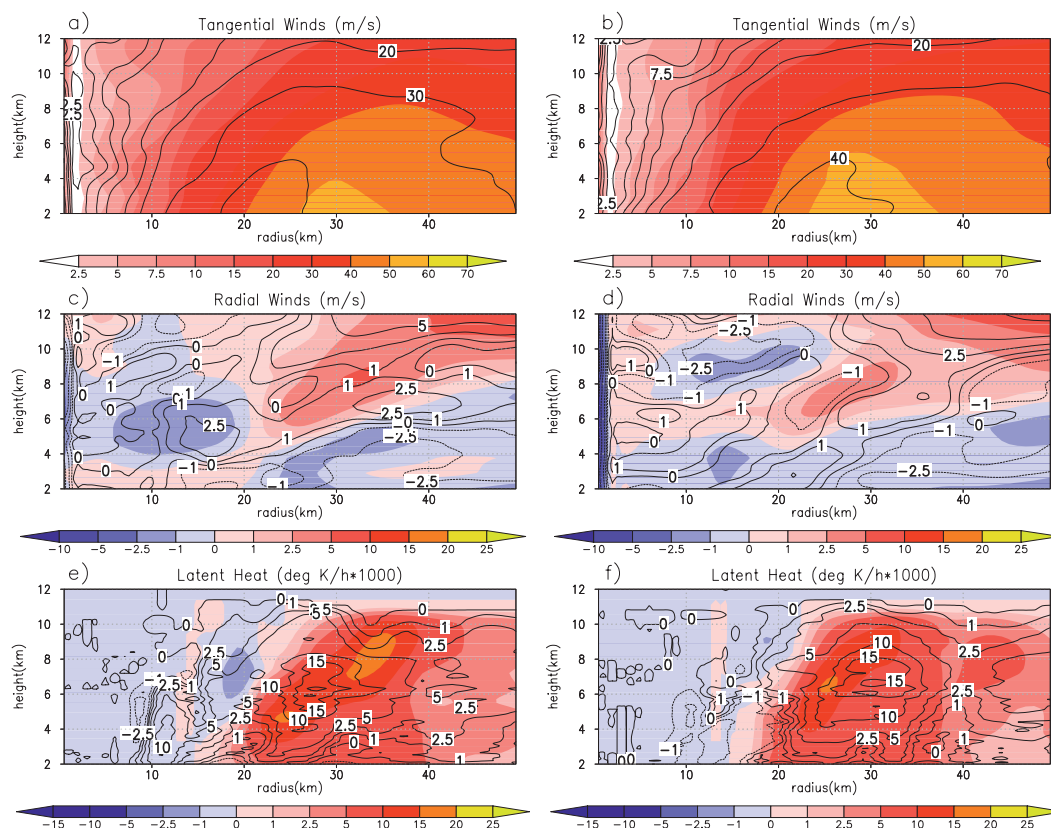


FIG. 12. Comparisons of the azimuthally averaged profiles between a model simulation (contours) using estimated parameters from DA1 and observations (shaded). Plots for (a),(b) tangential winds, (c),(d) radial winds, and (e),(f) latent heat for flight legs 5 (2117 UTC) and 9 (2333 UTC).

dominance of the horizontal wind observational data in the parameter estimates. This behavior can also be appreciated in the parameter estimates themselves, as seen in Figs. 10 and 11. The dominance of the horizontal wind field in SDA3 may be explained by the relative observational error associated with each type of observational data [see section 3a and Eq. (17)]; that is, the errors associated with the horizontal wind fields are lower than the errors in the latent heat observational field, leading the EnKF algorithm to rely more on the horizontal wind fields than the latent heat fields.

Figure 14 shows the  $L_2$  error norm of the SDA1–SDA3 simulations. As seen from the figure, SDA1 (the parameter estimates obtained using latent heat) produces lower errors than SDA2 and SDA3, which may be counterintuitive. Nevertheless, we should keep in mind that one of the main drivers in hurricane structure is latent heat release, which will also affect the horizontal wind fields. This is the main reason that, by estimating the parameters using latent heat fields, the simulation is able to obtain a very reliable structure of the hurricane, including the horizontal wind fields. Hence, the  $L_2$  error

norm of SDA1 is much smaller than that of SDA2 and SDA3. A key remark on the results shown in Fig. 14 is that of observation sensitivity, that is, the quality of information provided by the type of observations being used to estimate the parameters of interest. Not all observations provide the same quality of information. While some observations may be abundant and very reliable, their information may not be sufficient to estimate correctly important aspects of a hurricane. Such is the case with latent heat observations: they provide good information regarding the structure of the hurricane.

Though Fig. 14 demonstrates that SDA1 produces lower errors in the structure of the storm than SDA2 and SDA3, Fig. 15 illustrates that the intensity of SDA1 is significantly weaker than SDA3 and slightly weaker than SDA2. This finding may suggest that in order for SDA1 to reproduce accurately the intensity of Guillermo, additional observational data are needed below the range of the radar (approximately 1 km in height) or that other errors, such as numerical errors, are contributing to the intensity differences, that is, numerical diffusion and spurious evaporation associated with large numerical

