Determining key environmental 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

Earth and Environmental Sciences division/Space and Remote Sensing Group, Los Alamos National Laboratory, Los Alamos, New Mexico

NOAA/Cooperative Institute for Mesoscale Meteorological Studies, Norman, Oklahoma

Stephen R. Guimond

Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida

Jim Kao

Los Alamos National Laboratory, Los Alamos, New Mexico

* *Corresponding author address:* Jon M. Reisner, Los Alamos National Laboratory, MS D401, Los Alamos, NM 87545.

E-mail: reisner@lanl.gov

ABSTRACT

In this work we present the assimilation of dual-Doppler radar observations for rapidly intensifying hurricane Guillermo (1997) using the Ensemble Kalman Filter (EnKF) to determine key environmental parameters. 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 eyewall. To reproduce this observed structure within a hurricane model, background wind shear of some magnitude must be specified; as well as turbulence and surface parameters appropriately adjusted so that the impact of the shear on the simulated hurricane vortex can be realized. To help understand this complex nonlinear interaction between shear, surface fluxes and modeled turbulence, an ensemble of simulations have been conducted during which these key parameters were varied over a certain range and then sampled via a Latin hypercube approach. A sample of parameter combinations were chosen and used to compute an ensemble of 120 model simulations. Given the relatively large computational resources needed for a large ensemble the simulations were performed using the Jaguar XT5 suppercomputer available at Oak Ridge National Laboratory. A unique preconditioned matrix-free EnKF approach used herein enabled the large amount of dual-Doppler radar data to be readily incorporated during the assimilation stage: this development was also essential for an accurate estimation of the various turbulence parameters and environmental wind shear as shown in the results.

Acknowledgments.

Start acknowledgments here.

1. Introduction

Hurricanes are among the most destructive and costliest natural forces on Earth and many communities worldwide are affected by these extreme weather events. Hence, it is important to improve hurricane forecasts by solving the uncertainties in the numerical models.

Over the past decade, significant progress has been made in improving the forecast of hurricanes. Advances in radar technology allow oceanic hurricanes to be observed at very high resolution and together with data assimilation hurricane forecast has improved significantly. The studies of Zhang et al. (2009), Torn and Hakim (2009), and most recently Zou et al. (2010) have shown the improvement of hurricane forecast with data assimilation for various models. Nevertheless, many uncertainties remain unresolved in hurricane models. A significant source of uncertainty lie in parameters that influence a number of key physical behaviour of a hurricane simulation. The main objective of the present work is to accurately estimate key model parameters, through the use of the ensemble Kalman filter (EnKF), to improve hurricane forecasts. The parameters are estimated using dual-Doppler radar observations obtained from Hurricane Guillermo (1997).

Parameter estimation remains an important factor to improve hurricane simulations. In this study there are four parameters of interest that are estimated using EnKF data assimilation. The parameters are tunning coefficients for turbulent length scale associated with the turbulent kinetic energy model, wind shear that determines the shear impact on the simulation, surface friction which impacts the intensity and structure of the simulation, and surface moisture which controls the rate of intensification of the hurricane. These parameters have a significant influence in intensity and structure of all types of hurricanes, specially

shear hurricanes. It is therefore necessary to accurately estimates these parameters in order to improve hurricane predictability and analysis.

The use of EnKF for parameter estimation was first proposed by Anderson (2001) where the parameters were included in the state vector and are simultaneously estimated using data assimilation. Annan et al. (2005a) successfully applied the EnKF for estimating 12 parameters in a low-resolution coupled atmospheric-ocean model. In a later study they implemented their methodology to a realistic GCM model with identical twin experiments Annan et al. (2005b). However, the authors found that, unlike weather prediction, climate forecast depend strongly on parameterizations rather than initial conditions. Kivman (2003) found that the EnKF performed poorly for parameter estimation with the Lorenz model due to the inability of the Kalman filter theory to properly handle highly non-Gaussian probability distributions in the parameter space. Hacker and Snyder (2005) assimilated subsurface observations into a parameterized 1D PBL model to simultaneously estimate model state and parameters and suggested that the EnKF may help mitigate model error via parameter estimation. Aksoy et al. (2009) applied the EnKF method to simultaneous estimation of up to six parameters and the model state with a two-dimensional, hydrostatic, nonrotating, and incompressible sea-breeze model. They found that the estimation of single imperfect parameters with the EnKF is successful, while the quality of estimation deteriorates when the number of estimated parameters increases. Tong and Xue (2008) used the ensemble square root filter to estimate 5 microphysical parameters, related to rain, snow, hail/graupel and bulk densities of snow and hail/graupel, for a model-simulated supercell storm. The study simulated the radar data used for assimilation and simultaneous model state and parameter estimation. For this study the EnKF is used for parameter estimation in a hurricane model.

Due to the vast amounts of data to be assimilated we implemented a matrix-free EnKF were the relevant linear system is efficiently solved by a Sherman-Morrison linear solver.

The current study explores the use of EnKF for parameter estimation of hurricane models for analysis and prediction. The case to be examined is Hurricane Guillermo (1997), a shear storm. The assimilation will cover a 6-h time window in which Guillermo experienced rapid intensification from a category 3 to a category 4 hurricane. National Oceanic and Atmospheric Administration (NOAA) Hurricane Research Division (HRD) P-3 aircraft provided dual-Doppler airborne radar observations on 10 flight legs over much of the inner-core region (within 60 km of the eye) within this time period (Reasor et al. (2009), Sitkowski and Barnes (2009)). Making this data set unique and highly reliable to describe the most critical phase of a hurricane. Hurricane Guillermo started as a tropical depression on the 30th of 2100 UTC July 1997 which formed just west of the Mexican coast. This tropical depression originated from a well-defined tropical wave that emerged from the west coast of Africa on 16 July (Mayfield (1997)). Shortly after 0600UTC August, Guillermo underwent a rapid intensification stage with a 31 m s-1 increase of the 10-m maximum sustained winds and a 54-hPa decrease in sea level pressure in the following 30h (SB09), largely exceeding the RI criteria established by Kaplan and DeMaria (2003). The National Hurricane Center Best Track data estimated that Guillermo reached its peak intensity with 70 m s-1 winds and a 919 hPa minimum central pressure near 0000 UTC 5 August. As Guillermo moved over cooler water further to the west (128W, Fig. 1), the storm rapidly weakened to a tropical storm by 0600 UTC 8 August (Mayfield (1997)).

The paper is organized as follows. Section 2 describes the analytic equations set of the hurricane model, the discretization and a brief overview of the EnKF and its matrix-free

implementation. Section 3 describes the model setup as well as the ensemble simulation setup and results. It also presents the numerical results from the parameter estimation and a reinitialized simulation with the new parameters as well as a discussion of the results. A summary and final remarks are presented in section 4.

2. Paremeter estimation model

In this section the analytical and discrete equations utilized within the hurricane simulations will be presented. The model is comprised of both an equation set describing the gas phase, i.e., the Navier-Stokes, as well as an equation set describing the bulk cloud model.

a. Navier-Stokes equation set

Even though the discrete model is formulated in a three-dimensional (3-d) generalized coordinate frame, for ease of presentation the analytical equation set will be represented in 3-d Cartesian space. In this framework the momentum equations are expressed as follows

$$\begin{aligned}\frac{\partial(u^{i'}\rho)}{\partial t}+\frac{\partial(u^{j'}u^{i'}\rho)}{\partial x^{j'}} = &-\frac{\partial p'}{\partial x^{i'}}-g(\rho'+\rho_c+\rho_r+\rho_i+\rho_s+\rho_g)\delta_{i'3}\\ &-2\rho\epsilon_{i'm'o'}\Omega_{m'}(u^{o'}-u_e^{o'})+\Phi_{i'}+\frac{\partial(\kappa\rho\tau^{i'j'})}{\partial x^{j'}},\end{aligned} \tag{1}$$

where all indices range from 1 to 3; $u^{1'}$ and $u^{2'}$ are the Cartesian gas velocities in the horizontal, $x^{1'}$ and $x^{2'}$, directions; $u^{3'}$ is the Cartesian gas velocity in the vertical, $x^{3'}$, direction; $\Omega_{1'}=0$ with $\Omega_{2'}=2\Omega sin\varphi$, and $\Omega_{3'}=2\Omega cos\varphi$ being the $x^{3'}$ and $x^{2'}$ components of the Earth's rotation axis at the latitude φ; ρ is the gas density of the air, $\rho=\rho_d+\rho_v$ with ρ_d the dry air density and $\rho_v=\rho q_v$ the water vapor density and q_v the specific humidity; $p'=$

$p-p_e$ is the pressure perturbation with p the pressure of the gas computed using the ideal gas relationship, i.e., Eq. 11 in Reisner et al. (2005), and $p_e = p_e(z)$ the environmental pressure; $\rho' = \rho - \rho_e$ is the density perturbation with $\rho_e = \rho_e(z)$ the environmental density and the bouyancy force only being applied in the vertical direction, $\delta_{i'3}$, including the additional forcing due to cloud water (ρ_c), rain water (ρ_r), ice (ρ_i), snow (ρ_s), and graupel density (ρ_g) when the ECM is active; $u_e^{i'}$ are the environmental winds in all three directions; Φ are momentum exchange terms between the gas and particles that are active during LCM calculations; $\tau^{i'j'} = \frac{\partial u^{i'}}{\partial x^{j'}} + \frac{\partial u^{j'}}{\partial x^{i'}} - \frac{2}{3}\delta^{i'j'}\frac{\partial u^{s'}}{\partial x^{s'}}$ is the strain-rate tensor; and the parameters δ and ϵ being used to simplify the writing of the gravitational and Coriolis terms (see chapter 2 of Pielke 1984). Note, constants such as the earth's gravity, g, are defined in Table 1 of Reisner and Jeffery (2010).

