TROPICAL CYCLONE DATA ASSIMILATION: EXPERIMENTS WITH A COUPLED GLOBAL-LIMITED-AREA ANALYSIS SYSTEM

A Dissertation

by

CHRISTINA R. HOLT

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Istvan Szunyogh
Committee Members,	Robert Korty
	R. Saravanan
	Mikyoung Jun
Head of Department,	Ping Yang

May 2014

Major Subject: Atmospheric Sciences

ABSTRACT

This study investigates the benefits of employing a limited-area data assimilation (DA) system to enhance lower-resolution global analyses in the Northwest Pacific tropical cyclone (TC) basin. Numerical experiments are carried out with a global analysis system at horizontal resolution T62 and a limited-area analysis system at resolutions from 200 km to 36 km. The global and limited-area DA systems, which are both based on the Local Ensemble Transform Kalman Filter algorithm, are implemented using a unique configuration, in which the global DA system provides information about the large-scale analysis and background uncertainty to the limited-area DA system.

In experiments that address the global-to-limited-area resolution ratio, the limited-area analyses of the storm locations for experiments in which the ratio is 1:2 are, on average, more accurate than those from the global analyses. Increasing the resolution of the limited-area system beyond 100 km adds little direct benefit to the analysis of position or intensity, although 48 km analyses reduce boundary effects of coupling the models and may benefit analyses in which observations with larger representativeness error are assimilated. Two factors contribute to the higher accuracy of the limited-area analyses. First, the limited-area system improves the accuracy of the location estimates for strong storms, which is introduced when the background is updated by the global assimilation. Second, it improves the accuracy of the background estimate of the storm locations for moderate and weak storms. Improvements in the steering flow analysis due to increased resolution are modest and short-lived in the forecasts. Limited-area track forecasts are more accurate, on average, than global forecasts, independently of the strength of the storms up to five days. This forecast

improvement is due to the more accurate analysis of the initial position of storms and the better representation of the interactions between the storms and their immediate environment.

Experiments that test the treatment and quality control (QC) methods of TC observations show that significant gainful improvements can be achieved in the analyses and forecasts of TCs when observations with large representativeness error are not discarded in the online QC procedure. These experiments examine the impact of assimilating TCVitals SLP, QuikSCAT 10 m wind components, and reconnaissance dropsondes alongside the conventional observations assimilated by NCEP in real time. Implementing a Combined method that clips the special TC observations via Huberization when multiple observation types are unavailable, and keeping the TCVital observation when other special observations are present, showed significant systematic improvements for strong and moderate storm analyses and forecasts.

DEDICATION

I dedicate this dissertation to Mr. Jason Cole. You inspired me to set unreachable goals and make them a reality. Your advice and support has proven priceless. You have touched so many lives in such a small town and the world is a better place because of you.

ACKNOWLEDGEMENTS

I would like to thank our collaborators at AER, Inc., Ross Hoffman and Mark Leidner. The comments of two anonymous reviewers, as well as Ross Hoffman, were very helpful and made thought provoking suggestions that ultimately led to a better presentation of our findings in the first half of the work. The work was supported by ONR grant N000140910589.

My officemates Michael Herrera and Michael Battalio often provide much appreciated moral and technical support on a daily basis near the end of the process. Gyorgyi Gyarmati has provided an extraordinary amount of technical support through the years, and was there as a friend in the office whenever I needed it.

To the many graduate students who have come and gone while I've continued on, I appreciate all the help, guidance, and motivation you have knowingly, and unknowingly, provided. I owe my sanity to Kelly Keene and Christine Arnold, who are always there when it was time to make the hardest decisions, and when I needed a moment away from graduate school reality.

To my family, thank you for the continued support through the many many years I have been working to achieve my goals.

To Christopher Williamson, words cannot describe how grateful I am that you have been there with me through the entire journey. You are truly my best friend on this planet. I thank you for the sacrifices you have made for me, and the support you have provided that has led us down our path together. You are everything.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION*

Improvements for tropical cyclone (TC) track forecasts over the past decade have been attributed to increasing model resolution, improved data assimilation techniques, and the rapid increase in the number of routinely assimilated observations over oceans (Rappaport et al. 2009). This dissertation will outline the use of novel techniques for data assimilation that are consistent with those outlined in Rappaport et al. (2009) to improve TC track and intensity analyses and forecasts.

The data assimilation approaches for TCs that will be implemented in this study include 1) an ensemble based coupled global-limited-area analysis system, which is used to obtain a higher resolution analysis in a particular area of interest (specifically over a the NW Pacific TC basin), and 2) a method to make the initial ensemble robust to large observation error and observation innovation (the difference between an observation and the background estimate of the observed state). Both approaches will be implemented for the first time with an EnKF in a real world scenario.

The research analysis/forecast system described in Holt et al. (2013) to assimilate the observations is used and is based on the Local Ensemble Transform Kalman Filter (LETKF) algorithm, the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) model and Regional Spectral Model (RSM). The somewhat unusual choice of the RSM as the regional model is motivated by its consistency with the GFS in both dynamics and parameterizations. Since it is a perturbation model, it also benefits from receiving information from the global base fields at every grid point, instead of only through the lateral boundaries, like other

*Parts of Section 1 are reprinted with permission from "Can a Moderate-Resolution Limited-Area Data Assimilation System Add Value to the Global Analysis of Tropical Cyclones" by C. Holt, I. Szunyogh, and G. Gyarmati, 2013. Monthly Weather Review, 141,1866-1883, Copyright 2013 by the American Meteorological Society.

limited-area models.

While it may seem like an obvious choice, it is important to note that the superiority of a fully cycled limited-area analysis to a lower resolution global analysis is not self-evident. For instance, NCEP has struggled for many years to develop a limited-area data assimilation system that would provide a better analysis for their limited-area model than their global assimilation system. Their current limited-area data assimilation system is based on a partial cycling strategy: a cold start of the regional data assimilation system is started from a global analysis 12-h prior to the actual analysis time and cycled over four 3-h analysis time windows to produce the limited-area analysis (Rogers et al. 2009). The most likely explanation for the difficulties with fully cycling the limited-area analysis is that propagating information about the large scale flow through the lateral boundaries for an extended time is a challenging task. In addition, the gainful assimilation of satellite radiance observations, which has shown promising results for the global setting, is still an open problem of limited-area data assimilation, because, among other issues, no effective algorithm currently exists to estimate the bias in the radiance observations within the framework of a limited-area data assimilation system (Schwartz et al. 2012).

Our ensemble-based data assimilation approach provides several potential benefits for TC data assimilation. The coupling from the global to the limited-area component provides a simple framework to introduce information about the large-scale uncertainty into the limited-area data assimilation process. The benefits of flow dependent error statistics are often cited as one of the most important attributes of the ensemble-based approach for TCs, e.g. Torn and Hakim (2009); Torn (2010); Hamill et al. (2011); Zhang et al. (2011). Lastly, the ensemble-based approach is thought to be an efficient filter of imbalances that could be introduced by the assimilation of observations.

While improvements have been made to track forecasts for TCs, the improvement

of intensity forecasts has been much more modest, which is not surprising considering that the eyewall radius of a TC is about $25 - 50$ km (Kimball and Mulekar 2004) and, 4 to 10 grid points are needed to resolve a flow feature such as the eye of a TC (e.g., Fiorino and Elsberry (1989a,b); Grasso (2000); Skamarock (2004)). The 5 km nominal resolution that would be required, at minimum, to capture the structure of the eye of a TC is expected to remain unattainable in an operational real-time global analysis/forecast system until about 2020 (Källén 2012). Given that an appropriate limited-area-to-global analysis resolution ratio can be found, this challenge could be overcome without pushing the computational limits of significantly increasing global model resolution. Because the goal of their study is to obtain basin-wide coverage for a statistically significant sample of TCs, the global model resolution remains fixed at a low resolution and the limited-area model resolution varies with low-to-moderate grid spacing. At these resolution ranges, only a warm-core vortex can be expected, and as a result representativeness error will be significant in the experiments since the model grid cannot accurately resolve the flow features occurring in the atmosphere.

Both operational and research analysis systems typically employ algorithms to initialize and/or reposition the TC, however, the system in the current study uses only the assimilation of observations to update the TC position, intensity, and flow field. With potentially large representativeness error, along with large observation innovations that can result from poor estimates of background location, many observations that are meant to update the TC estimates of position and intensity are discarded in the online quality control procedure (an algorithm that checks the observations against the background estimate and discards statistical outliers). An offline QC procedure discards any observations that have NCEP flags large than 4 before they are subjected to the online QC procedure. In this study, efficient methods for the assimilation of observations with inherently large observation

innovations are tested. In particular, the goal is to obtain an efficient combination of the quality control procedures and the definition of the observation errors. The focus will be on three types of observations: QuikSCAT 10 m wind retrievals, Tropical Cyclone Vitals (TCVitals) minimum sea level pressure (SLP), and dropsondes from TC reconnaissance programs. Each of these observation types comes with its own set of challenges, but together they provide examples of observations of several atmospheric variables at a variety of vertical levels, from the surface through the mid-troposphere.

Experiments and results that focus on the limited-area-to-global resolution ratio are discussed in depth in Section 2, while Section 3 covers the advanced observation quality control techniques and results. Section 4 is a summary and discussion of the study.

2. A COUPLED-GLOBAL-LIMITED-AREA DATA ANALYSIS/FORECAST SYSTEM*

2.1 Experiment Design

2.1.a Analysis/Forecast System

The data assimilation system is an implementation of the LETKF algorithm (Ott et al. 2004; Hunt et al. 2007) on the NCEP GFS model (Szunyogh et al. 2005 and 2008) and the NCEP RSM (Merkova et al. 2011). The GFS is a spectral-transform model, and we integrate it using a triangular truncation with a cut-off wave number of 62 and 28 vertical sigma levels (T62L28).

The RSM is a nested limited-area version of the GFS model (Juang and Kanamitsu 1994), which uses a one-way nesting; that is, the global solution affects the regional solution, but the regional solution has no affect on the global solution. The vertical levels in the RSM are identical to the 28 sigma levels of the GFS. The horizontal resolution of the RSM in our experiments, which is defined by the grid spacing, varies from 200 km to 36 km. The particular implementation of the analysis/forecast system used here was first tested on winter storms over the U.S. and is described by Merkova et al. (2011). To obtain the LETKF analyses, all observations that were assimilated by NCEP in real-time in Summer 2004 are used, excluding satellite radiance observations and TCVitals information. (An example of the locations of such observations for a typical analysis cycle is shown in Fig. A.1.) In addition, the procedure to relocate the TCs in the background (first guess of the analysis), which

*Section 2 is reprinted with permission from "Can a Moderate-Resolution Limited-Area Data Assimilation System Add Value to the Global Analysis of Tropical Cyclones" by C. Holt, I. Szunyogh, and G. Gyarmati, 2013. Monthly Weather Review, 141,1866-1883, Copyright 2013 by the American Meteorological Society.

has been used in the operational systems of NCEP (Liu et al. 2000), is not used in our system. Hence, adjustment of the position of the storm by the data assimilation is due entirely to the assimilation of observations. In this Chapter, no consistent set of observations that provide direct information about the TC position or intensity are used. Occasionally reconnaissance dropsondes make it into the NCEP operational files and are assimilated when available.

2.1.b LETKF Data Assimilation

The LETKF, similar to other formulations of the Ensemble Kalman Filter, generates an ensemble of forecasts to estimate the background and covariance matrix of the background error. To be precise, the state update equation is

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{K}\delta\mathbf{y}, \tag{2.1}$$

where $\bar{\mathbf{x}}^a$ is the analysis, $\bar{\mathbf{x}}^b$ is the background, which is computed by taking the mean of the background ensemble, $\mathbf{K}$ is the ensemble-based estimate of the Kalman gain matrix $\mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}$, and $\delta\mathbf{y} = \mathbf{y}^o - H(\bar{\mathbf{x}}^b)$ is the observation innovation, where $\mathbf{y}^o$ is the vector of observations and H is the observation function that maps the model grid point variables into observables at the observation locations. The matrix $\mathbf{H}$ represents the linearization of H about $\bar{\mathbf{x}}^b$ and $\mathbf{R}$ is the observation error covariance matrix. The ensemble provides the estimate of both the $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$ matrix of background covariance for the observation locations and the $\mathbf{P}^b\mathbf{H}^T$ matrix of background covariance for the observation and grid point locations.

The unique features of the LETKF are that it assimilates all observations that may affect the estimate of a given state vector component simultaneously, and the analyses of the different state vector components are computed independently of each other (Hunt et al. 2007, Szunyogh et al. 2008, and Merkova et al. 2011). In practice,

the observations that may affect the analysis at the given grid point are selected by assimilating all observations from a prescribed local volume around the grid point. The definition of the local volume is a tunable parameter of the LETKF algorithm. In essence, the local volume has to be sufficiently small so that the ensemble can provide an efficient representation of the most important degrees of freedom in the space of the background errors, but sufficiently large, so the local volumes centered at neighboring grid points include similar subsets of the observations, which is necessary to ensure the smooth spatial variation of the analyzed fields.

2.1.c Configuration of the Data Assimilation System

In our configuration of the data assimilation system, each global ensemble member has a limited-area counterpart: the k-th member of the global analysis ensemble at the previous analysis time, t_{n-1}, provides the initial condition for the global forecast that (i) produces the k-th member of the global background ensemble at the analysis time, t_n and (ii) provides the boundary conditions and the large-scale forcing for the limited-area forecast that produces the k-th member of the limited-area background ensemble at t_n. The initial condition for the computation of the k-th member of the limited-area background ensemble at t_n is the k-th member of the limited-area analysis ensemble at t_{n-1}. The global data assimilation process at t_n produces the members of the global analysis ensemble at t_n, while the limited-area data assimilation process provides the members of the limited-area analysis ensemble at t_n. In what follows, we refer to the ensemble mean of the global analysis ensemble as the global analysis and the ensemble mean of the limited-area analysis ensemble as the limited-area analysis.

The tunable parameters of the LETKF are set to values that were found to provide a near-optimal performance of the analysis system given the available computational resources by the earlier studies with the same system (Szunyogh et al. 2008 and

Merkova et al. 2011). In particular, all experiments are carried out with $K = 40$ ensemble members and observations are assimilated within 800 km of each grid point in both the global and the limited-area system. In addition, covariance inflation is applied to the ensemble-based estimates of the background error covariance matrix to compensate for the effects of sampling errors (due to the finite size of the ensemble) and nonlinearities in the evolution of the state estimation errors, and to account for the effects of model errors. The covariance inflation in the global system is variable with height and with latitude. In the southern hemisphere and tropics, the inflation ranges from 1.25 at the surface to 1.2 at the top model level. There is a transition zone near 25° North where the inflation tapers from 1.35 at the surface to 1.3 at the top. At latitudes higher than 25° in the northern hemisphere, the inflation factor is 1.5 at the surface tapering to 1.3 at the top level. The covariance inflation factor in the regional system varies only with height and ranges from 1.5 at the surface to 1.3 in the top level.

Four daily analyses are obtained by assimilating observations from an observation time window of Δt = 6 h. This approach provides analyses at 0000 UTC, 0600 UTC, 1200 UTC, and 1800 UTC. Analyses are prepared for the period between 0000 UTC June 22, 2004 and 1800 UTC August 15, 2004. We generate the initial global ensemble for 0000 UTC June 22, 2004, which is necessary to start the cycling of the analysis, with the 0000 UTC operational GFS analyses truncated to T62 resolution from 40 different days during the summer of 2004. During the cycling period, there were eight TCs in the northwest Pacific basin. Of these eight, there were a variety of storm intensities ranging from Tropical Depression (< 34 kts) to Category 4 ($113 - 136$ kts) on the Saffir-Simpson Scale (Atangan et al. 2004). Basic information for each of these storms is included in Table B.1.

The domain for the limited-area calculations is chosen such that it includes the

tropical northwest Pacific, a large portion of eastern Asia, and the northeast Pacific ocean (Fig. A.2). The resolution of the limited-area system starts at roughly the same scale as the global analysis (200 km) and is increased in increments to 36 km, at which point it becomes too computationally expensive to further increase the resolution for the same domain.

Five-day deterministic forecasts from both the global and limited-area 48 km analyses are prepared every twelve hours over the course of the nearly two-month period we study.

2.2 Verification Methods

2.2.a Nominal Resolution

Analyzing and predicting the storm locations on a finite resolution grid introduces an inevitable error component into the position analyses and forecasts. Assuming that the four grid points surrounding the storm location are the vertices of a $\Delta x \times \Delta x$ square, the magnitude of this error component can take any value between zero and $\sqrt{2}\frac{\Delta x}{2}$ with equal probability. The discretization error formally enters the data assimialtion process by limiting the accuracy of the observation function H and the accuracy of the $\mathbf{H}$ linearization of H.

2.2.b Effective Resolution

Resolution also has a major effect on the accuracy of the estimate of the background error covariance matrices $\mathbf{HP}^b\mathbf{H}^T$ and $\mathbf{P}^b\mathbf{H}^T$. These effects are controlled, not by the grid spacing, but the much coarser *effective resolution* of the model, because the ensemble can estimate the background error covariances only at those scales that the model can resolve. The effective resolution of the analyses and forecasts is coarser than the Δx *grid spacing* (nominal resolution) of the model. The exact difference between the effective and the nominal resolution depends on a number of factors,

which include the discretization strategy, the truncation strategy (when a spectral discretization is used), and the parameterization schemes that represent the effect of processes at the sub-grid scales on the processes at the resolved scales.

The NCEP GFS model is a spectral transform model, which we integrate using a triangular truncation with cut-off wavenumber $T = 62$. The model uses an aliasing-free approach for the computation of the nonlinear terms, which requires that the number of grid points in the zonal direction, M, satisfies the condition $M \geq 3T + 1$, while the number of grid points in the meridional direction, Y, satisfies $Y \geq 3T/2$. To satisfy these conditions, we set the number of grid points to $M = 192$ and $N = 94$. In the region where the TCs are typically located, these choices for M and N provide a nominal resolution of $\Delta x \approx 200\ km$.

To provide an estimate of the effective resolution of our global analyses, we note that the LETKF analysis is obtained on the 192×94 grid, but then transformed to spectral space to obtain the spectral coefficients that provide the initial conditions for the T62 forecasts. This step amounts to a spectral filtering of the initial conditions; for instance, the cut-off wave number for the discrete Fourier transform of 192 grid-point variables in the zonal direction would be 96 (> 62), thus the spatial resolution associated with a cut-off wavenumber 62 is about 300 km. The effective resolution, however, is expected to be even lower than that: in an ensemble based Kalman filter scheme, the analysis is a linear combination of the background ensemble members, which implies that the analysis cannot have a higher resolution than that of the background forecasts, which usually have little energy at the tail end of the spectrum due to the diffusive effect of the parameterization schemes. The overall effect of the parameterization schemes must be diffusive, because the energy, which would otherwise accumulate at the tail end of the spectrum, has to be removed (e.g. Durran 2010).

An estimate for the effective resolution, which accounts for the effect of diffusion and is usually considered a conservative one, can be obtained by assuming that diffusion wipes out most kinetic energy for wave numbers larger than $2T/3$. Applying this estimation approach to our T62 resolution global analysis fields, leads to an estimate of 460 km for the effective resolution. Based on similar arguments, the effective resolution of our 100 km, 48 km, and 36 km limited-area simulations are about 230 km, 110 km, and 80 km, respectively.

Even more pessimistic estimates of the effective resolution are obtained when it is defined as the smallest scale where the model correctly captures the power spectrum of the kinetic energy distribution in the atmosphere. Following this approach, Källén (2012) concluded that the global model of the European Center for Medium-Range Weather Forecasts (ECMWF) at nominal resolutions 16 km, 10 km, and 5 km, had an effective resolution of 110 km, 70 km, and 30 km. Notice that this approach provides the same estimate of 110 km for the effective resolution at nominal resolution 16 km as our estimate based on the $2T/3$ rule at nominal resolution 48 km. This relationship between the two estimates indicate that while a 48 km resolution model can represent flow features with a characteristic spatial scale of 110 km, an efficient representation of the nonlinear interactions between those features would require a spatial resolution of 16 km.

2.2.c Estimation of the Track and Intensity Error

The Joint Typhoon Warning Center's (JTWC) Best Track data (Atangan et al. 2004) is used as the verification data set to assess the errors in the track and intensity analyses and forecasts. The Best Track is issued after the JTWC reassesses all available data once a storm has dissipated. The data set includes both six-hourly track and intensity information, but the intensity information has larger error. The

use of the Dvorak model is the main source of the discrepancy in intensity from real-time observations. Look-up tables based on the Dvorak intensity index have been derived from empirical data to provide corresponding estimates of minimum central pressure and maximum wind speed in given basins (Velden et al. 2006). Another discrepancy from other agencies' best track data sets arises from using the one-minute mean sustained wind speed. This procedure can lead to estimates that are higher than those provided by other agencies, which base their estimates on a ten-minute mean (Chu et al. 2002). Neither the Best Track data, nor the realtime version, called TCVitals (Keyser 2007), were assimilated in the experiments in this Chapter.

The Best Track data are all point estimates and are compared to our gridded analysis and forecast data. Storm locations in the analyses and forecasts are determined using a technique described by Suzuki-Parker (2012): first we identify the location of the storm as the grid point with the maximum positive vorticity at the 850 hPa level in the model, then determine the storm intensity as the lowest SLP found within 1° of the vorticity maximum. Once the cyclone tracks and intensities are extracted for all analyses or forecasts at times included in the Best Track data, root mean square errors (RMSE) are calculated over all the locations along the track to provide a measure of the track and intensity error for the entire lifetime of each TC.

2.2.d Discretization Error

Because we use a $2.5° \times 2.5°$ grid to verify the global analyses and forecasts, for which $\Delta x \approx 260\ km$, the maximum and the mean of the discretization error are about 180 km and 90 km, respectively. For the 100 km resolution limited-area grid, the maximum of the discretization error is 90 km, while the mean is 45 km. The same numbers for the 48 km grid are 34 and 17 km, while for the 36 km grid they

are 26 and 13 km.

2.2.e Statistical Significance Test for Auto-regressive Process

To test the statistical significance of the differences between the RMSE for the different analyses, we employ a method described in Szunyogh et al. (2008) and Aravéquia et al. (2011). The test is based on a paired t-test with a null hypothesis that the means of the two time series are the same. Assuming that the difference between two time series of analysis errors $\Delta(t)$ is described by a first-order autoregressive process, the test statistic is

$$z = \frac{\bar{\Delta}}{[V_\Delta / T']^{1/2}}, \tag{2.2}$$

where $\bar{\Delta}$ is the time average over $\Delta(t)$ and V_Δ is the variance of the time series $[\Delta(t), t = 1, 2, \cdots, T)]$. In Eq. 2, the effective sample size, T', is

$$T' = T\,\frac{1-r}{1+r}, \tag{2.3}$$

where the autocorrelation coefficient, r, is computed by

$$r = \frac{\sum_{t=1}^{T-1}\{[\Delta(t) - \Delta_{T1}][\Delta(t+1) - \Delta_{T2}]\}}{\{\sum_{t=1}^{T-1}[\Delta(t) - \Delta_{T1}]^2 \sum_{t=2}^{T}[\Delta(t) - \Delta_{T2}]^2\}^{1/2}}\,. \tag{2.4}$$

Here, Δ_{T1} is the average over the first $T-1$ time steps of $\Delta(t)$ and Δ_{T2} is the average over the last $T-1$ time steps of $\Delta(t)$.