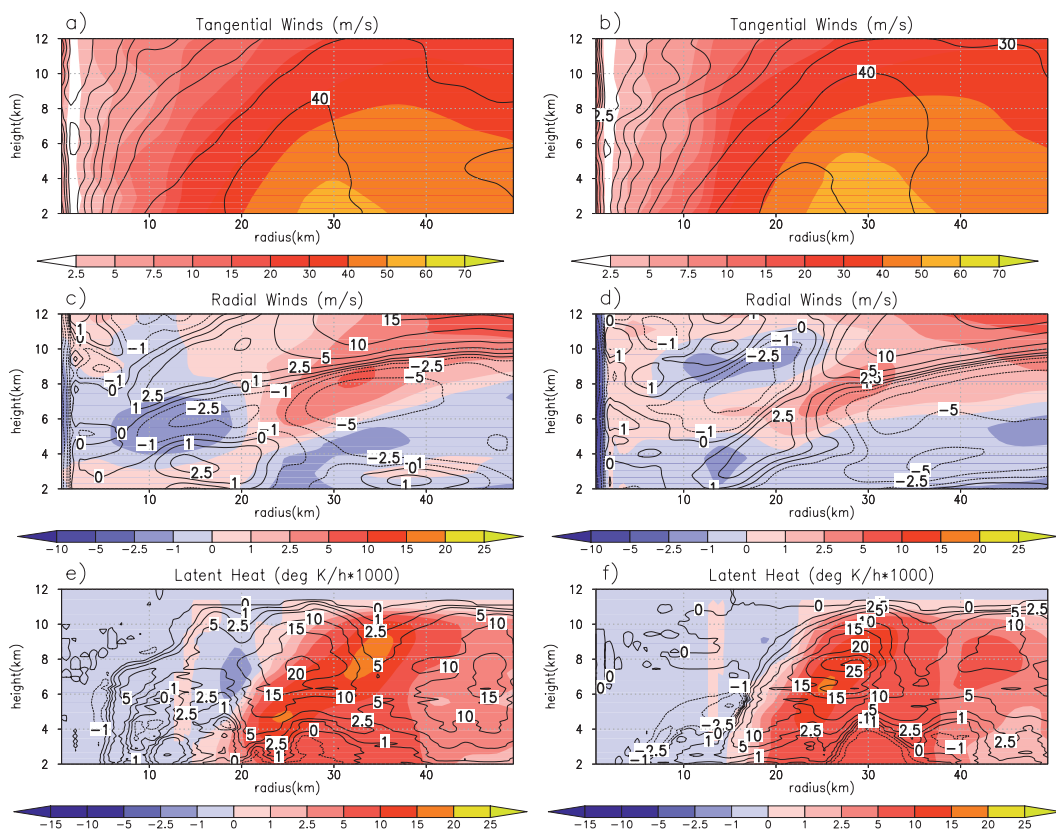


FIG. 13. As in Fig. 12, but from DA2 and observations.

errors found near cloud boundaries. For example, the bottom panels of Figs. 12 and 13 show areas of evaporative cooling occurring immediately to the left of regions of strong positive latent heat release that do not have an analog in the observed fields. But, though SDA1 appears to suffer from relatively large numerical errors that are also common to all hurricane models, the end result of the parameter estimation procedure is still a reduction in overall model forecast error for structure.

Another remark is that, like the control simulation, SDA2 produces more latent heat than the observed amount of latent heat (see Fig. 13), suggesting the simulation is indeed having to compensate for large amounts of spurious evaporative cooling. Also note that both SDA2 and SDA3 in both plots in Fig. 14 are very close to each other, suggesting that the parameters used in both simulations provide very similar results.

Similar to previous studies (Bryan 2012; Emanuel and Rotunno 2011), the primary sensitivity in intensity and structure is with regard to changes in surface friction and the turbulent length scale. The surface friction coefficient in the model leads to rather large changes in the size of the eyewall (structure) as well as the impact of shear on the intensity; that is, a smaller more rounded eyewall is

less susceptible to wind shear. Note, it is entirely possible to get the same intensity but completely different structures for two simulated hurricanes; that is, a large broken eyewall produced by a small surface friction parameter and a large turbulent length scale versus a small rounded eyewall produced by a large surface friction coefficient and a small turbulent length scale. Thus, the importance of using both sea level pressure and  $L_2$  error estimates with regard to ensuring that the estimated parameters are getting the right result (intensity) for the right reason (structure).

## 5. Summary and conclusions

This paper presented the estimation of key model parameters found within a hurricane model, through the use of EnKF data assimilation. The particular approach was to use the EnKF to estimate only the parameters at each time instance where observations were available. The advantage of this approach is that it avoids the combination of dynamic and nondynamic elements in the assimilation procedure, which introduces difficulties when estimating parameters. An efficient matrix-free EnKF data assimilation algorithm (Godinez and

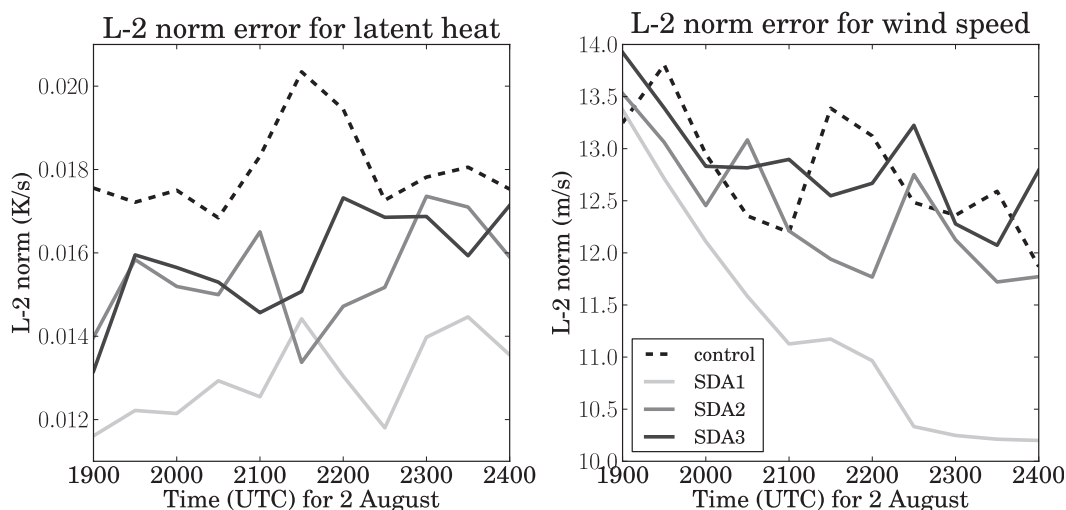


FIG. 14. Error estimates as a function of time computed using Eq. (20) of Reisner and Jeffery (2009) for the control simulation, SDA1, SDA2, and SDA3.

