TESTING A COUPLED GLOBAL-LIMITED-AREA DATA ASSIMILATION SYSTEM USING OBSERVATIONS FROM THE 2004 PACIFIC TYPHOON SEASON

A Thesis

by

CHRISTINA R. HOLT

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

MASTER OF SCIENCE

August 2011

Major Subject: Atmospheric Sciences

Testing a Coupled Global-limited-area Data Assimilation System Using Observations from the 2004 Pacific Typhoon Season

TESTING A COUPLED GLOBAL-LIMITED-AREA DATA ASSIMILATION SYSTEM USING OBSERVATIONS FROM THE 2004 PACIFIC TYPHOON SEASON

A Thesis

by

CHRISTINA R. HOLT

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

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Istvan Szunyogh
Committee Members,	Robert Korty
	Russ Schumacher
	Mikyoung Jun
Head of Department,	Kenneth Bowman

August 2011

Major Subject: Atmospheric Sciences

ABSTRACT

Testing a Coupled Global-limited-area Data Assimilation System Using Observations from the 2004 Pacific Typhoon Season. (August 2011)

Christina R. Holt, B.S., University of South Alabama

Chair of Advisory Committee: Dr. Istvan Szunyogh

Tropical cyclone (TC) track and intensity forecasts have improved in recent years due to increased model resolution, improved data assimilation, and the rapid increase in the number of routinely assimilated observations over oceans. The data assimilation approach that has received the most attention in recent years is Ensemble Kalman Filtering (EnKF). The most attractive feature of the EnKF is that it uses a fully flow-dependent estimate of the error statistics, which can have important benefits for the analysis of rapidly developing TCs.

We implemented the Local Ensemble Transform Kalman Filter algorithm, a variation of the EnKF, on a reduced-resolution version of the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) model and the NCEP Regional Spectral Model (RSM) to build a coupled global-limited area analysis/forecast system. This has been the first time, to our knowledge, that such a system is used for the analysis and forecast of tropical cyclones. We used data from summer 2004 to study eight tropical cyclones in the Northwest Pacific.

The benchmark data sets that we use to assess the performance of our system were the NCEP Reanalysis and the NCEP Operational GFS analyses from 2004. These benchmark analyses were both obtained by the Statistical Spectral Interpolation, which was the operational data assimilation system of NCEP in 2004. The GFS Operational analysis assimilated a large number of satellite radiance observations in addition to the observations assimilated in our system. All analyses are verified against the Joint Typhoon Warning Center Best Track data set. The errors are calculated for the position and intensity of the TCs.

The global component of the ensemble-based system showed improvement in position analysis over the NCEP Reanalysis, but showed no significant difference from the NCEP operational analysis for most of the storm tracks. The regional component of our system improves position analysis over all the global analyses. The intensity analyses, measured by the minimum sea level pressure, are of similar quality in all of the analyses. Regional deterministic forecasts started from our analyses are generally not significantly different from those started from the GFS operational analysis. On average, the regional experiments performed better for longer than 48 h sea level pressure forecasts, while the global forecast performed better in predicting the position for longer than 48 h.

DEDICATION

I would like to dedicate this thesis to the strong women who helped shaped my life. These women show me everyday that hard work, determination, and confidence can lead to success and happiness. My mom, my stepmom, and my best friend overcame great odds as single parents with their daughters in tow to achieve their goals. I have not had to face the same difficulties as these brave women, but I owe tremendous gratitude to all three of them for the inspirations they are and the courage they have given me. They are all my heroes.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Istvan Szunyogh, and my committee members, Dr. Robert Korty, Dr. Russ Schumacher, and Dr. Mikyoung Jun, for their guidance and support throughout the course of the research. Without the technical support of Gyorgyi Gyarmati, none of the research would have been possible. I would also like to thank our collaborators, Ross Hoffman and Mark Leidner at AER.

Thanks go to all the faculty and staff in the Department of Atmospheric Sciences, as well as my friends and colleagues for making Texas A&M my new home. My appreciation for the support of my family is insurmountable; I could not have made it through my long educational journey without them. My boyfriend, Christopher Williamson, has been my constant emotional support and source of encouragement throughout my college career, and for that I owe many thanks.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

1. INTRODUCTION

Tropical cyclone (TC) track forecasts have improved in recent years due to advances in numerical weather prediction (Rappaport et al. 2009). These improvements have been due to increased model resolution, improved data assimilation techniques, and the rapid increase in the number of routinely assimilated observations over oceans. Improvements in the analysis and forecasts of the intensity of TCs have been much more modest. This modest improvement is not surprising considering that the eyewall radius of a TC is about 25-50 km (Kimball and Mulekar 2004) and, as several authors (Fiorino and Elsberry 1989a,b; Grasso 2000; Skamarock 2004) have pointed out, 4 to 10 grid points are needed to resolve a flow feature such as the eye of a TC; the 5-10 km resolution that would be required to capture the structure of the eye of a TC is currently unattainable in an operational real-time global analysis/forecast system.

The data assimilation approach that has received the most attention in recent years is Ensemble Kalman Filtering (EnKF; Whitaker and Hamill 2002). The most attractive feature of the EnKF is that it uses a fully flow-dependent estimate of the error statistics, which can have important benefits for the analysis of a rapidly developing TC. One particular study that should be mentioned here is Hamill et al. (2011), which compared the performance of an implementation of EnKF on the National Centers for Environmental Prediction (NCEP) Global Forecast System

---

This thesis follows the style of *Weather and Forecasting*.

(GFS) model to that of the Grid-point Statistical Interpolation (GSI). The GSI is the current operational data assimilation component of the NCEP GFS (Kleist et al. 2009). Hamill et al. (2011) found that the track forecasts started from the EnKF analyses were more accurate than those from the GSI analyses, but they did not find a similar improvement in the intensity forecasts. The latter result was not unexpected considering the approximate 45 km resolution of the analysis/forecast system.

Torn and Hakim (2009) carried out EnKF data assimilation experiments using a 30 km resolution version of the Weather Research and Forecasting (WRF) model, with a 10 km nested grid in the TC region in some of the experiments. They found that while the EnKF provided accurate analysis of the TC position, the same was not true for the analysis of the intensity: the system on the coarser grid underestimated the intensity, while the finer grid overestimated the intensity. The overestimation of intensity was mainly due to the timing of intensification. Another interesting finding by Torn (2010) was that error in the intensity estimate increased with the intensity of the TC, most likely due to the fact that the analysis/forecast system has a greater difficulty resolving the steeper pressure gradient in the more intense TCs.

The present study, to the best of our knowledge, is the first investigation that employs a coupled global-limited-area EnKF data assimilation system to assimilate observations in a TC region. The particular region selected for this study is the Northwest Pacific (Fig. 1), where observations collected in summer 2004 are assimilated. The primary goal of the research presented here is to assess the analysis/forecast benefits of using a higher resolution limited-area model in the TC

region. A secondary goal is to study the effects of assimilating surface wind observations by scatterometers and targeted vertical soundings by dropsondes released from aircraft. Some experiments to compare the performance of the limited-area component for different horizontal resolutions are also presented.

The thesis is organized as follows. Section 2 gives an overview of the coupled global-limited-area data assimilation system. Section 3 describes analysis and forecast verification techniques employed by the study. Section 4 describes the characteristics of the 2004 Northwest Pacific Typhoons included in this study. The analysis verification results are presented in Section 5, while the forecast verification results are described in Section 6. Section 7 is a summary and discussion of the main results.

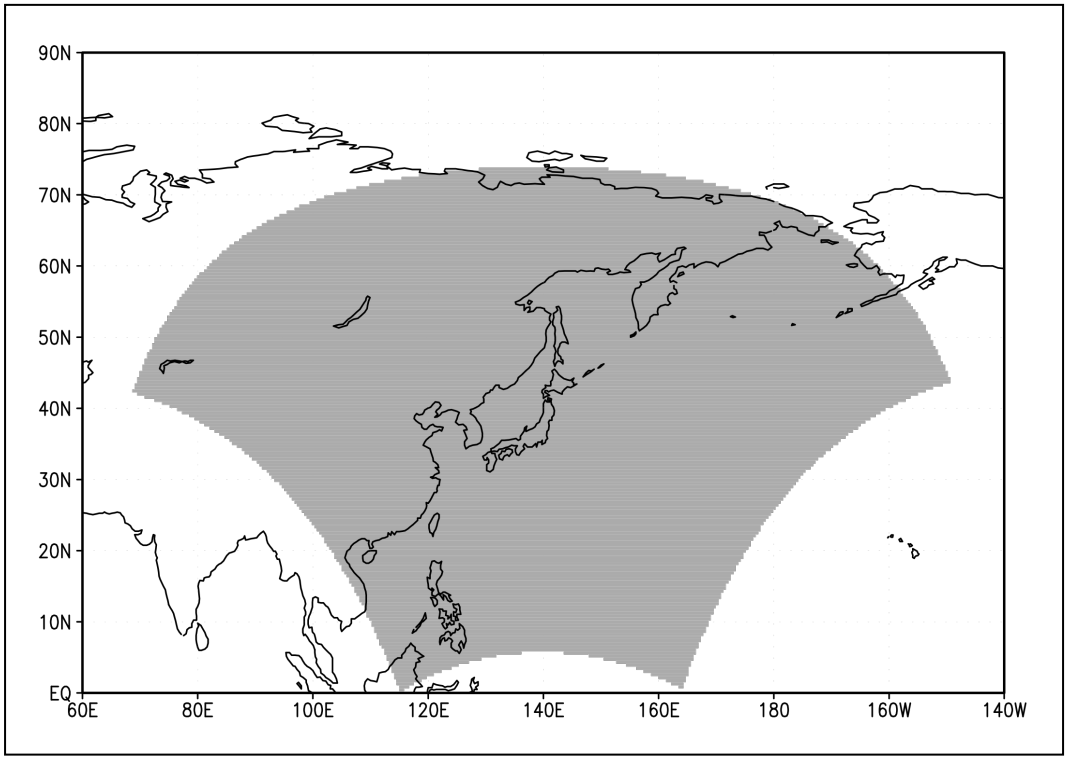

Fig. 1. Map indicating the domain of the RSM for all RSM ANL and FCST experiments.

2. ANALYSIS/FORECAST SYSTEM

The data assimilation system is an implementation of the Local Ensemble Transform Kalman Filter (LETKF) algorithm (Ott et al. 2004; Hunt et al. 2007) on the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) model (Szunyogh et al. 2005 and 2008) and the NCEP Regional Spectral Model (RSM). This implementation of the LETKF was first tested on winter storms over the U.S. and is described in 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. Neither "Tropical Cyclone relocation," nor bogusing techniques are employed in the system. The following describes the different components of the forecast/analysis system.

a. GFS Model

The GFS is a spectral-transform model and we use it at a cut-off wave number of 62 with 28 vertical sigma layers (T62L28). This spectral resolution corresponds to a grid spacing of approximately 210 km and the output is processed on a 2.5 x 2.5 degree grid. We generate the initial GFS ensemble for the LETKF with the 0000 UTC operational GFS analyses truncated to T62 resolution, from 40 different days during the summer of 2004.

b. Regional Spectral Model

The RSM is essentially a nested limited-area version of the GFS. To eliminate common lateral boundary issues faced with using limited-area grid-space models in conjunction with global spectral models, the RSM employs a technique of predicting the deviations from the global model forecast over the entire regional domain. These deviations are then combined with the global state solution to obtain the high-resolution forecasts (Juang and Kanamitsu, 1994).

The vertical levels in the RSM are identical to the 28 sigma layers of the GFS. The horizontal resolution of the RSM is 48 km in most of our experiments. In some experiments, the resolution is increased to 24 km without changing the vertical resolution. In the configuration we implement, the RSM does not propagate information back to the GFS. Having the RSM feed information back to the GFS during the data assimilation process was shown by Merkova et al. (2011) to produce mixed results compared to the configuration used in this study.

c. *The LETKF*

The LETKF is a data assimilation algorithm that, similar to other formulations of the ensemble-based Kalman Filter (EnKF) approach, uses an ensemble to generate an estimate of the forecast error covariance matrix. Thus, the error covariance evolves both between and within the analysis cycles. The main distinguishing features of the LETKF from other EnKF schemes are that it assimilates all observations that may affect the state estimate at a given grid point simultaneously and the analyses at the

different grid points are obtained in parallel, independently of each other grid-point analysis (Szunyogh et al. 2008).

An EnKF prepares a background (prior) estimate of the probability distribution of the state at analysis time t_n by evolving an ensemble of state estimates with the model from the previous analysis time, t_{n-1}. The prior estimate of the probability distribution of the state is then updated by assimilating the observations over the time window $[t_n - \Delta t/2,\ t_n + \Delta t/2]$.

The data assimilation process begins by creating a K-member background ensemble by an ensemble of forecasts

$$\mathbf{x}^{b(k)}(t_n) = \mathbf{M}\left[\mathbf{x}^{a(k)}(t_{n-1})\right], \qquad \text{k} = 1,\ldots,\text{K} \tag{1}$$

where $\mathbf{x}^{b(k)}$ is the k-th member of the background ensemble, $\mathbf{M}$ represents the model dynamics, and $\{\mathbf{x}^{a(k)}(t_{n-1}), \text{k} = 1,\ldots,K\}$ is the analysis ensemble at time t_{n-1}. The background ensemble mean $\overline{\mathbf{x}}^b(t_n)$ is calculated as

$$\overline{\mathbf{x}}^b(t_n) = K^{-1}\sum_{k=1}^{K}\mathbf{x}^{b(k)}(t_n) \tag{2}$$

and is used to construct the ensemble perturbation matrix, $\mathbf{X}^b(t_n)$, in which the k-th column is the k-th background perturbation, $\mathbf{X}^{b(k)}(t_n) = \mathbf{x}^{b(k)}(t_n) - \overline{\mathbf{x}}^b(t_n)$. The matrix of background perturbations is used in the calculation of the background error covariance matrix, $\mathbf{P}^b(t_n)$, such that

$$\mathbf{P}^b(t_n) = (K-1)^{-1}\mathbf{X}^b(t_n)\left[\mathbf{X}^b(t_n)\right]^T \tag{3}$$

The update step of an ensemble-transform-based scheme obtains the analysis mean by

$$\bar{\mathbf{x}}^a(t_n) = \bar{\mathbf{x}}^b(t_n) + \mathbf{X}^b(t_n)\mathbf{w}^a(t_n) \quad \textbf{(4)}$$

where $\mathbf{w}^a(t_n)$ is the particular choice of the weight vector **w,** which minimizes the quadratic cost function

$$J(\mathbf{w}) = (k-1)\mathbf{w}^T\mathbf{w} + \left(\mathbf{y}^o(t_n) - \mathbf{h}\left[\bar{\mathbf{x}}^b(t_n) + \mathbf{X}^b(t_n)\mathbf{w}\right]\right)^T \mathbf{R}^{-1}(t_n)\left(\mathbf{y}^o(t_n) - \mathbf{h}\left[\bar{\mathbf{x}}^b(t_n) + \mathbf{X}^b(t_n)\mathbf{w}\right]\right) \quad \textbf{(5)}$$

The components of the vector $\mathbf{y}^o(t_n)$ are the observations assimilated at time t_n and **h**(**x**) is a four-dimensional observation operator that maps the grid-point representation of the model state at t_n to the time and location of the observations, and converts model variables to observed quantities. The observation operator satisfies

$$\mathbf{y}^o(t_n) = \mathbf{h}\left[\mathbf{x}^t(t_n)\right] + \mathbf{e}(t_n), \quad \textbf{(6)}$$

where $\mathbf{x}^t(t_n)$ is the model representation of the true state at t_n and $\mathbf{e}(t_n)$ is a random noise vector that has a Gaussian distribution with mean **0** and covariance matrix $\mathbf{R}(t_n)$. To obtain the full ensemble of analyses, the analysis perturbations are added to the ensemble mean $\bar{\mathbf{x}}^a(t_n)$. These analysis perturbations are computed from the background perturbation by

$$\mathbf{X}^a(t_n) = \mathbf{X}^b(t_n)\mathbf{W}^a(t_n), \quad \textbf{(7)}$$

where $\mathbf{W}^a(t_n)$ is a weight matrix, in which each row is a weight vector for an ensemble member.