Under the assumption that the time series of errors is described by a first order autoregressive process, autocorrelation, r should fall within the range [0, 1]. While the intensity analysis and forecast error statistics generally satisfy this assumption,

the position analysis errors do not, because the analysis position error has a ‘‘short memory’’. Thus, in the computation of the statistical significance of the difference between the track errors of each experiment, we assume that the effective sample size T' is equal to the sample size T.

Once z is computed, the difference between the time series is deemed significant if the likelihood of obtaining this z is less than the significance level being tested. Differences between analysis time series of pressure or track error are sufficiently significant for our purposes if the probability of achieving the observed value of z, using a t-distribution with $T' - 1$ degrees of freedom, is greater than 90% (i.e. p-value <0.05 or p-value >0.95). An example of some of these metrics is provided in Table B.2 for the comparison of the global LETKF and the limited-area 100 km LETKF experiments.

2.2.f Steering Flow Error

A necessary condition for an accurate tropical cyclone track forecast, including the forecast that provides the background, is an accurate analysis and prediction of the steering flow. Here, we define the steering flow by the mass-weighted vertical average of the wind vector between the 850 hPa and the 250 hPa pressure level. We define the error in the deep layer wind, $STWE$, as

$$STWE = \sqrt{\overline{(u - u^t)^2 + (v - v^t)^2}}, \tag{2.5}$$

where u and v are the deep layer mean-weighted wind components in our analyses or forecasts, and u^t and v^t are the proxies for the true values of u and v and the overbar represents the time mean over all verification times. The two-dimensional field of STWE is calculated for the entire limited-area domain. We use u and v fields computed based on the NCEP operational analysis at 1° horizontal resolution

as the proxies u^t and v^t. The flow associated with the TCs is not removed for the calculation of the environment flow.

2.2.g Stratification by Storm Intensity

Verification statistics stratified by cyclone intensity are also prepared. This verification approach is, in part, motivated by Torn (2010), who found that the largest errors in the intensity analyses tended to occur for strong TCs (Category 3-5), while the largest track errors tended to occur for weak TCs (Tropical Depressions and Storms). The intensity and the position errors are separated based on Best Track wind speed into the three groups used in Torn (2010): Tropical Depression/Tropical Storm (< 63 kts), Category 1-2 ($63 - 96$ kts), and Category 3-5 (> 96 kts).

2.2.h NCEP Operational Analyses

To assess the quality of our LETKF analyses, we employ the NCEP operational analyses from 2004 as both benchmarks and verification data. For the computation of the verification statistics, the operational analyses are considered on a $2.5^\circ \times 2.5^\circ$ grid for the comparison with the global LETKF analyses, and on a $1^\circ \times 1^\circ$ (about 100 km resolution) grid for the comparisons with the limited-area analyses. The operational NCEP analysis used the same version of the model as we did, but with the Spectral Statistical Interpolation (SSI), a 3D variational analysis method (Parrish and Derber 1992), data assimilation system. The operational analyses have a higher, T254L64, resolution, which is roughly equivalent to a 50 km nominal resolution at 20°N. It also assimilated a large number of satellite radiance observations in addition to the observations assimilated in our system, and most importantly, it employed a TC relocation technique (Liu et al. 2000) based on the TCVitals. Thus the main sources of error in the operational analyses are the discretization error and the differences between the TCVitals and the Best Track data. The latter source

of error was examined by Trahan and Sparling (2012), who found that for the 2004 season, the frequency of the TCVitals position being different from the JTWC Best Track position by more than 40 km was just over 10 %. They also found, however, that in particular cases, the difference between the Best Track data and the TCVitals were surprisingly large. For instance, the difference between the JTWC Best Track data and the TCVitals can be larger than 100 km for weak storms.

2.3 Analysis Verification Results

2.3.a Verification of the Global LETKF Analyses

We start the discussion of the analysis verification results with the validation of the global LETKF system. The purpose of this validation exercise to show that the limited-area analyses are compared to a global analysis of reasonable quality.

The position RMSE for the different storms vary between about 150 km and 450 km (Fig. A.3.a), while the intensity RMSE varies between 5 hPa and 30 hPa for the LETKF analysis for each storm (Fig. A.3.c). When the position RMSE is compared for the global LETKF and the operational analysis at the resolution of the LETKF analysis, the difference between the two analyses is much smaller than the RMSE for the LETKF analysis. As expected, the differences between the intensity RMSE for the two systems are small, as neither of the analyses considered here has sufficient resolution to resolve the dynamics of the most intense phase of the lifecycle of a TC.

2.3.b Comparison of the Global and the Regional LETKF Analyses

We first compare the performance of the limited-area data assimilation system to that of the global LETKF system (Fig. A.4). Results are shown only for the 200 km and the 100 km resolution regional analyses. The comparison shows, as expected, that when the regional and global analyses have about the same resolution (200 km),

the global analysis is more accurate. While we expect this result to be rather general, we also expect the magnitude of the difference between the errors in the global and the limited-area analyses of equal resolution to be strongly dependent on such factors as the resolution, the size of the regional domain, the choice of data assimilation system, and the observation density.

When the resolution of the limited-area analysis is doubled to 100 km, the limited-area analysis has a statistically significant advantage over the global system for three of the eight storms and shows some significant systematic advantage. For the rest of the storms, the limited-area analyses is more accurate, but not at the 90 % confidence level. For all but two storms, the 100 km limited-area intensity analysis is also more accurate, but from a practical forecasting point of view, the advantage of the limited-area analysis is small.

2.3.c Comparison of the Different Resolution Regional Analyses

In Fig. A.5, the accuracy of the analyses is compared for each resolution to the accuracy of the analyses at the one step higher resolution (e.g. panels a and b compare the 200 km and the 100 km analyses, panels c and d compare the 100 km and the 48 km analyses, etc.). There is a clear improvement in the position analyses only when the resolution is increased from 200 km to 100 km, but further increases of the resolution show no further significant improvement. There is also a statistically significant improvement in the intensity analysis when the resolution is increased from 200 km to 100 km. While further increases of the resolution lead to further improvements in the intensity analysis (right panels), the magnitude of the improvement is small from a practical point of view. (The average reduction in the intensity RMSE due to increasing the resolution from 100 km to 48 km, or from 48 km to 36 km, is a mere 0.2 %.)

2.3.d Stratification of the Errors by Storm Intensity

The effect of storm intensity on the errors in the global and the limited-area position analyses is shown in Fig. A.6. The statistics shown in this figure are based on 15 data points for strong storms, 48 data points for moderate storms, and 155 data points for weak storms. The results suggest that storm intensity has a major effect on the distribution of the position error. First, the mean and the median of the errors for the strong storms are smaller than for the moderate and weak storms for all configurations of the analysis system. Second, large errors, which appear as statistical outliers in Fig. A.6 in the case of weak storms, are not nearly as large and do not occur with as great a frequency in the case of moderate and strong storms, as in the case of the weak storms. We note that the outliers for the weak storms are typically due to the given analysis not resolving a circulation at all. In addition, weaker systems often have less well-defined steering flow. Most likely, these two factors lead to the larger number of outliers for the weak storms.

The maximum error for the global LETKF analyses of strong storms is 246 km, while the mean error is 136 km. The maximum is larger by about 60 km, while mean by about 46 km, than what we would expect if the errors were solely due to discretization errors. The distribution of the errors, however, is very similar to that for the operational global analysis truncated to the resolution of the global LETKF analysis. That is, at that resolution, the LETKF can compensate for the advantage the operational system has due to the use of TC relocation.

The limited-area analyses of the positions of the strong storms are clearly more accurate than the global LETKF analyses of the storms. According to a Kolmogorov-Smirnov two-sample test for distribution (e.g., Massey (1951)), the distributions of the errors for both the 100 km and the 36 km resolution systems, but not for the

48 km system, are significantly different from the distribution of position errors for the global LETKF at a 90 % confidence level. Neither the distributions, nor the means of the errors are different for the three different resolution limited-area analyses. Nevertheless, the maximum of the position error (excluding outliers) indicates a slight decrease of the position error with increasing resolution: the maximum of the position error for the 100 km, 48 km, and 36 km resolution limited-area analyses is, respectively, 201 km, 177 km, and 94 km.

The mean (85 km) and the maximum (201 km) of the error for the 100 km resolution system is consistent with our estimate of the effective resolution (230 km) for that grid spacing, but the mean and the maximum for the 48 km and 36 km resolution systems are larger than what would be expected based on the estimates of the effective resolution. This suggests that at resolutions finer than 100 km, the analysis system cannot take advantage of the smaller discretization errors and the higher effective resolution. This result suggests that either there is no sufficient observed information to take advantage of the higher-resolution grid and/or the quality of the estimates of the background covariances do not improve sufficiently to lead to a better use of the available observations. The operational analysis, which has a nominal resolution about the same as that of the 48 km resolution limited-area analysis, can achieve a higher accuracy than the limited-area analyses (a mean error of 58 km and a maximum error of 99 km, excluding outliers). The higher accuracy of the operational analysis is most likely primarily due to the use of TC relocation: the TCVitals are based on a large amount of observed information, which is not assimilated in our system. In addition, the TC relocation procedure does not rely on the estimates of the background covariances for the correction of the storm location, thus its accuracy is limited by the grid spacing and the accuracy of the TCVitals, rather than the effective resolution. Finally, the assimilation of a large number of

radiance observations in the operational system, which are not assimilated in our system, may lead to a more accurate analysis of the steering flow.

The superiority of the higher-resolution limited-area analyses to the global LETKF analysis is most apparent for the weak storms: the distributions and means of the errors are significantly different from the global distribution of errors for all limited-area resolutions at the 90 % confidence level.

For completeness, we also show the stratification of the intensity errors by storm intensity (Fig. A.7). The most striking feature of the intensity errors is that they are larger for the strong storms than for the moderate storms, and for the moderate storms than for the weak storms. The differences between the performance of the global systems, the performance of the global and limited-area systems, and the performance of different resolution limited-area systems are small and not statistically significant. This result suggests that, as expected, analyzing the intensity is equally challenging for all investigated configurations of the system.

2.3.e Stratification of the Analysis Increments by Storm Intensity

Next, we investigate the effect of storm intensity on the relationship between the background and the analysis errors. The results of our investigation are summarized in Fig. A.8. In this figure, a dot over a diagonal indicates a case in which the assimilation of the observation improved the state estimate provided by the background, while a dot below the diagonal indicates a degradation by the assimilation of observations. The one-sentence segue of the results is that the assimilation of observations tends to help when the background error is large. In particular, in the case of strong storms, data assimilation has a positive effect on the accuracy of the state estimate only for the 200 km resolution data assimilation system, which is the one configuration that tends to provide a low-accuracy background for strong storms. The assimilation

of the observations leads to the largest improvement in the 100 km, 48 km, and 36 km resolution analyses of moderate storms. There is a similar, but somewhat less dramatic improvement in the state estimates for the weak storms.

2.3.f Errors in the Analysis of the Steering Flow

The accuracy of the analysis of the steering flow can have a major effect on the accuracy of the analysis of the location of the storms at later times through the background estimate of the storm locations. The geographical distribution of the RMSE of the difference between the analyses of the steering flow in the different limited-area experiments is shown in Fig. A.9. (We recall that the proxy for the true steering flow in the computation of the error is the operational NCEP analysis at 1° resolution.) In the extratropics, the limited-area analyses of the steering flow are generally more accurate than the global analysis in all but the 200 km resolution limited-area analysis. In the tropics the advantage of the limited-area system is less obvious. The only regions where the limited-area analyses are clearly less accurate than the global analysis are near the western and eastern lateral boundaries in the extratropics. The magnitude of the degradation near the boundaries decreases with increasing resolution. This result is most likely due to the fact that in the higher-resolution system more grid points are used for the relaxation of the limited-area model solution to the global model solutions near the lateral boundaries.

Is the improved analysis of the steering flow in the 100 km (and higher) resolution limited-area analyses due to an improved background and/or to a larger improvement of the state estimate in the update step? First, we compare the background errors for the different resolution analyses (Fig. A.10). For the interpretation of the results, we recall that the background is the ensemble mean of the ensemble of 6-hour forecasts. The results show that except for the 200 km resolution configuration, the background

error is smaller for the limited-area systems, than the global system. In addition, while increasing the resolution from 200 km to 100 km helps significantly, the further increase of the resolution cannot further enhance the quality of the limited-area background. This result suggests that the 100 km and finer resolution analyses are more accurate than the global analysis, in part, because the higher-resolution limited-area model provides a more accurate background. To see whether the higher-resolution limited-area analysis can also more efficiently assimilate the observations, we compare the magnitude of the corrections made by the limited-area and the global analyses to the background (Fig. A.11). The results show that the limited-area systems more efficiently reduce the background error.

There are two potential factors that can contribute to the more efficient use of the observations by the limited-area data assimilation systems: the more accurate estimation of the background covariance and the more accurate computation of the observation operator. In a system where the observation operator implements only spatial and temporal interpolations, as is the case in our study, a higher-resolution grid is a guarantee for a more accurate observation operator. A more accurate observation operator can result in a more accurate analysis because it leads to more accurate computation of the innovation, $\delta\mathbf{y}$. It can also lead to a more accurate weighting of the background and the observed information in the analysis, by allowing for a more accurate estimation of $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$. Finally, it spreads the information more efficiently from the observation locations to the model grid points with the help of a more accurate estimate of $\mathbf{P}^b\mathbf{H}^T$. The small improvement found in the analysis of the steering wind when the resolution is increased beyond 100 km suggests that once that resolution is reached, the interpolation errors become secondary compared to the errors in the background and the estimates of the background error statistics and/or the limited availability of observed information.

2.4 Forecast Verification Results

2.4.a Mean Forecast Error

We compare the error in the global forecasts started from the global LETKF analyses with the error in the 48 km forecasts started from the 48 km limited-area analyses. The mean error of the position forecasts for the first five forecast days is reduced independent of the storm strength at analysis time (Fig. A.12). Are these forecasts improved due to the better analysis of the position of the storms or to the improvements in the analysis of the steering flow? First, in the domain averaged sense, the advantage of the limited-area forecast of the steering flow is short-lived; about 18 hours (Fig. A.13). A more careful investigation of the error in the steering flow shows, however, that in the immediate vicinity of some of the storms the improvements in the wind forecasts survive up to about 72-h forecast time (Fig. A.14). Thus, we conjecture that the forecast improvements are due to the more accurate analysis of the storm locations and the better representation of the interactions between the storms and their immediate environment.

2.4.b Stratification of Forecast Errors by Storm Intensity

We find statistically significant improvements only in the forecasts of the weak storms (Fig. A.15). (The results for the moderate and the strong storms are not shown.) Even though the errors can be quite large in the forecasts of these storms, the forecast improvement in the mean error by the limited-area system is statistically significant up to the 72-h forecast time. Since the errors for the weak storms are larger than for the moderate and strong storms, their weight in the overall average of the errors is also larger. The combination of these relatively large improvements in the forecast of weak storms and the small (statistically not significant) improvements in the forecasts of the moderate and strong storms lead to the clear improvement in

the overall forecast errors shown in Fig. A.12.

3. THE ASSIMILATION OF TC OBSERVATIONS

Over the past decade, the increase in direct sampling activities of tropical cyclones (TCs) has led to a host of data assimilation challenges. One particularly challenging aspect of assimilating observations of TCs is the proper interpretation of large observation innovations (the difference between an observation of the state and the background estimate of the state). Aside from gross errors in the observations, large innovations can arise in two main ways for TC assimilation. One cause is representativeness error, which occurs when observations are strongly affected by flow features that are not resolved in the model. Because representativeness error depends on both the atmospheric state and its model representation, it is difficult to categorically quantify it for a given observation type. A second source of large innovations for TCs is not necessarily a result of large errors in the observation or background representation of the locations of the storms: a small displacement in the background location of the TC from the observed location can result in large innovations for observations of fields with large gradients and/or spatial variability. Consider the scenarios illustrated in Fig. A.16. A wind observation taken in the northeast quadrant of the TC is combined with a relatively accurate background estimate of the storm such that the resulting analysis would reflect the true atmospheric state that is being sampled. In scenario B, the background estimate of the TC is displaced slightly from the observed location, leading to a large observation innovation for the same observation. In Scenario B, the observation would either be rejected as a bad observation, or it would affect the analyzed fields in an unrealistic way.

In this Section, we search for efficient methods for the assimilation of observations

associated with potentially large observation innovations. In particular, we seek an efficient combination of the quality control procedures and the definition of the observation errors. We focus on three types of observations: QuikSCAT 10 m wind retrievals, Tropical Cyclone Vitals (TCVitals) minimum sea level pressure (SLP), and dropsonde vertical profiles from TC reconnaissance programs. The system that will be implemented is identical to the analysis/forecast system implemented for the experiments in the previous Section. One beneficial aspect of the ensemble-based approach for the aforementioned types of experiments is that the EnKF is thought to be an efficient filter of imbalances that could be introduced by the assimilation of observations.

Experiments testing assimilation methods of the special observations will be carried out in three stages. First, we will investigate several approaches for each observation type in a series of single-update experiments. Second, we will implement the methods that have been found the most promising in the single-update experiments to carry out cycled experiments over the lifetime of Typhoon Sinlaku (2008). Because Sinlaku is among the most intensely observed storms on record, it provides a unique opportunity for applying experimental data assimilation techniques. Sinlaku was one of the strongest typhoons to occur during a highly organized, multi-organization field campaign targeting all stages of the life cycle of a tropical cyclone from origination as a tropical wave through extra-tropical transition: the THORPEX (The Observing System Research and Predictability Experiment) Pacific-Asian Regional Campaign (T-PARC) and Tropical Cyclone Structure 2008 (TCS-08) field campaigns covered much of the Northwestern Pacific basin and included observations from diverse platforms, many of which were operationally available in real time. Typhoon Sinlaku was meteorologically interesting in that the operational centers had difficulty forecasting several important periods during the lifetime of the cyclone: first, a rapid

intensification over a period of two days from 45 kts to 120 kts, then the re-curvature to the northeast, and finally a second rapid intensification from 20 to 70 kts over 18 hours, which was not captured by the model forecasts or satellite observations, but was recorded by the dropsonde observations (Cooper et al. 2008). Because of the analysis and forecast challenges posed by the storm and the abundance of observations, Sinklaku provides an ideal setting for the cycled experiments. Lastly, we will consider the impact of the special observations for a sample of 21 TCs from two separate seasons in the Northwest Pacific basin.

3.1 Special Observations

3.1.a QuikSCAT Observations

QuikSCAT 10 m wind observations are thought to be potentially highly useful for tropical cyclone analysis and forecasting because the near surface winds play an important role in the transfer of heat, moisture, and momentum in the TC environment (Leidner et al. 2003). While investigating the potential impact of QuikSCAT observations may seem irrelevant considering that the instrument is no longer active, understanding the potential benefits of having similar observations in the future is important for the planning of the next generation of observing systems. Several studies have discussed the impact of the assimilation of scatterometer observations on TC analyses. Kunii et al. (2012) used an ensemble-based technique for the estimation of the analysis impact of scatterometer observations, while others used 3D- and 4D-VAR data assimilation systems to assimilate scatterometer observations (Isaksen and Janssen 2004; Chen and Snyder 2007; Leidner et al. 2003) and found positive impacts on both track and intensity forecasts.

The challenges associated with assimilating QuikSCAT retrieval observations make it one of the most complicated data sets to assimilate. Not only must we

address the causes of large observation innovation, but the observations have an inherent representativeness error associated with their gridding process, and finding a proper definition of the observation function is highly challenging.

The Level2B processing of the QuikSCAT observations starts by separating single backscatter pulse measurements into $8 - 12$ slices, each approximately 6 x 25 km. Since pulses overlap, there are multiple slices of measurements for the same physical area. The satellite measurement swath is divided into a square grid of Wind Vector Cells (WVCs), which are 25 x 25 km for the 25 km dataset. The slices of each pulse are assigned to the grid of WVCs based on the location of the centroid of the slice. Because of the overlap of pulses, many slices fall into each WVC (Lungu 2006). The wind speed and direction are chosen by an inversion process that calculates the maximum likelihood estimates of the polarized measurements within each WVC. In essence, each wind observation on the grid is the weighted average of the winds within a 12.5 km radius of the grid point location and is done over a sufficiently short amount of time that the observation may be considered to be instantaneous for our purposes. The spatial averaging, however, may introduce a significant amount of representativeness error in the presence of the strong winds ($\geq$ 30 ms^{-1}) associated with TCs.

The observation function maps the background estimate of the observation to the time and location of the observation. The observation function as we define it for the QuikSCAT retrievals, assumes that the boundary layer is well-mixed and neutrally stable. The virtual potential temperature remains constant with height in a well-mixed boundary layer. Since QuikSCAT winds are derived from satellite measurements as neutral-stability winds, and given the amount of vertical mixing associated with a TC, both assumptions are appropriate.

The extrapolation of the 10 m model-equivalent background wind is performed

using a basic hydrostatic approximation,

$$\frac{dp}{dz} = -\bar{\rho}g, \tag{3.1}$$

to determine the height difference, dz, and pressure difference, dp, between the first sigma level and 10 m, where $\bar{\rho}$ is the average density of the air in the layer, and g is acceleration due to gravity. The log-wind relation,

$$V_{10} = V\frac{log_{10}\left(\frac{10.0}{z_0}\right)}{log_{10}\left(\frac{z_1}{z_0}\right)} \tag{3.2}$$

makes it possible to reduce the model wind, V, at the first sigma level to a wind at a height of 10 m, V_{10}, where z_1 is the height of the first sigma level in meters, and z_0 is the surface roughness length defined through empirical methods in Hoffman (2011). These equations are underdetermined without an estimate for the average density of the layer, which can be addressed with the following relationships. The virtual potential temperature, Θ_v, is

$$\Theta_v = \Theta(1 + 0.61w), \tag{3.3}$$

where w is the mixing ratio, and the potential temperature, Θ, is defined as

$$\Theta = T\left(\frac{p_o}{p}\right)^{\chi} \tag{3.4}$$

where T is temperature, p is pressure, p_0 is the pressure at a reference level of 1000 mb, and $\chi = R/c_p$ is a constant. Since Θ_v remains constant with height under our assumptions about the layer, we may remove the pressure and the temperature height dependence from Eqn. 3.4. The average layer pressure $\bar{p} = (p_{sfc} + p_1)/2$,

and average temperature, $\bar{T} = \bar{p}/(\bar{\rho}R)$, between the surface and first sigma level are substituted into Eq. 3.4. With these substitutions, Eqn. 3.3 then becomes

$$\Theta_v = \frac{\bar{p}}{\bar{\rho}R}\left(\frac{p_o}{\bar{p}}\right)^{\chi}(1 + 0.61w). \tag{3.5}$$

The resulting average density then, is

$$\bar{\rho} = \frac{\bar{p}}{\Theta_v R}\left(\frac{p_0}{\bar{p}}\right)^{\chi}(1 + 0.61w). \tag{3.6}$$

which determines a solution for Equations 3.1 and 3.2.