Moulton 2012) is used to assimilate the derived data fields—namely, horizontal wind or latent heat—available for Hurricane Guillermo. Likewise, upon completion of a 120-member ensemble that reasonably reproduced observations, the parameter estimation experiments show that a large number of data points are indeed required within the current approach to provide a reasonable estimate of the model parameters. Nevertheless, the parameter estimation procedure presented in this work can be easily applied to other models and datasets.

A unique aspect of this work was the utilization of derived fields of latent heat to estimate key model parameters. Unlike latent heat, which can be directly linked to a fundamental physical process occurring within a hurricane model, that is, condensation, utilization of other data fields, such as radar reflectivity, requires the model to capture faithfully physical processes that are not yet well understood, that is, collision-coalescence, and that are also not the primary driver for hurricane structure and intensification, potentially leading to large errors in parameter estimates.

Our experimental results show that the estimates obtained using derived latent heat fields produced lower model forecast errors in structure than a simulation using parameter estimates obtained from horizontal wind fields or radar reflectivity alone (not shown). Furthermore, the results with latent heat produced a hurricane that has a reasonable intensity, although not as close to the observed minimum sea level pressure as the results with latent heat and horizontal wind velocity fields. This may be explained by the presence of numerical errors in the model, such as numerical diffusion and spurious evaporation. This result also suggests the parameters

associated with the primary component of hurricane intensification, condensation of water vapor into cloud water, should also be included in the current parameter estimation procedure. It is important to note that deriving latent heat fields requires accurate vertical velocity measurements, which in most cases are not available. The availability of dual-Doppler radar data, for Hurricane Guillermo, made the computation of latent heat possible. Such a dataset might not be easily acquired for other hurricanes, but one of the contributions of this work is to demonstrate the value of such a dataset for parameter estimation.

An important finding of this research is the influence that a small number of parameters has upon the

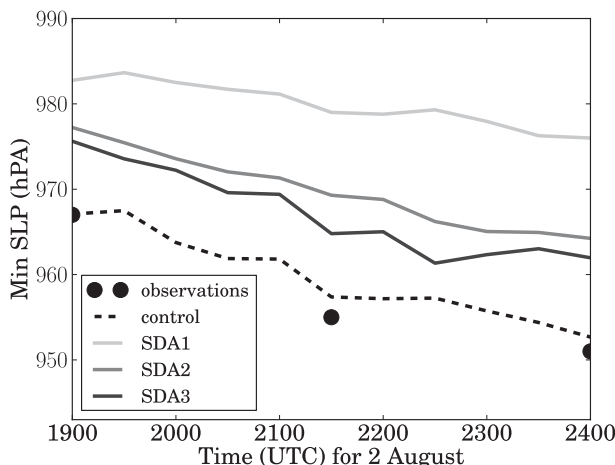


FIG. 15. Minimum sea level pressure for the control simulation, SDA1, SDA2, and SDA3 along with the observed pressure from Hurricane Guillermo.

TABLE 5. Base parameter values for sensitivity analysis experiments. Values were chosen after a series of trial-and-error experiments were performed to find a set of values that would agree reasonably well with observed minimum sea level pressure for Hurricane Guillermo (1997). Objective was to obtain a strong sheared hurricane whose vortex was asymmetric.

Parameter	Base value
Surface moisture ( $\text{kg kg}^{-1}$ )	0.20
Wind shear ( $\text{m s}^{-1}$ )	0.1
Turbulent length scale (m)	1.0
Surface friction ( $\text{m}^2 \text{s}^{-1}$ )	5.0

evolution and formation of a simulated hurricane. The various structural aspects of hurricanes can only be explored in a limited way by a few parameters. Because of this, we see a discrepancy in structure and intensity from the different parameter estimates using the different data fields (latent heat and wind velocity fields). Improving both hurricane structure and intensity, by estimating parameter using latent heat

observations, will be further explored in our future research.

Another subtle aspect suggested by this paper is that in order for a given hurricane model to reproduce both a realistic latent heat field and the correct intensity, numerical errors, especially near cloud edges, must be small. Currently, all hurricane models produce large numerical errors near cloud boundaries with these errors possibly inducing significant amounts of spurious evaporation. Hence, future work is needed to help reduce the impact of cloud-edge errors via either calibrating a tuning coefficient employed within an evaporative limiter—see Eq. (A24) in Reisner and Jeffery (2009)—using the current EnKF procedure or replacing Eulerian cloud modeling approaches with a potentially more accurate Lagrangian approach (Andrejczuk et al. 2008).