The energy equation is expressed as

$$\frac{\partial(\theta\rho)}{\partial t} + \frac{\partial(u^{i'}\theta\rho)}{\partial x^{i'}} = f_{energy} + \frac{\partial F_\theta^{i'}}{\partial x^{i'}}, \tag{2}$$

where θ is the potential temperature; $\theta = T(\frac{p_o}{p})^{\frac{R_d^*}{C_p^*}}$, with T the temperature of the gas, $R_d^* = (1 + 0.61q_v)R_d$, and $C_p^* = (1 + 0.94q_v)C_p$; $f_{energy} = \frac{\theta L}{TC_p^*} fdensity$ represents the release of energy during phase conversion associated with summation over all particle types, $fdensity$, produced by the bulk microphysical model and the diffusional flux of potential temperature fluctuations being defined, e.g., in the $x^{1'}$-direction as $F_\theta^{1'} = \rho\kappa\frac{\partial\theta}{\partial x^{1'}}$.

The diffusion coefficient, κ, found in Eqs. 1 and 2, as well as all other scalar equations, is determined from a TKE model with this equation being expressed as

$$\frac{\partial(TKE\rho)}{\partial t} + \frac{\partial(u^{i'}TKE\rho)}{\partial x^{i'}} = \kappa\rho\tau^{i'j'}\frac{\partial u^{i'}}{\partial x^{j'}} + g\kappa\frac{\partial\rho}{\partial x^{3'}} - \phi^{dissip}\frac{TKE^2\rho}{L_s} + \frac{\partial F_{TKE}^{i'}}{\partial x^{i'}}, \tag{3}$$

with the first term on the right hand side being the shear generation of turbulence, the second

term the buoyancy generation of turbulence, the third term the dissipation of turbulence, and the last term the diffusion of turbulence where $\kappa = 0.09 L_s \sqrt{TKE}$ with L_s the turbulence length scale.

The conservation equation for water vapor density, ρ_v, can be written as follows

$$\frac{\partial(\rho_v)}{\partial t} + \frac{\partial(u^{i'}\rho_v)}{\partial x^{i'}} = -fdensity + \frac{\partial F^{i'}_{\rho_v}}{\partial x^{i'}}. \tag{4}$$

Note, the conservation equation for ρ is nearly identical to Eq. 4, except that the turbulence diffusion term is not present.

b. Bulk microphysical model

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

$$\frac{\partial(\rho_{part})}{\partial t} + \frac{\partial[(u^{i'} - wfall_{part}\delta_{i'3})\rho_{part}]}{\partial x^{i'}} = fdensity_{part} + \frac{\partial F^{i'}_{\rho_{part}}}{\partial x^{i'}}, \tag{5}$$

whereas the conservation equation for either cloud droplet number (N_c) or ice particle number (N_i), $N_{part} = N_c, N_i$, can be written as follows

$$\frac{\partial(N_{part})}{\partial t} + \frac{\partial[(u^{i'} - wfall_{part}\delta_{i'3})N_{part}]}{\partial x^{i'}} = fnumber_{part} + \frac{\partial F^{i'}_{N_{part}}}{\partial x^{i'}}, \tag{6}$$

where $wfall_{part}$, $fdensity_{part}$, and $fnumber_{part}$ 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 (2010) 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).

c. Discrete model

The discrete model for the Navier-Stokes equation set and the bulk microphysical model closely follows what was described in section 2c of RJ. This discrete equation set formulated on an A-grid 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 (QUICKEST, Leonard and Drummond 1995) with these quantities having the possibility of being limited by a flux-corrected transport procedure (Zalesak 1979).

d. Parameters of interest

Examining vertical profiles from ECMWF 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 achieved the initialization of horizontally homogeneous potential temperature, water vapor, and total gas density fields for all Guillermo simulations. Though some uncertainty exists within the thermodynamic fields with regard to the actual environment versus the perturbed environment obtain from the ECMWF soundings, the impact of this uncertainty was deemed to be smaller than that associated with the momentum fields, i.e., it is not obvious the actual shear impacting Guillermo. So to incorporate this uncertainty into a given Guillermo simulation, the horizontal velocity fields, $u^{1'}$ and $u^{2'}$, were initialized

as follows

$$u^{1'}(x^{3'}) = \phi_{shear}[ecmwf_u(x^{3'}) + 1.5], \quad (7)$$

$$u^{2'}(x^{3'}) = \phi_{shear}[ecmwf_v(x^{3'}) - 1.5], \quad (8)$$

where ϕ_{shear} is a tuning coefficient that determines the shear impacting hurricane Guillermo within a range of 0 and 1, and $ecmwf_u$ and $ecmwf_v$ represent composite soundings calculated by averaging various soundings obtained from the ECMWF data file.

Given the delicate balance in nature that is needed for a sheared hurricane to intensify, it is not entirely obvious whether numerical models, that are necessarily limited in resolution, can accurately represent boundary 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

$$\kappa = \kappa_{surfacefriction} \tanh\left(\frac{\mathbf{V_h}}{80}\right), \quad (9)$$

where $\kappa_{surfacefriction}$ is a tuning coefficient that ranges from 0.1 to 10 m^2 s^{-1} and $\mathbf{V_h}$ is the horizontal wind speed. A no-slip boundary condition was utilized in the horizontal momentum equations ($u^{1'}$ =$u^{2'}$=0) with the magnitude of $\kappa_{surfacefriction}$ 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.

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

$$q_v^s = q_{vs}(0.75 + q_{vsurface} \tanh\left(\frac{\mathbf{V_h}}{30}\right), \tag{10}$$

where q_{vs} is the saturated vapor pressure over water and $q_{vsurface}$ is a tuning coefficient that ranges in value from 0.0 to 0.2. This term enters into surface diffusional flux calculations in discrete form as follows

$$F_{q_v}^{3'} = \kappa \frac{q_v^1 - q_v^s}{0.5\Delta x^{3'}}, \tag{11}$$

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

$$L_s^m = \phi_{turb} L_s, \tag{12}$$

where the tuning coefficient, ϕ_{turb}, ranged from 0.1 to 10.

e. Parameter estimation method: EnKF

The ensemble Kalman Filter is a Monte Carlo approach of the Kalman filter which estimates the covariances between observed variables and the state through an ensemble of model forecasts. The EnKF was first introduced by Evensen (1994) and is discussed in detail in Evensen and van Leeuwen (1996), and in Houtekamer and Mitchell (1998). A practical approach of the EnKF algorithm is given in Evensen (2003).

The EnKF for parameter estimation was first proposed by Anderson (2001). He suggested estimating the parameters by including them as part of the model state and computing the analysis of an augmented state vector. Using this augmented state vector, Tong and Xue (2008) estimated microphysical parameters of the Advanced Regional Prediction System (citation needed) through the Ensemble Square Root Filter.

For our study, only the parameters will be estimated, not the state vector of the model. Since the state is not being estimated the EnKF is directly applied to the parameters, this is, the state vector contains only the parameter values. Nevertheless, the model covariance matrix is still required for the innovation with observations.

Let $\mathbf{x}^f \in \mathbb{R}^n$ be the model state vector and $\mathbf{y}^o \in \mathbb{R}^{n_{obs}}$ a set of n_{obs} observations. Since only a limited set of imperfect observations are available, the true state of the system $\mathbf{x}^t$ cannot be determined precisely. Therefore it is more convenient to consider $\mathbf{x}^t$ as a random variable and estimate $p(\mathbf{x}^t)$, the probability distribution function (pdf), for a given set of observations. The Kalman filter provides formulas to estimate $p(\mathbf{x^t}|\mathbf{y}^o)$ given the assumptions that the model is linear, $\mathbf{y}^o$ is linearly related to $\mathbf{x}^t$, and that the background forecast of $p(\mathbf{x}^t)$ and observational error are Gaussian. Under these assumptions the Kalman filter estimates for the mean $\bar{\mathbf{x}}^a$ and covariance $\mathbf{P}^a \in \mathbb{R}^{n\times n}$ satisfy the equations

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\bar{\mathbf{x}}^f\right) \tag{13}$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}^f \tag{14}$$

where

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}, \tag{15}$$

is the Kalman gain matrix. Here $\mathbf{P}^f \in \mathbb{R}^{n\times n}$ is the model forecast covariance matrix,

$\mathbf{R} \in \mathbb{R}^{n_{obs} \times n_{obs}}$ is the observation covariance matrix, and $\mathbf{H} \in \mathbb{R}^{n_{obs} \times n}$ is an observation operator matrix that maps state variables onto observations.