3.1.b Tropical Cyclone Vitals

Tropical Cyclone Vitals (TCVitals) are a database of real-time estimates of TC parameters, which can include the minimum SLP, position, maximum wind speed, shape, and size just to name a few (Trahan and Sparling 2012). Because the main purpose of using the TCVitals observations in this study is to deepen and reposition the cyclone through the regular data assimilation process, we will use only the information about the value and the location of the minimum SLP. (Since our analysis system does not employ TC relocation or cyclone bogusing techniques, the analyzed position and intensity of the cyclone is entirely due to the assimilation of observations.) The smoothing that occurs as a result of using an ensemble mean to estimate the most likely background and analysis state can be detrimental to the analysis of both the intensity and position of the TC if there is measurable uncertainty in the position of the TC. Chen and Snyder (2007) assimilated only the position of the vortex in an idealized ensemble Kalman filter (EnKF) experiment with limited success and found that when the difference between the background location and the observed location of the TC was larger than the vortex size, the EnKF performed poorly. They

also showed that assimilating intensity reduces the ensemble spread of the vortex positions, leading to quicker convergence of the ensemble tracks, and minimizing the negative effects of assimilating positions with large innovations. This is a promising result for the simultaneous assimilation of position and intensity. Torn and Hakim (2009) chose not to assimilate minimum SLP in their low-resolution (30 km) outer limited-area domain, because they believed it would lead to degraded track and intensity forecasts since the resolution could not accurately simulate TC structure. In a later study, Torn (2010) showed that there was an increase in acoustic and gravity waves by as much as 30 % when intensity observations with large innovations were forced into the analysis, which essentially resulted in the rejection of the observation as the model quickly returned to a state that could be resolved at the given grid spacing. They did, however, assimilate the TC advisory position in another study (Torn and Hakim 2009), and found similar results as Chen and Snyder (2007): the position observations had a negative affect when they were more than a few grid points from the background estimate of the location. The results of these studies suggest that assimilating both the intensity and location of the vortex could be promising for a moderate resolution system, as long as the background estimate of location is not too far from the observed position.

The TCVitals serve as an excellent set of observations to test the effect of observation error definition on the analysis. The single SLP is estimated only at the synoptic analysis times, and surface pressure is already among the types of observations assimilated in the LETKF, so errors are minimal compared to other observation types. Because the observations of SLP tend to have a large representativeness error component, especially for strong TCs, the experiments assimilating TCVitals will address whether it is beneficial to account for representativeness error within the definition of the observation error.

Several studies suggest that the observation error for the TCVitals position is between 10 and 20 km (Trahan and Sparling 2012; Torn and Hakim 2009; Chen and Snyder 2007). Because the current implementation of our data assimilation system does not allow for an explicit definition of the errors in the position of the observations, the TCVitals position errors enter the data assimilation process as errors in the TCVitals surface pressure observations. We test multiple options for the definition of observation standard error for the TCVitals surface pressure observations. The first option employs the same climatological departure of 3 hPa as was used by Torn (2010). A second, sliding scale for error is defined by linearly fitting the intensity error in the NCEP operational analysis and scaling it to 10% to limit the maximum value of error to a reasonable estimate less than about 10 hPa. The NCEP operational analysis has a similar grid spacing as our limited-area analyses, therefore it is expected to be a reasonable proxy for the representativeness error in a state-of-the-art analysis system with moderate resolution. A more sophisticated method, which could be used is to define a sliding scale TCVital SLP error as a function of grid spacing, observed central pressure, and ambient pressure following an exponential function (Holland 1980), but for the initial testing of the potential benefits of a sliding scale approach, our simpler approach should be sufficient.

A third option is to uniformly apply a very low observation error ($\leq$ 1 hPa, comparable with background standard deviation in the ensemble) to all minimum SLP observations. The potential drawback of this approach is that when the observations have weights nearly equal to the background estimate in the analysis, the likelihood of introducing gross imbalances into the analyzed state significantly increases.

3.1.c Reconnaissance Dropsondes

Weissmann et al. (2011) found that assimilating reconnaissance dropsondes in

3D-Var systems with lower resolution can improve track forecasts by as much as 40%, while other studies found that forecasts from global 3D-Var systems were improved by $10-30\%$ (Aberson 2003, 2008; Wu et al. 2007a,b; Pu et al. 2008; Chou and Wu 2008). Findings of both Weissmann et al. (2011) and Chou et al. (2011) indicated that the assimilation of the same observations by 4D-Var operational systems was less beneficial, because these systems already exhibited lower errors due to more efficient assimilation of the other observations, including satellite radiances. This result indicates that the improvement of the analysis due to the assimilation of dropsonde observations strongly depends on the quality of the background and the data assimilation system. Torn and Hakim (2009) and Torn (2010) used an EnKF system with the Weather Research and Forecasting (WRF) model to study TCs, but they refrained from assimilating reconnaissance dropsondes in the outer domains with resolutions of $30-36$ km because of the degraded forecast skill shown by Aberson (2008) at similar resolutions.

In this study, we will use dropsondes to address observations that result in large innovations at multiple levels. Because the estimates of the observation errors and observation function are well-defined for these types of observations, this is by far the least challenging type of observation of those considered in this study from a data assimilation point of view. The main challenge is with the large innovation, so we can investigate different quality control techniques associated with direct measurements of TCs.

3.2 Experiment Design

3.2.a Analysis/Forecast Experiments

The analysis experiments discussed here include both single-update and cycled experiments. The control experiment, from which most other experiments will be

started initially, assimilates all the conventional observations assimilated at NCEP in real time excluding satellite radiances and TCVitals.

All experiments in this Section are carried out with the same T62L28 resolution for the GFS, while the RSM horizontal resolution remains fixed at 48 km. Any in situ observations from the TPARC/TCS-08 campaign that made it into the NCEP operational system are assimilated for the control, including dropsondes. The data assimilation process was cycled for a sufficient amount of time to eliminate transient effects before the TC observations experiments were started.

3.2.b Quality Control by Huberization

Our implementation of the LETKF performs two main quality control (QC) checks of the observations. An offline QC is done first by rejecting observations that have an NCEP quality control flag of 4 or greater. Then, in an online QC procedure, the LETKF compares the remaining observations to the background mean estimate of the state. If the observation innovation exceeds a prescribed threshold value, c, then the observation is discarded. In past studies, c was set to at least five times either the background ensemble standard deviation, or the observation error, whichever is greater.

For this choice of the threshold, however, the QC algorithm discards most direct observations of TCs because the observation innovation is often large due to large errors in the background and/or a large representativeness error in the observations that are not accounted for by the standard estimates of the observation error variance in the data assimilation system. This experience motivates us to consider alternative approaches for online quality control.

3.2.c Huberization for Special Observations

The European Center for Medium-range Weather Forecasting (ECMWF) has implemented a QC method that reduces the weight of the observations associated with large innovations instead of rejecting them (Tavolato and Isaksen 2010). The procedure is called the Huber norm QC because it is formally based on a robustification of the 4D-Var cost function to statistical outlier observation errors. Roh et al. (2013) defined a similar robustification method for an EnKF. In practical terms, the approach clips the unusually large innovations to a prescribed maximum. The Huberized EnKF analysis, $\hat{\mathbf{x}}^{\mathbf{a}}$, is

$$\hat{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{K}G(\delta\mathbf{y}), \tag{3.7}$$

where the Huber function, $G(\delta\mathbf{y})$, is defined by

$$G(\delta\mathbf{y}) = \begin{cases} \delta\mathbf{y} & \text{if } |\delta\mathbf{y}| < c, \\ c & \text{if } \delta\mathbf{y} \geq c, \\ -c & \text{if } \delta\mathbf{y} \leq -c, \end{cases} \tag{3.8}$$

where c is a prescribed clipping height. While Roh et al. (2013) described several methods for the selection of the clipping height, they all involve calculations with clean data. In the present study, we find a proper value of the clipping height by numerical experimentation using the real observations rather than clean data experiments, which would require the generation of simulated observations with prescribed error statistics.

3.3 Results

3.3.a Single Update Experiments

First, a series of single update experiments was performed in order to assess the appropriate treatment of each of the observation types. A summary of the naming convention for the experiments is provided in Table B.3. These experiments explore the effects of a single observation type on the analysis, as well as the interactions of combinations of observation types. The single-update experiments that produce the most promising results are tested more rigorously in cycled experiments over many analysis cycles. Although many tests were performed, the single-update experiments with sub-optimal, or insignificant results are not discussed.

We first assess the assimilation of the TCVital SLP. We find that in the single update experiments the SLP observation can generate the largest analysis impact of any of the special observation types assimilated in this study, yet it is the most likely to be rejected by the original online QC procedure. Figure A.17 shows the results of various approaches of TCVital SLP assimilation. The Kept+Slide experiment, in which the SLP observation is kept with no QC with a standard error of 0.44 hPa -- the result of using the Sliding method described in Section 3.13.1.b -- leads to the greatest SLP impact, deepening the cyclone by as much as 2 hPa compared to the original QC experiment. This is the deepest storm achieved in any of our single update experiments, which is expected given that the observation error is very similar to the background standard deviation and the full observation innovation affects the analysis solution. In the Constant experiments (3 hPa error), there is minimal impact by the TCVital observation since its error is much larger than the background standard deviation of SLP at that location, rendering it essentially ineffective at updating the intensity of the cyclone in the analysis, even though it is not discarded

by QC. At this analysis time, the observed cyclone is not so deep that we would expect serious problems to arise by assimilating the TCVital SLP with a very low error value, however a stronger storm might generate unwanted imbalances when keeping an observation with a very large representativeness error component. Huberizing the TCVital SLP innovation when the observation error is low (Clip+Slide) provides a compromise between the two extremes, resulting in a deeper storm than the Kept + Constant (large observation error) experiment, yet not as strong as the Kept + Slide (small observation error) experiment. The SLP observation not only acts to correct the intensity analysis, but also acts to update the position analysis, even in the absence of a significant pressure analysis increment. This is likely a result of reducing the ensemble spread in the background position estimate of the TC during the update step, and is especially important in the analysis of weak storms where the signal-to-noise ratio can be much higher than that of stronger storms.

Even though the storm position shifts as a result of TCVital SLP assimilation, there are still some concerns associated with the analyzed wind field, and is not necessarily the fix-all observation that, at first sight, it seems to be. Chen and Snyder (2007) and Torn and Hakim (2009) both show that a location observation can have negative impacts on the analysis by generating a secondary circulation instead of moving the vortex to the correct location. Evidence of a similar phenomenon is seen in the wind field of our TCVital only experiments whenever there is a major intensity change compared to the background. Figure A.18 shows a north-south elongation in the flow for experiments that only assimilate TCVitals SLP, a feature that is seen in neither the background nor the NCEP analysis. Neither the QC method, nor the observation error has, any affect on the shape of the flow when only TCVitals SLP is assimilated: a problem that may be eliminated only by improving the background estimate of location when this is the only available observation. In

the presence of other observations, however, the pattern of the circulation could see gainful improvements, too.

Experiments that assimilate only dropsonde or QuikSCAT observations show marked improvement in the pattern of circulation over the TCVital-only experiments as long as the observations are not subjected to the original QC method. The low-level winds in these single-observation-type experiments (Fig. A.19) show that the wind field is more consistent with what we would expect to see in an analysis at a moderate resolution: a tighter inner core of relatively high-speed winds around a more accurately defined center of circulation. The speed of the winds in the circulation is also increased when a QC method other than the traditional QC is used, but not to the degree seen in TCVitals-only experiments. The dropsonde observations play a more significant role in increasing the wind speed than the QuikSCAT observations, likely because their impact is through a deeper layer, but both types of observations are valuable in relocating the TC circulation.

For both QuikSCAT and dropsonde observations, as with any observing platform, there are occasionally truly erroneous observations that make it past the offline QC. While there is evidence that these observations should be trusted more than in our original QC method, keeping the observations without performing an online QC would likely introduce other problems into the analysis. The results hitherto described suggest that Huberization may offer the needed compromise for an online QC method for the QuikSCAT and the dropsonde observations, while the results are less clear for the TCVitals.

Experiments that utilize all three special observation types support the findings of the single-observation type experiments (Fig. A.20). When the special observations are either Huberized or Kept, the TCVital SLP observation serves the main purpose of intensifying the TC, while the QuikSCAT and dropsonde observations act to

update the circulation pattern and constrain the impact of the TCVital observation on maximum analysis intensity; all of the observation types play a role in the position update. With these results, it becomes obvious that the original QC method impedes the potential accuracy in the analysis of TCs in our analysis system.

Based on the experience of the single update experiments, we devised a combined experiment with the following parameters:

- dropsonde and QuikSCAT observations are assimilated by using Huberization with a clipping height which is three times the original threshold;
- TCVitals SLP is kept with no QC;
- the error in TCVitals SLP is set to 0.5 hPa for all observations.

Since the QuikSCAT observations seem to impose a constraint on the potential intensity of the TC, the TCVital observation can be assimilated without QC when QuikSCAT observations are also assimilated. When QuikSCAT observations are not available, however, the TCVitals should be also be Huberized in order to reduce the potential distortion of the wind field, and limit the impact of observations of strong storms.

The results of deterministic forecasts initialized from each of the single-update analysis experiments suggest that both the experiment that Kept all observations with no QC and the Combined experiment provide reasonable initial conditions suitable for improving TC forecasts. Track and intensity forecasts from both of these analyses perform better than the forecasts started from the control (original QC) experiment and the All Obs Clipped experiment for the first 60 hours. The Combined experiment shows a slight track improvement over the others from 60 - 84 hours, (Fig. A.21), but the intensity forecast is degraded over the same time period. After 60 hours, the initial conditions are no longer affecting the forecasts.

3.3.b Cycled Experiments

3.3.3.1 Sinlaku Analyses

Because the effects of using only a TCVital SLP observation were not immediately obvious in a single update analysis, the discussion in this section will briefly touch on two TCVital only experiments , as well as the superiority of the Combined experiment over the other methods discussed in teh previous section.

Each of these experiments was started on September 8, 2008 at 1200 UTC from the background ensemble of the control experiment, and cycled over 35 6-hour time cycles terminatng the experiment on September 17 at 0000 UTC. This time period was chosen to cover the first rapid intensification and the subsequent recurvature of Typhoon Sinlaku. Naming conventions for the experiments in this section are the same as in the Single Update experiments (see Table B.3) with one exception: combined is altered so that TCVitals SLP is kept with no QC only when QuikSCAT observations are present within a 5 degree radius, otherwise, the SLP is Huberized, along with the other TC observations, at a clipping height of $3c$.

The statistics presented here are based on the analyzed position and intensity at each analysis time. The method for determining these metrics is the same as in Holt et al. (2013): identify the location of the storm as the grid point with the maximum positive vorticity at the 850 hPa level in the model, then determine the storm intensity as the lowest SLP found within 1° of the vorticity maximum.

Figure A.22a shows the time evolution of the minimum SLP in the analyses of the different experiments. The RSM Control experiment with conventional QC and no special TC observations performs poorly and even degrades the global LETKF analysis at times. While the LETKF experiments somewhat capture the same trend as the NCEP Operational analysis, none of them, including the NCEP analysis, captures

the Best Track trend or intensity. The average track analysis for the LETKF global and RSM Control experiments also show the poorest results.

The TCVonly Kept + Slide experiment improves the simulated TC intensity early on, and then again at the end of the cycling period, but does poorly during the most intense period of the TC development - a direct reflection of the observation error as we have defined it. When the observation error is low (storm is weak), the analysis is more accurate, and vice versa, which is the same result we observed in the single update experiments for an analysis time when the storm was of relatively low intensity. Although the position error is drastically improved over the LETKF control experiments, the experiment is still not the best choice among the top contenders. The Slide experiment shows that defining representativeness error as a component of the observation error is not the most effective assimilation method, especially for strong storms, for either intensity or position analysis. The results at the beginning and the end of the cycle, however, support the choice to define a consistently small TCVital observation error during the duration of the cycle.

Keeping the TCVital as the only observation type while also definining a very low observation error (Keep TCVonly 0.5) shows an example of why the TCVital observation can be detrimental to the analysis when used alone. While this experiment most accurately captures the rapid intensification of Sinlaku, the maximum intensity is unrealistically over-estimated and the average analysis position error over the cycling time-period reflects a sub-optimal use of the single SLP observation.

The Combined experiment has the overall best performance: it captures both a rapid intensification and the general trend in the intensity analysis, as well as provides the most accurate average position analysis. On average, the Combined experiment improves the control intensity analyis by nearly 40% for Sinlaku over these cycling times, and outperforms the NCEP analysis by 25%. In addition, the

position analysis is improved by an outstanding 62% over the Control, and is 34% better than the NCEP operational analysis with its advanced methods.

In theory, a better analysis should lead to a more accurate forecast, but because many factors affect a TC forecast, an accurate analysis of position and intensity is never a guarantee of a better position or intensity forecast. The next section addresses the impact of observation use on the forecasts.

3.3.3.2 Sinlaku Cycled Forecasts

Five-day deterministic forecasts were started every six hours from the cycled analysis experiments discussed in the previous section. At each analysis time, the time-averaged forecast error is calculated and compared for the different analyses. The statistical significance of the difference between the forecasts from the different experiments is tested using the autocorrelation method discussed in detail in (Holt et al. 2013) at a 95% confidence level.

A major concern reiterated throughout this paper is the possibility of exciting strong imbalances between the wind and the mass state variables. A rapid decrease of the gain in accuracy achieved at analysis time in the forecasts would be an indication of the presence of strong imbalances.

Figures A.23 and A.24 show that the Combined 0.5 experiment performs much better than the control forecast at most analysis times through 3 days for both track and intensity forecasts, as was the result for a few of the other experiments that assimilated multiple observation types with the experimental approaches (not shown). At 4 and 5 day lead times, the forecasts of both intensity and position are still improved over the control by a large magnitude, but the difference is not as often statistically significant.

The results of cycled forecasts suggest that the improvements in the initial

conditions provided by carefully assimilating TC observations are a key factor in improving the forecasts in our analysis/forecast system. Many of the forecasts improvements were of similar magnitude, but as was already mentioned, the best overall analysis was provided by the Combined experiment, therefore, it will be the focus of a seasonal analysis/forecast experiment.

3.3.c Seasonal Experiments

For a more statistically significant analysis of the Combined 0.5 method for the QC of direct observations of TCs, statistics from 21 combined storms from 2004 and 2008 seasons are considered. For these experiments, the TCs are binned by Best Track wind speed into bins containing 371 weak, 84 moderate, and 45 strong TCs. The Combined 0.5 experiment is compared to the control analysis, the global LETKF analysis, and the NCEP operational analysis.

The Combined 0.5 experiment makes statistically significant improvements to the regional LETKF analysis in both the position analysis (Fig. A.25) and the intensity analysis (Fig. A.26) of strong storms, and is systematically consistent with the NCEP operational position analysis for all storm strengths. The Combined method even performs better than the NCEP operational analyses on the intensity of the moderate and strong storms. The systematic improvement over the traditional QC method suggests that the Huberization method of QC for observational data sets known to have large innovation errors is an effective alternative for QC within the framework of an EnKF to improve the analysis of TCs.

Five-day forecasts were started every 12 hours from the global and regional LETKF analyses. The results from these experiments were also binned according to Best Track intensity estimates at verification time. The Combined experiment improves intensity forecasts for strong storms (Fig. A.27) through the first five

forecast days and moderate intensity storms through the first 3 forecast days. The Combined analysis improves the long lead time regional track forecasts of moderate and strong storms, as well (Fig. A.28). There is neither degradation nor improvement in the track and intensity forecasts of weak storms, most likely due to the higher signal-to-noise ratio - a problem exacerbated by the fact that weak TCs do not always warrant the issuance of a TCVital observation. The results for the strong and moderate storms, however, are encouraging.

4. CONCLUSIONS

First, we investigated the benefits of employing a limited-area data assimilation system to enhance the lower-resolution global analyses in the Northwest Pacific TC basin. While several authors used ensemble-based data assimilation systems in the past to downscale information about the background uncertainty in limited-area systems employing multiple nests in the TC regions (Torn and Hakim 2009 and Zhang et al. 2011), to the best of our knowledge, ours is the first study to use the approach in a setting where the outermost nest is a global model. The resolutions considered here were much lower than what would be necessary to resolve the inner core of the TCs and observations of the inner core were not assimilated. Since studies by others have shown that such capabilities are necessary to achieve significant reductions in the analyses and the ensuing forecasts of the intensity (e.g., Zhang et al. 2011), we did not expect to find significant differences between the quality of the limited-area and the global intensity analyses.

We found that the limited-area data assimilation system enhanced the accuracy of the analysis of the position of the storms, but the benefits of increasing the resolution beyond 100 km were limited. (The particular value of the critical resolution would most likely change with the resolution of the global model and the observational data sets assimilated.) Two factors contributed to the higher accuracy of the limited-area analysis:

- in the case of strong (Category 3-5) storms, the assimilation of observations in the global system often degrades the accuracy of the analysis, while the effect of the assimilation of observations in the limited-area system is closer to neutral (last column of Fig. A.8);

- in the case of the moderate (Category 1-2) and weak (Tropical storms and depressions) storms, the use of the limited-area system greatly reduces the number of unusually large (statistical outlier) errors (Fig. A.6), because the assimilation of the observations can efficiently correct the large background errors in the limited-area systems (bottom three rows in the first and middle columns of Fig. A.8).

We also found that the limited-area system improved the prediction of the storm tracks for the first five forecast days. Our analysis shows that the forecast improvement is due to the more accurate analysis of the position of the storms and the better representation of the interactions between the storm and their immediate environment.

Can we expect, based on the results of the present study, that the global analysis would benefit from feeding back information from the limited-area system using the joint states approach of Yoon et al. (2012)? First, the higher accuracy of the analysis of the steering flow in the limited-area analyses suggests that the higher resolution of the limited-area analysis leads to a better interpretation of the observations in the case of moderate and weak storms. This result suggests that the information provided by the limited-area system about the position for the global system is potentially of lower value for the strong storms, for which the analysis error is closer to the discretization error than for the weaker storms.