*Acknowledgments.* This work was supported by the Laboratory Directed Research and Development Program of the Los Alamos National Laboratory, which is

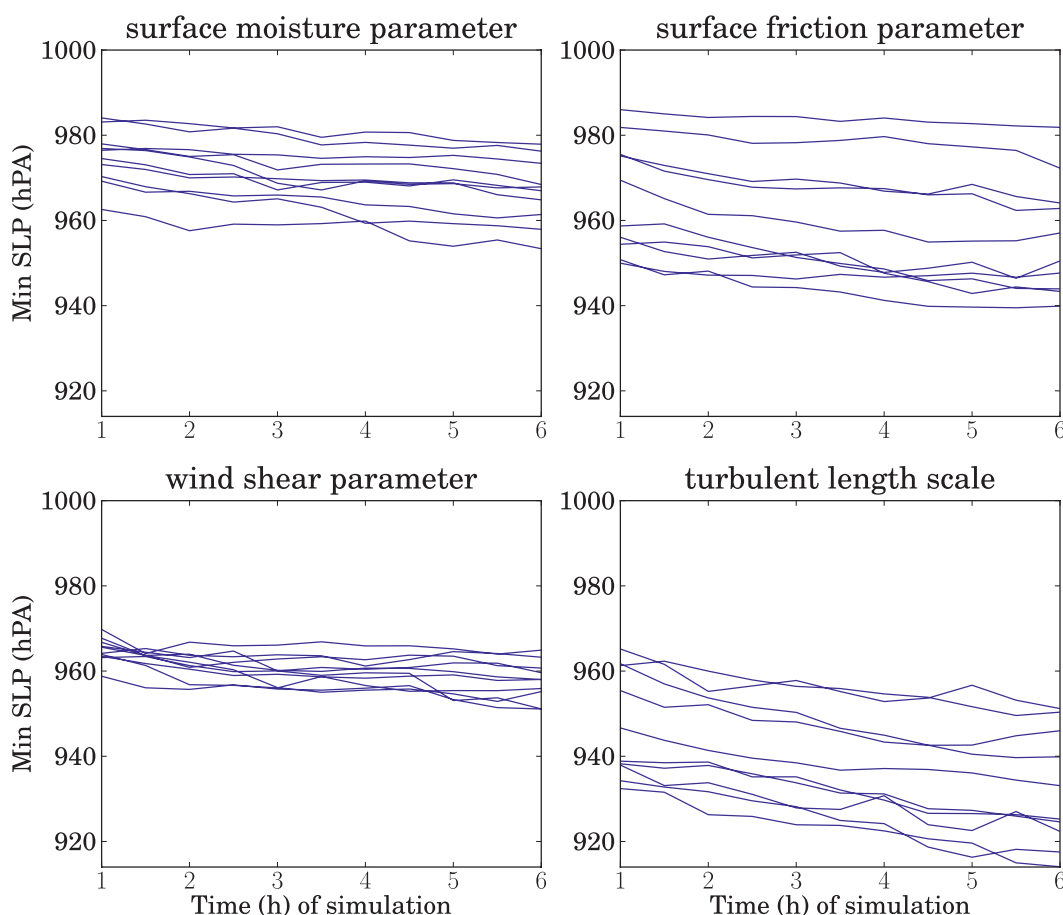


FIG. A1. Minimum sea level pressure for each ensemble member corresponding to the perturbation of (top left) surface moisture, (top right) surface friction, (bottom left) wind shear, and (bottom right) turbulent length scale. Minimum sea level pressure is computed from the seventh to the twelfth hour of simulation.