In the coupled global-limited-area data assimilation system, each of 40 global ensemble members has a limited-area counterpart. First, the global analysis is prepared and then each member of the global background ensemble is used to define the global component (boundary conditions and large scale forcing) of a high-resolution limited-area background ensemble member. The initial conditions for the limited-area background ensemble at time t_n are the members of the limited-area analysis ensemble at time t_{n-1}.

For this study, each analysis is obtained by cycling the system for six hours with an observation time window of $\Delta t = 6$ h. This approach provides four analyses per day at 0000 UTC, 0600 UTC, 1200 UTC, and 1800 UTC. 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 sizes of the ensemble) and nonlinearities, and to account for the effects of model errors. The covariance inflation ranges from 1.5 at the equator down to 1.3 in roughly the northern half of the domain. Observations are assimilated within an 800 km radius of each grid point.

3. VERIFICATION METHODS

a. Benchmark Analyses

The benchmark analyses we use to assess the quality of the LETKF analyses are the NCEP/NCAR (National Center for Atmospheric Research) Reanalysis (Kalnay et al. 1996) and the NCEP operational analyses from 2004. The NCEP/NCAR Reanalysis was chosen for comparison because it is also based on the NCEP global operational model and was prepared at the same T62L28 resolution as our global analysis. The difference from our analysis/forecast system is that the Reanalysis used an older version of the model and the Spectral Statistical Interpolation (SSI), a 3D variational analysis method (Parrish and Derber 1992). The operational NCEP analysis (referred to as NCEP Oper ANL) used the same version of the model we use and the SSI, but was originally prepared at T254L64 resolution, which is roughly equivalent to 55 km resolution. It also assimilated a large number of satellite radiance observations in addition to the observations assimilated in our system, and employed TC relocation techniques. Each of the global benchmark analyses is compared to the experimental analyses on a 2.5 x 2.5 degree grid.

b. Analysis Verification

The Joint Typhoon Warning Center's (JTWC) Best Track data (Atagan et al. 2004) is used as the verification data set to assess the errors in the track and intensity analyses and forecasts. Since the Best Track information was not assimilated in our

experiments, it provides independent verification information. The Best Track is issued after the JTWC reanalyzes all available data once a storm has dissipated. The data set includes six-hourly track and intensity information. The intensity estimates, derived from the Dvorak model, are based on one-minute mean sustained wind speed. This procedure leads to estimates that are higher than those provided by other agencies, which base their estimates on a ten-minute mean (Chu et al. 2002). The derived one-minute mean wind is then interpolated to provide a minimum central surface pressure. The data from this data set are all point estimates and will be compared to gridded model data. The vast difference in resolution when comparing the two is a large source of error. Scaling the best track point estimates to the resolution of the model output would be a more even comparison, but is beyond the scope of the current research.

Analyses are prepared using the coupled LETKF system for three base cases: a global analysis (GFS LETKF ANL), a regional analysis (RSM LETKF ANL), and a regional analysis that includes QuikSCAT satellite-derived surface wind observations (RSM LETKF ANL QS) in addition to the observations assimilated in the GFS LETKF ANL and RSM LETKF ANL experiments. Analyses are prepared for the period between 0000 UTC June 22, 2004 and 1800 UTC August 15, 2004. These analyses, along with the NCEP Reanalysis and NCEP Oper ANL, are verified against the JTWC data set.

Because there is a strong variability in intensity among the eight TCs that occurred in the northwest Pacific during the period we consider, the method by which

the storm center is located is based on storm intensity. After visually assessing the sea level pressure and low-level (lowest model level) vorticity fields for each case, the center-finding method was determined: if the model indicates that there is a closed low-pressure center near the Best Track location, the location of the minimum model pressure in that area serves as the storm position at the time. When the model indicates no closed low-pressure center near the Best Track, which occurs in the case of the least intense TCs, the position of the TC is taken to be the location of the vorticity maximum. The minimum pressure, then, is taken to be the sea level pressure at the vorticity maximum grid point. The two methods produce slightly different results when applied to closed low-pressure centers. Choosing the location of the lowest model pressure to mark the track, whenever possible, is motivated by previous studies (e. g. Torn 2010), which followed the same approach.

Once the cyclone tracks and intensities for all analyses (including NCEP benchmarks) are extracted, root mean square errors (RMSE) are calculated to provide a single number as a measure of the error in the track for the entire lifetime of each TC. To test the significance of the differences between the RMSE for the different configurations of the analysis/forecast system, 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 two time series are the same, versus the alternative that they are different. Assuming that the difference, $\Delta(x,t)$, between two time series of analysis or forecast errors is described by a first-order autoregressive process, the test statistic is

$$z(x) = \frac{\overline{\Delta}(x)}{\left[V_{\Delta}(x)/T'(x)\right]^{1/2}}, \tag{8}$$

where $\overline{\Delta}(x)$ is the time average over $\Delta(x,t)$ and $V_{\Delta}(x)$ is the variance of the time series ($\Delta(x,t)$: t=1,2,…T). In Eq. 8, the effective sample size, T', is

$$T'(x) = T\left[1 - r(x)\right]\left[1 + r(x)\right]^{-1}, \tag{9}$$

where the autocorrelation coefficient, $r(x)$, is computed by

$$r(x) = \frac{\sum_{t=1}^{T-1}\left\{\left[\Delta(x,t) - \Delta_{T1}(x)\right]\left[\Delta(x,t+1) - \Delta_{T2}(x)\right]\right\}}{\left\{\sum_{t=1}^{T-1}\left[\Delta(x,t) - \Delta_{T1}(x)\right]^2 \sum_{t=2}^{T}\left[\Delta(x,t) - \Delta_{T2}(x)\right]^2\right\}^{1/2}}. \tag{10}$$

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

Under the assumption that the time series of errors is described by a first order autoregressive process, autocorrelation should fall within the range [0, 1]. While most analyses and forecast error statistics satisfy this assumption, the track analysis errors do not, because the position error has a "short memory". Thus, in the computation of the statistical significance of the difference between the errors, we assume that the effective sample size T' is equal to the sample size T.

Once $z(x)$ is computed, the difference between the time series is deemed significant if the likelihood of obtaining this $z(x)$ is less than the significance level being tested. In this study, significance at the 90% confidence level is sufficient. Differences between analysis time series of pressure or track error are significant at the 90% level if $|z| \geq 1.65$ using a standard normal distribution. We also consider the

significance of the differences based on the t-distribution. The t-distribution has heavier tails to account for smaller sample sizes. In this case, the probability of obtaining our tested value given a true null (p-value) is calculated. Any values less that 0.1 are considered significant at the 90% confidence level.

Verification statistics stratified by cyclone intensity are also prepared. This verification approach is motivated by Torn (2010), who found that the largest errors in intensity analysis tended to occur in strong TCs (Category 3-5), while the largest track errors tended to occur in weak TCs (Tropical Depressions and Storms). Observations of minimum SLP and position error are separated based on Best Track wind speed into the three groups used in Torn (2010): Tropical Depression/Tropical Storm (< 64 kts), Category 1-2 (64-95 kts), and Category 3-5 (> 95 kts).

The distributions of ensemble mean analysis errors stratified by TC Category are analyzed with the use of box plots. This type of graph shows the distribution of the data by displaying three quartiles, the minimum, and the maximum on a rectangular box. The lower edge is the first quartile, while the upper edge is the third quartile. The center line indicates the median of the data set. The lines that extend away from the top and bottom of the box are often called “whiskers” and extend to the largest data point within 1.5 interquartile ranges from the respective quartile. Data points plotted beyond the “whiskers” are outliers. The shape can tell us the variability (height of box), departure from symmetry, and center of the data (position of median in box). The disposition of the box from zero tells us the general magnitude of the data (Montgomery and Runger 2007).

c. Special Cases

Besides the surface wind observations by scatterometer, it is useful to investigate other factors that may affect the performance of the analysis/forecast system. In a select few observation time windows, there are in situ measurements from the Dropwindsonde Observations for Typhoon Surveillance near the Taiwan Region (DOTSTAR) program (Wu et al., 2005). These observations have been specifically targeted to maximize TC predictability and will be included when available as a separate analysis (RSM LETKF ANL DS).

The DOTSTAR program obtains in situ measurements by utilizing the Airborne Vertical Atmosphere Profiling System (AVAPS) and flying an ASTRA aircraft in a pre-determined flight path around the TC. This path is chosen based on short-term model forecasts and targeted observation products. Dropwindsondes are released from an altitude of 14 km for five hours over the course of the mission (Wu 2005). Three individual missions were conducted for Typhoon Mindulle over three days, just before the storm underwent rapid intensification.

The effect of the resolution of the regional component of the system is another factor we investigate because resolution is known to play a role in the maximum intensity, or minimum pressure, that a modeled cyclone can obtain. In order to keep the computational cost manageable, 24 km resolution analyses were prepared only for two separate time periods with slightly different domain positions, but identical domain sizes. The first time period and domain (Fig. 2) takes advantage of the fact that the first two storms, Mindulle and Tingting, occur fairly close to each other in

space and time. The second (not shown) is chosen to cover the full track of Typhoon Namtheun, providing more verification data for an intense storm. We refer to these cases collectively as RSM LETKF ANL 24. Choosing these two periods allows for the use of the analysis techniques we described in the previous section, including stratification by Best Track wind intensity, since the intense storms went through moderate and weak stages as well. There are 121 data points available for the verification of RSM LETKF ANL 24. We compare the RSM LETKF ANL 24 case with a subset of corresponding RSM LETKF ANL QS cases, because RSM LETKF ANL 24 also assimilates the scatterometer observations.

d. Forecast Verification

The quality of short-term forecasts is an important measure of the quality of the analyses used as the initial conditions of the forecasts. Deterministic forecasts were prepared for the cases where the analysis verification results suggested relatively large differences between the qualities of the analyses with the different analysis/forecast systems. Each forecast is initialized from its respective ensemble analysis mean, that is, the GFS LETKF ANL is the initial condition of the forecast GFS LETKF FCST. The naming convention for each forecast will be the analysis name with FCST substituted for ANL. We verify the deterministic forecasts against the Best Track information to assess the errors in track and intensity produced by our analysis/forecast system and compare the results of each experimental design to determine the relative performance of each system component.

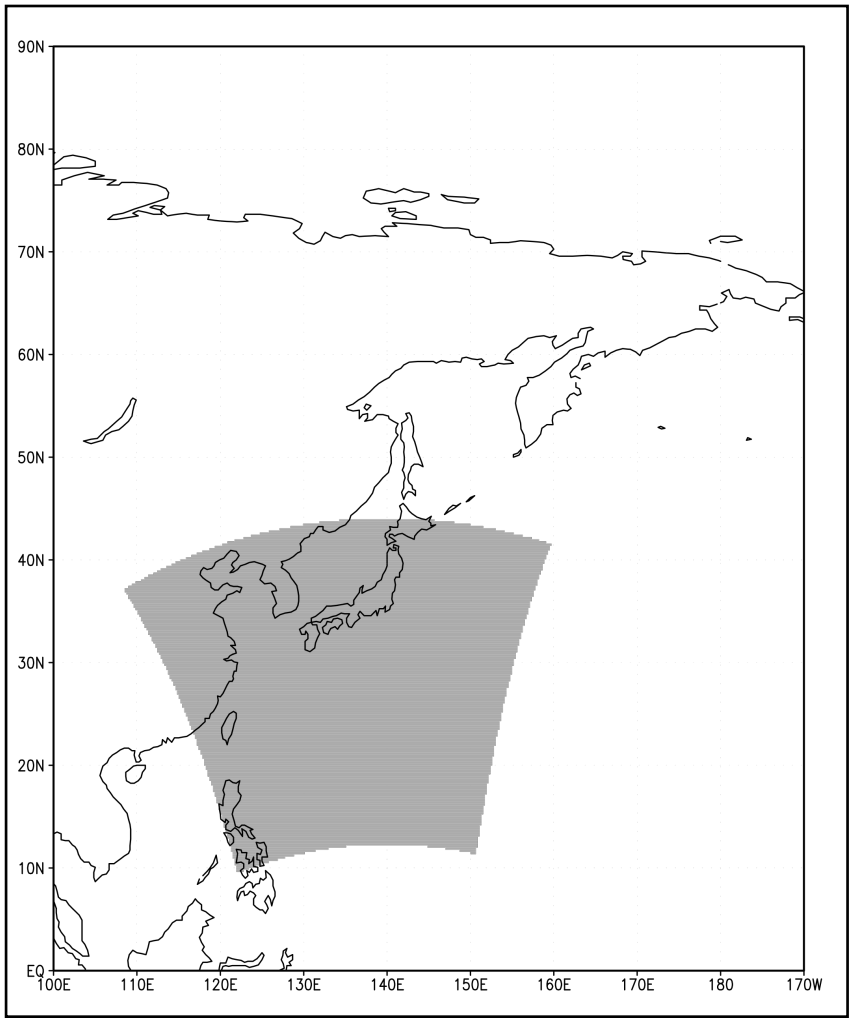

Fig. 2. Map indicating the increased resolution domain for the Typhoons Mindulle and Tingting.

4. SUMMER 2004 NORTHWEST PACIFIC TYPHOONS

During the summer of 2004 between June 22 and August 15, there were eight TCs in the northwest Pacific basin. Of these eight, there were a variety of storm intensities ranging from Tropical Depression (< 63 kts) to Category 4 (113-155 kts) on the Saffir-Simpson Scale (Atagan et al. 2004). During this time, there were also reconnaissance flights providing targeted in situ measurements for one of the most intense storms, Mindulle. Basic information for each of these storms is included in Table 1.

Table 1. Tropical cyclone name, number, minimum sea level pressure at maximum intensity, Saffir-Simpson Category, number of observations, and time period for the cyclones studied. Data is taken from Atagan, et al. 2004. (TS = Tropical Storm)

Name	Storm Number	Minimum SLP at Max Intensity (hPa)	Maximum Saffir-Simpson Category	Number of Observation Times Included	Time Period Included (mm/dd/hh)
Mindulle	10	916	Cat. 4	50	06/22/00Z – 07/04/06Z
Tingting	11	963	Cat. 1	38	06/24/18Z – 07/04/00Z
Kompasu	12	991	TS	18	07/12/12Z – 07/16/12Z
Namtheun	13	927	Cat. 4	33	07/24/06Z – 08/01/06Z
Malou	14	954	Cat. 2	27	08/02/18Z – 08/09/06Z
Meranti	15	997	TS	14	08/02/00Z – 08/05/06Z
Rananim	16	954	Cat. 2	25	08/07/00Z – 08/13/00Z
Malakas	17	997	TS	13	08/10/06Z – 08/13/06Z

5. ANALYSIS VERIFICATION RESULTS

a. Main Analyses

The analysis track error averaged over the eight storms (Fig. 3) shows that both the global and regional LETKF systems performed better than the NCEP reanalysis. The GFS LETKF ANL is about 60 km more accurate, while the regional analyses are about 80 km more accurate than the NCEP Reanalysis. NCEP Oper ANL is the best of the global analyses on average, performing similarly to the regional track analyses. The variability among the average minimum sea level pressure (SLP) errors in the different systems (Fig. 4) is much smaller than the variability among track errors. They are, however, in good agreement with the results based on verifying the track on the relative quality of the different systems. There is also some indication that the assimilation of QuikSCAT data improves the intensity analysis. These results also show that the ensemble-based analysis systems performed adequately without using bogusing or TC relocation techniques.