In Section two, our focus turned toward the treatment and assimilation of direct observations of TCs. Several approaches for assimilating TCVitals SLP, QuikSCAT 10 m wind components, and reconnaissance dropsondes were studied within our coupled ensemble-based global-limited-area analysis/forecast system. These observations have an inherently large representativeness error. Experiments were conducted to determine

how best to address these errors, exploring options in which the representativeness error is defined as a component of the observation error, or through the QC of the observations. Findings indicate that the best analysis is achieved when all observations are assimilated with an online QC method that does not discard valuable observations, and not necessarily by defining observation error with a representativeness error component.

To the best of our knowledge this is the first application of Huberization method of Roh et al. (2013) to a real-world EnKF analysis. Huberization clips observation innovations to a prescribed maximum so that observations that are deemed outliers may still contribute to the update step of the assimilation process, but not with their full potential impact. While the optimal clipping height may be estimated by simulated observation experiments, as outlined in Roh et al. (2013), we found a reasonable value by numerical experimentation with real observations.

Implementing a Combined method that clips the special TC observations via Huberization when multiple observation types are unavailable, and keeping the TCVital observation when other special observations are present, showed drastic systematic improvements for strong and moderate storm analyses and forecasts. Intensity analyses and forecasts saw large improvements relative to our control experiment, which assimilated no special TC observations, and even outperformed the NCEP analysis benchmark in some cases. While the track improvements were more modest, there were definite improvements for position error in both the analyses and forecasts. There was also evidence that the Combined experiment utilized the observation in such a way that the pattern of circulation around the center of the TC was improved, which could have a secondary impact on improving the TC forecasts.

The systematic improvements seen in the analyses and forecasts of moderate and strong TCs are encouraging for the application of the Huberization method within an

EnKF system for observations with large observation innovations, such as those that directly sample tropical cyclones.

REFERENCES

Aberson, S. D., 2003: Targeted observations to improve operational tropical cyclone track forecast guidance. *Mon. Wea. Rev.*, **131**, 1613--1628.

Aberson, S. D., 2008: Large forecast degradations due to synoptic surveillance during the 2004 and 2005 hurricane seasons. *Monthy Weather Review*, **136**, 3138--3150.

Aravéquia, J. A., I. Szunyogh, E. J. Fertig, E. Kalnay, D. Kuhl, and E. J. Kostelich, 2011: Evaluation of a strategy for the assimilation of satellite radiance observations with the local ensemble transform Kalman filter. *Monthy Weather Review*, **60**, 103--118.

Atangan, J. F., A. Preble, A. C. Bryant, and Y. Pitts, 2004: 2004 annual tropical cyclone report. Tech. rep., U.S. Naval Pacific Meteorology and Oceanography Center/ Joint Typhoon Warning Center, Pearl Harbor, Hawaii.

Chen, Y. and C. Snyder, 2007: Assimilating vortex position with an ensemble Kalman filter. *Monthy Weather Review*, **135**, 1828--1845.

Chou, K.-H. and C.-C. Wu, 2008: Typhoon initialization in a mesoscale model - combination of the boguesed vortex and the dropwindsonde data in DOTSTAR. *Mon. Wea. Rev.*, **136**, 865--879.

Chou, K.-H., C.-C. Wu, P.-H. Lin, S. D. Aberson, M. Weissmann, F. Harnisch, and T. Nakazawa, 2011: The impact of dropwindsonde observations on typhoon track forecasts in DOTSTAR and T-PARC. *Monthy Weather Review*, **139**, 1728--1743.

Chu, J., C. R. Sampson, A. S. Levine, and E. Fukada, 2002: The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945-2000. JTWC Report NRL/MR/75540-02-16, Joint Typhoon Warning Center.

Cooper, G. A., R. J. Falvey, and M. Vancas, 2008: 2008 annual tropical cyclone

report. Tech. rep., U. S. Naval Pacific Meteorology and Oceanography Center/ Joint Typhoon Warning Center, Pearl Harbor, Hawaii.

Durran, D. R., 2010: *Numerical Methods in Fluid Dynamics with Applications to Geophysics*. 2d ed., Springer.

Fiorino, M. and R. L. Elsberry, 1989a: Some aspects of vertical structure related to tropical cyclone motion. *Journal of Atmospheric Sciences*, **46**, 975--990.

Fiorino, M. and R. L. Elsberry, 1989b: Some aspects of vertical structure related to tropical cyclone motion. *JAS*, **117**, 721--727.

Grasso, L. D., 2000: The differentiation between grid spacing and resolution and their application to numerical modeling. *Bulletin of the American Meteorological Society*, **81**, 579--580.

Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Monthy Weather Review*, **139**, 668--688.

Hoffman, R. N., 2011: Neutral stability height corrections for winds. Tech. rep., AER, Inc. ArXiv:1107.1416 [physics.ao-ph].

Holland, G. J., 1980: An analytic model of the wind and pressure profiles in hurricanes. *Monthy Weather Review*, **108**, 1212--1218.

Holt, C. R., I. Szunyogh, and G. Gyarmati, 2013: Can a moderate-resolution limited-area data assimilation system add value to the global analysis of tropical cyclones? *Monthy Weather Review*, **141**, 1866--1883.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112--126.

Isaksen, L. and P. A. E. M. Janssen, 2004: Impact of ERS scatterometer winds in ECMWF's assimilation system. *Q. J. R. Meteorol. Soc.*, **130**, 1793--1814.

Juang, H.-M. and M. Kanamitsu, 1994: The NMC nested regional spectral model. *Monthy Weather Review*, **122**, 3--26.

Källén, E., 2012: Progress in medium-range weather forecasting at ECMWF. *Aksel Wiin-Nielsen Symposium*, New Orleans, LA. Available online at [http://ams.confex.com/ams/92Annual/flvgateway.cgi/id/20368?recordingid=20368].

Keyser, D., 2007: Format of tropical cyclone vital statistics records "TCVITALS". Tech. rep., NCEP Environmental Modeling Center. URL `http://www.emc.ncep.noaa.gov/mmb/data_processing/tcvitals_description.htm`.

Kimball, S. K. and M. S. Mulekar, 2004: A 15-year climatology of North Atlantic tropical cyclones. Part I: Size parameters. *Journal of Climate*, **17**, 3555--3575.

Kunii, M., T. Miyoshi, and E. Kalnay, 2012: Estimating the impact of real observation in regional numerical weather prediction using an ensemble Kalman filter. *Monthy Weather Review*, **140**, 1975--1987.

Leidner, S. M., L. Isaksen, and R. N. Hoffman, 2003: Impact of NSCAT Winds on tropical cyclones in the ECMWF 4DVAR assimilation system. *Montlhy Weather Review*, **131**, 3--26.

Liu, Q., T. Marchok, H.-L. Pan, M. Bender, and S. Lord, 2000: Improvements in hurricane initialization and forecasting at NCEP with global and regional (GFDL) models. Office Note #472, NWS, http://www.nws.noaa.gov/om/tpb/472body.htm.

Lungu, T., 2006: *QuikSCAT Science Data Product User's Manual: Overview and geophysical data products*. Jet Propulsion Laboratory California Institute of Technology, 3d ed.

Massey, F. J., 1951: The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, **46 (253)**, 68--78.

Merkova, D., I. Szunyogh, and E. Ott, 2011: Strategies for coupling global and limited-area ensemble Kalman filter assimilation. *Nonlinear Processes in Geophysics*, **18**,

415--430.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, and E. J. Kostelich, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415--428.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthy Weather Review*, **12**, 1747--1763.

Pu, Z., X. Li, C. S. Velden, S. D. Aberson, and W. T. Liu, 2008: The impact of aircraft dropsonded and satellite wind data on numerical simulations of two landfalling tropical storms during the Tropical Cloud Systems and Processes Experiment. *Weather and Forecasting*, **23**, 62--79.

Rappaport, E. N., et al., 2009: Advances and challenges at the National Hurricane Center. *Weather and Forecasting*, **24**, 395--419.

Rogers, E., et al., 2009: The NCEP north american mesoscale modeling system : Recent changes and future plans. *Preprints*, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE [Available online at http://ams.confex.com/ams/pdfpapers/154114.pdf.], American Meteorological Society, 2A4.

Roh, S., M. G. Genton, M. Jun, I. Szunyogh, and I. Hoteit, 2013: Observation quality control with a robust ensemble Kalman filter. *Monthly Weather Review*, **141**, 4414--4428.

Schwartz, C., Z. Liu, Y. Chen, and X.-Y. Huang, 2012: Impact of assimilating microwave radiances with a limited-area ensemble data assimilation system on forecasts of Typhoon Morakot. *Weather and Forecasting*, **27**, 424--437.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Monthy Weather Review*, **132**, 3019--3032.

Suzuki-Parker, A., 2012: *Uncertainties and Limitations in Simulating Tropical Cyclones*, chap. 2. Springer-Verlag Berlin Heidelberg.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, **60A**, 113--130.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke, 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the NCEP global model. *Tellus*, **57A**, 528--545.

Tavolato, C. and L. Isaksen, 2010: Huber norma qulaity control in the IFS. *ECMWF Newsletter*, **122**, 27--31.

Torn, R. D., 2010: Performance of a mesoscale ensemble Kalman filter (EnKF) during the NOAA high-resolution hurricane test. *Monthy Weather Review*, **138**, 4375--4392.

Torn, R. D. and G. J. Hakim, 2009: Ensemble data assimilation applied to RAINEX: Observations of Hurricane Katrina (2005). *Monthy Weather Review*, **137**, 2817--2829.

Trahan, S. and L. Sparling, 2012: An analysis of NCEP tropical cyclone vitals and potential. *Weather and Forecasting, in press.*

Velden, C., et al., 2006: A satellite-based method that has endured for over 30 years. *Bulletin of the American Meteorological Society*, 1195--1209.

Weissmann, M., et al., 2011: The influence of assimilating dropsonde data on typhoon track and midlatitude forecasts. *Monthy Weather Review*, **139**, 908--920.

Wu, C.-C., J.-H. Chen, P.-H. Lin, and K.-H. Chou, 2007a: Targeted observations of tropical cyclone movement based on the adjoint-derived sensitivity steering vector. *Journal of Atmospheric Sciences*, **64**, 2611--2626.

Wu, C.-C., K.-H. Chou, P.-H. Lin, S. D. Aberson, M. S. Peng, and T. Nakazawa, 2007b: The impact of dropwindsonde data on typhoon track forecasts in DOTSTAR. *Weather and Forecasting*, **22**, 1157--1176.

Yoon, Y., B. Hunt, and I. Szunyogh, 2012: Ensemble regional data assimilation using joint states. *Tellus, in print, available online at http://arxiv.org/abs/1108.0983.*

Zhang, F., Y. Weng, J. Gamache, and F. Marks, 2011: Performance of cloud-resolving hurricane initialization an prediction during 2008-2010 with ensemble data assimilation of inner-core airborne dopper radar observations. *Geophysical Research Letters*, **38**, L15 810.

APPENDIX

FIGURES

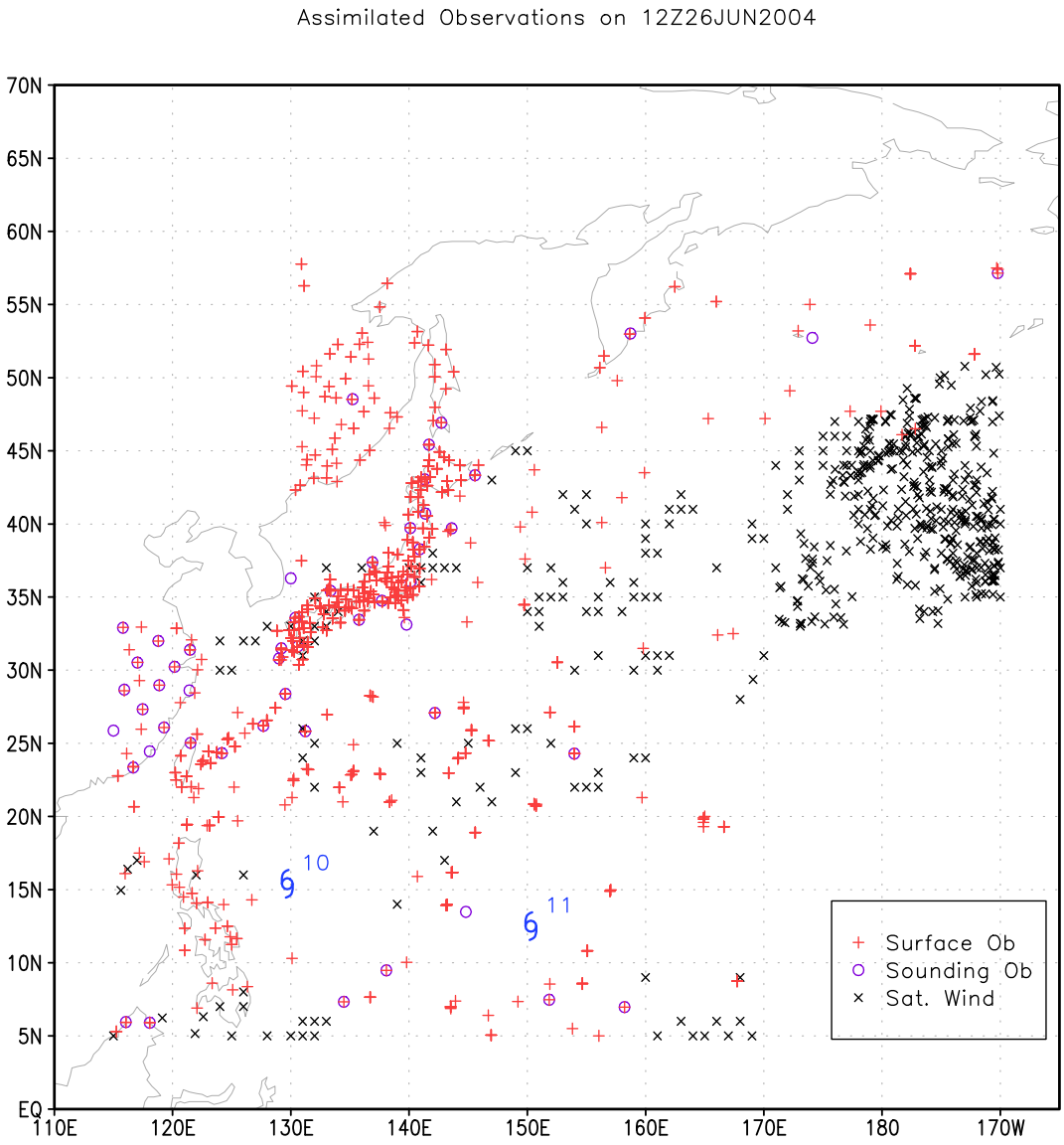

FIG. 1. Illustration of the observation coverage in the vicinity of the TC basin from June 26, 2004 at 1200 UTC. The surface and sounding observations are much more dense over land and were not included in this figure for clarity. The Best Track locations of Typhoon Mindulle (10) and Typhoon Tingting (11) are indicated by TC symbols.

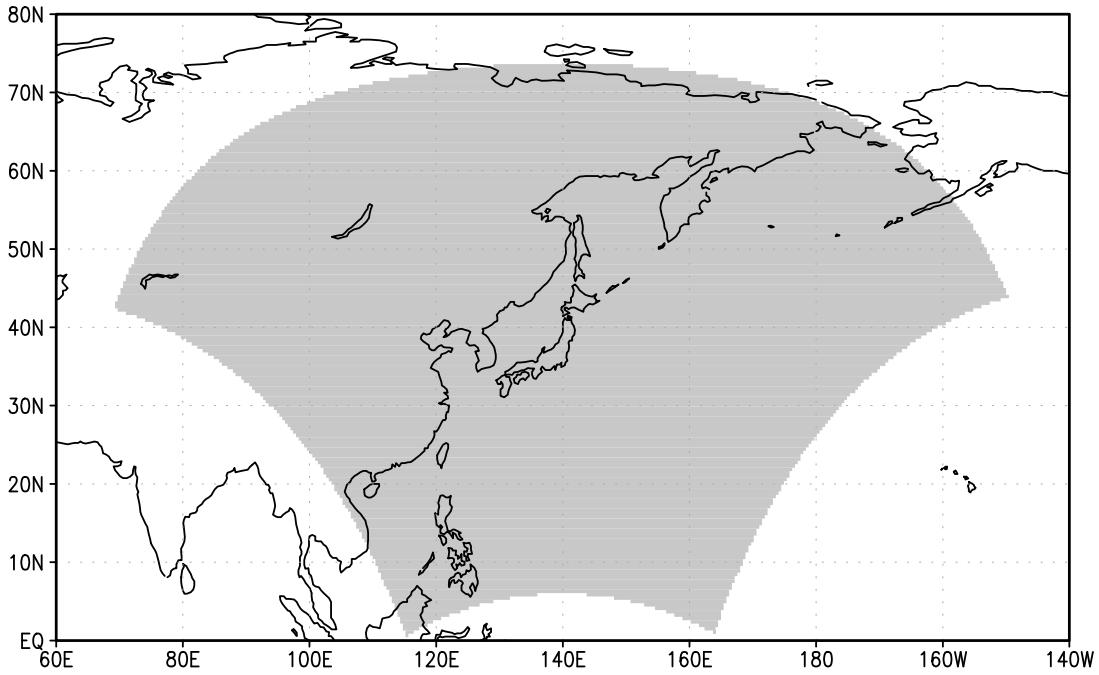

FIG. 2. Limited-area domain used for the limited-area analyses.

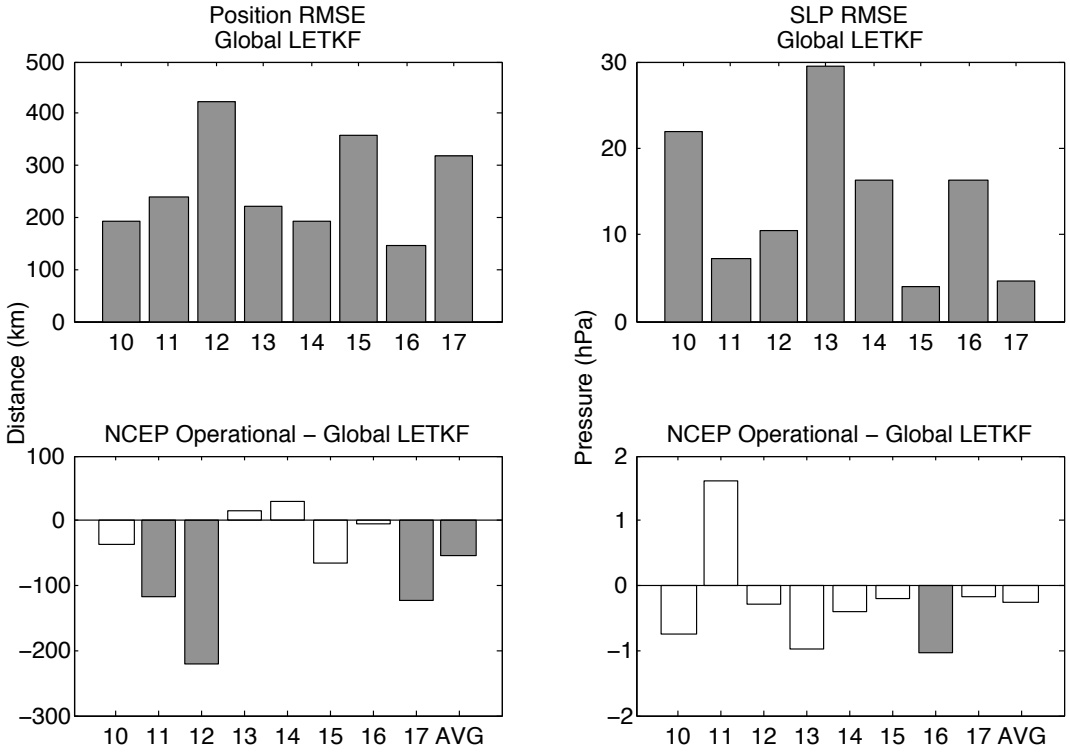

FIG. 3. Position and intensity root mean square error (RMSE) for the global LETKF analysis (a and b, respectively) and the differences between the position (panel c) and intensity (panel d) errors in the truncated operational analysis and the LETKF analysis. Each bar shows the RMSE error or the difference between the RMSE for a particular storm (numbered), where the mean is computed over the life cycle of the TC or the RMSE error difference averaged over all times for all storms (AVG in panels c and d). Shaded bars in panels c and d indicate storms for which the difference is statistically significant. For the numbering of the storms, see Table B.1. A positive value indicates that the LETKF analysis is more accurate than the reanalysis or the truncated operational analysis.

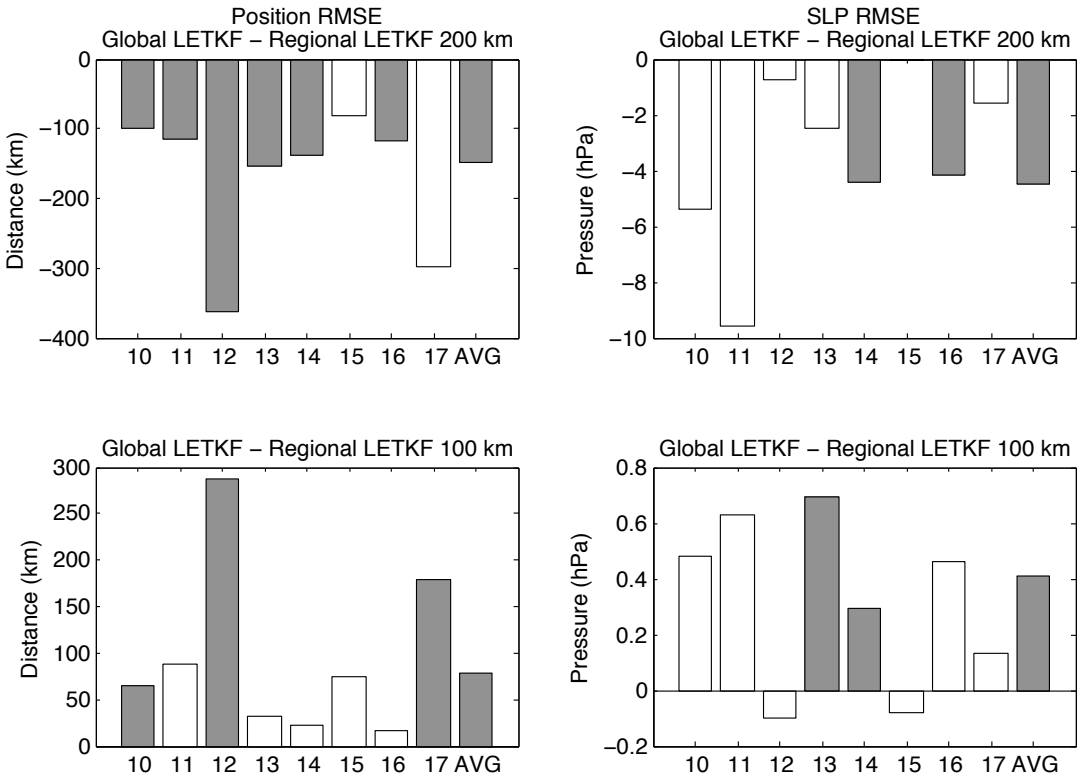

FIG. 4. Comparison between the errors in the coarse resolution limited-area LETKF analyses and the global LETKF analysis. The format of the figure is the same as that of Fig. A.3. A positive value indicates that the limited-area analysis is more accurate.