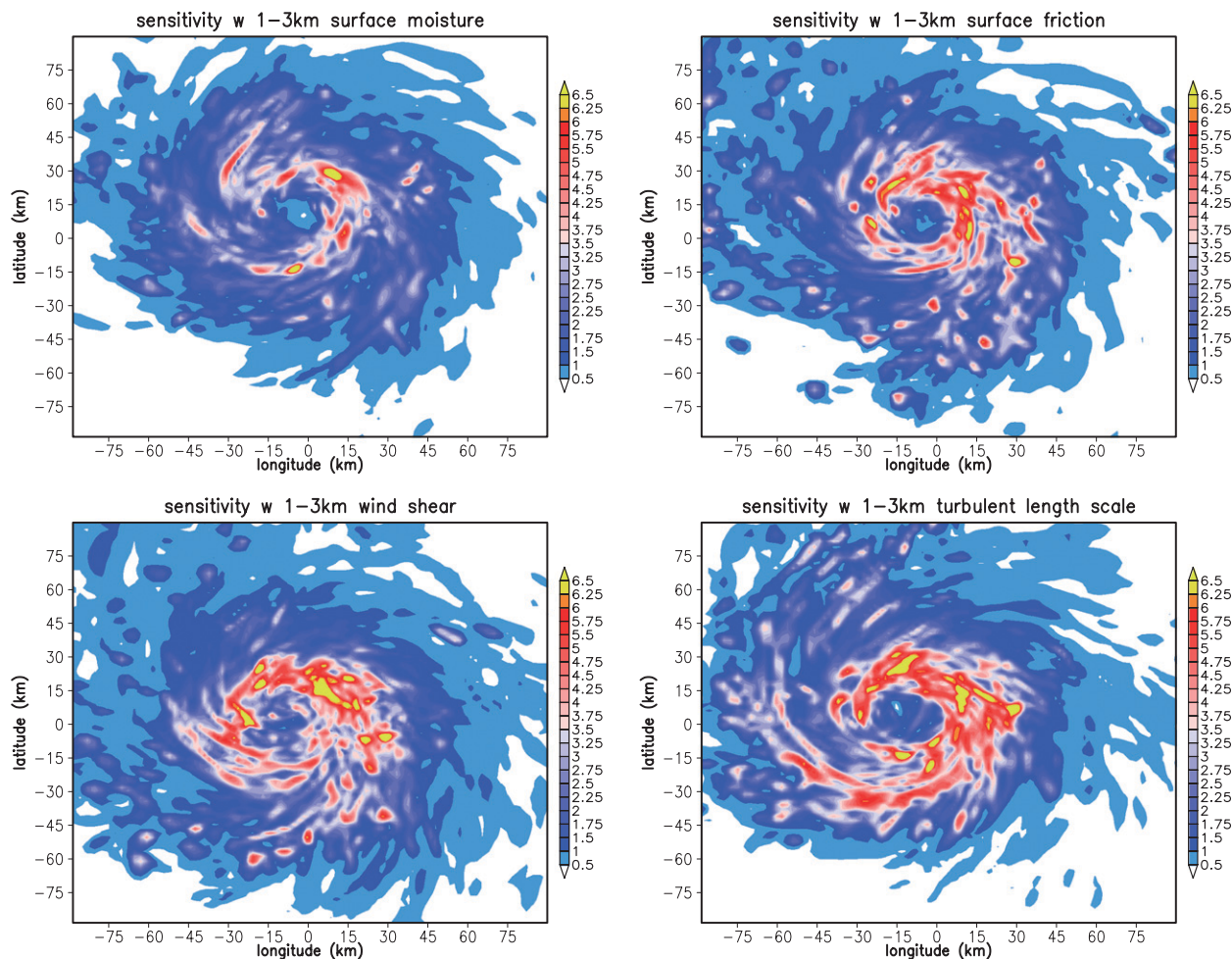


FIG. A2. Vertical wind speed sensitivity to perturbations in (top left) surface moisture, (top right) surface friction, (bottom left) wind shear, and (bottom right) turbulent length scale. Vertical wind speed sensitivity fields are shown at the eleventh hour into the simulation and averaged over 1–3 km of height.

under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy under DOE Contracts W-7405-ENG-36 and LA-UR-10-04291. Computer resources were provided both by the Computing Division at Los Alamos and the Oak Ridge National Laboratory Cray clusters; approved for public release, LA-UR-11-10121.

## APPENDIX A

### Parametric Sensitivity Experiments

A series of sensitivity experiments, using an ensemble of model simulations, is performed for the model parameters of interest. The main objective of this sensitivity study is to determine the impact that the four model parameters independently have on the structure and intensity of Hurricane Guillermo. As detailed in

section 2a(3), the four model parameters affect the intensity and/or structure of a given hurricane; hence, the sensitivity analysis will provide information of their individual impact.

#### a. Experiment setup

In this ensemble-based sensitivity study, a series of simulations were undertaken by varying a given parameter's value while fixing all other model parameters to prespecified values (Table 5). Specifically, the sensitivity of each individual parameter is studied over an ensemble of 10 members where each parameter is sampled using a Latin hypercube method over a prespecified interval (Table 1) to generate its corresponding ensemble. As with the 120-member ensemble, the hurricane model is initialized as described in section 3b, and model sensitivity is analyzed over the 6–12-h period. The sensitivity is computed by taking the