For non-linear models, Evensen (1994) introduced the ensemble Kalman filter (EnKF) which uses an ensemble of forecast to estimate $\mathbf{P}^f$ and evolve it through time. Following Evensen (2003), let $\mathbf{x}_i^f$ be the n-dimensional model state vector forecast for the i^{th} ensemble member of an m-member ensemble, and define $\mathbf{y}_i^o$ as

$$\mathbf{y}_i^o = \mathbf{y}^o + \varepsilon_i \tag{16}$$

where ε_i is normaly distributed with zero mean. The Kalman filter analysis $\mathbf{x}_i^a$ for each ensemble member is given by

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}\left(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f\right) \tag{17}$$

$$\mathbf{K} = \hat{\mathbf{P}}^f\mathbf{H}^T\left(\mathbf{H}\hat{\mathbf{P}}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{18}$$

where the model covariance matrix is computed as

$$\hat{\mathbf{P}}^f = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{x}_i^f - \bar{\mathbf{x}}^f\right)\left(\mathbf{x}_i^f - \bar{\mathbf{x}}^f\right)^T. \tag{19}$$

System (17)-(18) can be reformulated as

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \hat{\mathbf{P}}^f\mathbf{H}^T\mathbf{z} \tag{20}$$

where the linear system

$$\left(\mathbf{H}\hat{\mathbf{P}}^f\mathbf{H}^T + \mathbf{R}\right)\mathbf{z} = \mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f \tag{21}$$

is solved for each ensemble member. Since the matrix in (21) is symmetric, the conjugate gradient algorithm is implemented to solve the linear system for all ensemble members.

In the case of hurricane models $n \gg n_{obs}$, where $n \sim O(10^7)$, making $\hat{\mathbf{P}}^f$ too large and impractical to compute and hold in memory. However this matrix is never explicitly formed. Instead the Kalman formulation (20)-(21) requires the action of $\mathbf{H}\hat{\mathbf{P}}^f\mathbf{H}^T$, $\hat{\mathbf{P}}^f\mathbf{H}^T$, and $\mathbf{R}$ over a vector which can be done in a matrix-free computation. Notice that the matrices can be written as

$$\hat{\mathbf{P}}^f\mathbf{H}^T = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{x}_i^f - \bar{\mathbf{x}}^f\right)\left(\mathbf{H}\mathbf{x}_i^f - \mathbf{H}\bar{\mathbf{x}}^f\right)^T \tag{22}$$

$$\mathbf{H}\hat{\mathbf{P}}^f\mathbf{H}^T = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{H}\mathbf{x}_i^f - \mathbf{H}\bar{\mathbf{x}}^f\right)\left(\mathbf{H}\mathbf{x}_i^f - \mathbf{H}\bar{\mathbf{x}}^f\right)^T. \tag{23}$$

For models where there are a vast amount of observations (large n_{obs}) solving the linear system (21) repetitively can be inefficient. To overcome this difficulty an efficient preconditioner is implemented to solve system (21) which takes advantage of the form of the matrix to efficiently compute the solution of the system. The preconditioner was developed by Maponi (2007) as a linear solver based on the Sherman-Morrison identity. In experiments with $n_{obs} \sim O(10^5)$ the conjugate gradient with the Sherman-Morrison preconditioner was observed to converge in only 2 to 3 iterations per ensemble member, making it extremely efficient to assimilate vast amounts of data.

3. Ensemble simulation and parameter estimation

a. Guillermo model 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, hurricane

Guillermo simulations have been undertaken in which a mean wind of 1.5 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

$$\Delta x_c^{1'} = 6000\sin^2(\phi_g x^*) + 1000, \quad (24)$$

$$\Delta x_c^{2'} = 6000\sin^2(\phi_g y^*) + 1000, \quad (25)$$

where $\phi_g = \frac{\pi}{Ngp_{i'}}$ determines how quickly the grid spacing changes from 7 km near the model edges to 1 km near the center with $Ngp_{i'}$ the number of grid points in either direction and x^*, y^* represent grid values for a normalized grid with a domain employing $0.5Ngp_{i'}$ grid points away from a center location in which $x^* = y^* = 0$. Like the horizontal direction, the vertical direction also employs stretching with highest resolution near the ocean boundary, approximately 50 m, and coarsest near the model top, 500 m, with 86 vertical grid points being utilized to resolve a domain extending upwards 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.

To initialize the horizontal momentum fields associated with Guillermo the dual-Doppler data obtained at 1855 UTC 2 August 1997 was interpolated to the grid with these fields being incorporated into the model via nudging over a one hour time period. Additionally, an interpolated field of latent field derived using a procedure by Guimond (**need full reference**) was also incorporated into the energy equation via nudging over the same one hour window. Next, given values of ϕ_{shear}, $\kappa_{surfacefriction}$, $q_{vsurface}$, and ϕ_{turb} generated by the Latin hypercube sampling technique the simulations are run for a total of 12 hours with 6

hours for initialization and 6 hours for comparison against the dual-Doppler radar data.

b. Ensemble Kalman Filter setup

A matrix-free implementation of the EnKF is used for the assimilation of the hurricane data for parameter estimation. Since only the parameters are to be estimated, the state vector comprises only of the parameter values for each ensemble and not the model state. The model state is used for the innovation vector for comparison with the observations.

To avoid filter divergence and compute the time distribution of the parameters it was decided to assimilate each time period without restarting the ensemble with the new analysis parameters. The reasoning behind this technique is to treat each time period where observations are available as an independent set of observations from which the parameters are estimated. In this way the time distribution of the parameters are independent.

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 assimilation.

Through prior experiments it was observed that the main driver in the structure and intensity of the hurricane model were the four key environmental parameters. It is therefore reasonable to limit the generation of the ensemble on the perturbation of these parameters. To create an ensemble with sufficient initial spread the set of parameters were perturbed using a Latin Hypercube sampling strategy with a uniform distribution. The interval for each parameter was carefully chosen based on previous experiments and from a physical intuition of their interaction with the model. Table 1 shows the initial interval for each

parameter.

The ensemble is then integrated for 6 hours to develop an approximately realistic background error covariance structure before the first observations are assimilated.

c. Data manipulation and selection procedure

The dual-Doppler data set provides 10 snapshots of dBZ reflectivity values for a 6-h window at 30 minute intervals.

- Latent heat and wind data more appropriate for the parameters to be estimated.
- Describe procedure in which latent heat and wind data was obtained from dBZ reflectivity (Guimond)
- Why are we using latent heat and wind data instead of dBZ reflectivity? Include a couple of citations and references.

In order to remove spurious data points a six-stencil weighting procedure is performed for each data point where dBZ reflectivity is greater that zero. Furthermore, the data is selected such that its absolute values is above a tolerance of $1.0e-5$.

d. Ensemble spread and structure

The spread and appropriate model statistics are critical for the success of assimilation with EnKF since the method depends on proper model and observation statistics (such as both and 1st moments?) to optimally determine the true parameter values.

Before focusing on the simulated storm structure per say, it is important to determine whether the model is able to reproduce the observed storm intensity using standard metrics. Figure 3 shows the simulated pressure traces for the 120 ensemble members (blue lines), the ensemble average (black line) and the 3-hourly observations from the NHC advisories (red line). Note that for the remainder of this work, the times shown are relative to the starting time of the analysis, namely 21600s or 6 h.The ensemble-averaged pressure trace is in remarkably good agreement with observations with a difference less than 5 hPae. The pressure trace ensemble spread around the mean about +/-25 hPa, whih is sufficient (**ref?-maybe explain why in this context, HUMBERTO**) to describe useful model statistics for data assimilation with EnKF. As mentioned earlier, likely the most interesting and unusual aspect of this hurricane case is the rapid intensification of the system while being embedded in a sheared environment, which has been long known to be detrimental to storm intensification or even formation (e.g., Gray, 1968). Because the storm was embedded in an environment characterized with vertical wind speed shear, its eyewall horizontal structure was asymmetric with a dominant wavenumber 1 mode (e.g., Reasor et al. (2009), Sitkowski and Barnes (2009)).

Figure 4 shows the ensemble averaged layer averaged vertical velocities at two different layers to highlight the vertical variation of the inner core asymmetries. On average, the model is able to reasonably capture the observed inner core asymmetries location-wise and magnitude-wise. The ability of the code to produce a good ensemble averaged provide some confidence that the initial parameter spread captures the correct dynamics of the hurricane. Because Figure 4 only provides a snapshot of the ensemble averaged horizontal structure and because the magnitudes of the storm inner core asymmetries change with height, the vertical

velocities were averaged in time (for a 2 h time period centered on the time of figure 4) and within a cylindrical ring delimited by radii 15 and 40 km. This procedure was adopted from Reasor et al. (2009).

A Fourier spectral decomposition was then carried out on this data which results are shown in figure 5. Clearly, the storm wavenumber 1 asymmetry magnitude is evident with a maximum magnitude of about half of that of the primary symmetric mode (or wavenumber zero). The simulated wavenumber 0 magnitude is in excellent agreement with the one reported by Reasor et al. (2009) who also found a value near 1.4 m/s. They found, however, a larger magnitude of the wavenumber 1 component with a value reaching 1.2 m/s. Similar analysis was carried out for vertical vorticity (not shown) and the results are, again, in very good agreement with Reasor et al. (2009) who showed that the symmetric mode was largely dominant with the remaining wavenumbers (i.e., ¡4) having maximum magnitudes less than 15% of that of the symmetric mode.