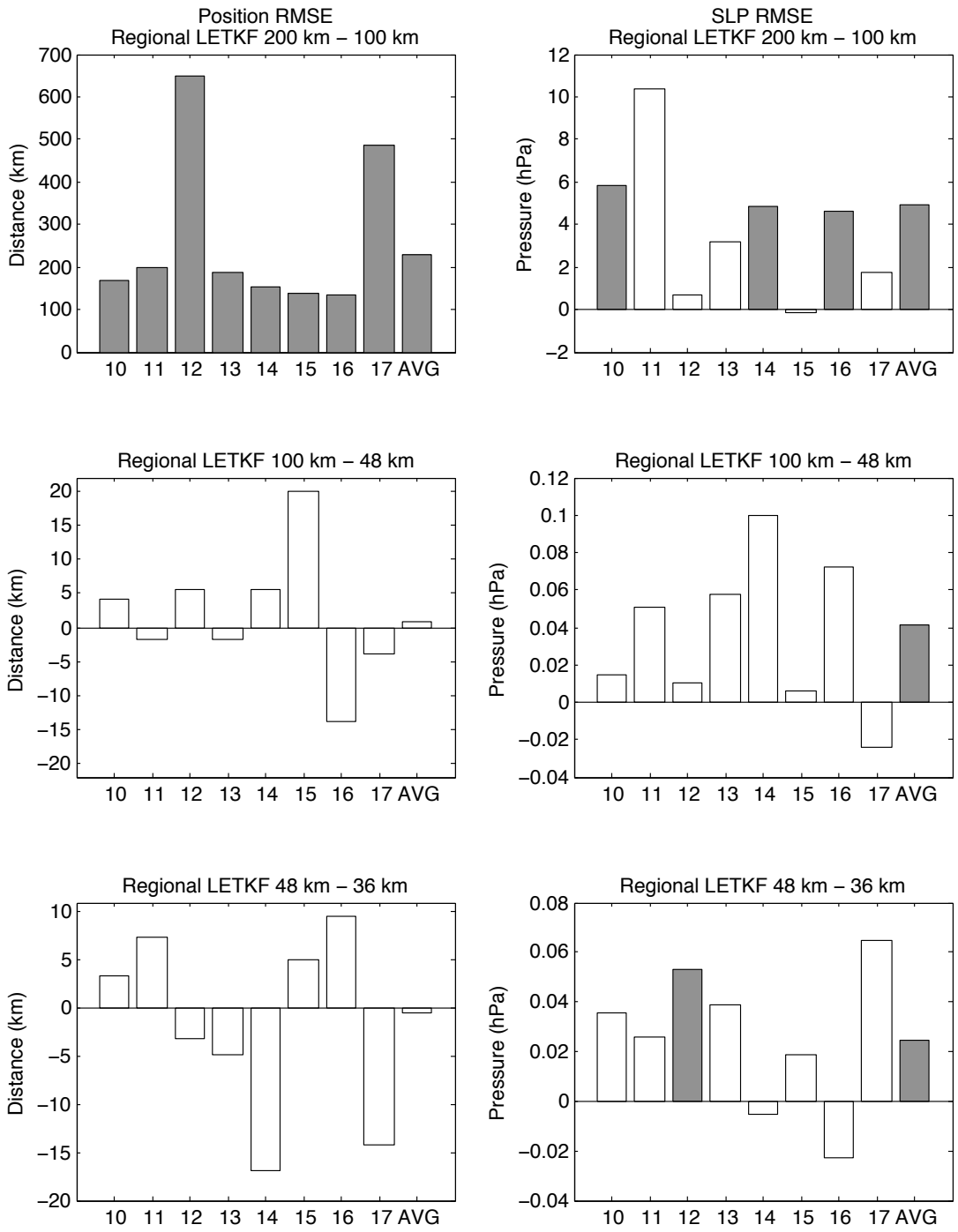

FIG. 5. Comparison between the errors for the limited-area analyses of different resolutions using the format of Fig. A.3. A positive value indicates that the higher resolution analysis is more accurate. The resolution increases from top to bottom.

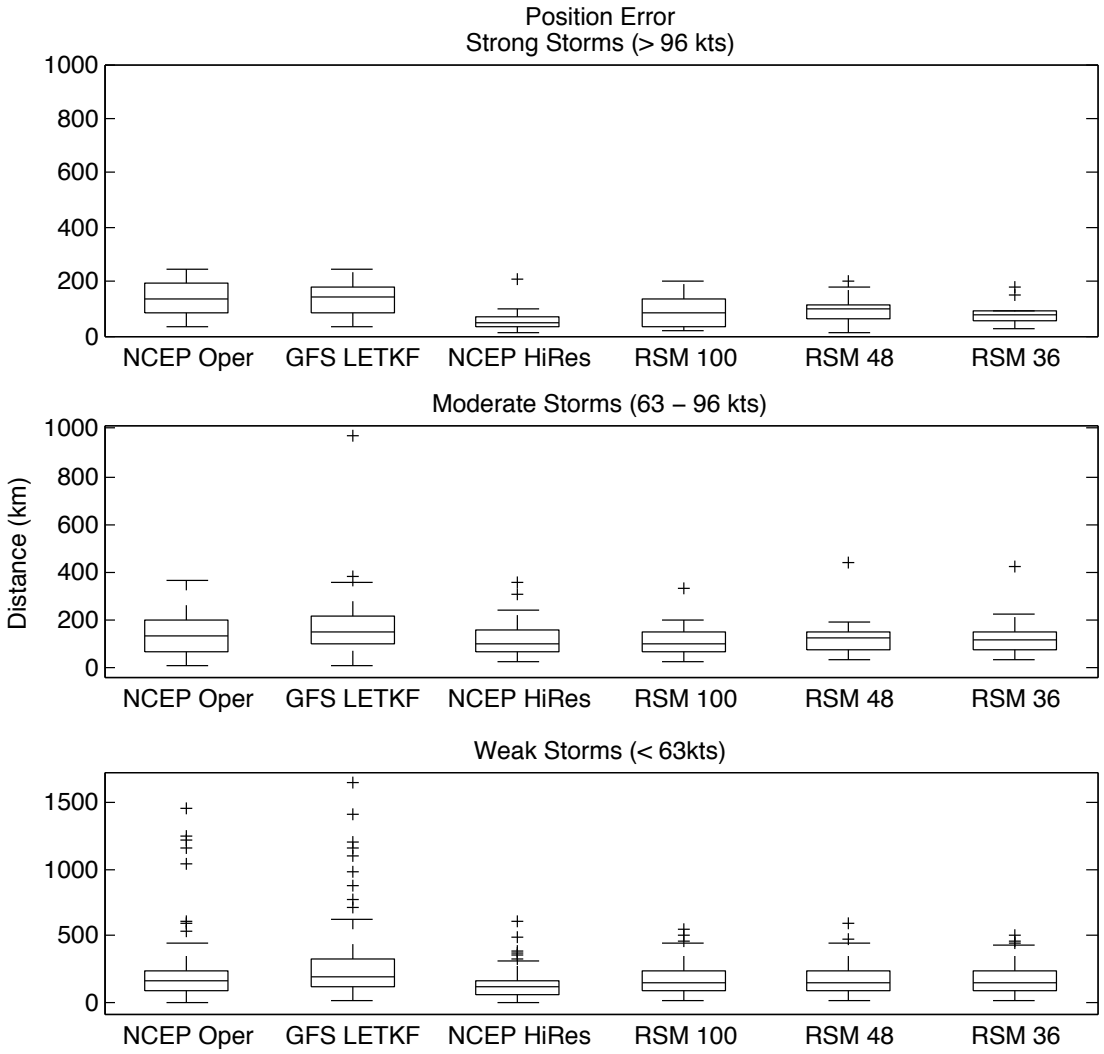

FIG. 6. Stratification of the distribution of the position analysis errors by storm intensity. The distributions are obtained by grouping errors for all Category 3 and 4 cyclones (top), Category 1 and 2 cyclones (middle), and tropical storms and depressions (bottom). Each box plot represents the distribution of the analysis errors for a different analysis system.

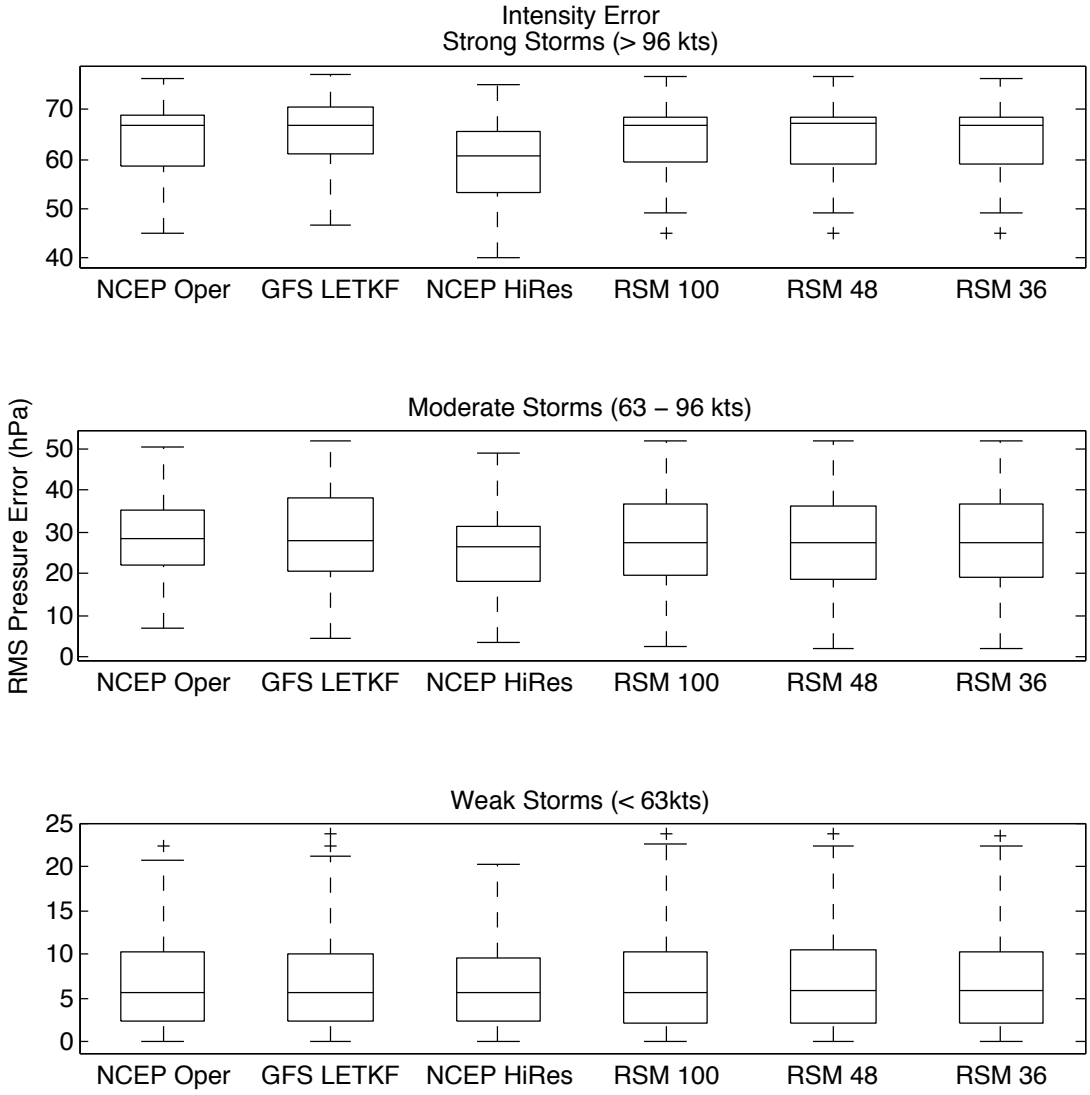

FIG. 7. Same as Fig. A.6 except for the intensity errors.

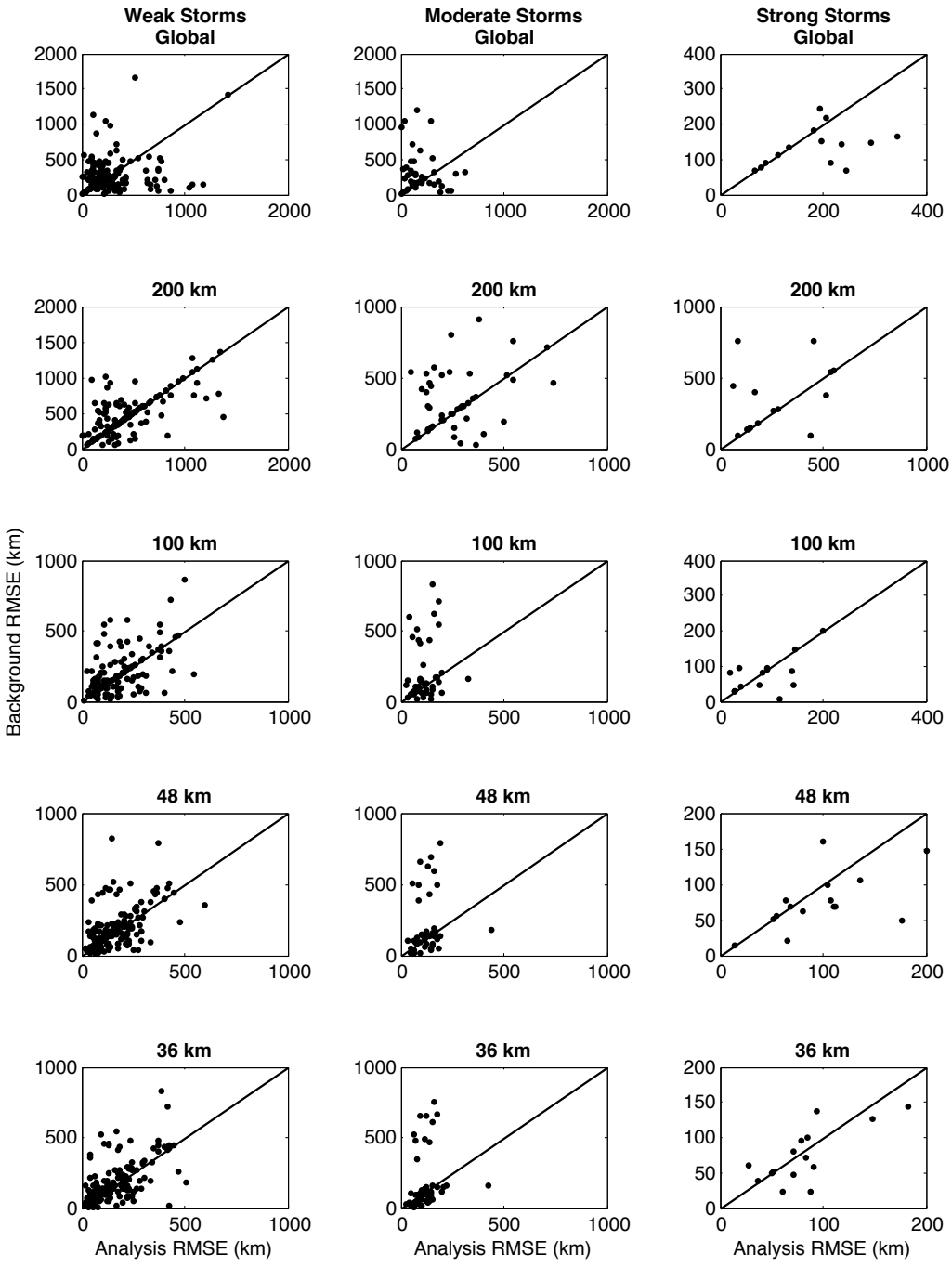

FIG. 8. Background versus analysis root-mean-square position error. Each dot indicates a pair of background and analysis errors for a single analysis. Results are shown for different configurations (rows) of the LETKF and for different storm intensities (columns).

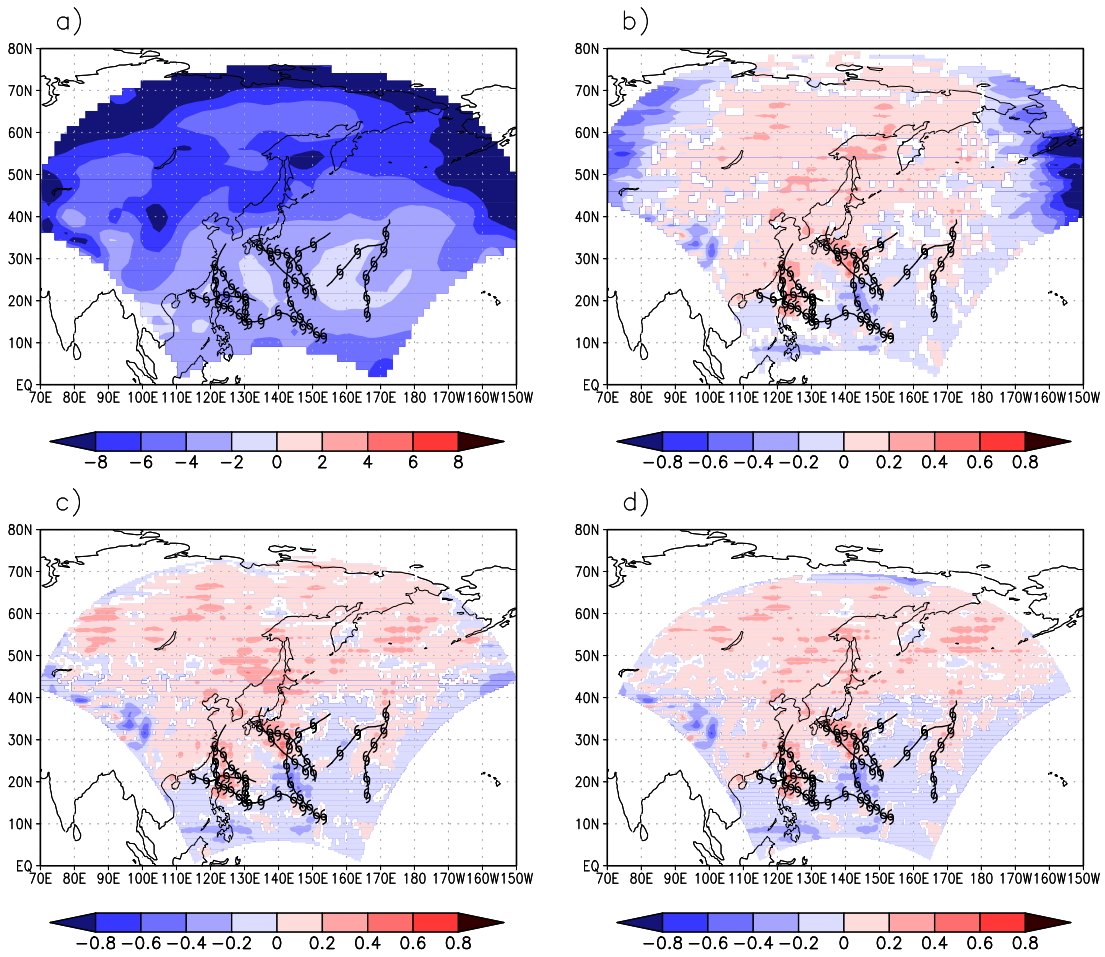

FIG. 9. Difference between the root mean square steering flow errors for the global analysis and the limited-area analyses at horizontal resolutions of a) 200 km, b) 100 km, c) 48 km, and d) 36 km. The mean in the computation of the root-mean-square error is taken over all verification times. Positive values indicate a superior regional analysis performance. The typhoon tracks have been marked. Note the difference in the shading interval between panel a) and the rest of the panels. Shading is only plotted for those values that indicate a significant difference between the mean of the time series at each grid point.