We find that for Category 1 Typhoons and stronger, the minimum SLP error is inversely proportional to the Best Track minimum central pressure. Fig. 5 shows an illustration of this result for Typhoon Mindulle. At the peak intensity, the RSM analyses are more accurate by about 2 hPa over the duration of the most intense observations. The nearly one-to-one inverse relationship between intensity and error is most likely related to the low horizontal resolution of our analysis/forecast system.

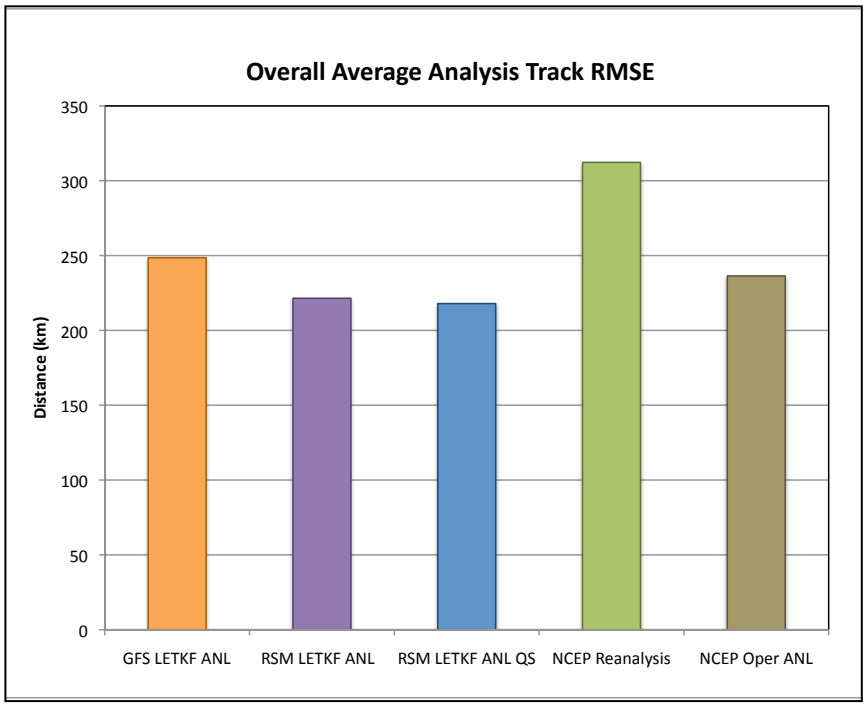

Fig. 3. Average Track RMSE over all 218 observations of eight tropical cyclones.

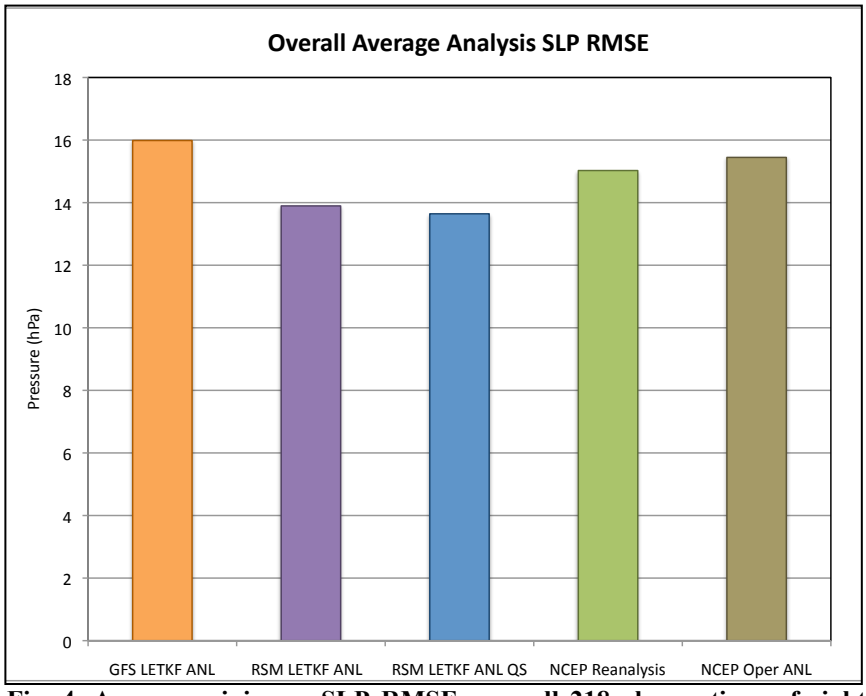

Fig. 4. Average minimum SLP RMSE over all 218 observations of eight tropical cyclones.

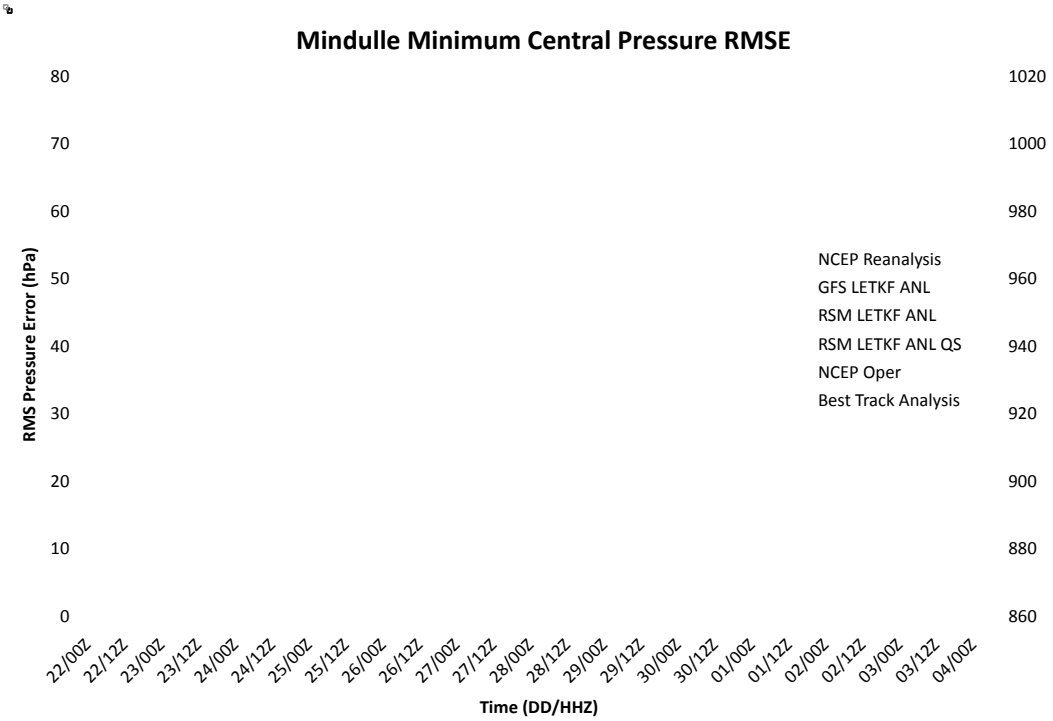

Fig. 5. Minimum SLP RMSE time series for Typhoon Mindulle. The Best Track Analysis pressure (black) corresponds to the secondary (right) axis, while the error for each analysis is plotted in the color indicated by the legend and corresponds to the primary (left) axis.

Fig. 6 shows that although the minimum sea level pressure errors grow as cyclones intensify, the trend in the pressure time series captures the intensification. With much higher-resolution simulations, it may be possible to significantly reduce the SLP error observed in the current experiments, but the simulations would be much more computationally expensive.

Resolution also produces some error in the position of the storm. There is no way to locate the center of the storm with higher precision than the size of the grid box it is located in. Fig. 7 shows the tracks for the global and regional analyses for Typhoon Mindulle as an example.

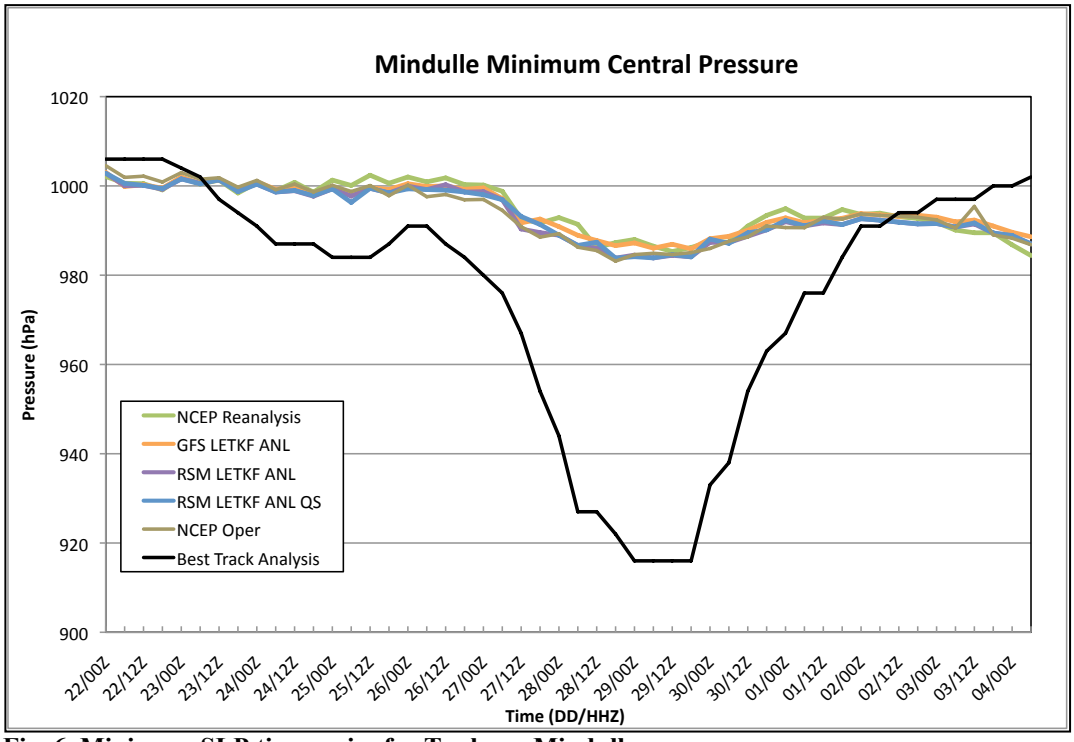

Fig. 6. Minimum SLP time series for Typhoon Mindulle.

Assessment of the performance for each model configuration was done on a storm-by-storm basis using the two-sample *t*-test to test for the statistical significance of the difference between the errors for the different configurations. Five comparisons are made: NCEP Reanalysis vs. GFS LETKF ANL; NCEP Oper ANL vs. GFS LETKF ANL; GFS LETKF ANL vs. RSM LETKF ANL; and RSM LETKF ANL QS vs. RSM LETKF ANL. Tables 2 - 7 show the actual sample sizes, the effective sample sizes, *z*-scores, and p-values from the t-distribution for minimum SLP and positions errors. For SLP comparisons, a small percentage of time series differences are very slightly anti-correlated, returning small negative autocorrelation values that are

Table 2. Sample size, effective sample size, and Z score for minimum SLP error for each global comparison of each TC. Shaded boxes indicate values significant at the 90% confidence level.

Storm Name	Time Steps	NCEP - GFS		NCEP Oper - GFS		GFS - RSM	
		T′	Z	T′	Z	T′	Z
Mindulle	50	20.3	3.03	16.5	2.51	4.7	1.06
Tingting	38	18.2	0.40	12.7	-2.15	10.8	0.91
Kompasu	18	3.9	-1.52	9.3	1.32	3.6	2.74
Namtheun	33	4.6	-1.33	8.4	2.29	16.5	4.85
Malou	27	4.4	1.94	19.5	1.43	26.0	2.67
Meranti	14	6.4	-0.63	7.8	-0.01	13.0	0.64
Raninim	25	7.9	-1.60	6.9	1.25	2.7	1.38
Malakas	13	12.0	3.05	12.0	1.48	2.9	1.23

Table 3. Sample size, effective sample size, and Z score for minimum SLP error for each regional comparison of each TC. Shaded boxes indicate values significant at the 90% confidence level.

Storm Name	Time Steps	NCEP Oper - RSM		RSM QS - RSM	
		T′	Z	T′	Z
Mindulle	50	21.9	-1.19	30.1	0.49
Tingting	38	4.2	1.13	35.6	1.15
Kompasu	18	3.7	2.66	17.0	0.60
Namtheun	33	12.4	-1.11	27.4	0.61
Malou	27	21.8	-0.07	26.0	0.52
Meranti	14	9.9	0.21	11.1	-0.57
Raninim	25	9.5	-0.47	20.2	1.55
Malakas	13	12	-0.76	10.0	0.41

Table 4. Sample size, effective sample size, and p-value from the t-distribution for minimum SLP error for each global comparison of each TC. Shaded boxes indicate values significant at the 90% confidence level.

Storm Name	Time Steps	NCEP - GFS		NCEP Oper - GFS		GFS - RSM	
		T'	p-value	T'	p-value	T'	p-value
Mindulle	50	20.3	0.008	16.5	0.023	4.7	0.201
Tingting	38	18.2	0.363	12.7	0.047	10.8	0.251
Kompasu	18	3.9	0.117	9.3	0.160	3.6	0.031
Namtheun	33	4.6	0.148	8.4	0.041	16.5	0.000
Malou	27	6.8	0.072	19.5	0.142	26.0	0.001
Meranti	14	6.4	0.304	7.8	0.385	13.0	0.315
Raninim	25	7.9	0.110	6.9	0.170	2.7	0.125
Malakas	13	12.0	0.010	12.0	0.132	2.9	0.150

Table 5. Sample size, effective sample size, and p-value from the t-distribution for minimum SLP error for each regional comparison of each TC. Shaded boxes indicate values significant at the 90% confidence level.

Storm Name	Time Steps	NCEP Oper - RSM		RSM QS - RSM	
		T'	p-value	T'	p-value
Mindulle	50	21.9	0.192	30.1	0.349
Tingting	38	4.2	0.182	35.6	0.204
Kompasu	18	3.7	0.034	17.0	0.324
Namtheun	33	12.4	0.207	27.4	0.326
Malou	27	21.8	0.393	26.0	0.368
Meranti	14	9.9	0.379	11.1	0.326
Raninim	25	9.5	0.343	20.2	0.121
Malakas	13	12	0.288	10.0	0.354

Table 6. Sample size and Z-score for position error of each TC. Shading indicates values significant at the 90% confidence level.

Storm Name	Analysis Time Steps	Z				
		NCEP - GFS	NCEP Oper - GFS	GFS - RSM	RSM QS - RSM	NCEP Oper - RSM
Mindulle	50	1.82	1.71	1.81	-1.51	0.44
Tingting	38	0.62	2.57	0.38	-0.20	-1.20
Kompasu	18	2.63	-2.26	1.47	1.73	2.74
Namtheun	33	1.54	-0.53	2.85	0.27	2.66
Malou	27	2.99	0.44	2.00	-0.53	1.36
Meranti	14	1.95	0.70	-1.33	-1.33	-1.64
Raninim	25	1.73	-1.29	-0.26	2.04	1.13
Malakas	13	-0.58	1.82	1.58	-1.92	-0.81

Table 7. P-value for position error of each TC based on the t-distribution. Shading indicates those values that are significant at the 90% confidence level.