To provide a more comprehensive view of the simulated storm structure, the simulated storms ensemble averaged azimuthal structure was compared to observations and are presented for two flight legs in Figure 6 in a similar fashion than Zou et al. (2010). Overall, the averaged data exhibit reasonable agreement with observations while however notable differences ought to be highlighted. For instance, the simulated eye size is smaller by about 5-10 km with a shallower secondary circulation owing in the model to the imposition of a Rayleigh damping layer at and above 16 km AGL. The exact causes behind the disparity in eyewall size arise from many factors. Recent simulations using a higher-fidelity Lagrangian microphysical model showed that for example enhanced evaporation between the eye and eyewall had a large influence on the storms intensity and structure as pointed out in **Fierro**

et al. (2011), (citation needed, ALEX) for the case of Hurricane Rita (2005). The simulated eyewall slope, magnitudes of azimuthally-averaged tangential, radial winds and latent heating are in good agreement with observations. Additional plots were made for the remaining 8 flight legs (not shown) and displayed similar attributes.

Because the above analysis is valid for the ensemble averaged data, a similar comparison with individual members was warranted: For this comparison, the member showing closer agreement with the observed pressure trace, namely member 44 (maybe show figure of pressure traces?), was chosen and the results are shown in figure 7. While the ensemble averaged data seem to show better overall agreement with observation, this figure highlights the existing spread of simulated structures that sometime exhibit large differences with observations: The latent heat fields in particular exhibit a double maximum structure in contrast to observations. Clearly the simulation produces also a stronger secondary circulation and upper level outflow for the same intensity as defined by minimum surface pressure.

In the next step of the analysis, namely the assimilation, the observations were thinned out by including those observations whose latent heat absolute values were greater or equal to a certain tolerance. Table 2 show the number of observations considered for different tolerances at 6 hours.

An assimilation experiment was performed for each tolerance resulting in the assimilation of different number of observations. Figure 8 shows the evolution of the average ensemble parameter estimate after assimilation as a function of number of observations assimilated at 14100 seconds. The blue dots represent the ensemble average parameter value and the green lines represent the standard deviation, with the x-axis showing the number of observations assimilated and the y-axis the parameter value. The first value on each plot, zero value on

the x-axis, represents the ensemble average parameter value of the initial spread.

It is clear from figure 8 that as the number of assimilated observations increases there is a transition where the analysis veers off course. As the assimilation reaches information saturation with observations the analysis resumes a more stable course for the approximation of the parameters. This behaviour can be attributed to either the critical role that the number of observations can play in the assimilation or the potential impact that particular observations have on the assimilation. In either case an analysis greatly benefits from assimilating a large enough number of observations that capture the essential information needed for an accurate estimation of the parameters.

e. EnKF performance and analysis parameters

The EnKF experiments begin assimilation at $t = 0$h (relative to $t = 6$h), corresponding to the first instance of data available, and assimilates another 9 time periods at 30 min increments. An analysis parameter is obtained for each of the 10 time periods without restarting the ensemble.

For comparison the parameter values were estimated using derived 3D latent heat fields and derived 3D wind data. Figure 9 shows the time distribution of the ensemble-averaged parameter values after assimilation of the observations every 30 min. The observation mask for both latent heat and wind data is defined by including the observation locations where absolute latent heat is greater or equal to 1.0e 5 deg K/ h. This ensures that the same observation location is selected for all the data sets available.

Figure 10 shows the average and standard deviation of the time distributed analysis

parameter values for latent heat data (blue dot) and wind data (red dot). These were computed from the ensemble average analysis parameter. The vertical lines indicate the time variance of each ensemble estimate for each data set. Both estimates are within variance of each other, indicating that the analysis parameters from the assimilation of latent heat data and wind data are withing a range of tolerance. Moreover, because latent heat can be used as a surrogate/proxy for lightning, assimilation of lightning data obtained by either VLH/VHF sensors can provide accurate estimates of key parameter values and model calibration for hurricanes. Not only lightning can be more readily detected over a wider range with high temporal resolution than radar reflectivity or wind, but the latter can be used to fill in the void many platforms such as radars suffer from. This advantage is especially desirable over data-parsed areas over oceans, where all hurricanes form and intensify.

Three deterministic forecasts are initialized using the resulting parameter values of the EnKF analysis for each assimilated type of data set, namely, wind, latent heat and radar reflectivity.

The time average of the analysis parameters is computed over the 10 time periods of assimilation for latent heat, wind, and dBZ data. A new simulation for the three sets of estimated parameters was carried out to obtain an improved model solution for its respective data.

The results are shown in Figures 11 and 12 which compares azimuhtal structures between the observations and restarted model simulation (contours), using estimated parameters from assimilation of latent heat and wind fields, respectively. The assimilation of latent heat results in an overall weaker storm with azimuthally averaged tangential winds rarely exceeding 50 m/s in contrast to observations (Fig. 11). Clearly better results are obtained

when using wind data (Fig. 12). Assimilating wind data, in contrast to latent heat, favors a stronger storms with a more well defined stronger secondary circulation and tangential wind profile, In particular the azimuthal location of the maximum latent heating is well captured by the wind field...

4. Summary and conclusions

APPENDIX

REFERENCES

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part I: Storm-Scale Analyses. *Mon. Wea. Rev.*, **137**, 1805–1824.

Anderson, J., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Annan, J. D., J. C. Hargreaves, N. R. Edwards, and R. Marsh, 2005a: Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Modelling*, **8**, 135 – 154.

Annan, J. D., D. J. Lunt, J. C. Hargreaves, and P. J. Valdes, 2005b: Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter. *Nonlin. Processes Geophys.*, **12**, 363–371.

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.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, **53**, 343–367.

Evensen, G. 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.

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.

Kaplan, J. and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea. Forecasting*, **18**, 1093–1108.

Kivman, G. A., 2003: Sequential parameter estimation for stochastic systems. *Nonlinear Processes in Geophysics*, **10 (3)**, 253–259.

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. Num. Meth. Fluids*, **20**, 421–442.

Maponi, P., 2007: The solution of linear systems by using the Sherman–Morrison formula. *Linear Alg. Appl.*, **420**, 276–294.

Mayfield, M., 1997: Preliminary report: Hurricane Guillermo 30 July-15 1997, NHC/NOAA.

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.

Pielke, R., 1984: *Mesoscale Meteorological Modeling.* Acedemic Press, San Diego.

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 J. Jeffery, 2010: A smooth cloud model. *Mon. Wea. Rev.*, **137**, 1825–1843.

Reisner, J., V. Mousseau, A. Wyszogrodzki, and D. Knoll, 2005: A fully implicit 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 microphyiscs 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). *Monthly Weather Review*, **137**, 2817–2829.

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.

List of Tables

parameter	interval
surface moisture parameter	$[0.05, 0.2]$
wind shear parameter	$[0.1, 1.0]$
turbulent length scale	$[0.1, 10.0]$
surface friction parameter	$[0.1, 10.0]$

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

TABLE 2. Number of observations for a given minimum tolerance of the absolute value of latent heat for observations at 6 hours.

tolerance	Number of obs.
0.080	94
0.070	291
0.060	516
0.050	1051
0.040	2131
0.030	4395
0.020	10942
0.018	13355
0.016	16609
0.014	20790
0.012	26475
0.010	33957

List of Figures

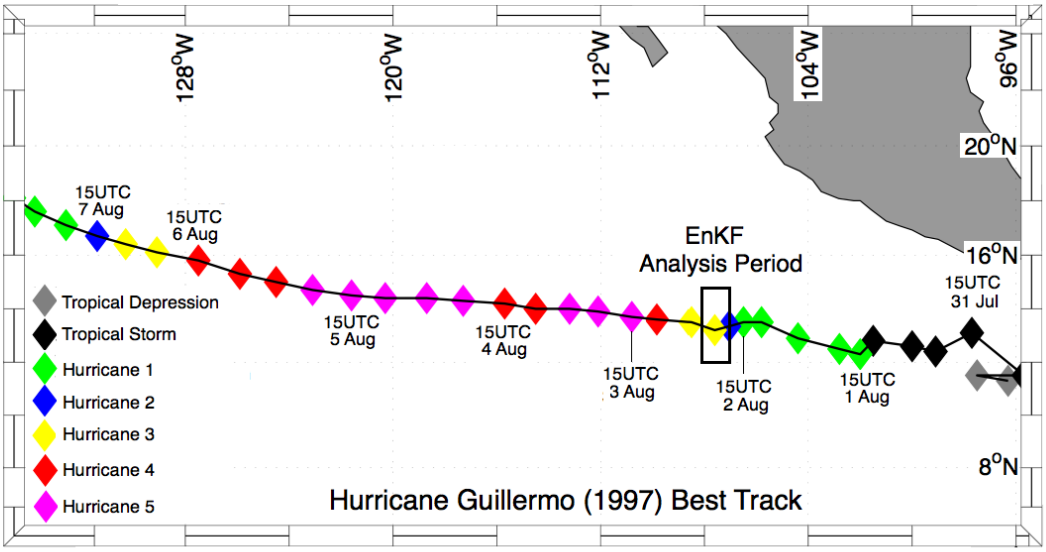

FIG. 1. Best Track for Hurricane Guillermo (1997). Hurricane Intensity is color-coded based on the Saffir-Simpson scale with legend shown on the bottom left of the figure. Data courtesy of the Tropical Prediction Center (TPC), NOAA. The EnKF analysis period is denoted by the small black rectangle.