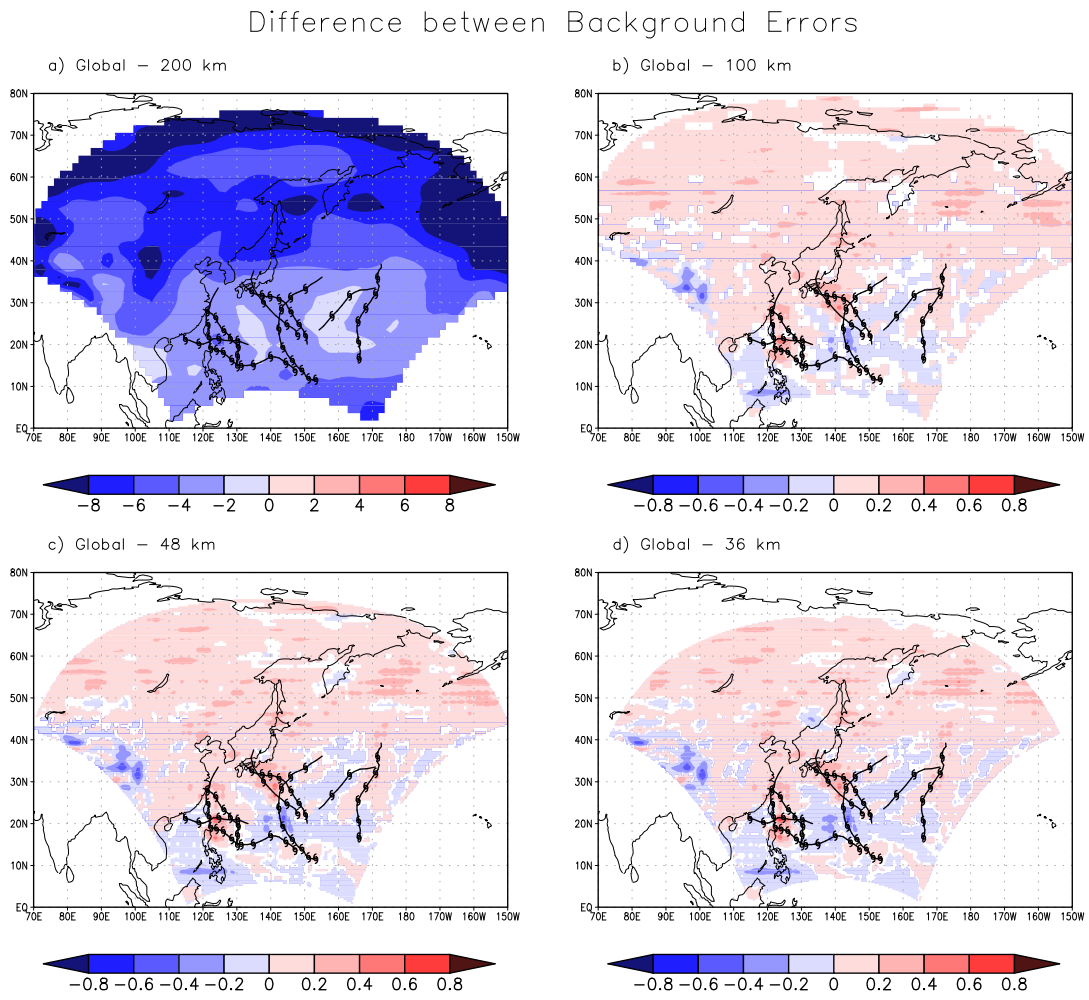

FIG. 10. Difference between the root mean square steering flow background errors for the global and limited-area backgrounds at horizontal resolutions of a) 200 *km*, b) 100 *km*, c) 48 *km*, and d) 36 *km*. Computation is the same as in Fig. A.9

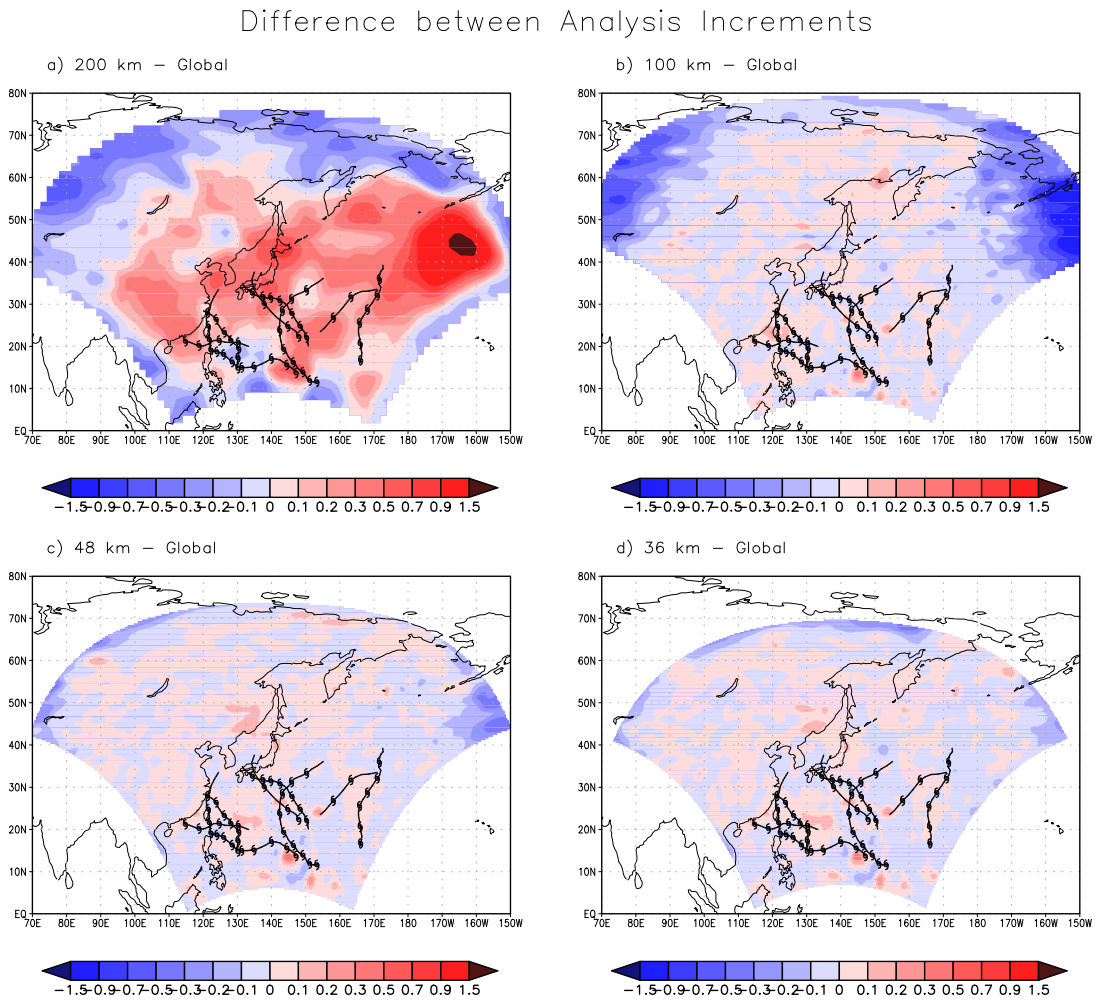

FIG. 11. The difference between the magnitudes of the background error reduction by the assimilation of observations in the global and the limited-area analyses. Results are shown for the limited-area analyses of a) 200 km, b) 100 km, c) 48 km, and d) 36 km resolution. Positive values indicate that the limited-area system reduced the background error more than the global system.

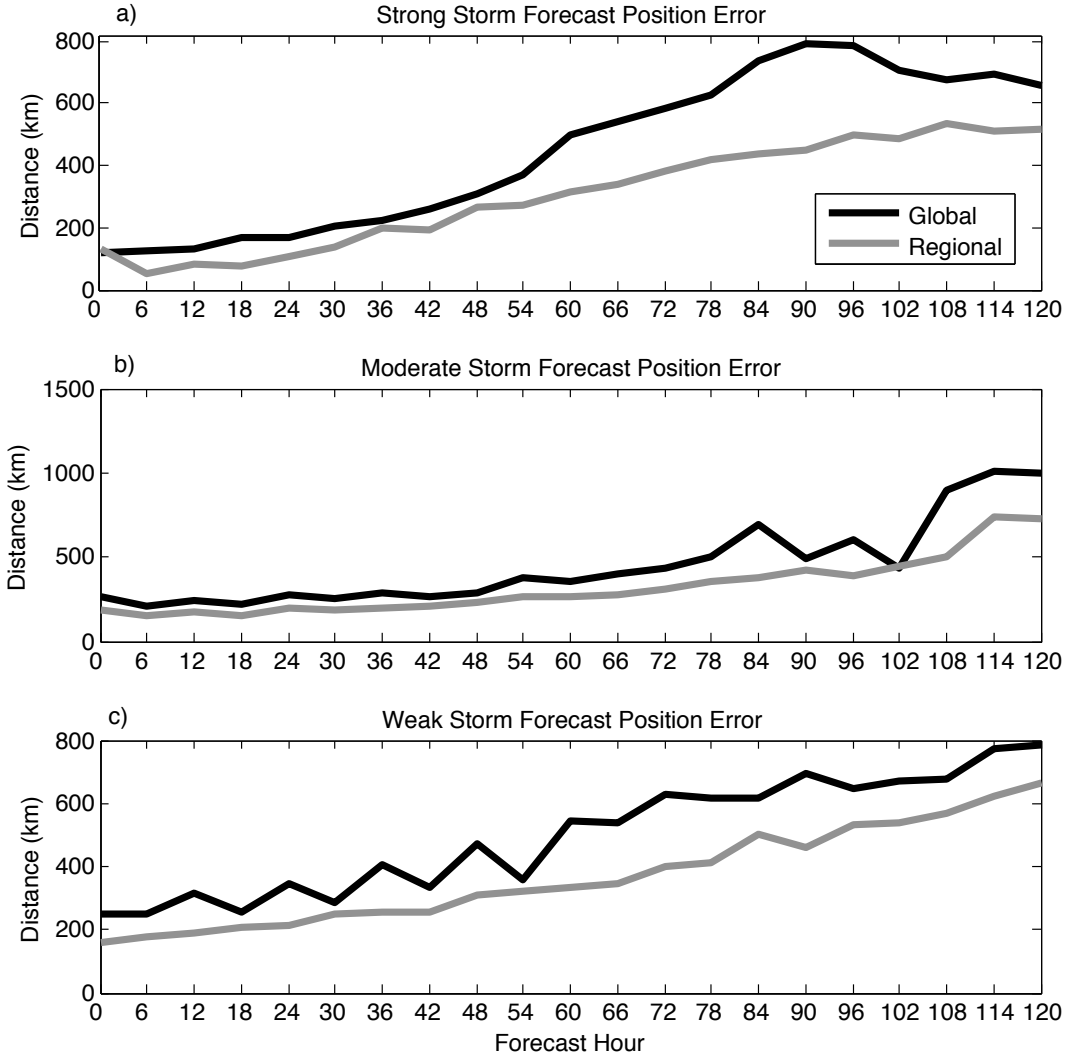

FIG. 12. Comparison between the evolution of the root-mean-square error in the global (black) and limited-area (gray) forecast position. Results are shown for the initially a) strong, b) moderate, and c) weak storms.

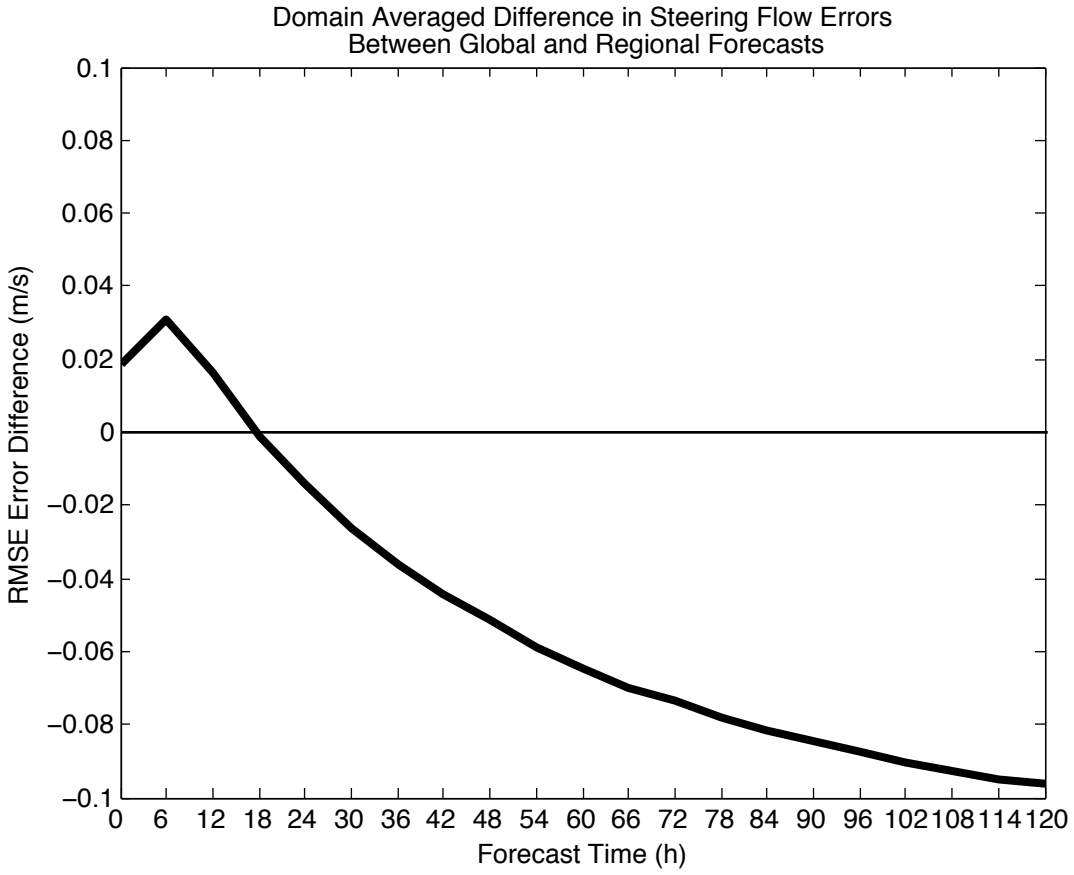

FIG. 13. Domain-averaged difference between the RMS errors in the global and limited-area forecasts of the steering flow. Positive values indicate a superior regional analysis performance.

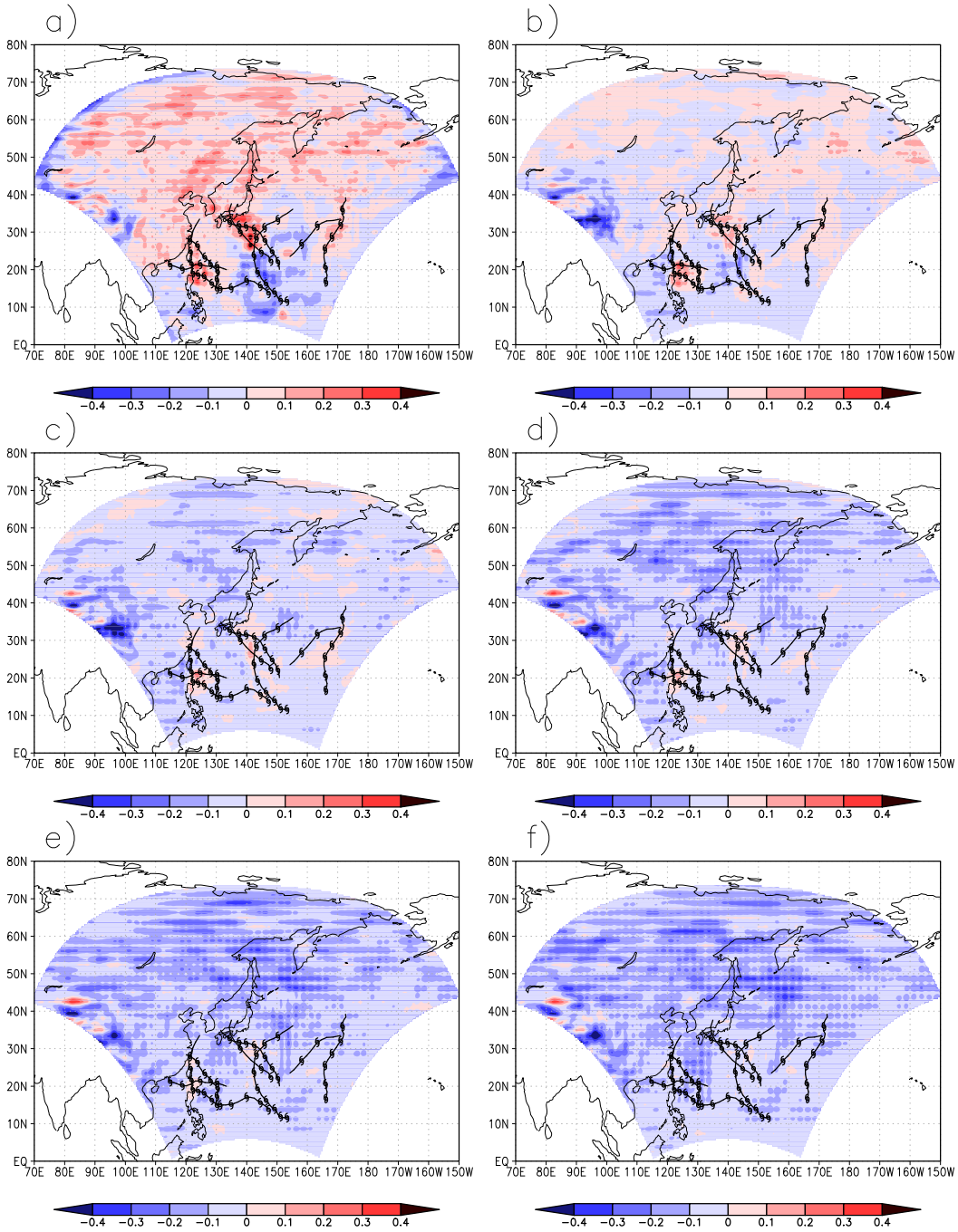

FIG. 14. Difference between the root mean square errors in the global and the limited-area forecasts of the steering flow at forecast times a) 0 hr, b) 24 hr, c) 48 hr, d) 72 hr, e) 96 hr, and f)120 hr. The mean in the computation of root-mean-square error is taken over all verification times. Positive values indicate a superior regional analysis performance. The typhoon tracks have been marked.

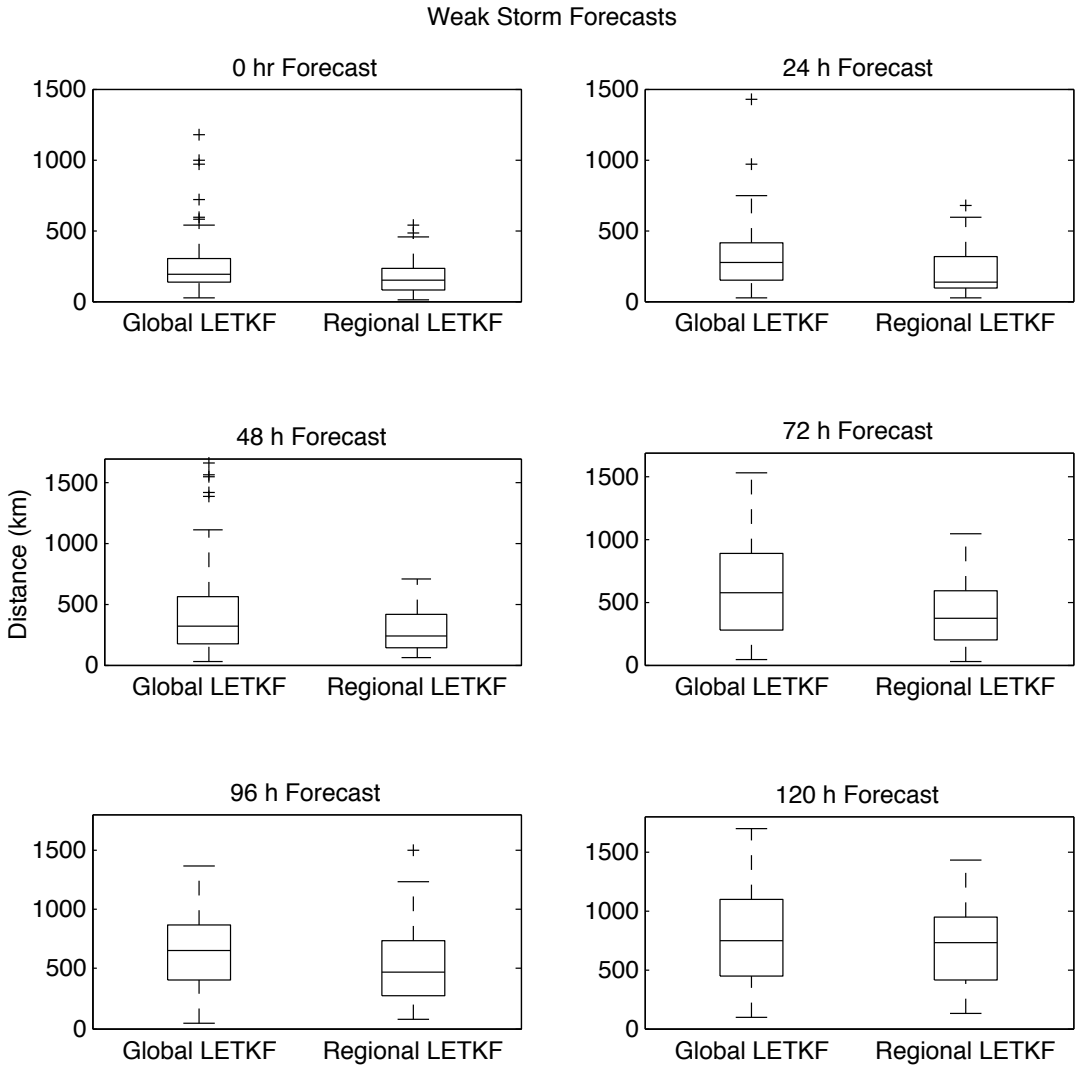

FIG. 15. Distributions of the errors in the forecast of the position of the initially weak storms.

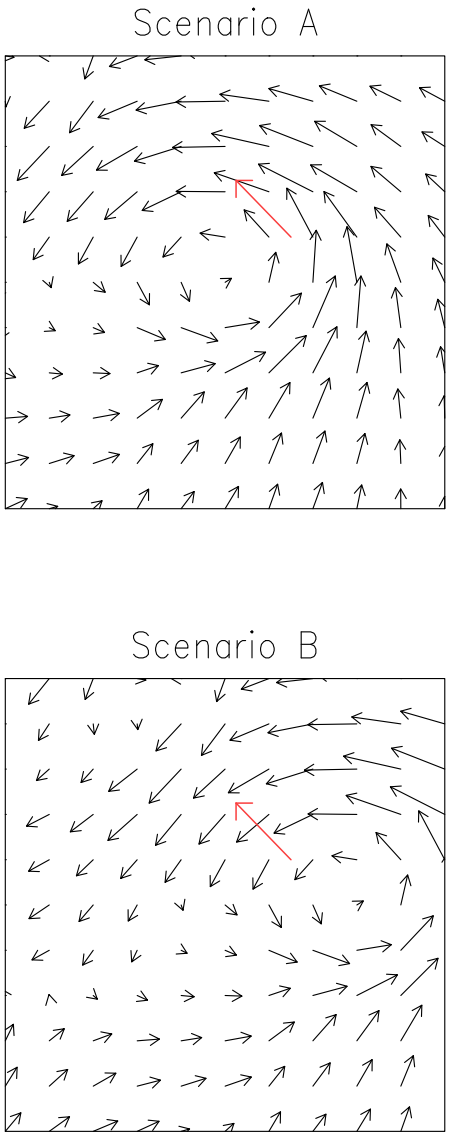

FIG. 16. Schematic of scenarios for background estimates of the wind field (black vectors) and a single wind observation (red vector).