Storm Name	Analysis Time Steps	p-value				
		NCEP - GFS	NCEP Oper - GFS	GFS - RSM	RSM QS - RSM	NCEP Oper - RSM
Mindulle	50	0.078	0.093	0.079	0.126	0.360
Tingting	38	0.325	0.017	0.368	0.388	0.192
Kompasu	18	0.018	0.037	0.135	0.092	0.015
Namtheun	33	0.122	0.343	0.010	0.382	0.015
Malou	27	0.007	0.358	0.058	0.342	0.156
Meranti	14	0.065	0.302	0.160	0.161	0.105
Raninim	25	0.091	0.172	0.381	0.054	0.206
Malakas	13	0.326	0.080	0.115	0.069	0.275

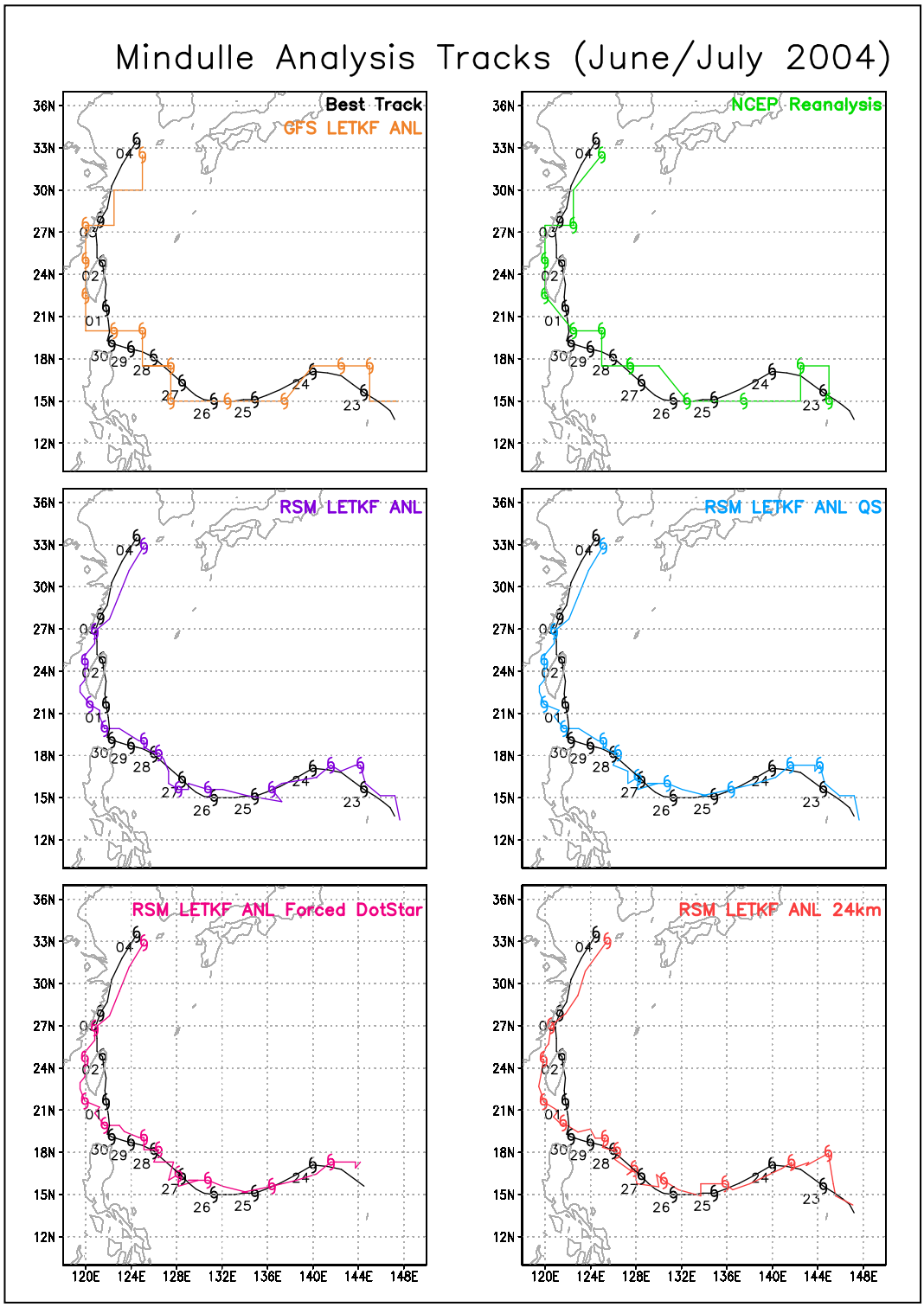

Fig. 7. Tracks for Typhoon Mindulle from each of the 6 available model configurations. In each panel black represents the JTWC track with TC symbols at 0000 UTC each day.

typically an order of magnitude smaller than positive autocorrelation values. For these instances, the autocorrelation was set to zero so that the effective sample size reduced to the number of observations in the time series. The differences in track analysis are not autocorrelated like the SLP analysis, so there is no need to calculate an effective sample size for the test statistic for track error differences.

Fig. 8 shows $\bar{\Delta}$(track) and $\bar{\Delta}$(SLP), where $\bar{\Delta}$ is the time average of NCEP Reanalysis error minus GFS LETKF ANL error. Positive values indicate that the GFS LETKF ANL performs better. Storms 11, 13, and 17 show differences in track error that are not significant at the 90 % confidence level. Of the remaining five storms, the GFS LETKF ANL produces a significantly more accurate analysis, ranging from 30-180 km better than NCEP Reanalysis. The $\bar{\Delta}$(SLP) analysis shows that only three comparisons produce significant differences. Each of these three shows an enhanced minimum SLP analysis by the GFS LETKF ANL. The scatter plot in Fig. 9 shows that there are a large number of cases where NCEP Reanalysis has much higher position error than the GFS LETKF ANL. In cases where the GFS LETKF ANL has larger error, the difference between the magnitudes of the errors tends to be larger. The scatter plot of minimum SLP (Fig. 10) of the two analyses shows smaller differences, but there are a larger number of NCEP Reanalysis points with greater error, especially for errors less than 30 hPa indicating an enhancement of minimum SLP analysis by the GFS LETKF ANL.

Fig. 11 is the same as Fig. 8 except that $\bar{\Delta}$ is defined as NCEP Oper ANL minus GFS LETKF ANL. Positive values indicate that the GFS LETKF ANL

performs better. The lack of significance among several of track comparisons indicates that the GFS LETKF ANL performs similarly to the NCEP Oper ANL for about half of the storms. The NCEP Oper ANL performs track analysis significantly better in three of the four storms. The scatter plot of NCEP Oper ANL error vs. GFS LETKF ANL error supports the findings of the average $\bar{\Delta}$(track) shown in Fig. 12. Here, large number of high-error data points indicate lower errors for GFS LETKF ANL than NCEP Oper ANL. Fig. 13 shows the scatter plot for minimum SLP, indicating no difference between the two analyses for errors less than 30 hPa. For larger errors, GFS LETKF ANL performs better than NCEP Oper ANL.

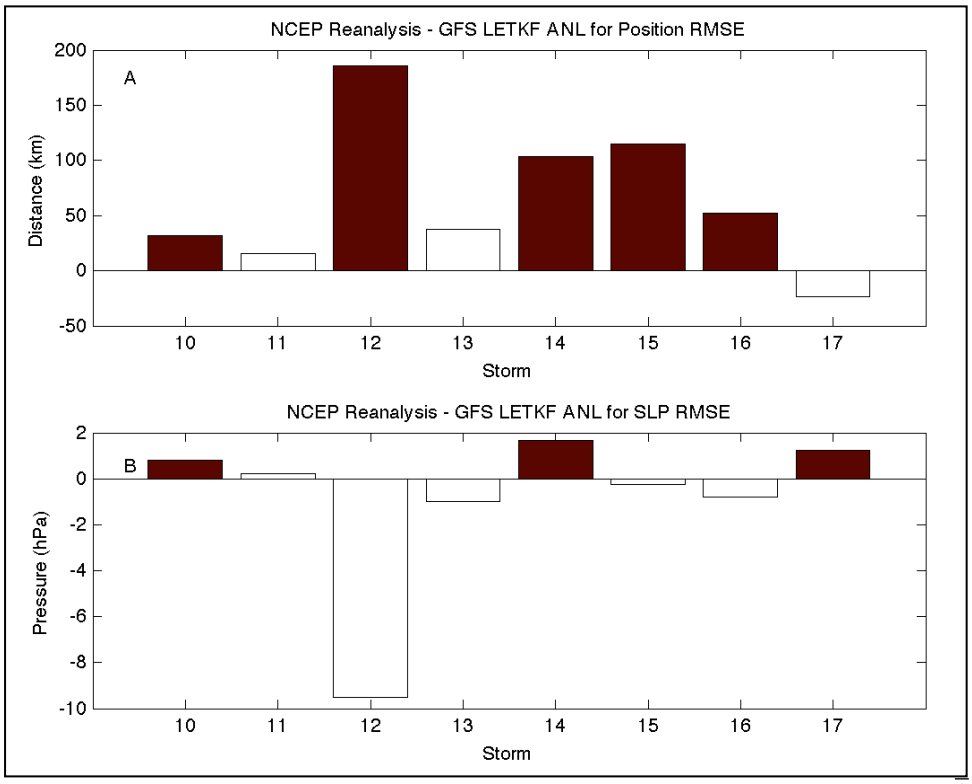

Fig. 8. NCEP Reanalysis minus GFS LETKF ANL root mean square error for A) $\bar{\Delta}$ (track) and B) $\bar{\Delta}$(SLP) for each of the eight cyclones. Shaded boxes indicate the storms in which $\bar{\Delta}$ is significant.

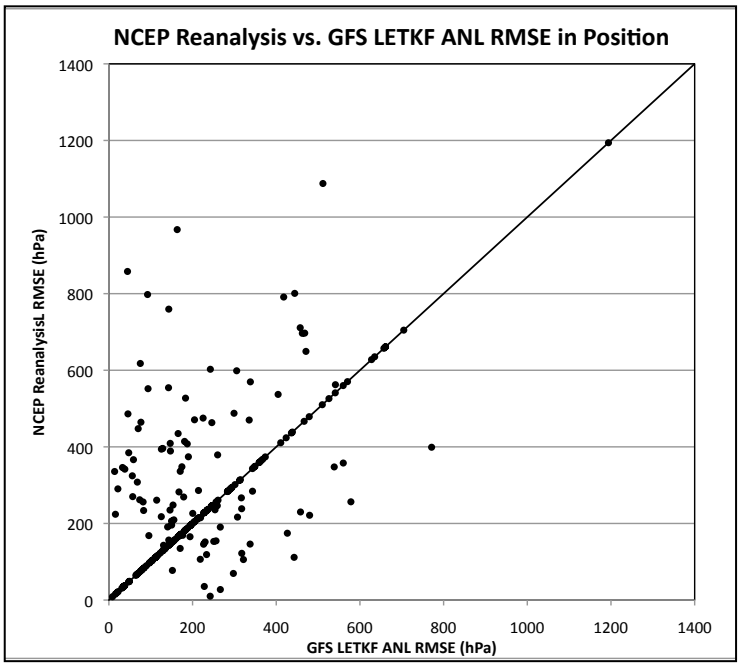

Fig. 9. RMSE in position analysis for NCEP Reanalysis vs. GFS LETKF ANL for all data points. The scatter plot also shows the 45° line between the errors.

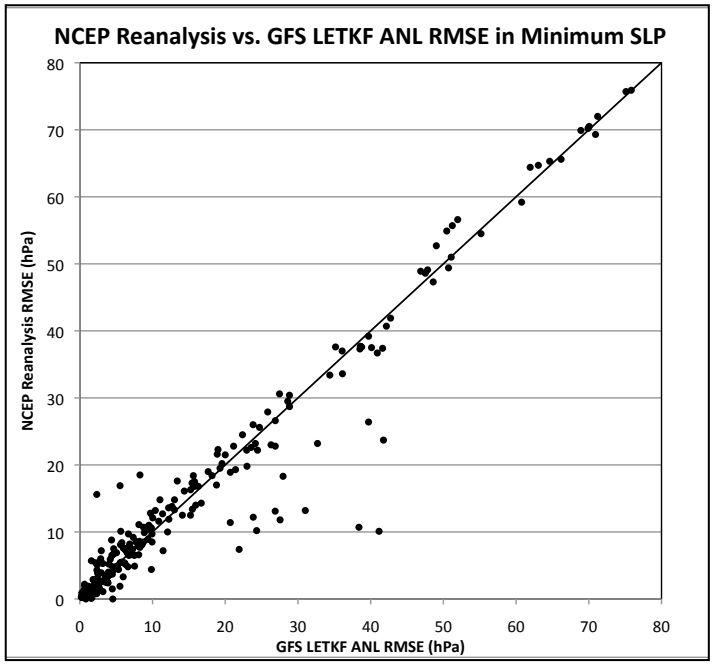

Fig. 10. RMSE in minimum SLP analysis for NCEP Reanalysis vs. GFS LETKF ANL for all data points. The scatter plot also shows the 45° line between the errors.

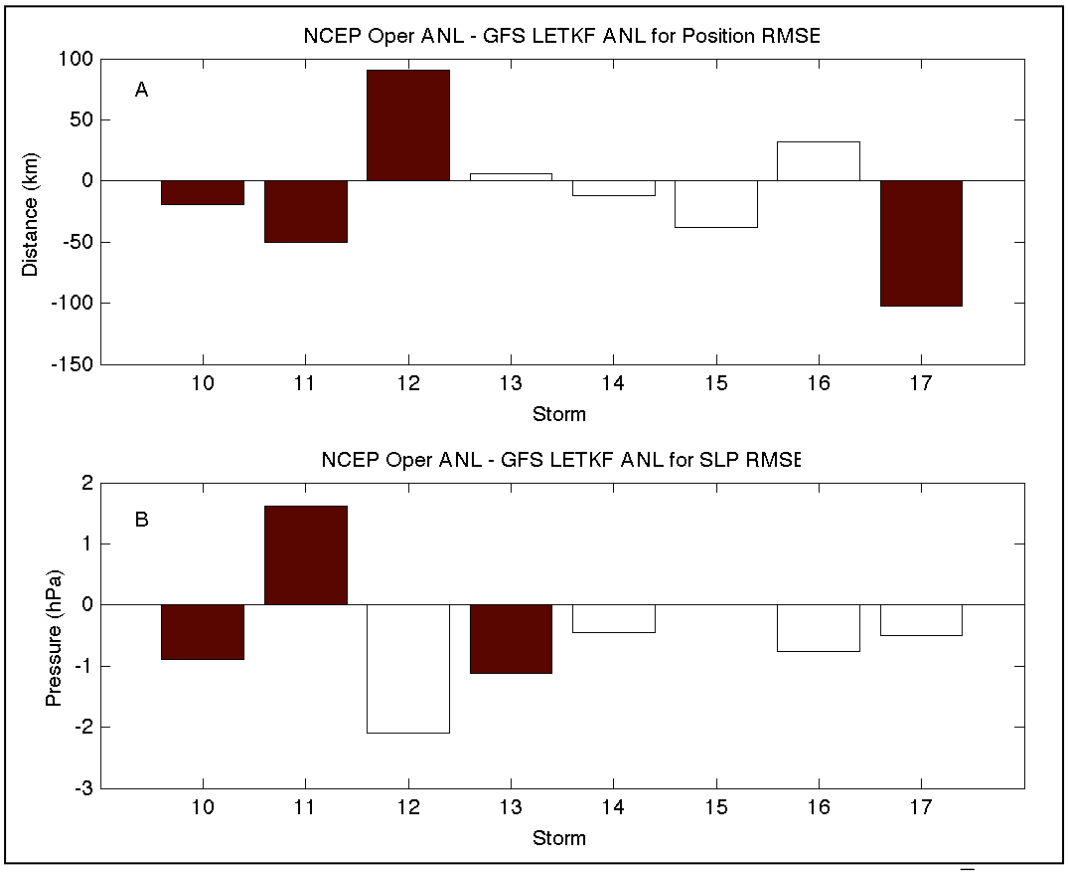

Fig. 11. NCEP Oper ANL minus GFS LETKF ANL root mean square error for A) $\bar{\Delta}$(track) and B) $\bar{\Delta}$(SLP) for each of the eight cyclones. Shaded boxes indicate the storms for which $\bar{\Delta}$ is significant.