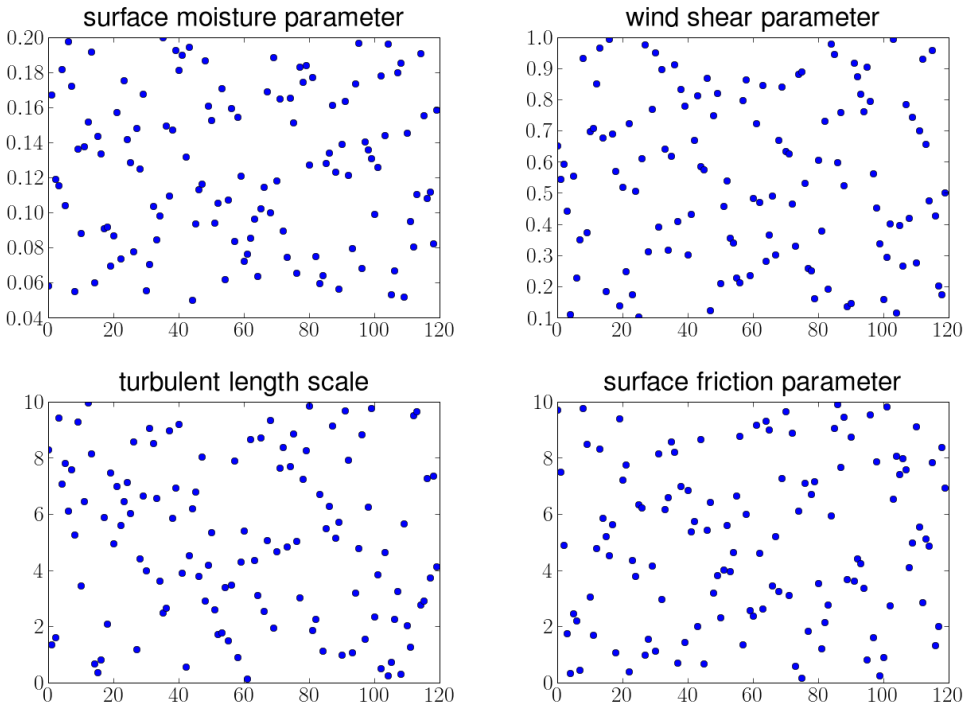

FIG. 2. Initial parameter spread for 120 samples obtained with the Latin Hypercube sampling strategy using the limits in table 1

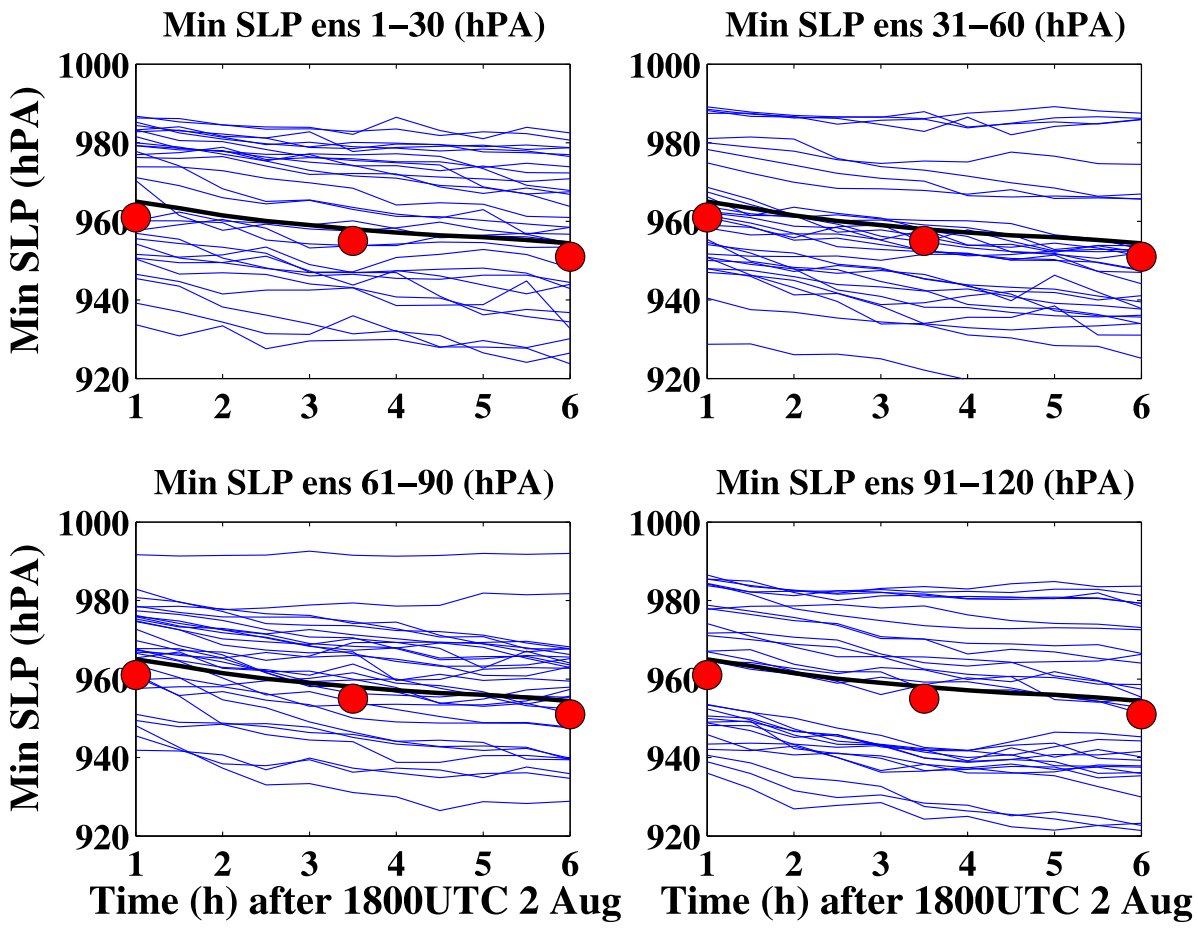

FIG. 3. Minimum sea level pressure for each ensemble member (blue), ensemble average (black), and observations (red)

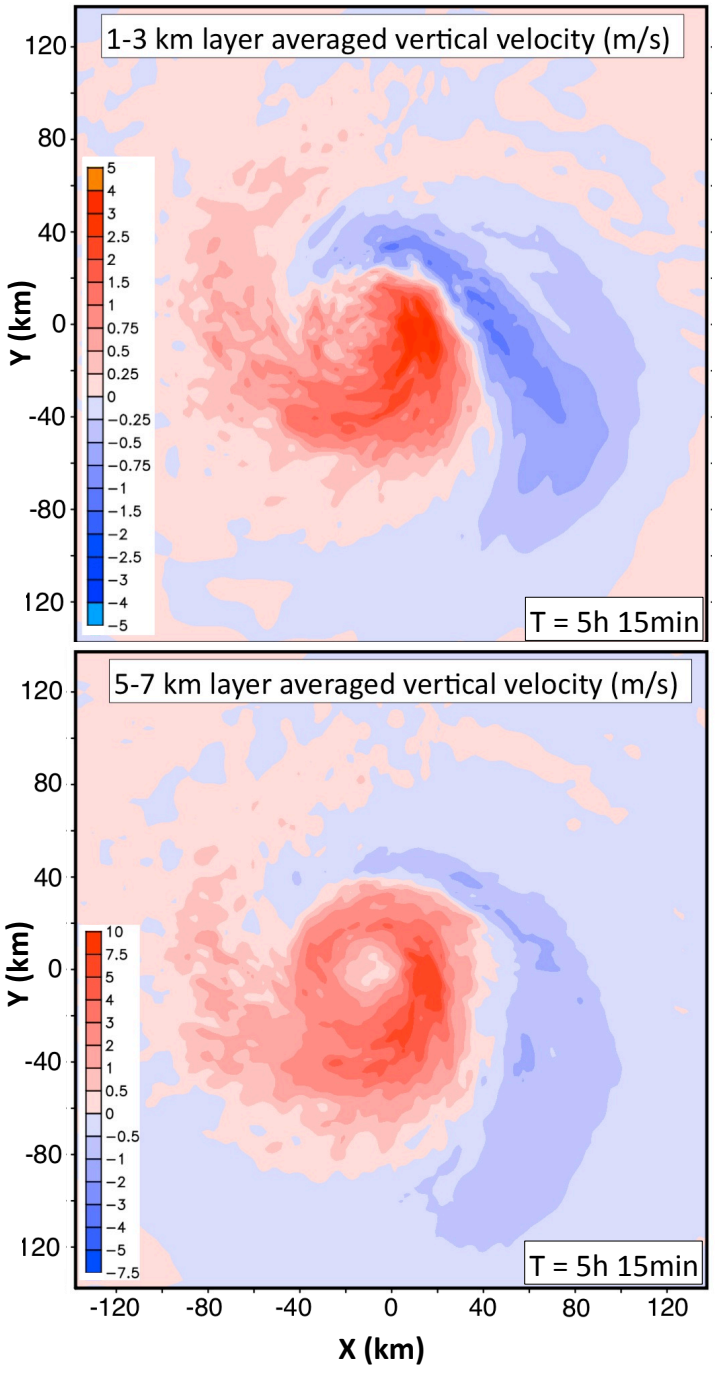

FIG. 4. Ensemble average W field at 39600 seconds.