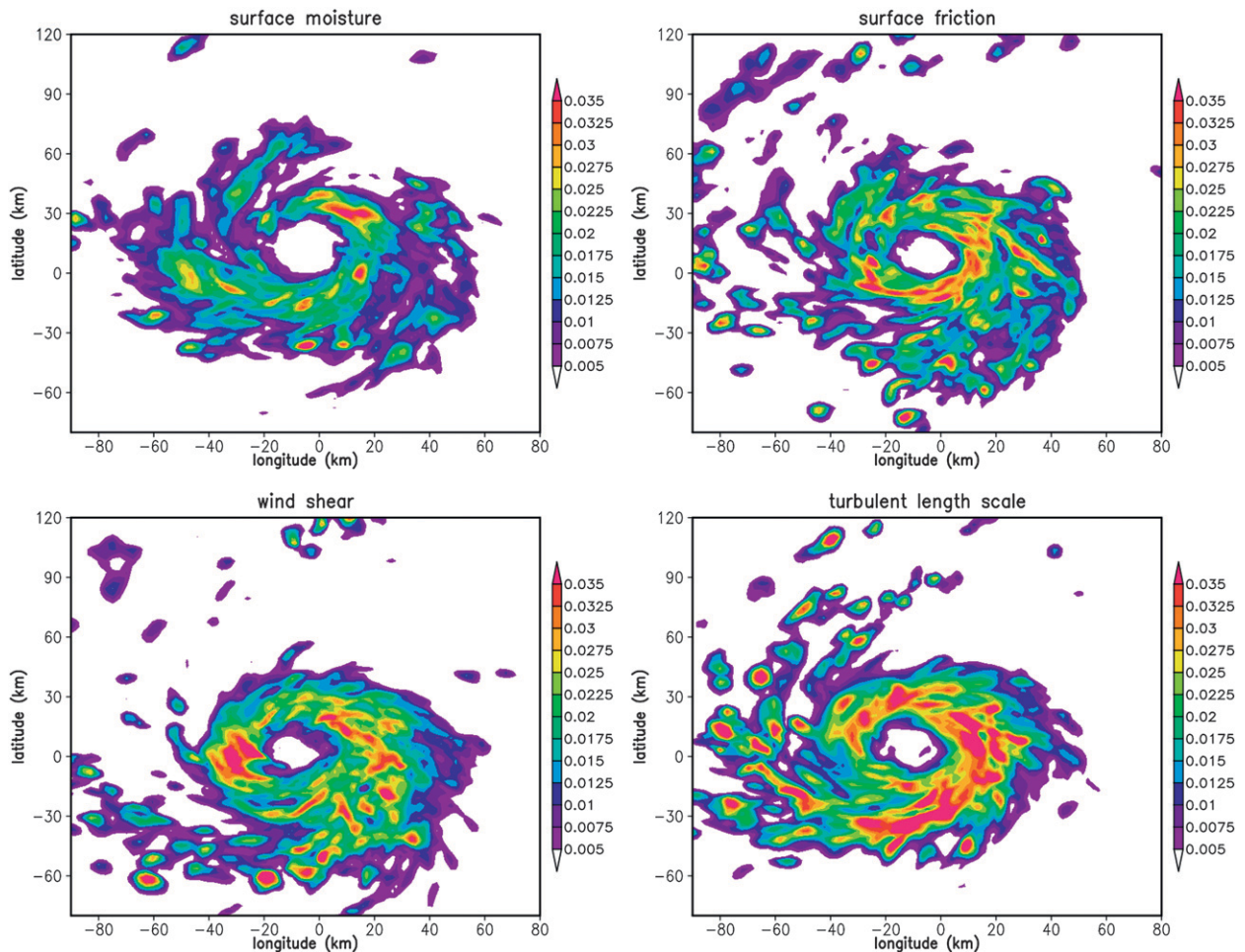


FIG. A3. Horizontal view of the latent heat sensitivity to perturbations in (top left) surface moisture, (top right) surface friction, (bottom left) wind shear, and (bottom right) turbulent length scale. Sensitivity is shown at the eleventh hour into the simulation and averaged over 0.5–12 km of height.

ensemble standard deviation of the wind speed and latent heat fields. Additionally, the minimum sea level pressure is computed for each ensemble member for analysis.

### b. Sensitivity results

Figure A1 shows the minimum sea level pressure for each of the 10 ensemble members corresponding to each of the four parameters. From the plots it is evident that the turbulent length scale and surface friction parameters have a high impact on the intensity of the storm, with this result being consistent with that of Bryan (2012) and Emanuel and Rotunno (2011). The hurricane's intensity appears to be less sensitive to the other two parameters (surface moisture and wind shear), but this may be attributed to the fact that the specific hurricane being simulated is relatively intense. However, this does not imply that these parameters do not have an impact on the structure of the simulated hurricanes.

Figure A2 shows the vertical wind speed sensitivity to the four parameters. As can be seen, the structure of the hurricane shows high sensitivity to perturbations in both surface friction and turbulent length scale, as well as the wind shear parameter, indicating that while the wind shear parameter does not significantly influence intensity, the parameter does play a rather significant role in determining the correct structure of the hurricane, and hence it must not be neglected.

Figures A3 and A4 show the sensitivity of the latent heat fields to the four parameters. As with the vertical wind speed sensitivity, the dominant parameters for latent heat are the surface friction and turbulent length scale parameters. The sensitivity to the turbulent length scale parameter is more pronounced in the bottom-right corner of the horizontal plot (Fig. A3) and the left portion of the vertical plot (Fig. A4), indicating a strong asymmetrical reaction to the