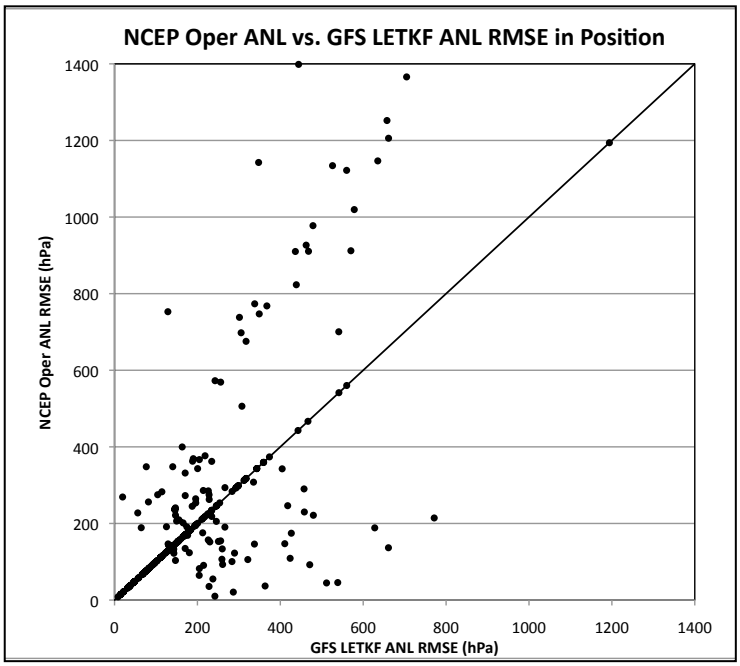

Fig. 12. RMSE in position analysis for NCEP Oper ANL vs. GFS LETKF ANL for all data points. The scatter plot also shows the 45° line between the errors.

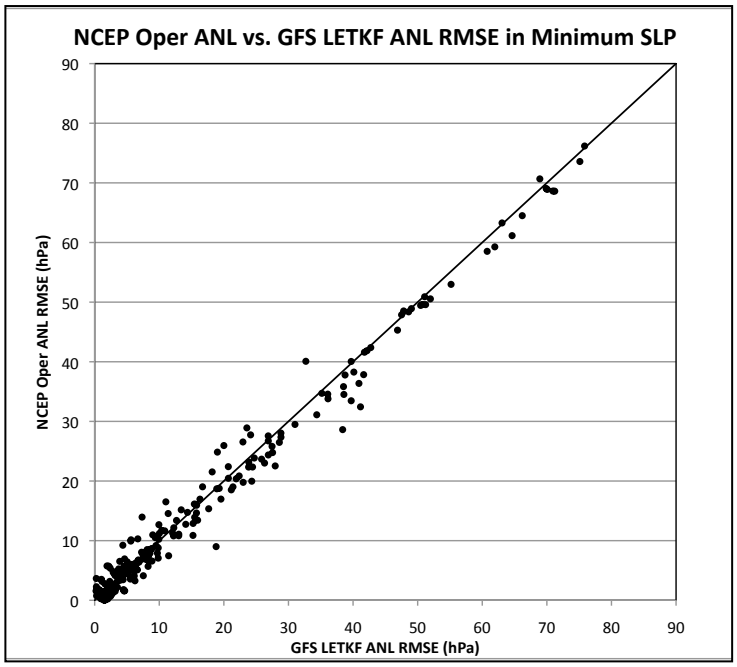

Fig. 13. RMSE in minimum SLP analysis for NCEP Oper ANL vs. GFS LETKF ANL for all data points. The scatter plot also shows the 45° line between the errors.

Fig. 14 is the same as Fig. 8 except that $\bar{\Delta}$ is defined as NCEP Oper ANL minus RSM LETKF ANL. Statistically significant track results indicate that the RSM LETKF ANL performs better in the analysis of position of the cyclone than the NCEP Oper ANL for two storms. The lack of significant differences in SLP for most of the storms indicates that our regional component performs similarly to the NCEP Oper ANL. The scatter plots for the position (Fig. 15) and minimum SLP (Fig. 16) errors also support a slightly better analyses by the RSM LETKF ANL.

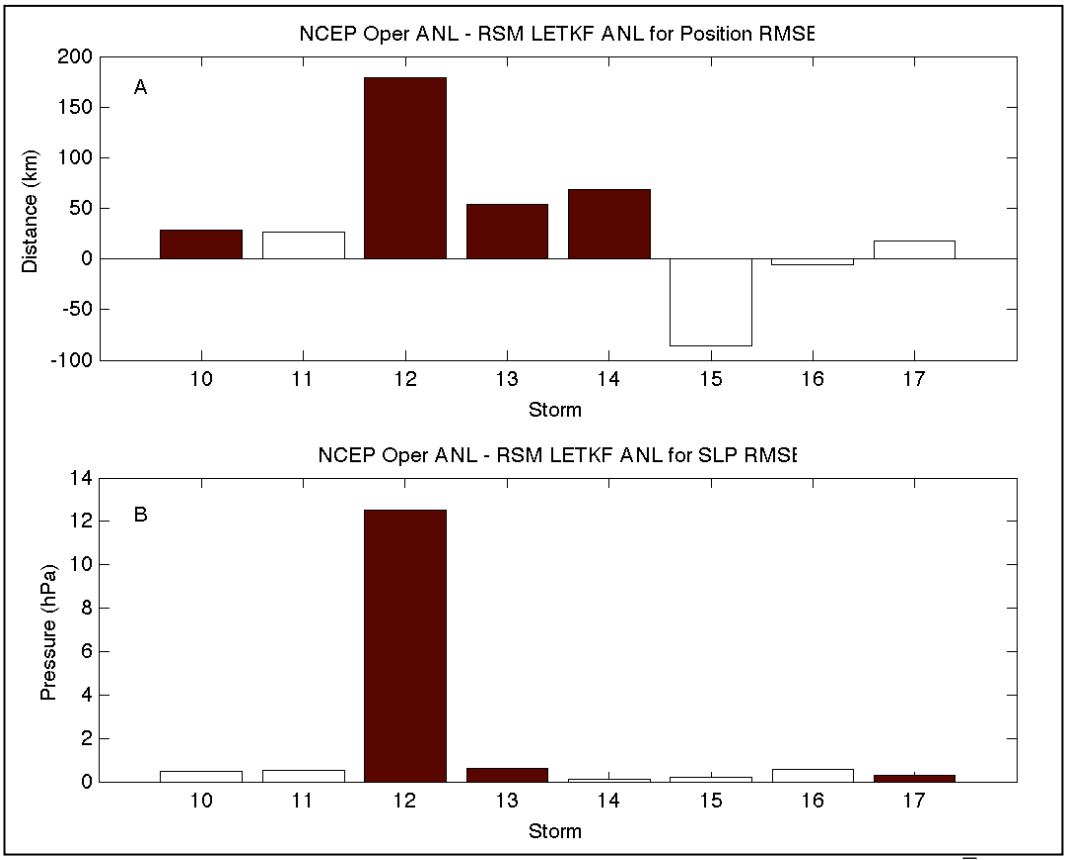

Fig. 14. NCEP Oper ANL minus RSM LETKF ANL root mean square error for A) $\bar{\Delta}$(track) and B) $\bar{\Delta}$(SLP) for each of the eight cyclones. Shaded boxes indicate the storms for which $\bar{\Delta}$ is significant.

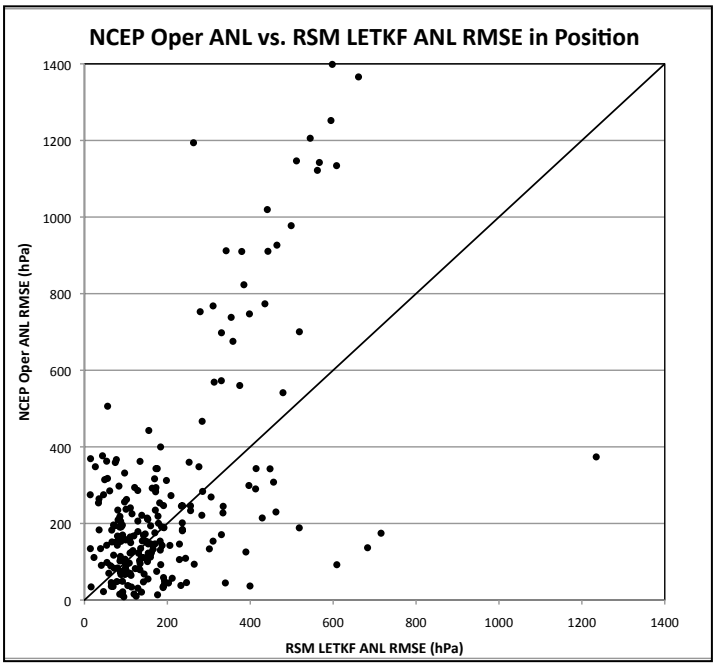

Fig. 15. RMSE in position analysis for NCEP Oper ANL vs. RSM LETKF ANL for all data points. The scatter plot also shows the 45° line between the errors.

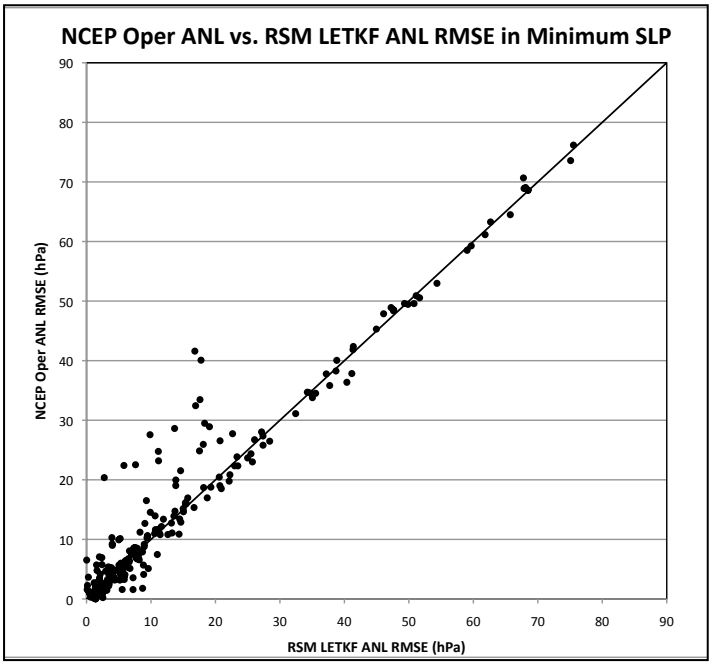

Fig. 16. RMSE in minimum SLP analysis for NCEP Oper ANL vs. RSM LETKF ANL for all data points. The scatter plot also shows the 45° line between the errors.

Fig. 17 is the same as Fig. 8 except that $\bar{\Delta}$ is defined as GFS LETKF ANL minus RSM LETKF ANL, where positive values indicate a superior RSM LETKF ANL. For $\bar{\Delta}$(track) error comparison, only storms 1, 4, and 5 have significant mean differences, all of which indicate a better RSM LETKF ANL that is more accurate by about 25 – 75 km at a 90% confidence level. The $\bar{\Delta}$(SLP) analysis provides significant differences for only three storms, all of which indicate that the RSM LETKF ANL is the better intensity analysis. This result provides further support for increasing resolution to more accurately capture the true minimum SLP, as well as using a second data assimilation step in the regional component of the system. Fig. 18 shows that RSM LETKF ANL exhibits lower errors than GFS LETKF ANL for position analysis. The scatter plot for minimum SLP (Fig. 19) indicates that the RSM performs better than the GFS LETKF ANL in cases where the error is smaller than about 40 hPa, while the two systems perform similarly for larger error cases.

In Fig. 11, Fig. 14, and Fig. 17, the minimum SLP error differences are the largest for TS Kompasu (TC 3). The GFS LETKF ANL recognized the flow feature only in the vorticity field. Because the pressure of the cyclone and the pressure of the surroundings were so similar for this weak storm, the pressure errors in the GFS were much higher than in the benchmarks, where the surrounding SLP was lower, and in the RSM LETKF ANL which captured a low-pressure center, as well as a vorticity maximum.

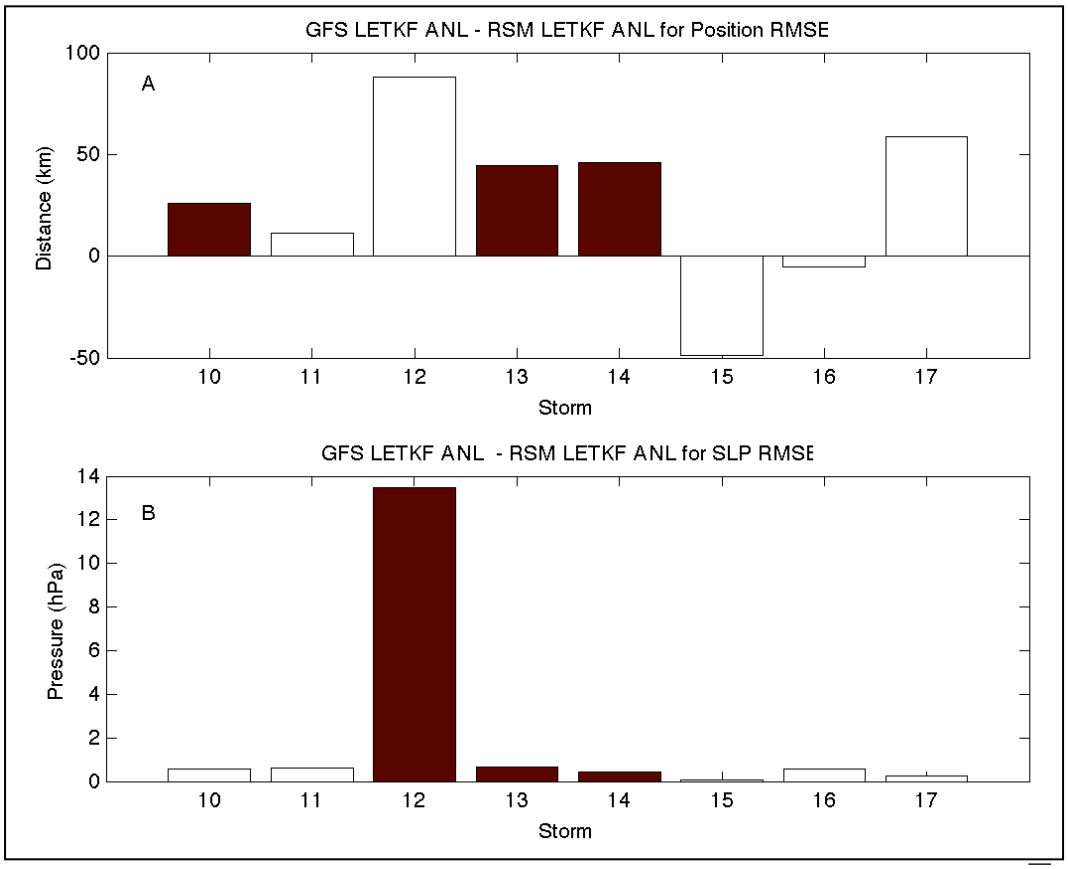

Fig. 17. GFS LETKF ANL minus RSM LETKF ANL root mean square error for A) $\overline{\Delta}$ (track) and B) $\overline{\Delta}$ (SLP) for each of the eight cyclones. Shaded boxes indicate the storms for which $\overline{\Delta}$ is significant.