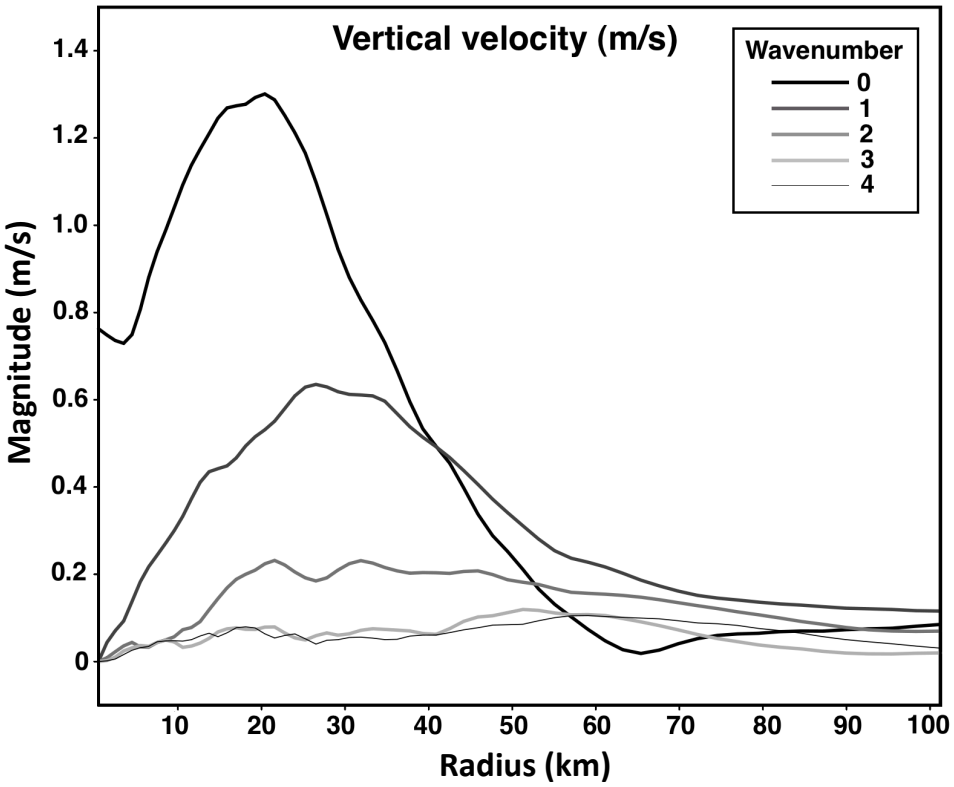

FIG. 5. Fourier analysis ensemble average of W field

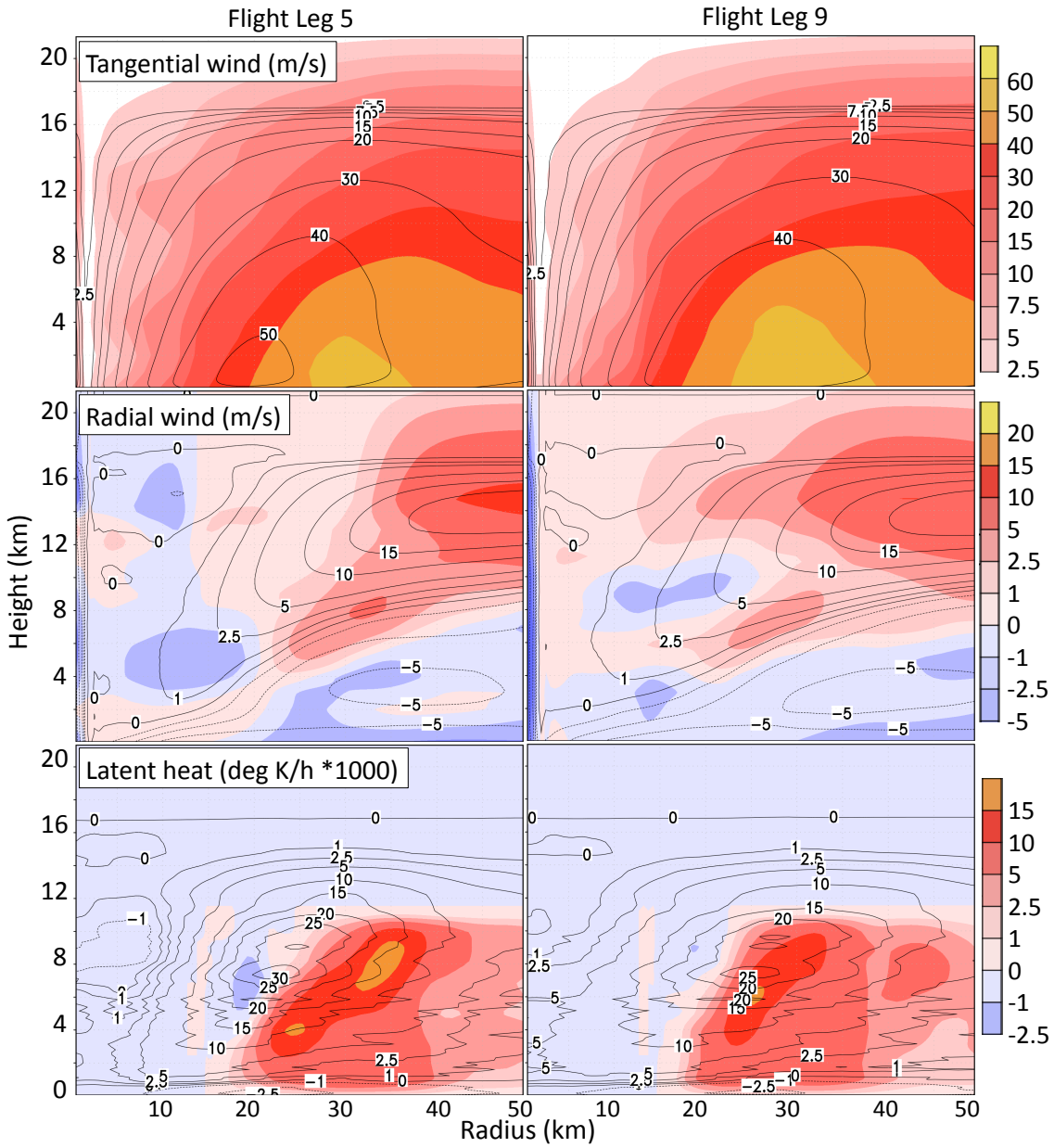

FIG. 6. Comparison between the ensemble-averaged azimuthally-averaged profile and observations for tangential winds (top), radial winds (center), latent heat (bottom) for 21600 seconds (left panels) and 28800 seconds.

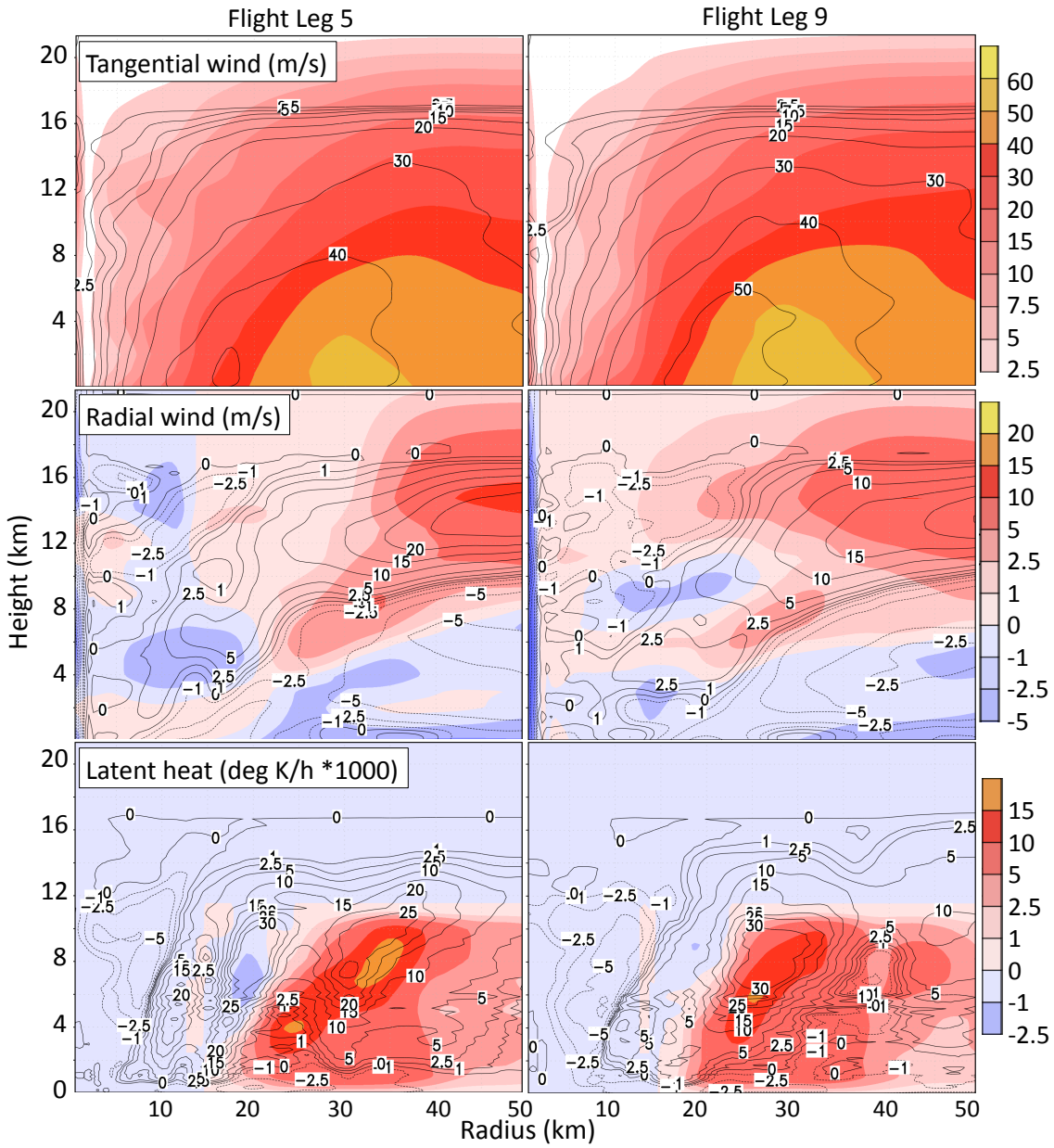

FIG. 7. Comparison between ensemble member with the best minimum SLP, member 44, and latent heat and winds observations. Azimuthally-averaged profile and observations for tangential winds (top), radial winds (center), latent heat (bottom)