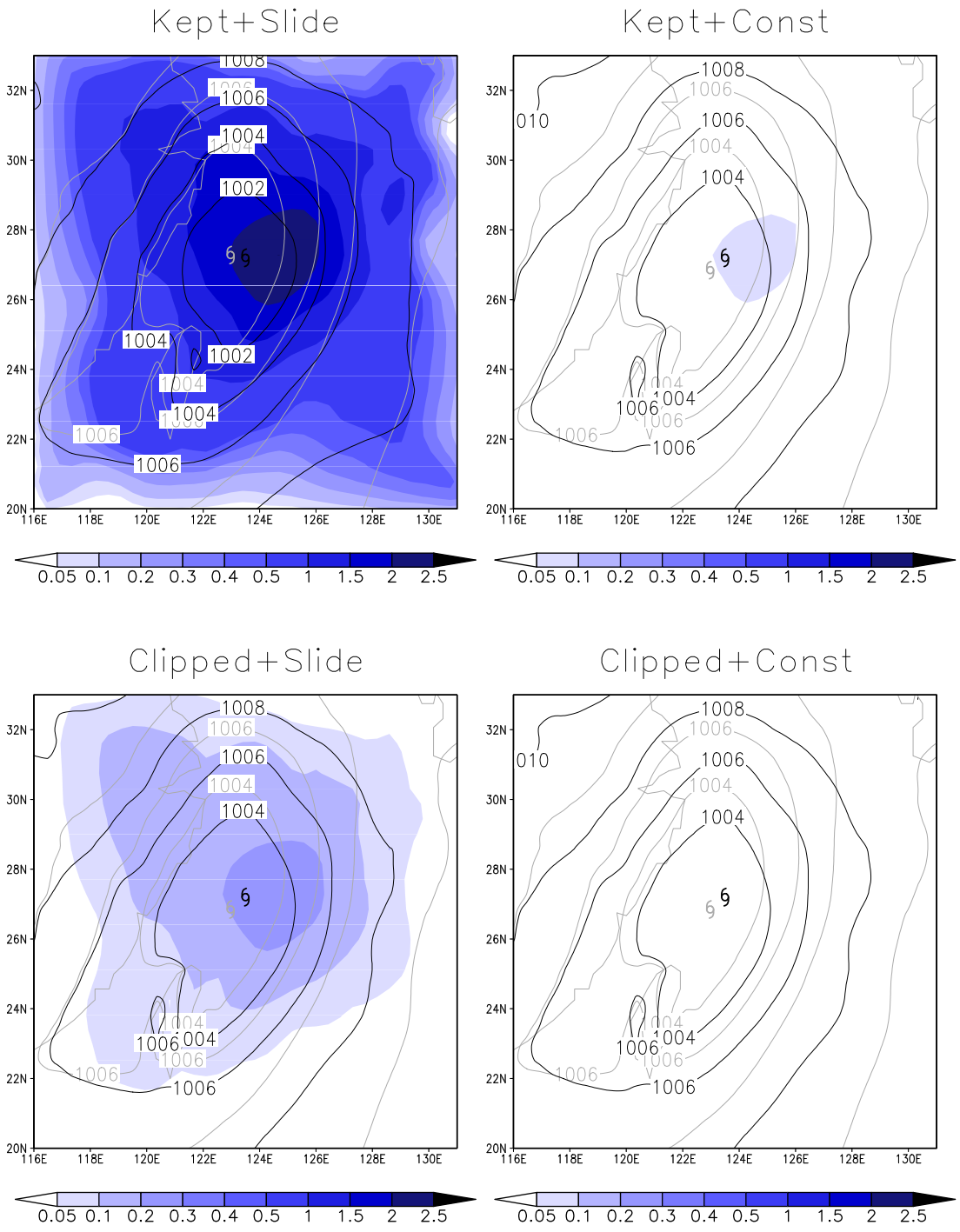

FIG. 17. Mean SLP for experiments assimilating TCVitals MSLP estimates. Shading is SLP observation impact (hPa), defined as experiment analysis minus control analysis. Contours represent the SLP analysis (black) and the background (gray). The tropical cyclone centers indicate the Best Track (black) and experiment (gray) positions.

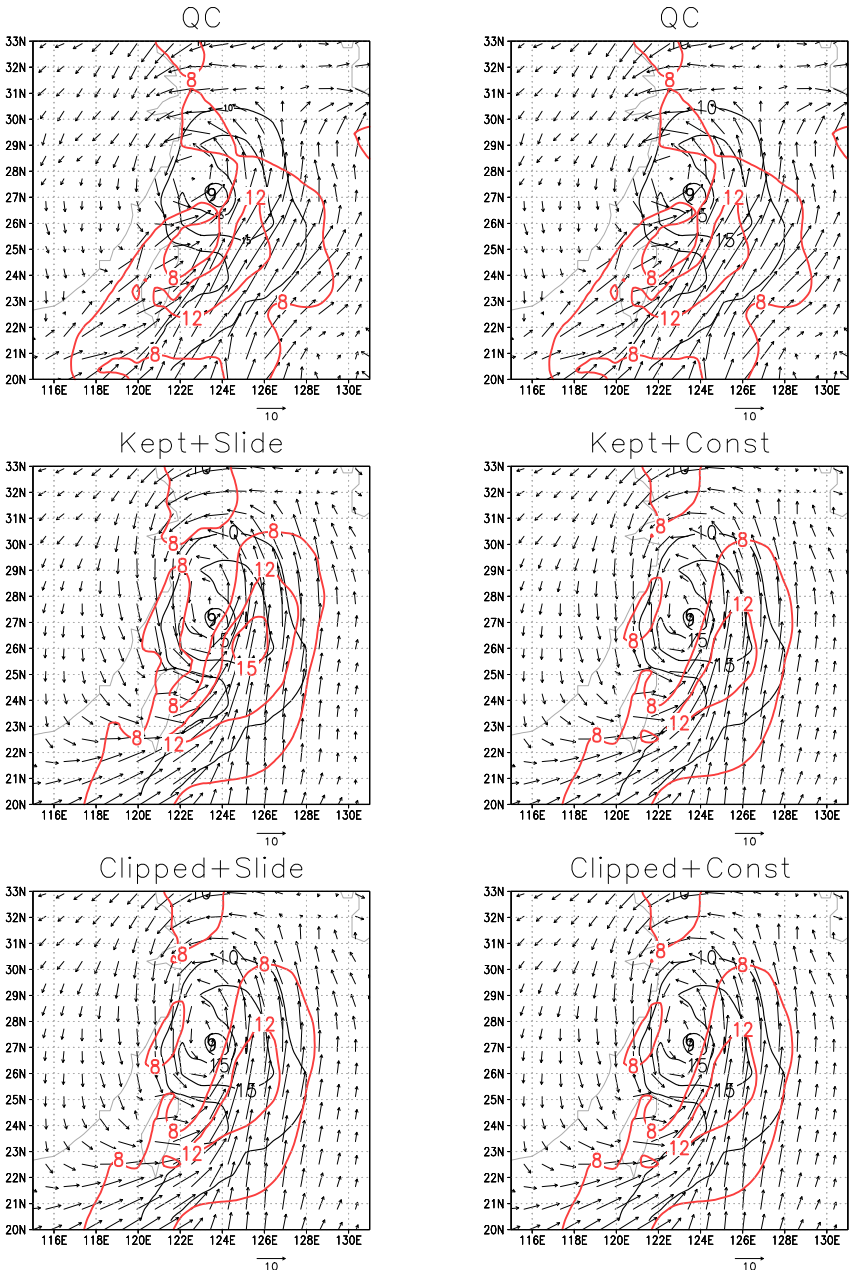

FIG. 18. First sigma level wind field (vectors) and speed (contoured; red is experiment, black is NCEP operational analysis). The top row is the traditional online quality control (QC), while the second row is the result of forcing the observations into the analysis with no QC (Kept), and finally the last row is the Huberized (Clipped) analysis. The left column assimilates TCVitals observations with Slide error (0.44 hPa), while the right column assimilates TCVitals observations with 3 hPa error. The tropical cyclone symbol indicates Best Track location.

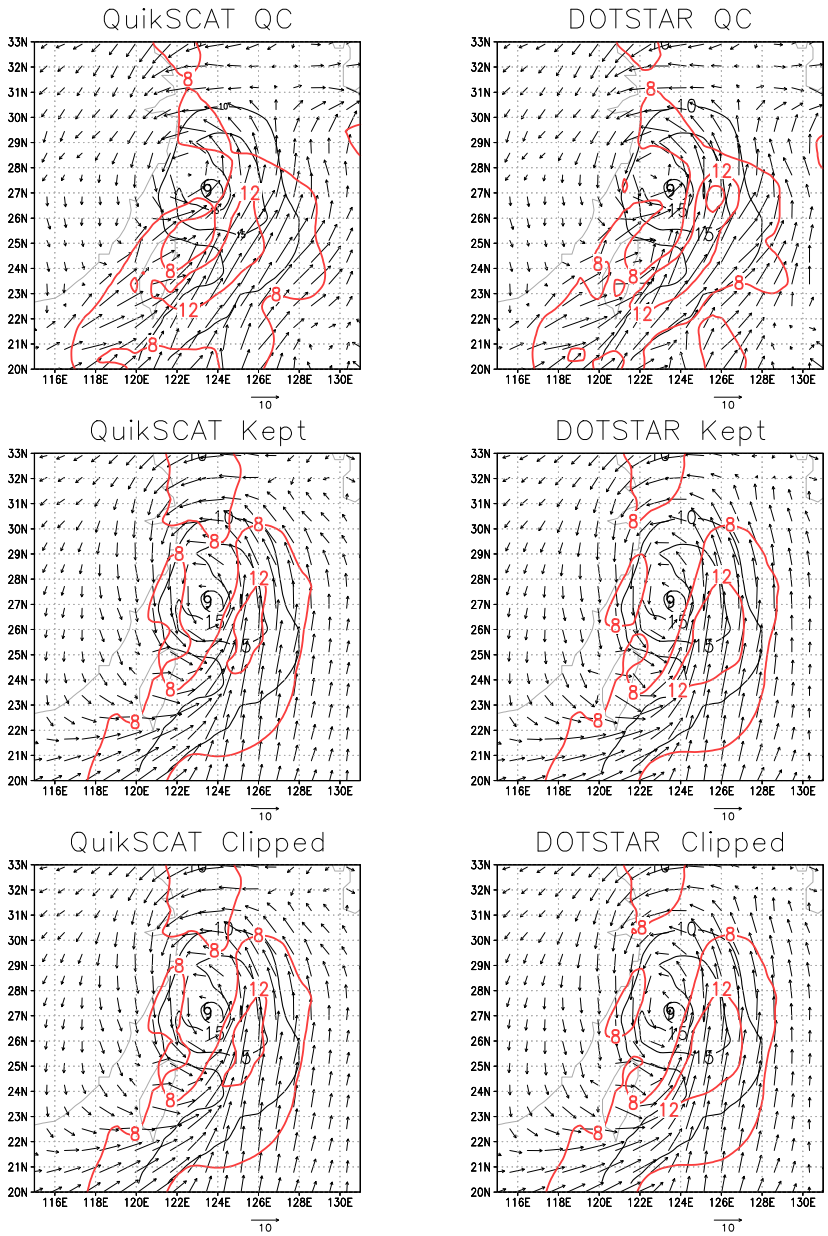

FIG. 19. First sigma level wind field (vectors) and speed (contoured; red is experiment, black is NCEP operational analysis). The top row is the traditional online quality control (QC), while the second row is the result of forcing the observations into the analysis with no QC (Kept), and finally the last row is the Huberized (Clipped) analysis. The left column assimilates only QuikSCAT special observations, while the right column assimilates only DOTSTAR special observations. The tropical cyclone symbol indicates Best Track location.

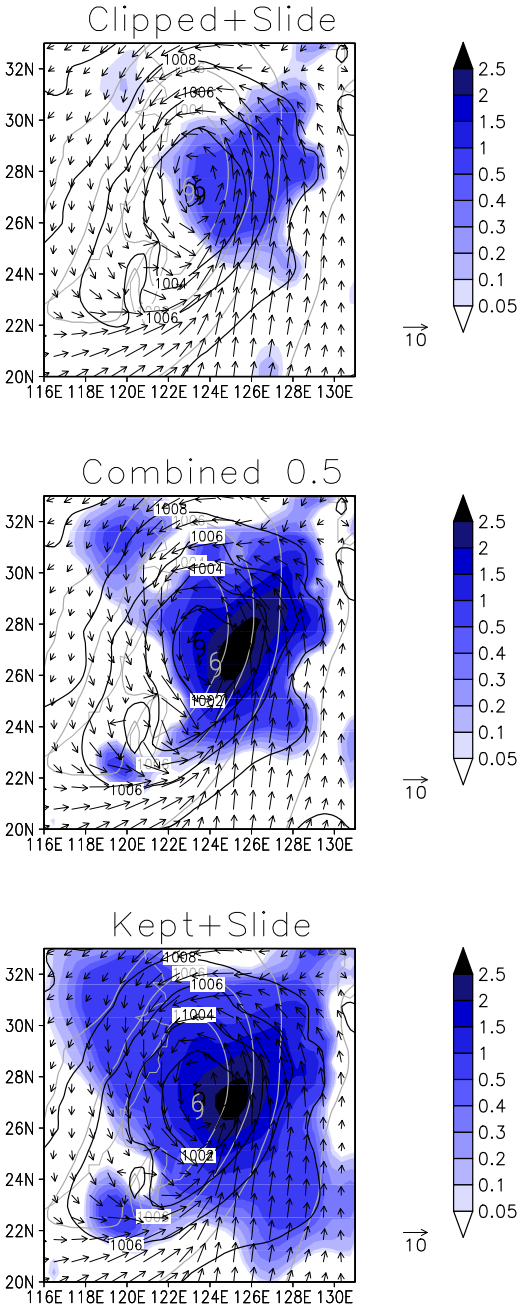

FIG. 20. MSLP and wind field for experiments assimilating all three types of TC observations. The top and bottom panels assimilate all observations with either Huberization (Clipped) or no QC (Kept) applied to each observation and Slide TCVital SLP error (0.44 hPa), while the middle panel applies the Combined method of QC for the observations. The countours, shading, and vectors are the same as in Fig. A.17.

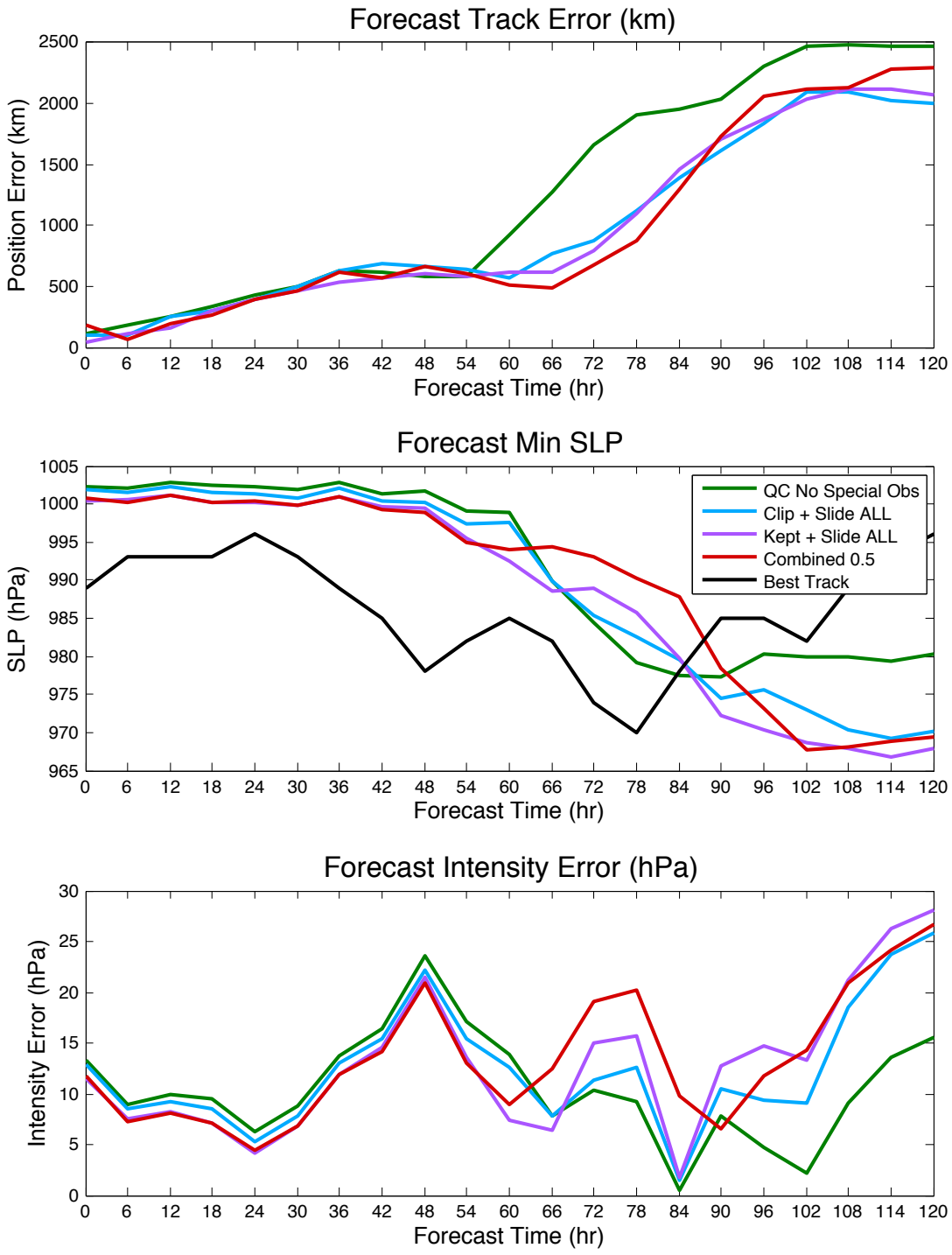

FIG. 21. Forecast track error (top), minimum SLP (middle), and minimum SLP error (bottom) for single update experiments.

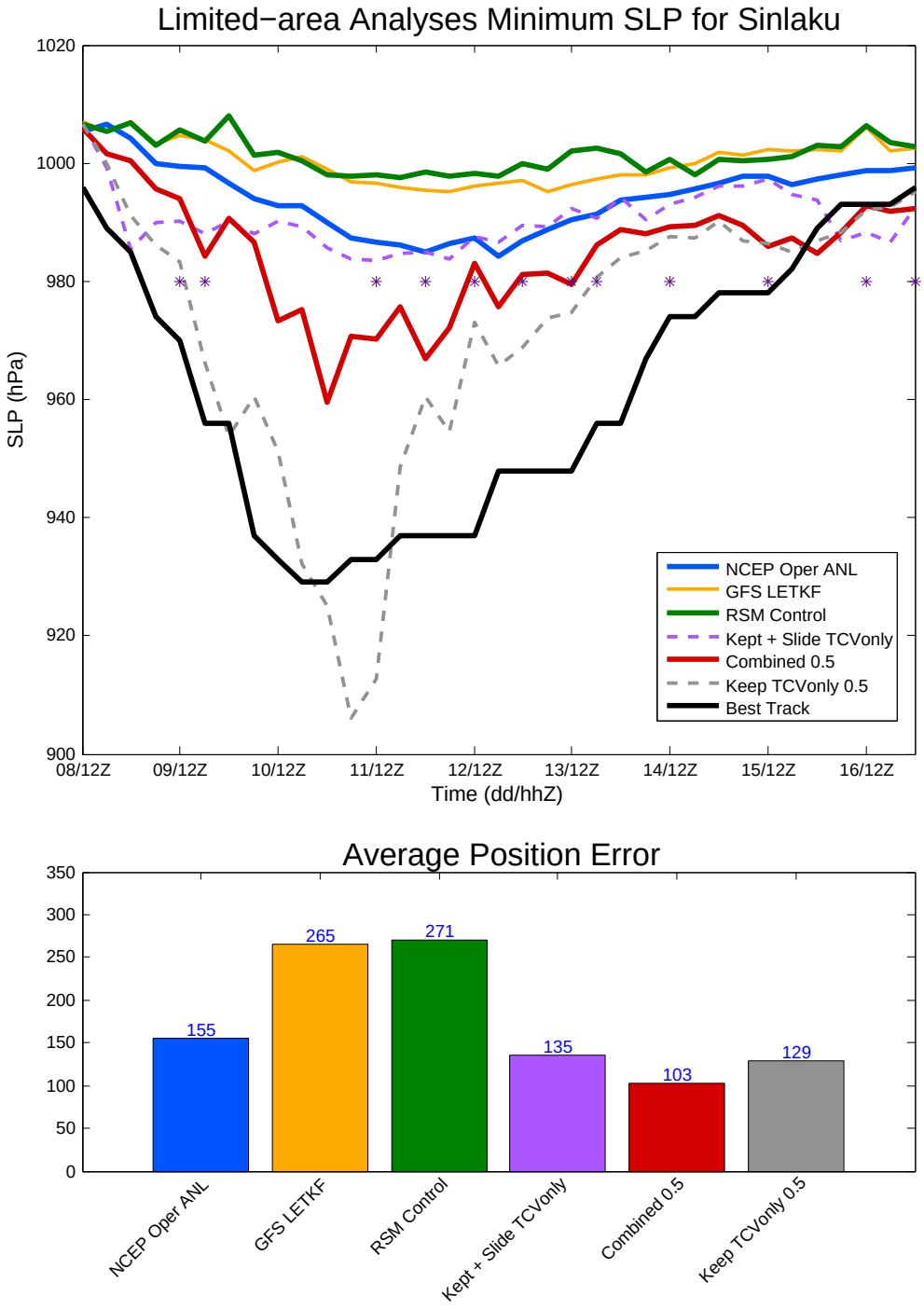

FIG. 22. Analyzed minimum SLP (top) and average position error over all analysis cycles (bottom) for Typhoon Sinlaku. Stars in the top panel indicate the times at which QuikSCAT observations were available near the TC.

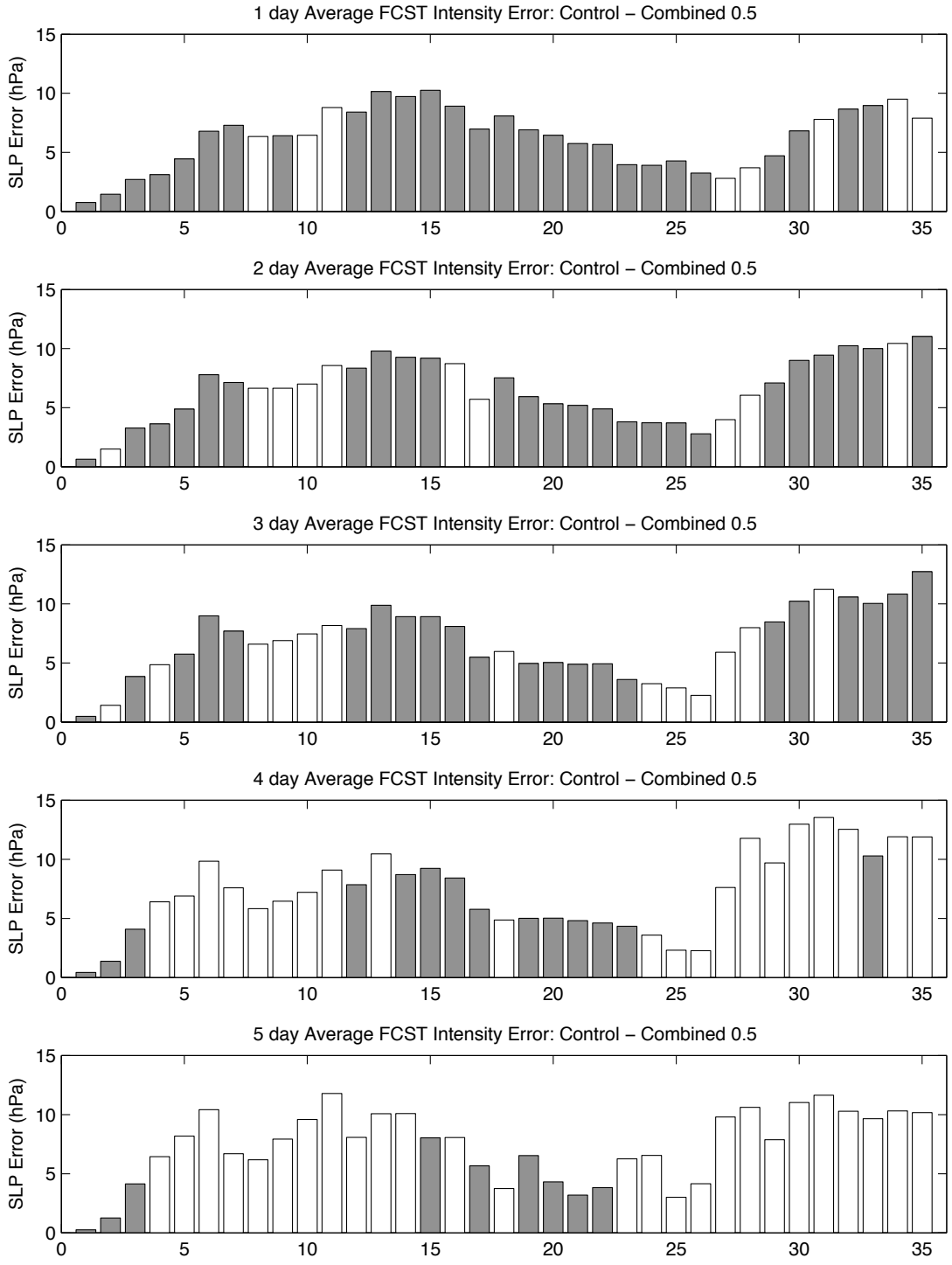

FIG. 23. Difference between daily forecast intensity error averages of the Control and Combined 0.5 experiments. Gray shading indicates that the difference is statistically significant at the 95% confidence level.