RSM LETKF ANL QS and RSM LETKF ANL are compared in Fig. 20, where positive differences indicate a better analysis by the RSM LETKF ANL QS. These results are quite different from the previous relationships. None of the minimum SLP differences are significant, which means the two analyses produce almost identical minimum SLP estimates. This result is supported by the very linear one-to-one relationship shown in the scatter plot of Fig. 21. Only one of the storms shows that track analysis is significantly better for the RSM LETKF ANL QS, two are

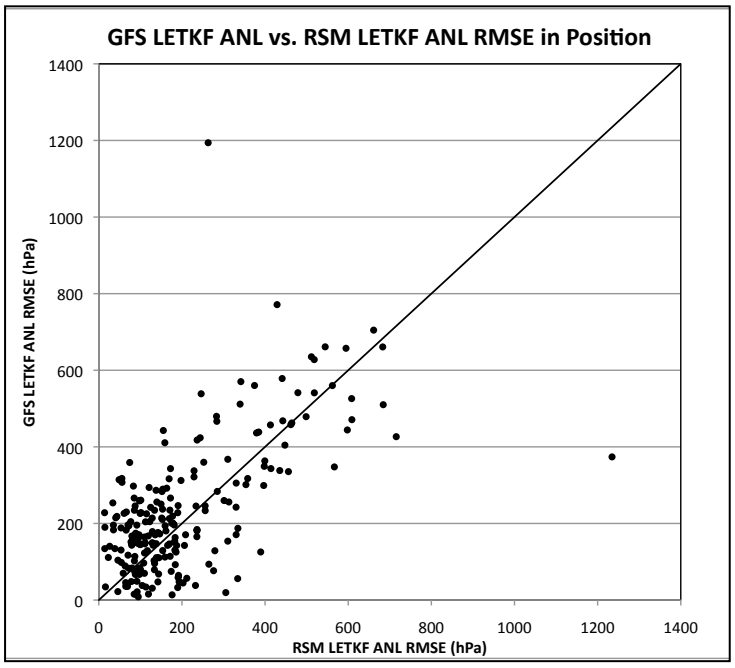

Fig. 18. RMSE in position analysis for GFS LETKF ANL vs. RSM LETKF ANL for all data points. The scatter plot also shows the 45° line between the errors.

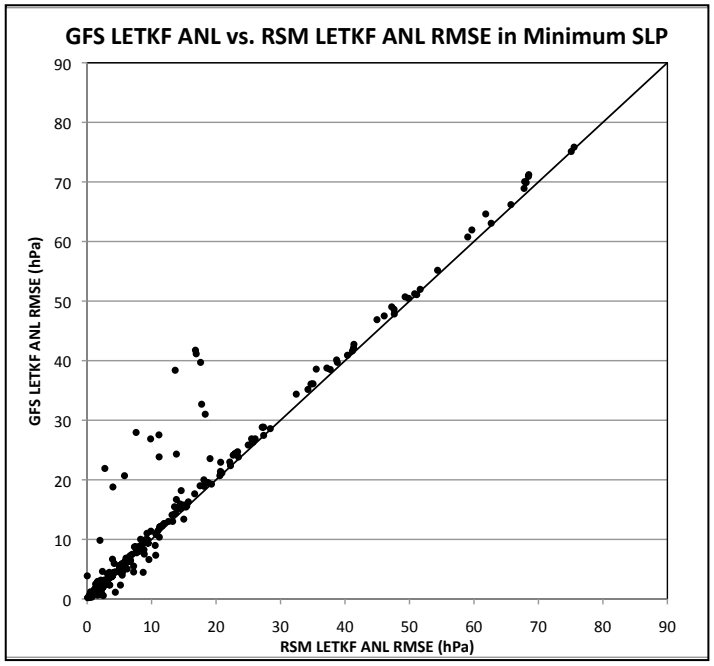

Fig. 19. RMSE in minimum SLP analysis for GFS LETKF ANL vs. RSM LETKF ANL for all data points. The scatter plot also shows the 45° line between the errors.

significantly better for RSM LETKF ANL, while the rest provide insignificant differences. The scatter plot for this comparison (Fig. 22) of position analysis also supports that the RSM LETKF ANL may be slightly better. These results indicate that assimilating the QuikSCAT data only in the regional component of our system has little effect on overall track and intensity analysis.

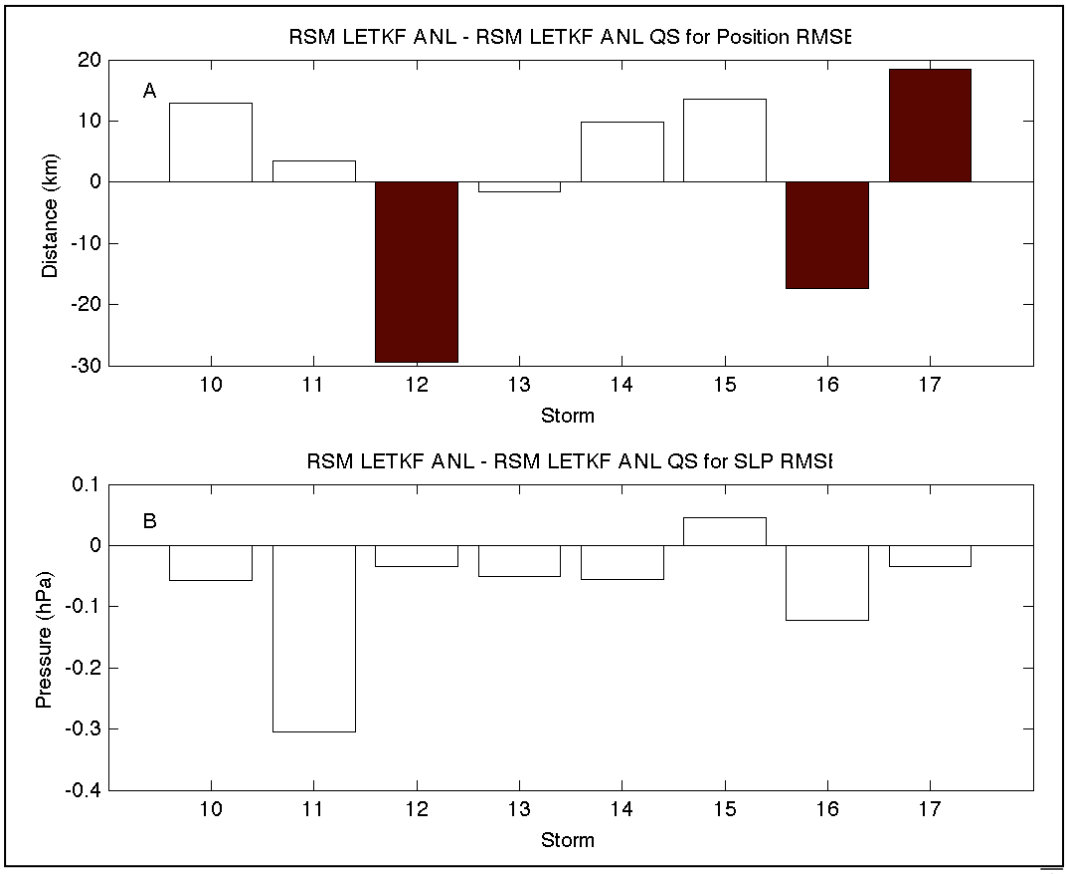

Fig. 20. RSM LETKF ANL QS minus RSM LETKF ANL root mean square error for A) $\overline{\Delta}$ (track) and B) $\overline{\Delta}$ (SLP) for each of the eight cyclones. Shaded boxes indicate the storms for which $\overline{\Delta}$ is significant.

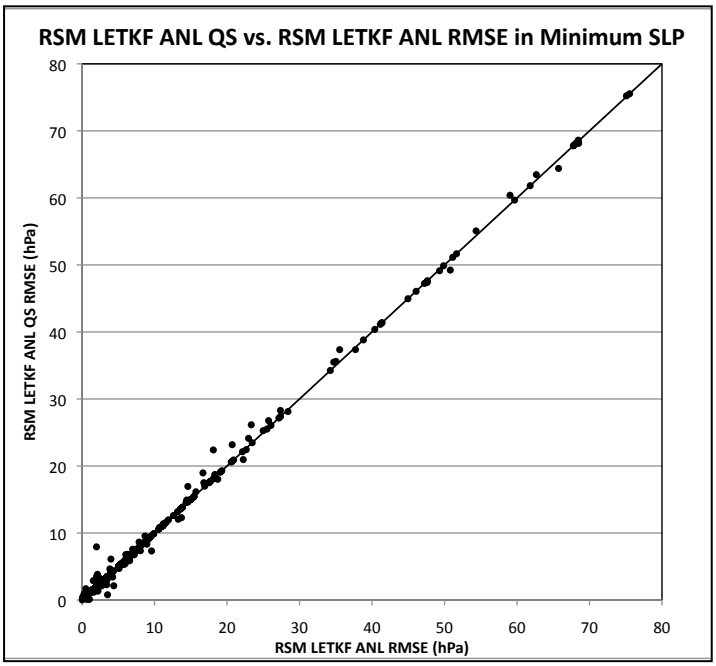

Fig. 21. RMSE in minimum SLP analysis for RSM LETKF ANL QS vs. RSM LETKF ANL for all data points. The scatter plot also shows the 45° line between the errors.

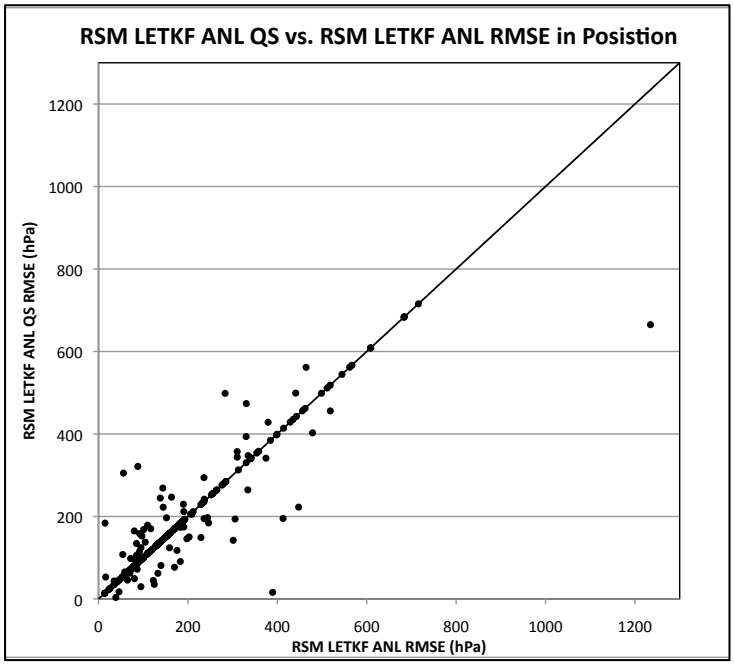

Fig. 22. RMSE in position analysis for RSM LETKF ANL QS vs. RSM LETKF ANL for all data points. The scatter plot also shows the 45° line between the errors.

The peak magnitudes of the SLP errors vary widely from storm to storm, increasing almost linearly with storm intensity. The peak magnitudes of track errors, however, have the opposite response and tend to be lower on average for intense cyclones. Grouping the individual point observations by JTWC Best Track wind speed intensity for strong, moderate, and weak cyclones and analyzing the distribution of errors for each explains the general response. Fig. 23 and Fig. 24 show intensity and track error box plots as a function of Best Track wind speed. In each model configuration, there are 15 strong TCs, 48 moderate TCs, and 155 weak TCs.

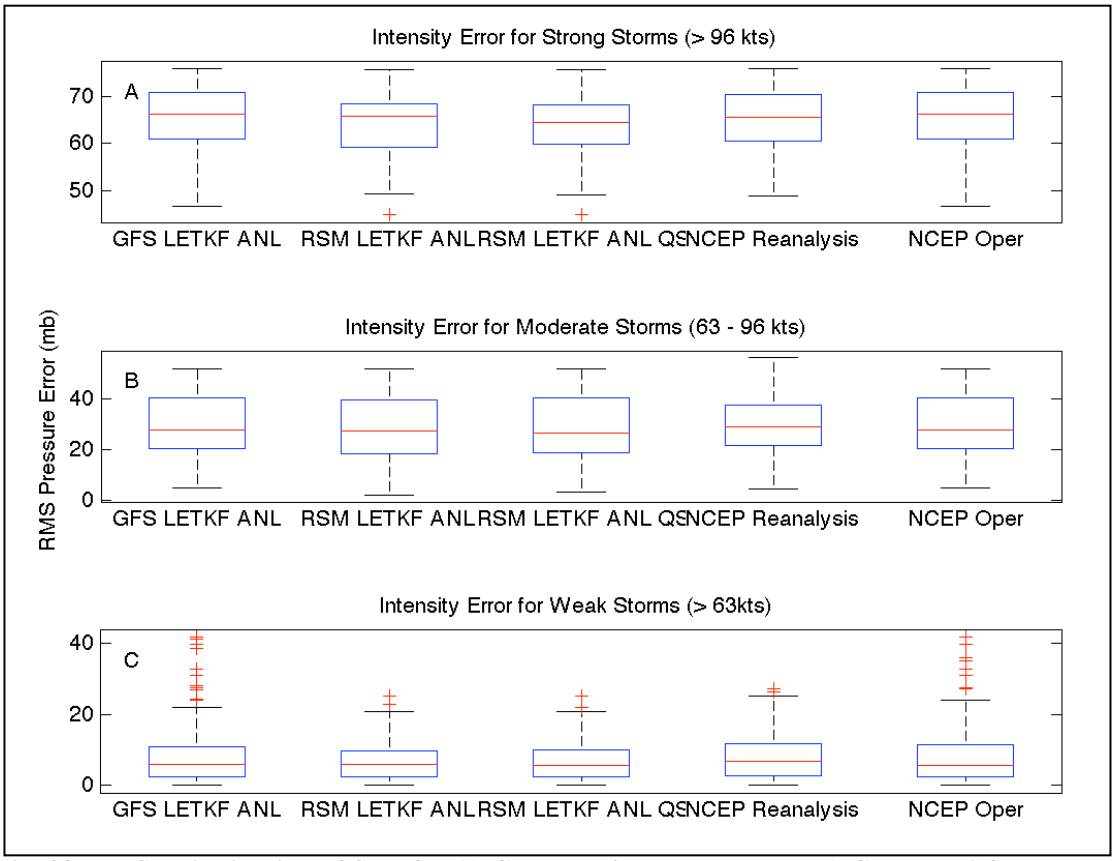

Fig. 23. RMSE distribution of SLP for A) Category 4 and 5 cyclones, B) Category 1-3 cyclones, and C) tropical storms and depressions. Each box represents the errors associated with a different component or configuration of the system.