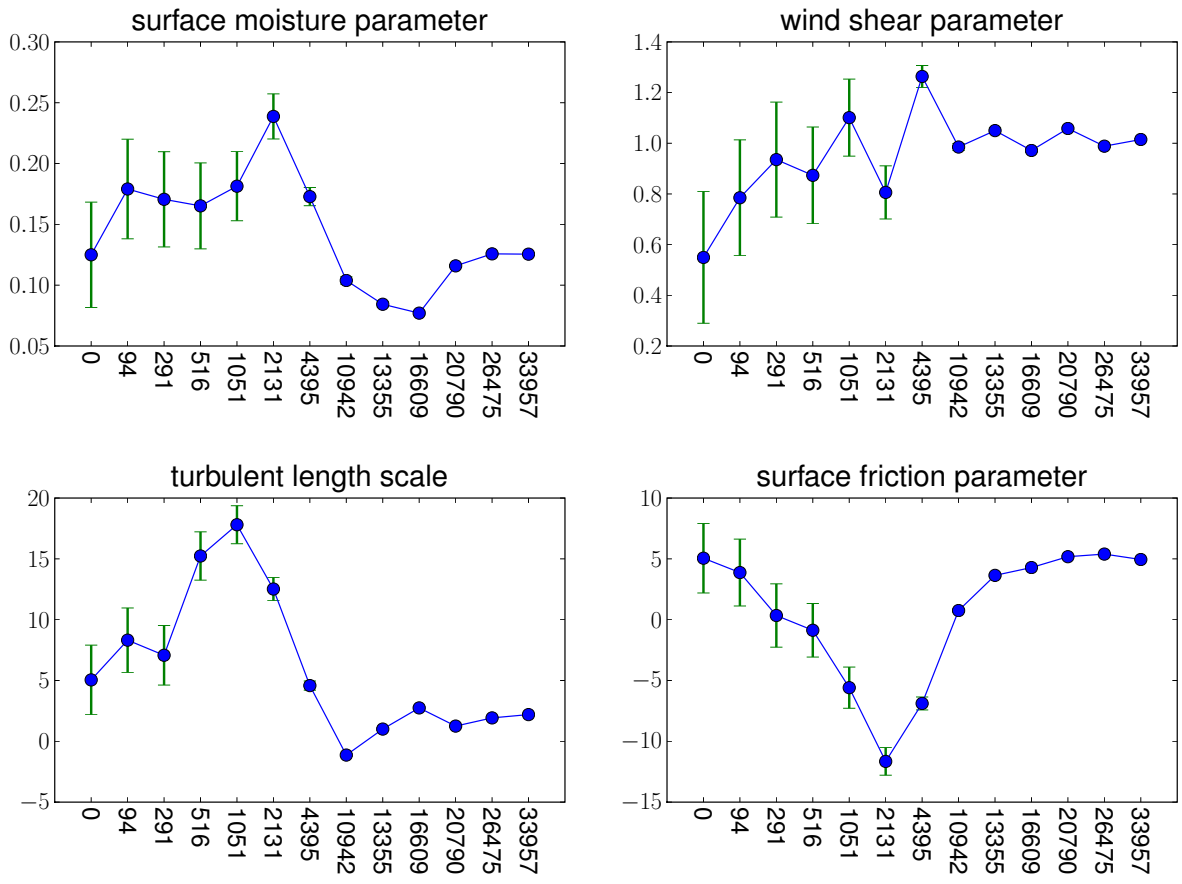

FIG. 8. Parameter analysis with EnKF as a function of number of observations assimilated at 14100 seconds. The blue dots are the ensemble average parameter and the green vertical lines indicate the ensemble variance of the parameters.

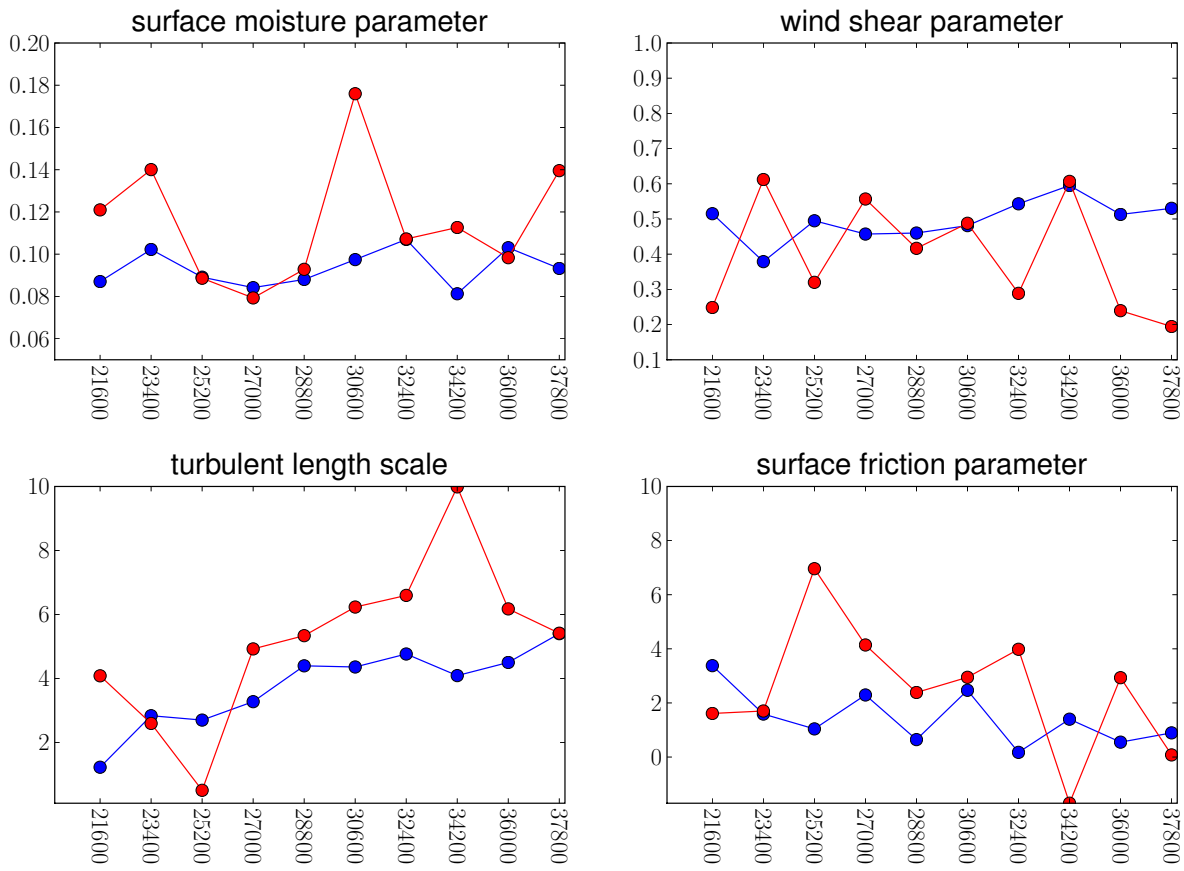

FIG. 9. Time distribution of ensemble average parameter estimates without analysis restart, where the blue line represents estimates obtained with assimilating latent heat data and red with wind data.

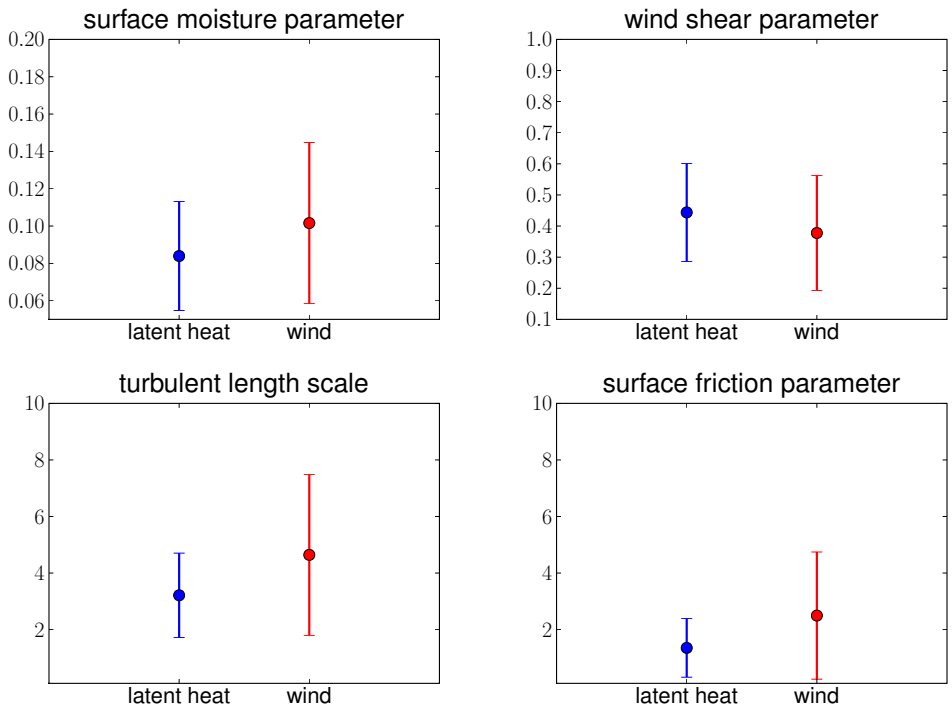

FIG. 10. Analysis parameters averaged over time for each ensemble member for heat data (blue dot) and wind data (red dot). The vertical lines from the blue and red dots indicate the variance of the analysis within each parameter estimate.