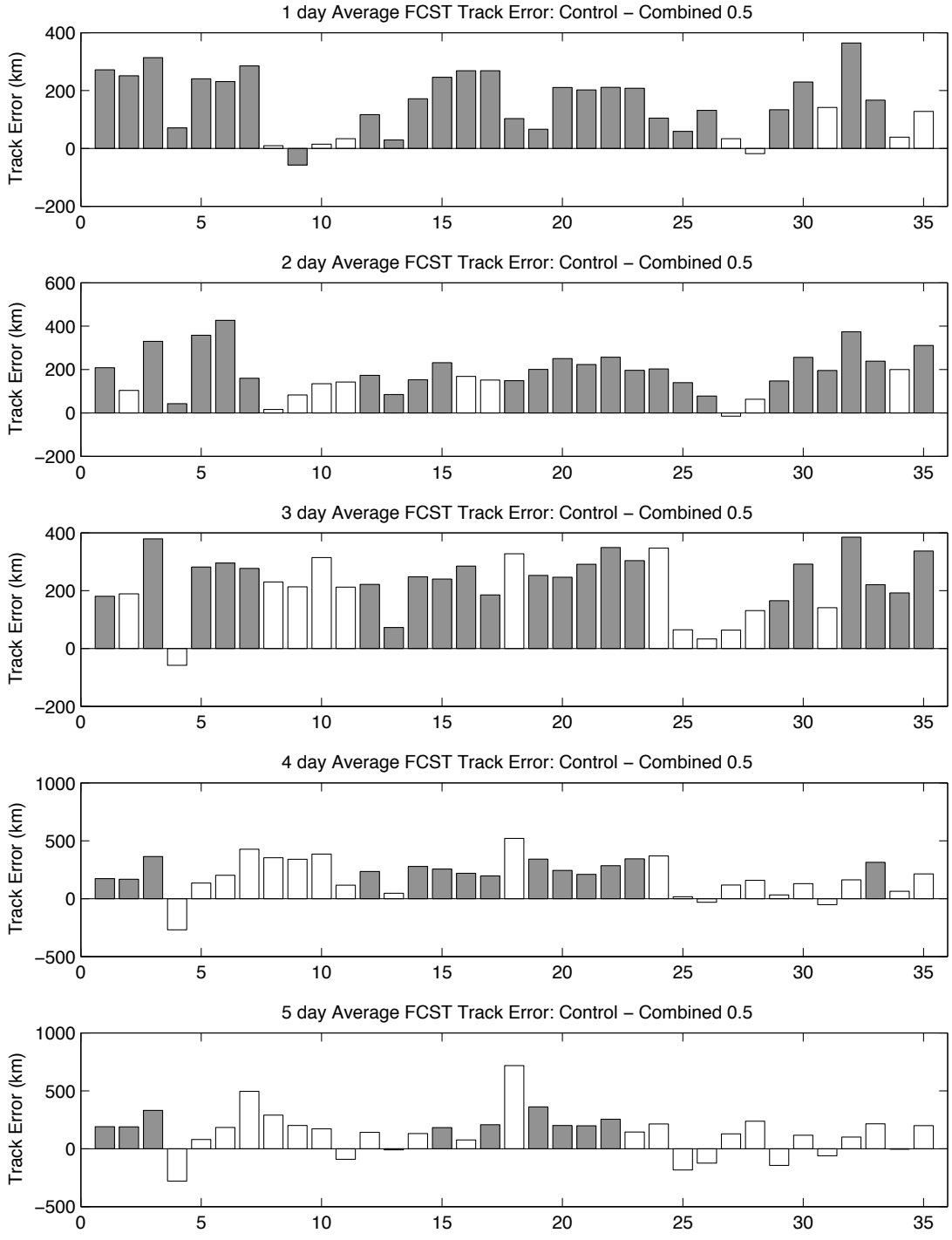

FIG. 24. Difference between daily forecast track error averages of the Control and Combined 0.5 experiments. Gray shading indicates that the difference is statistically significant at the 95% confidence level.

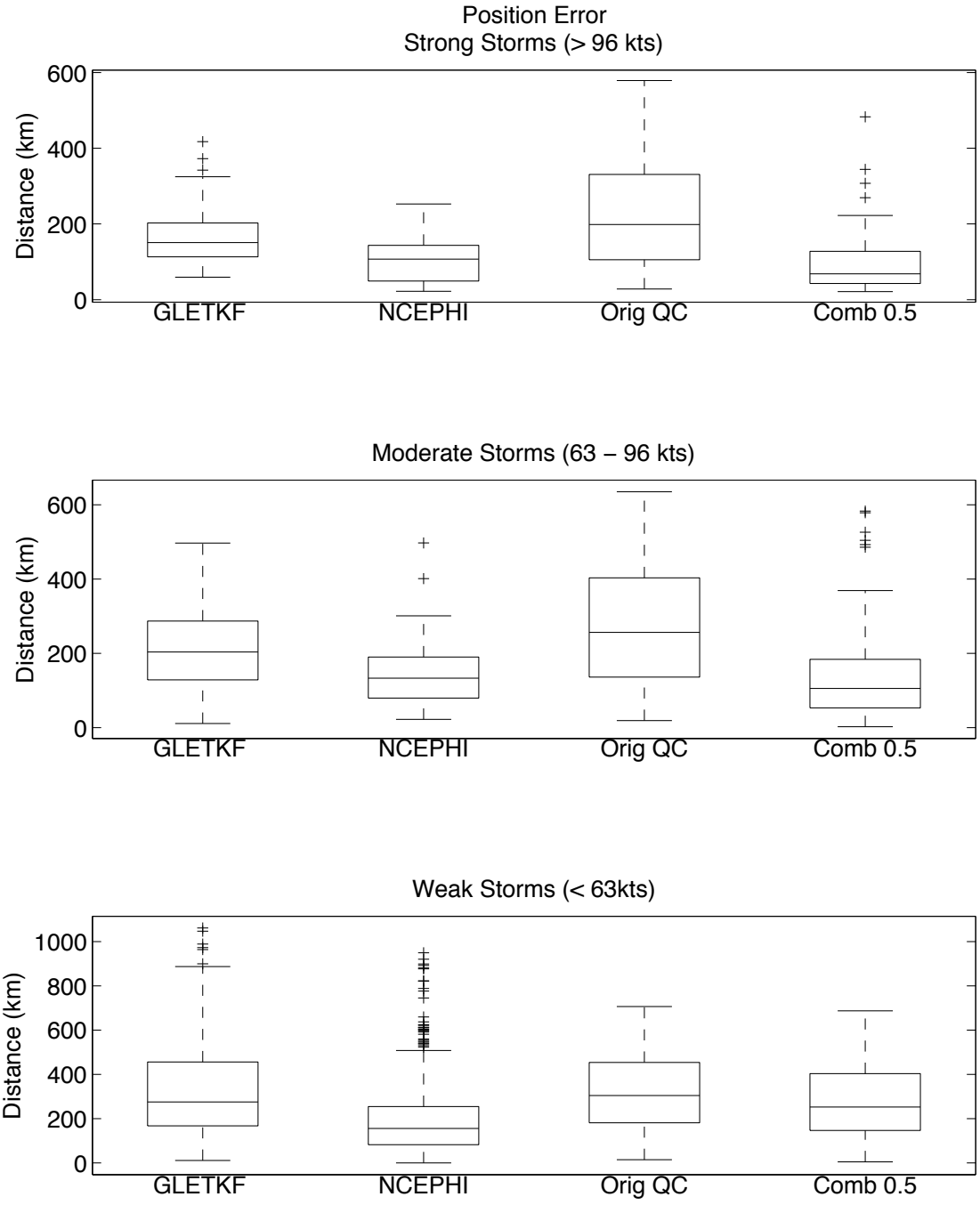

FIG. 25. Distributions of analysis position errors binned by TC intensity. GLETKF is the global LETKF analysis. NCEPHI is the NCEP operational analysis at 1° resolution. Orig QC is the control experiment. Comb 0.5 is the Combined experiment with TCVitals SLP error defined as a constant $0.5hPa$.

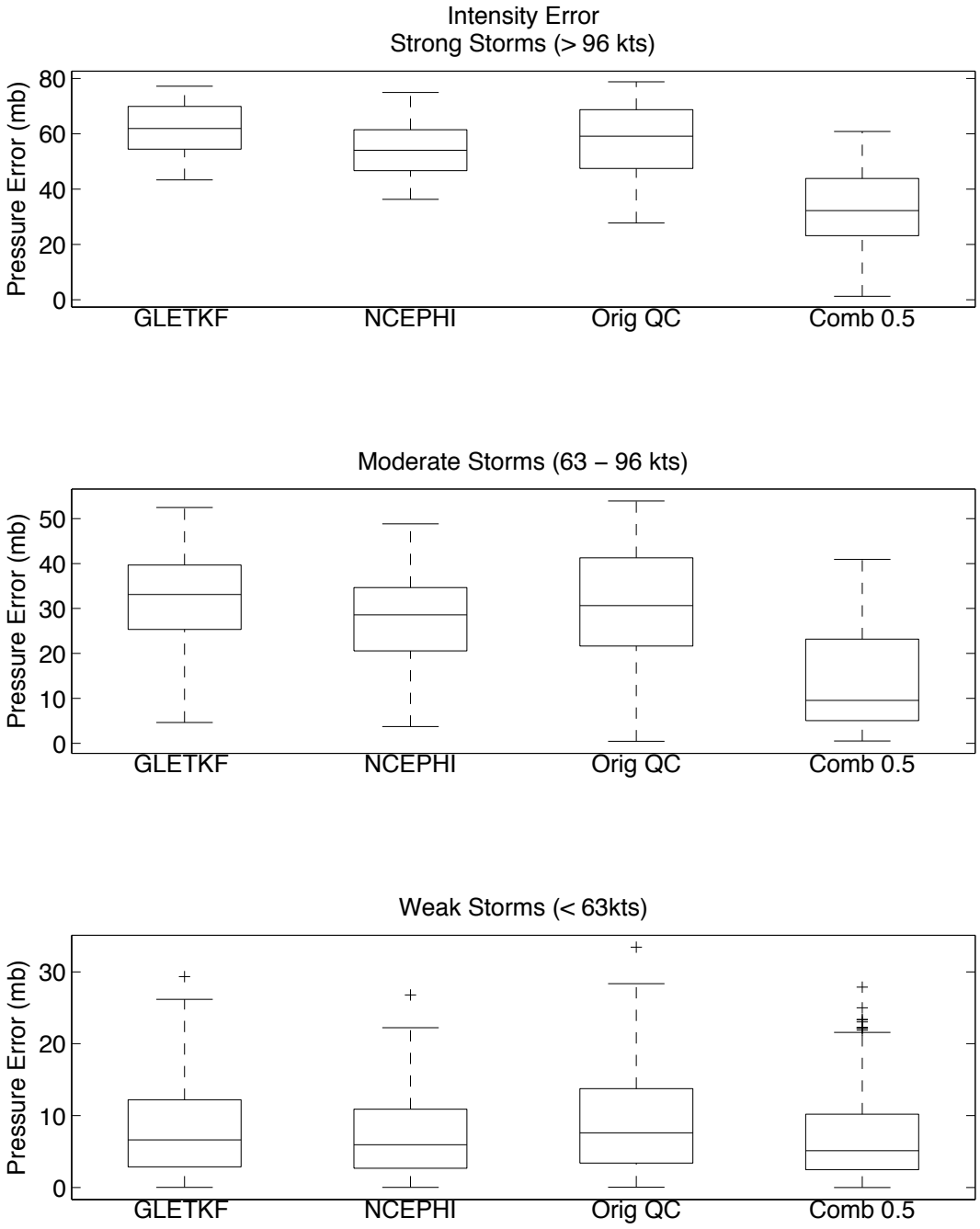

FIG. 26. Distributions of analysis intensity errors binned by TC intensity.

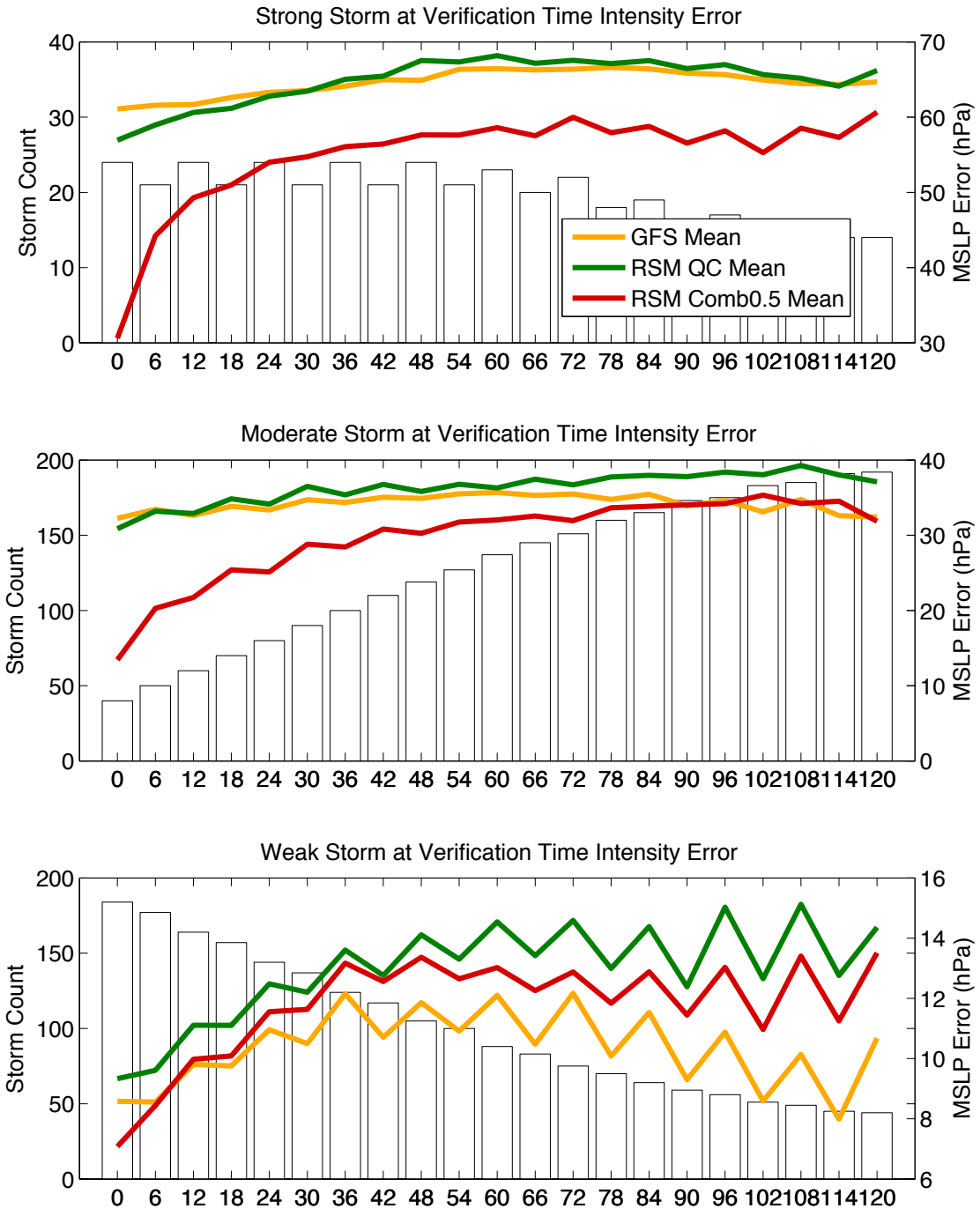

FIG. 27. Time series of average intensity error binned according to Best Track intensity at verification (corresponding to right y-axis). Bars represent the number of observations used to calculate the average at each verification time, and correspond to the left y-axis.

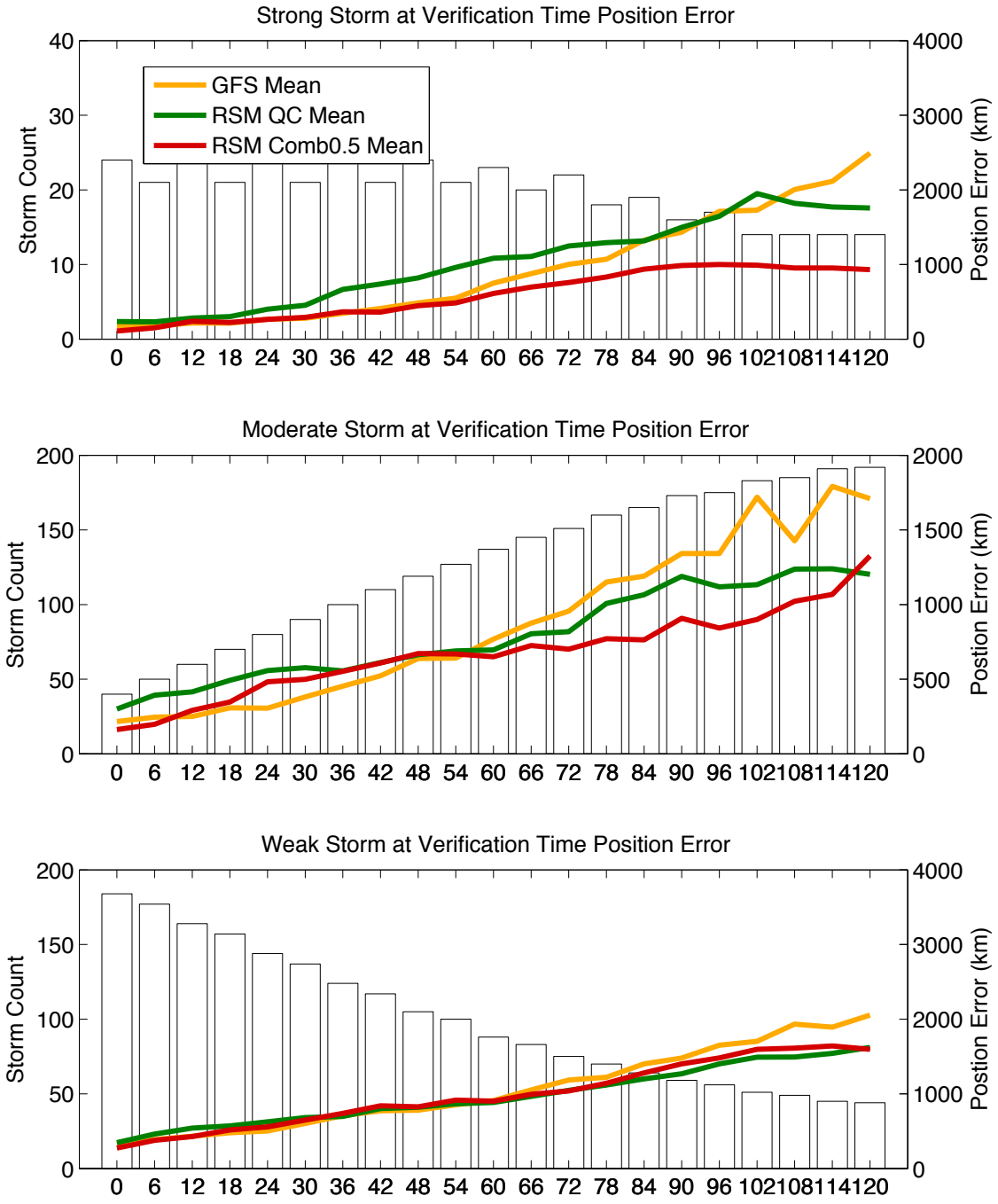

FIG. 28. Time series of average track error binned according to Best Track intensity at verification (corresponding to right y-axis). Bars represent the number of observations used to calculate the average at each verification time, and correspond to the left y-axis.

APPENDIX

TABLES

TABLE 1. 2004 Typhoons and Tropical Storms included in this study. Storm number indicates the order in which the storm was named in the 2004 season. Data is taken from Atangan et al. (2004). (TS = Tropical Storm).

Name	**Storm Number**	**Minimum SLP at Max Intensity (hPa)**	**Maximum Saffir-Simpson Category**	**Time Period Included (mm/dd/hh)**
Mindulle	10	916	Cat. 4	06/22/00Z - 07/04/06Z
Tingting	11	963	Cat. 1	06/24/18Z - 07/04/00Z
Kompasu	12	991	TS	07/12/12Z - 07/16/12Z
Namtheun	13	927	Cat. 4	07/24/06Z 08/01/06Z
Malou	14	954	Cat. 2	08/02/18Z - 08/09/06Z
Meranti	15	997	TS	08/02/00Z - 08/05/06Z
Rananim	16	954	Cat. 2	08/07/00Z - 08/13/00Z
Malakas	17	997	TS	08/10/06Z - 08/13/06Z

TABLE 2. Number of time steps, T, autocorrelation coefficient, r, effective sample size, T', and p value of the test statistic for each of the storms from the comparison of the global LETKF and RSM 100 km experiments for TC intensity.

Storm	T	r	T'	p
10	50	0.82	4.73	0.23
11	38	0.47	13.27	0.19
12	18	0.62	4.04	0.36
13	33	0.00	32.00	0.00
14	27	0.00	26.00	0.01
15	14	0.00	13.00	0.27
16	25	0.75	3.41	0.10
17	13	0.72	2.0	0.30

TABLE 3. Naming conventions and descriptions of experiments.

Descriptor	**Definition**
QC	Discard over threshold, c (Traditional)
Clipped	Huberized at clipping height, c
Kept	No QC. Observation is kept.
Slide	TCVital Error is on a sliding scale as described in Section 3.13.1.b. For single update experiments, error is 0.44 hPa.
Const	TCVital Error is fixed at 3hPa for single update experiments.
Combined ##	Combines QC methods as described in Section 3.33.3.a. ## indicates the value of the TCVital SLP error used.