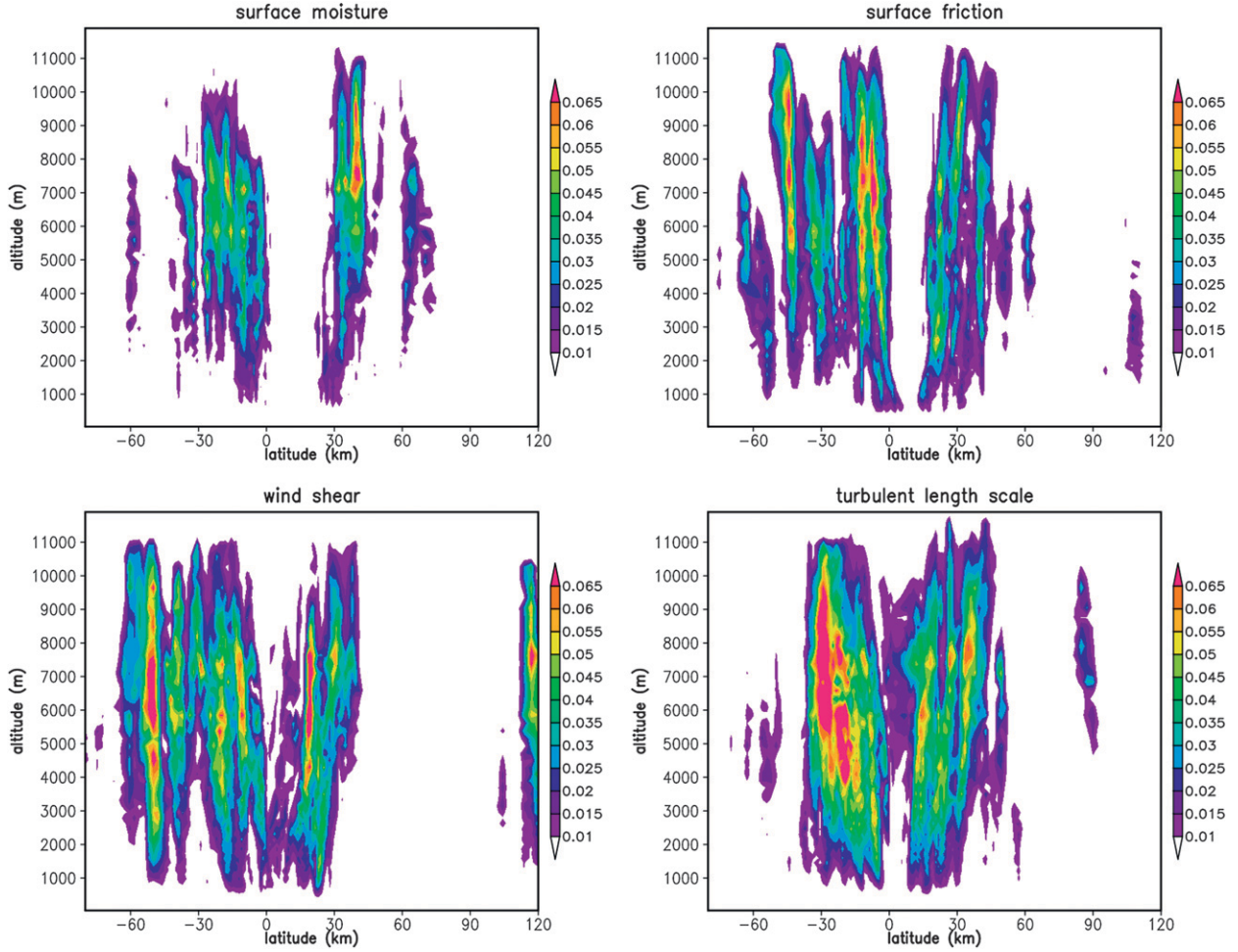


FIG. A4. Vertical view of latent heat sensitivity to perturbations in (top left) surface moisture, (top right) surface friction, (bottom left) wind shear, and (bottom right) turbulent length scale. Sensitivity shown at eleventh hour into the simulation; the longitude is set to zero (center of storm).

latent heat field from perturbations in the turbulent parameter.

Thus, this ensemble suggests that the intensity is sensitive to changes in turbulent length scale, surface friction, and surface moisture parameters, while the structure is very sensitive to changes in surface friction, turbulent length scale, and the wind shear parameters. Given that one group of parameters will influence the intensity of the storm and another (not completely disjoint) group of parameters will influence the structure, it is highly desirable to estimate simultaneously all four parameters. As noted by Yussouf and Stensrud (2012), the simultaneous estimation of all parameters will provide a more accurate set of parameter values, in the presence of model error, compared to single-parameter estimation experiments with the EnKF. This notion is consistent with the findings from the identical twin experiment.

## APPENDIX B

### EnKF Equations for Parameter Estimation

Many studies have used the EnKF data assimilation to simultaneously estimate model state and parameters. This can be achieved by using an augmented state vector where the parameters are appended at the end of the vector. In this appendix we review the EnKF equations for the simultaneous state and parameter estimation and extract the necessary equations for parameter estimation.

Let  $\mathbf{p} \in \mathbb{R}^\ell$  be a vector holding the model parameters and let  $\mathbf{x}^f \in \mathbb{R}^n$  be the model state forecast. We define the augmented state vector as

$$\mathbf{w} = \begin{pmatrix} \mathbf{x}^f \\ \mathbf{p} \end{pmatrix} \in \mathbb{R}^{n+\ell} \quad (\text{B1})$$

and let  $\mathbf{w}_i$  for  $i = 1, \dots, N$  be an ensemble of model state forecast and parameters. For a vector of  $m$  observations  $\mathbf{y}^o \in \mathbb{R}^m$  the EnKF analysis equations are given by

$$\mathbf{w}_i^a = \mathbf{w}_i + \tilde{\mathbf{K}}(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f), \quad i = 1, \dots, N \quad (\text{B2})$$

$$\tilde{\mathbf{K}} = \begin{pmatrix} \mathbf{P}^f & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{H}^T \\ 0 \end{pmatrix} (\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}, \quad (\text{B3})$$

where  $\mathbf{P}^f \in \mathbb{R}^{n \times n}$  is the model forecast covariance matrix,  $\mathbf{B} \in \mathbb{R}^{\ell \times \ell}$  is the parameter covariance matrix,  $\mathbf{C} \in \mathbb{R}^{n \times \ell}$  is the cross-correlation matrix between the model forecast and parameters,  $\mathbf{R} \in \mathbb{R}^{m \times m}$  is the observations covariance matrix,  $\mathbf{H} \in \mathbb{R}^{m \times n}$  is an observation operator matrix that maps state variables onto observations, and  $\mathbf{y}_i^o$  is a perturbed observation vector. The parameter correlation matrix is given by

$$\mathbf{B} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{p}_i - \bar{\mathbf{p}})^T, \quad (\text{B4})$$

and the cross-correlation matrix is given by

$$\mathbf{C} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i^f - \bar{\mathbf{x}}^f)(\mathbf{p}_i - \bar{\mathbf{p}})^T, \quad (\text{B5})$$

where  $\bar{\mathbf{x}}^f$  and  $\bar{\mathbf{p}}$  are the ensemble average of the model forecast and parameters, respectively.

The system of Eqs. (B2) and (B3) can be written as

$$(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})\mathbf{z}_i = (\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f) \quad (\text{B6})$$

$$\mathbf{w}_i^a = \mathbf{w}_i + \begin{pmatrix} \mathbf{P}^f & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{H}^T \\ 0 \end{pmatrix} \mathbf{z}_i, \quad (\text{B7})$$

where the vector  $\mathbf{z}_i \in \mathbb{R}^m$  is the solution of Eq. (B6). The augmented matrix in Eq. (B7) can be simplified as

$$\begin{pmatrix} \mathbf{P}^f & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{H}^T \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{P}^f\mathbf{H}^T \\ \mathbf{C}^T\mathbf{H}^T \end{pmatrix},$$

so we have the following analysis update equations for the model forecast and parameters:

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{P}^f\mathbf{H}^T\mathbf{z}_i, \quad (\text{B8})$$

$$\mathbf{p}_i^a = \mathbf{p}_i + \mathbf{C}^T\mathbf{H}^T\mathbf{z}_i. \quad (\text{B9})$$

The update Eq. (B9), together with Eqs. (B5) and (B6), form a system that estimates the model parameters for a given dataset. This is the system used in our current study.