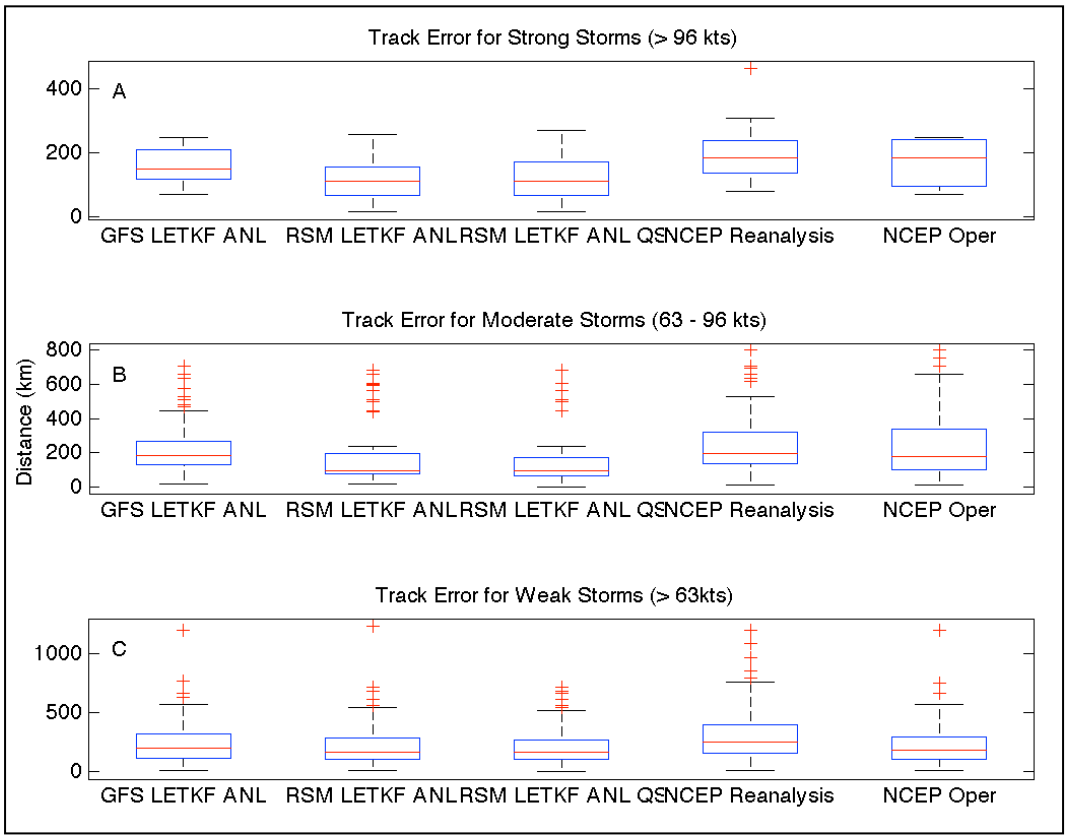

Fig. 24. RMSE distribution of position for A) Category 4 and 5 cyclones, B) Category 1-3 cyclones, and C) tropical storms and depressions. Each box represents the errors associated with a different analysis.

Fig. 23A shows the distribution of SLP errors for Category 4 and 5 TC observations. The magnitude of SLP errors are nearly twice those seen in moderate cyclones (Category 1-3) in Fig. 23B, which are close to three times those of the weak cyclones (Tropical Storms and Depressions) (Fig. 23C). The inter-quartile ranges for the moderate cyclones are approximately double those for other storm strength categories. Recall that a larger inter-quartile range means there is more variability in the data. The increased variability of the minimum SLP found in the moderate category is possibly due the inverse relationship seen in intensity and error. For the

well-analyzed lower Category SLP observations included in this bin, the errors are small, whereas the minimum SLP errors are much larger for the Category 3 storms in the same bin. This is also reflected in the fact that the distributions are slightly skewed toward higher errors (i.e. the median is closer to the lower edge of the box), which is not as pronounced for the other intensity bins. The number of outliers is greatest for the weak cyclones, indicating that the analysis of minimum SLP is least reliable when a strong closed low is not present.

The distributions of position errors in Fig. 24 for each of the storm strength categories show similar magnitudes for the inter-quartile ranges, all ranging from about 100 to 200 km. The smallest, however, tend to occur in the case of strong cyclones indicating a more accurate, less varying track analysis. The median tends to be highest (around 200 km) in the weak storms (Fig. 24C), indicating a lack of skill in determining the position for each weak TC. There are only a few outliers for the 155 weak observations in each of the analyses, indicating that although there is higher error, there is some degree of predictability. Relatively, there are more outliers for the moderate storms (Fig. 24B) for which there are only 48 observations. This indicates that although position errors are typically lower, they are also more variable than for strong or weak cyclones. The outliers for the weak and moderate storms range between 400 and 800 km, which can be a very large, and potentially deadly track analysis error. Again, for the moderate storms, the distribution of errors is skewed toward higher errors, whereas the strong and weak cyclone track error distributions are fairly symmetric.

b. DOTSTAR

Although the DOTSTAR data leads to only an average of less than 1 hPa improvement over the analysis lifetime of Typhoon Mindulle, the *t*-test indicates that the small improvement is significant at the 90% confidence level. The RSM LETKF ANL DS did not indicate significant improvement for track analysis on average.

The distribution of the RSM LETKF ANL DS errors in Fig. 25 indicates the strong similarity in the two cases for both minimum SLP and track. Minimum SLP distributions for both cases are skewed heavily toward higher errors. The distributions of the two experiments are almost identical. This assessment of the difference between the two cases provides no obvious conclusion as to which is more accurate.

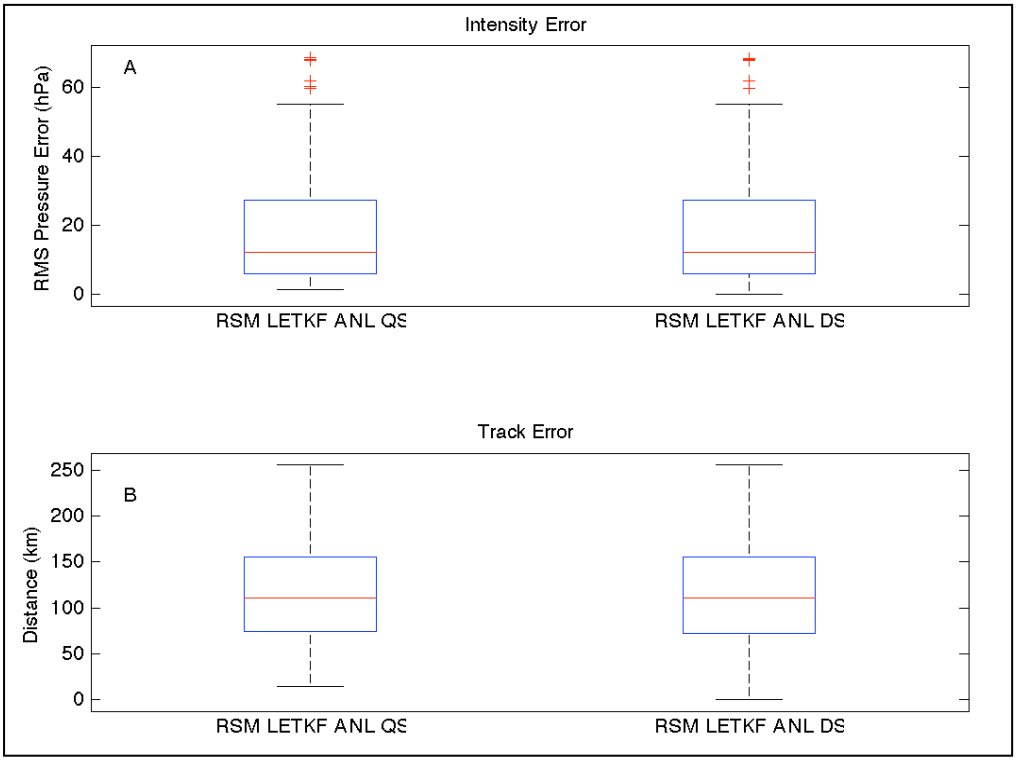

Fig. 25. RMSE distribution for A) minimum SLP and B) position for the RSM LETKF ANL QS and RSM LETKF ANL DS cases.

c. *Increased Resolution*

The RSM LETKF ANL 24 does not show a significant improvement over RSM LETKF ANL QS in capturing minimum SLP. Fig. 26 shows the average difference between minimum SLP error between the two experiments over the course of each of the three cyclones. None of the cyclone analyses indicate a significant difference in minimum SLP error and the errors shown in Fig. 27 are all fairly clustered and small in magnitude. One cyclone produced a $\overline{\Delta}$(track) value that was significant, showing that the RSM LETKF ANL 24 degrades that track analysis, which is confirmed by Fig. 28.

The box plot shown in Fig. 29A indicates that the distribution of SLP errors is nearly identical for the RSM LETKF ANL QS and RSM LETKF ANL 24 km experiments. Fig. 29B shows slight differences between distributions of track errors. The LETKF ANL 24 case shows more variability of track error with a slightly larger inter-quartile range, but fewer outliers.

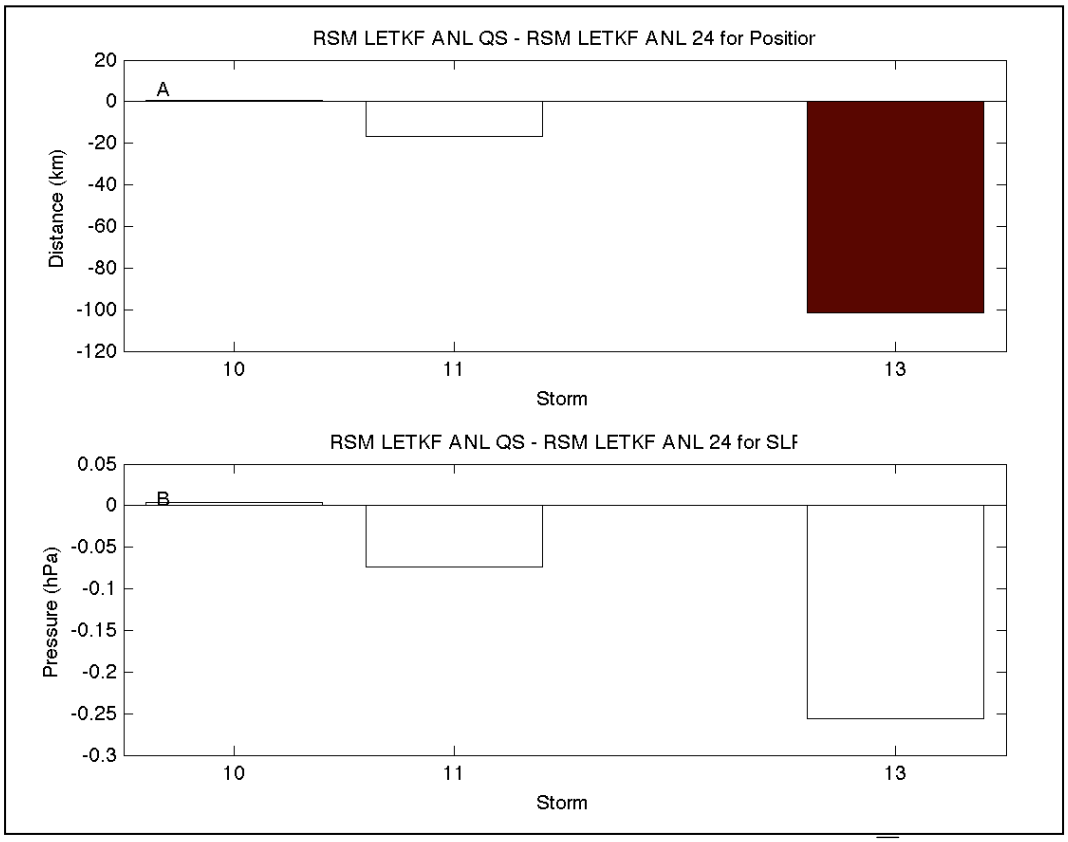

Fig. 26. RSM LETKF ANL QS minus RSM LETKF ANL 24 RMSE for A) $\overline{\Delta}$(track) and B) $\overline{\Delta}$(SLP) for each of the three cyclones. Shaded boxes indicate the only significant differences for either $\overline{\Delta}$(track) or $\overline{\Delta}$(SLP).

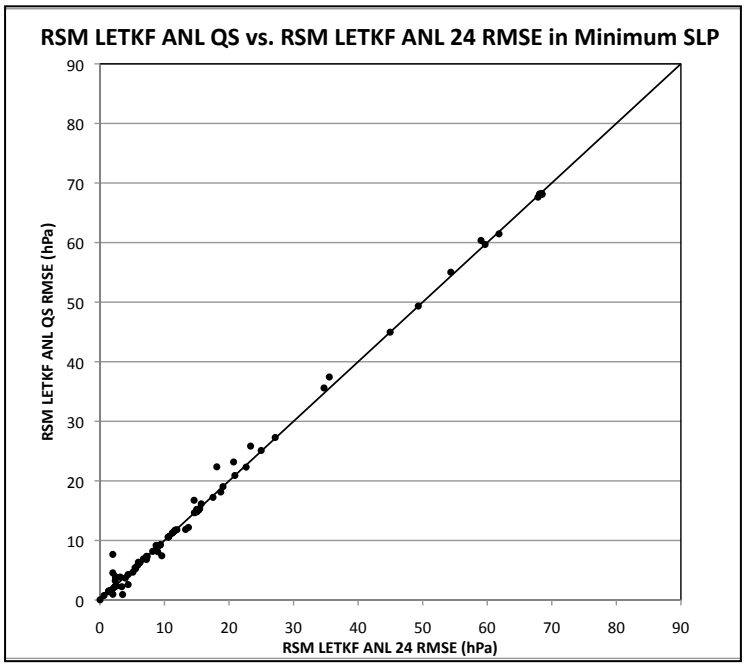

Fig. 27. RMSE in minimum SLP analysis for RSM LETKF ANL QS vs. RSM LETKF ANL 24 for all data points. The scatter plot also shows the 45° line between the errors.

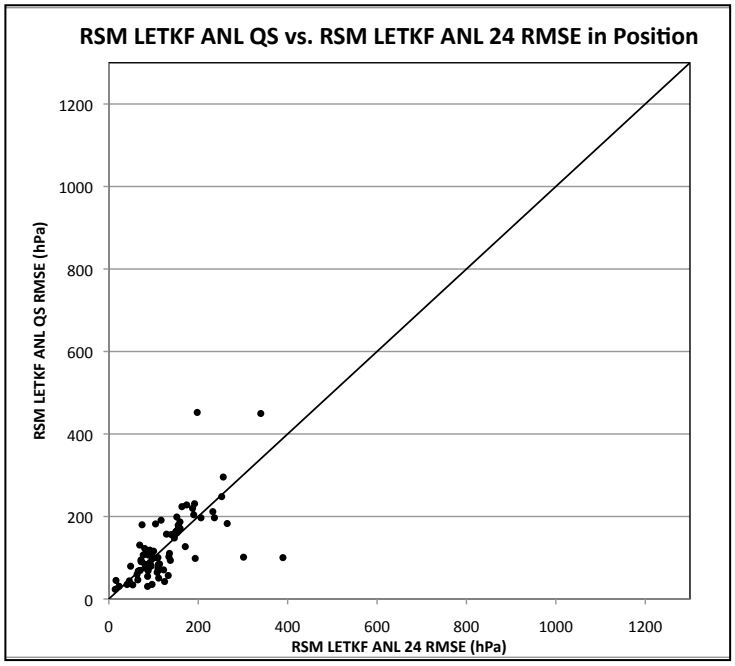

Fig. 28. RMSE in position analysis for RSM LETKF ANL QS vs. RSM LETKF ANL 24 for all data points. The scatter plot also shows the 45° line between the errors.