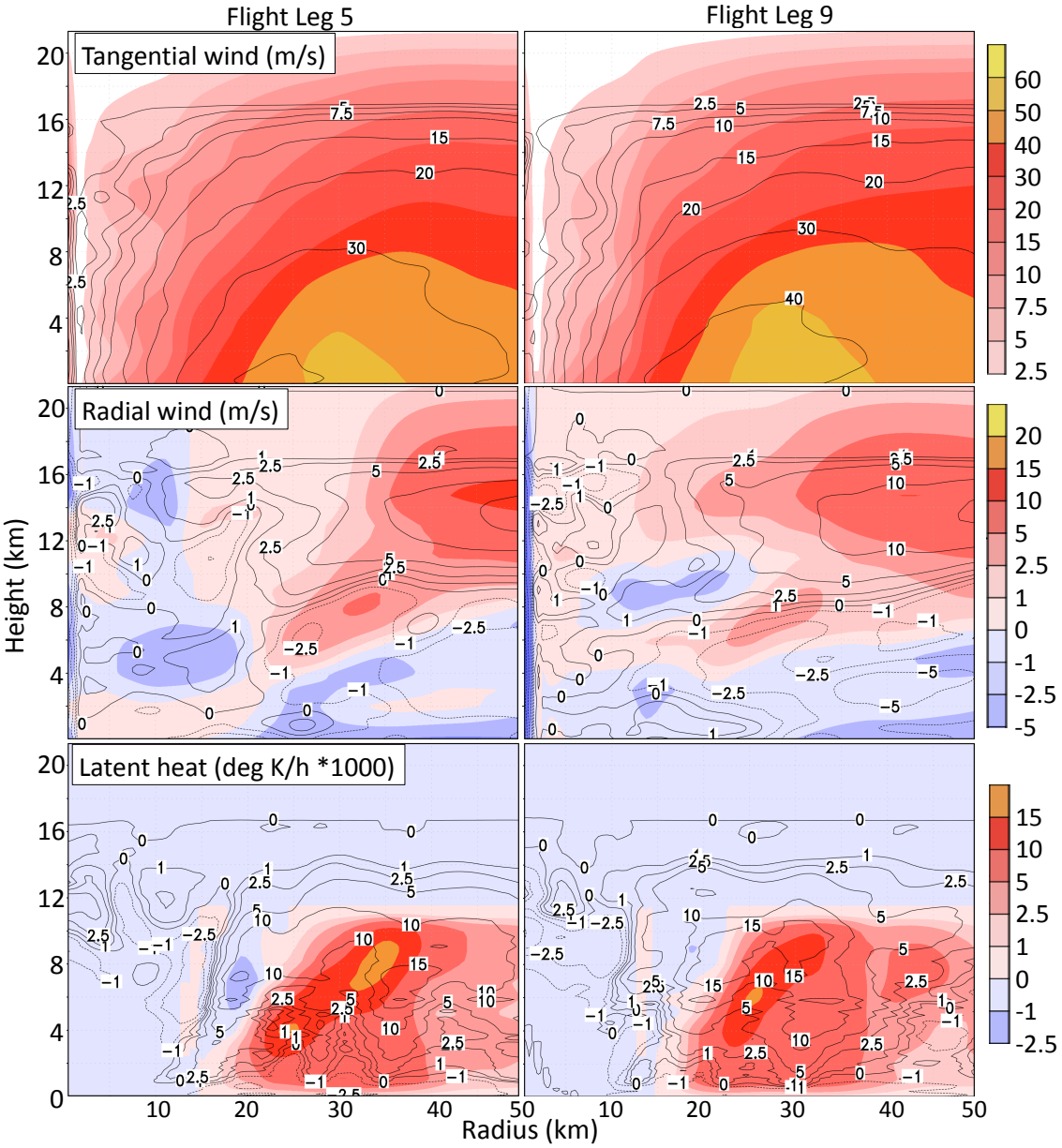

FIG. 11. Comparison of the azimuthally-averaged profile between restarted model simulation (contours), using estimated parameters from latent heat, and observations (color shaded). Plots for tangential winds (top), radial winds (center), latent heat (bottom) for 21600 seconds (left panels) and 28800 seconds (right panels).

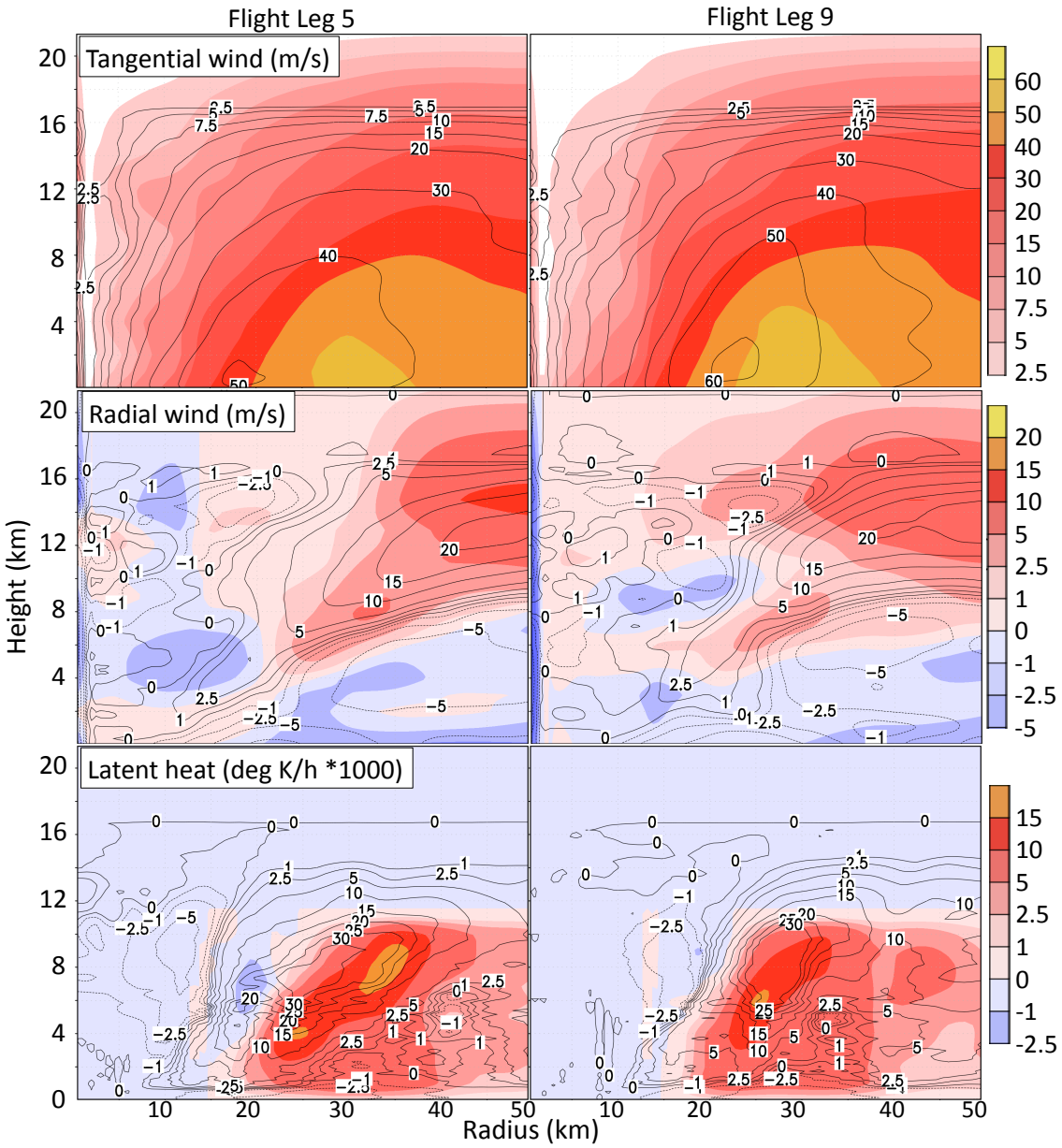

FIG. 12. Comparison of the azimuthally-averaged profile between restarted model simulation (contours), using estimated parameters from wind data, and observations (color shaded). Plots for tangential winds (top), radial winds (center), latent heat (bottom) for 21600 seconds (left panels) and 28800 seconds (right panels).