## REFERENCES

- Aksoy, A., F. Zhang, and J. Nielsen-Gammon, 2006a: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Wea. Rev.*, **134**, 2951–2970.
- , —, and —, 2006b: Ensemble-based simultaneous state and parameter estimation with MM5. *Geophys. Res. Lett.*, **33**, L12801, doi:10.1029/2006GL026186.
- Anderson, J., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Andrejczuk, M., J. Reisner, B. Henson, M. Dubey, and C. Jeffery, 2008: The potential impacts of pollution on a nondrizzling stratus deck: Does aerosol number matter more than type? *J. Geophys. Res.*, **113**, D19204, doi:10.1029/2007JD009445.
- Annan, J. D., J. C. Hargreaves, N. R. Edwards, and R. Marsh, 2005: Parameter estimation in an intermediate complexity Earth system model using an ensemble Kalman filter. *Ocean Modell.*, **8**, 135–154.
- Arakawa, A., and V. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods Comput. Phys.*, **17**, 173–265.
- Bryan, G. H., 2012: Effects of surface exchange coefficients and turbulence length scales on the intensity and structure of numerically simulated hurricanes. *Mon. Wea. Rev.*, **140**, 1125–1143.
- Emanuel, K., and R. Rotunno, 2011: Self-stratification of tropical cyclone outflow. Part I: Implications for storm structure. *J. Atmos. Sci.*, **68**, 2236–2249.
- 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.
- , and P. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85–96.
- Gao, J., M. Xue, A. Shapiro, and K. Droegemeier, 1999: A variational method for the analysis of three-dimensional wind fields from two Doppler radars. *Mon. Wea. Rev.*, **127**, 2128–2142.
- Godinez, H., and J. Moulton, 2012: An efficient matrix-free algorithm for the ensemble Kalman filter. *Comput. Geosci.*, **16**, 565–575, doi:10.1007/s10596-011-9268-9.
- Guimond, S., M. Bourassa, and P. Reasor, 2011: A latent heat retrieval and its effects on the intensity and structure change of Hurricane Guillermo (1997). Part I: The algorithm and observations. *J. Atmos. Sci.*, **68**, 1549–1567.
- Hacker, J. P., and C. Snyder, 2005: Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL. *Mon. Wea. Rev.*, **133**, 3260–3275.
- Houtekamer, P., and H. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Hu, X.-M., F. Zhang, and J. W. Nielsen-Gammon, 2010: Ensemble-based simultaneous state and parameter estimation for treatment of mesoscale model error: A real-data study. *Geophys. Res. Lett.*, **37**, L08802, doi:10.1029/2010GL043017.
- Leonard, B., and J. Drummond, 1995: Why you should not use ‘hybrid’, ‘power-law’ or related exponential schemes for convective modeling—There are better alternatives. *Int. J. Numer. Methods Fluids*, **20**, 421–442.
- McFarquhar, G., and R. Black, 2004: Observations of particle size and phase in tropical cyclones: Implications for mesoscale modeling of microphysical processes. *J. Atmos. Sci.*, **61**, 777–794.
- Nielsen-Gammon, J., X. Hu, F. Zhang, and J. Pleim, 2010: Evaluation of planetary boundary layer scheme sensitivities

- for the purpose of parameter estimation. *Mon. Wea. Rev.*, **138**, 3400–3417.
- Reasor, P., M. Eastin, and J. Gamache, 2009: Rapidly intensifying Hurricane Guillermo (1997). Part I: Low-wavenumber structure and evolution. *Mon. Wea. Rev.*, **137**, 603–631.
- Reisner, J., and C. Jeffery, 2009: A smooth cloud model. *Mon. Wea. Rev.*, **137**, 1825–1843.
- , A. Mousseau, A. Wyszogrodzki, and D. Knoll, 2005: An implicitly balanced hurricane model with physics-based preconditioning. *Mon. Wea. Rev.*, **133**, 1003–1022.
- Sitkowski, M., and G. Barnes, 2009: Low-level thermodynamic, kinematic, and reflectivity fields of Hurricane Guillermo (1997) during rapid intensification. *Mon. Wea. Rev.*, **137**, 645–663.
- Thompson, G., R. Rasmussen, and K. Manning, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115.
- Tong, M., and M. Xue, 2008: 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**, 1649–1668.
- Torn, R. D., and G. J. Hakim, 2009: Ensemble data assimilation applied to RAINEX observations of Hurricane Katrina (2005). *Mon. Wea. Rev.*, **137**, 2817–2829.
- Yussouf, N., and D. Stensrud, 2012: Comparison of single-parameter and multiparameter ensembles for assimilation of radar observations using the ensemble Kalman filter. *Mon. Wea. Rev.*, **140**, 562–586.
- Zalesak, S., 1979: Fully multidimensional flux-corrected transport algorithm for fluids. *J. Comput. Phys.*, **31**, 335–362.
- Zhang, F., Y. Weng, J. Sippel, Z. Meng, and C. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2105–2125.
- Zou, X., Y. Wu, and P. S. Ray, 2010: Verification of a high-resolution model forecast using airborne Doppler radar analysis during the rapid intensification of Hurricane Guillermo. *J. Appl. Meteor. Climatol.*, **49**, 807–820.