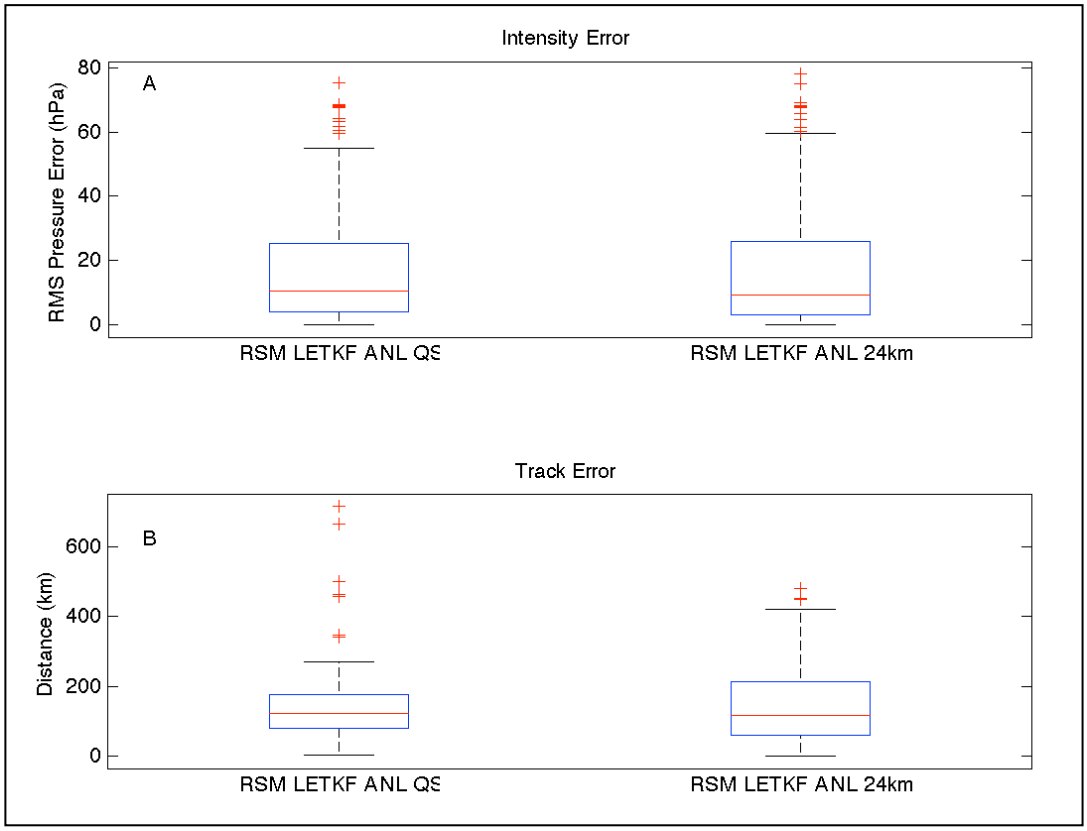

Fig. 29. RMSE distribution for A) minimum SLP and B) position for the RSM LETKF ANL QS and RSM LETKF ANL 24 cases.

6. FORECAST VERIFICATION RESULTS

Deterministic forecasts were started from three analysis times and ran for 120 h. Forecast 1 was started on June 26 at 0000 UTC using GFS LETKF ANL, RSM LETKF ANL, RSM LETKF ANL QS, and the GFS operational analysis truncated to T62 resolution (RSM T62-O). Typhoons Mindulle (10) and Tingting (11) were observed during this time. Just after initialization, both underwent rapid intensification. Forecast 2 was started on July 27 at 0000 UTC using the same set of five cases as initial and boundary conditions. Typhoon Namtheun was observed at its peak intensity at the time of initialization. Lastly, Forecast 3 was started on August 9 at 1800 UTC. Typhoon Rananim was observed during the first 84 h of the forecast, thus only 14 time steps were analyzed for this case versus the 21 of the other initialization times.

Again, testing the performance of each system component with the *t*-test for correlated data, we find that results are slightly ambiguous with the smaller sample size. First, the system is tested against a deterministic operationally based forecast with $\overline{\Delta}$ defined as RSM LETKF FCST minus RSM T62-O FCST. None of the forecasts resulted in $\overline{\Delta}$(track) that is significant at the 90% confidence level. Namtheun is the only forecast to provide statistically significant $\overline{\Delta}$(SLP), which showed an average 0.4 hPa increase of skill by the RSM T62-O FCST.

Testing the significance of the regional component of the system is done by using $\overline{\Delta}$ defined as GFS LETKF FCST minus RSM LETKF FCST. The analysis of $\overline{\Delta}$

(SLP) shows that significant differences exist for both Mindulle and Tingting, where the RSM LETKF FCST performs better by about 0.7 hPa. The results of the $\overline{\Delta}$(track) comparison show that none of the storm tracks are significantly improved or degraded by the RSM LETKF FCST.

The differences between RSM LETKF FCST and RSM LETKF FCST QS cases provide mixed results. None of the $\overline{\Delta}$(SLP) values is significant, however $\overline{\Delta}$ (track) is significant for Typhoon Rananim. Although the RSM LETKF FCST QS shows only a slight improvement of about 18 km, it is significant. In the case of Typhoon Namtheun, both $\overline{\Delta}$ comparisons were several orders of magnitude smaller than the values observed for the other cyclones. This is likely due to the absence of QuikSCAT data within the regional domain at the time of initialization. The analyses at initialization show slight differences in minimum SLP, but the overall perturbation is not enough to result in different forecasts.

Time evolution of forecast errors is as important, if not more important, than the average error when analyzing the value of system components. Fig. 30 shows the forecast time series of minimum SLP error from Typhoon Mindulle over 120 h starting on June 26 at 0000 UTC. During the first half of the forecast period, the RSM T62-O performed most poorly. This continued until Mindulle reached its maximum strength, at which time all of the forecasts provided similar intensities, with no evident skill-leader. Fig. 31 is the same as Fig. 30 for Typhoon Tingting. At t = 60 h, all cases begin performing similar and continue to do so until the last 24 h. At Day 5, the RSM LETKF FCST QS has the lowest minimum SLP errors. The Namtheun forecast of

minimum SLP (Fig. 32) is much more conclusive. For short-term forecasts, each experiment has very similar error. At t = 48 h, the RSM T62-O FCST begins to consistently provide the superior forecast. For the last 72 h of the period, RSM LETKF FCST (identical to RSM LETKF FCST QS for this initialization) performs better than GFS LETKF FCST. Finally, Rananim (Fig. 33) shows a similar pattern as Namtheun in that the forecasts are all similar for short lead times, but after t = 36 h, GFS LETKF FCST consistently has the highest amount of errors.

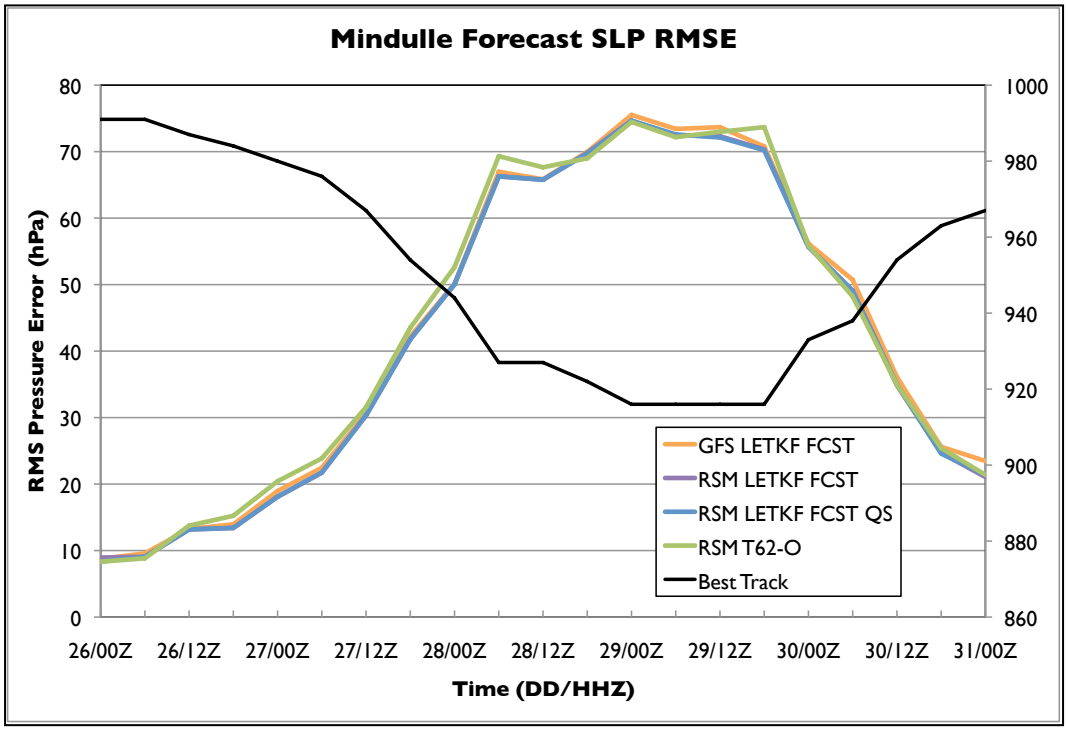

Fig. 30. RMSE for minimum SLP forecast initialized on June 26 0000 UTC for Typhoon Mindulle. Best Track minimum SLP (black) corresponds to the secondary axis, while all other series are error in the forecast and correspond to the primary axis.

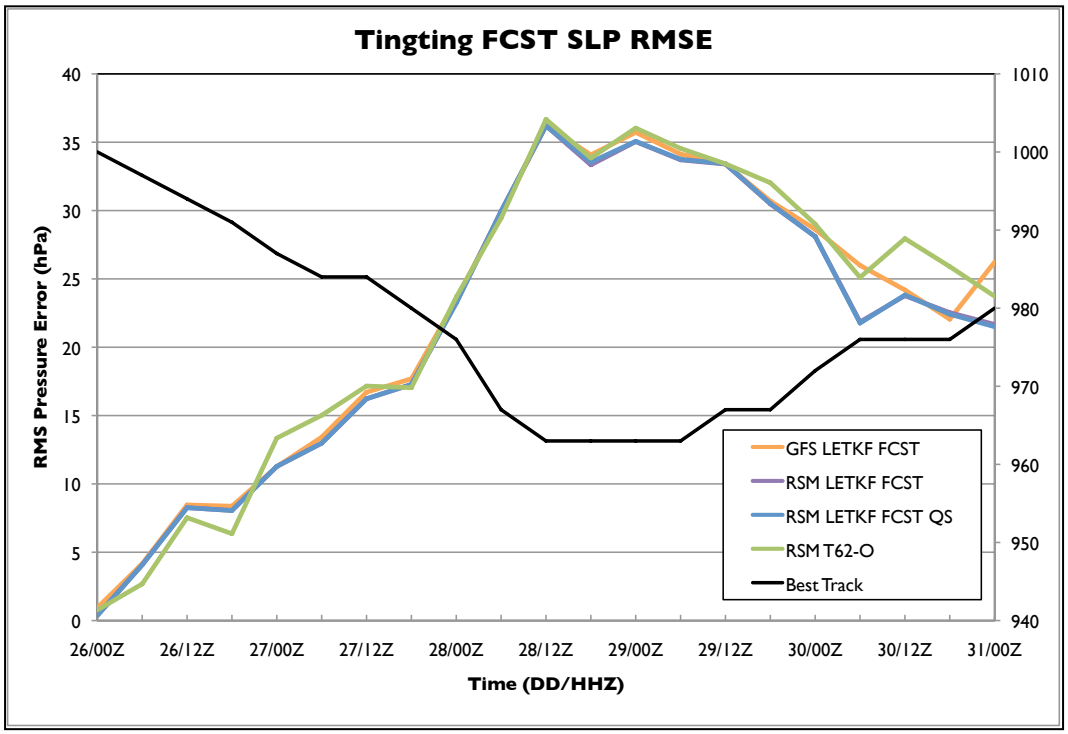

Fig. 31. Same as Fig. 30 except for Typhoon Tingting.

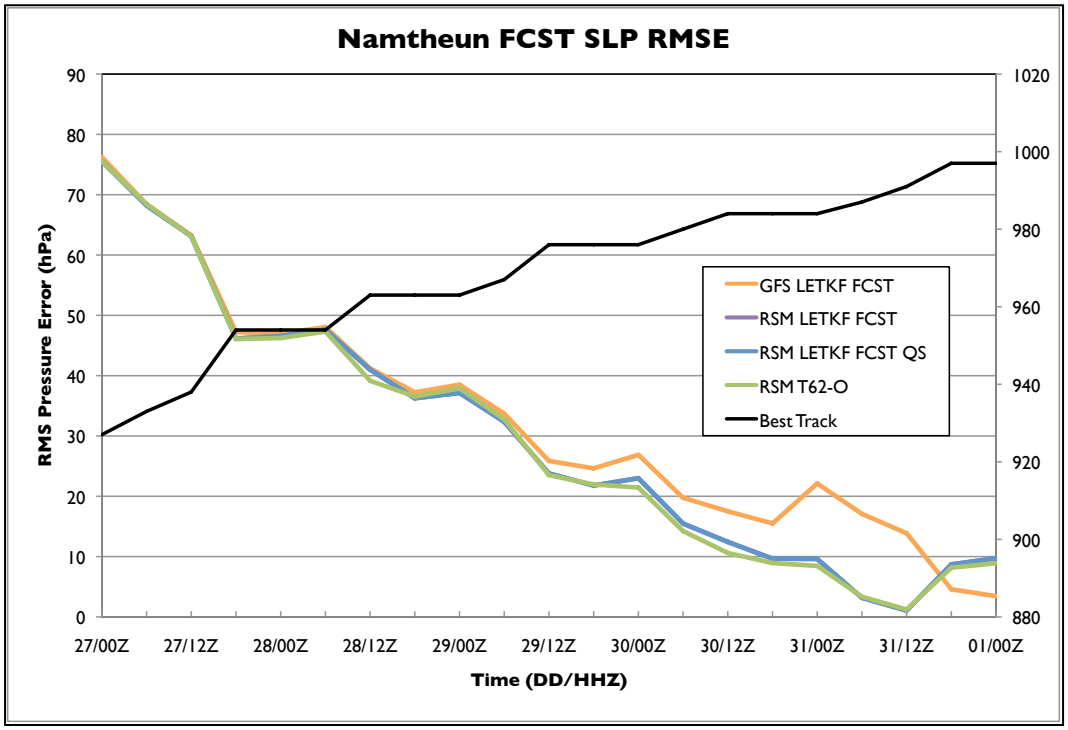

Fig. 32. Same as Fig. 30 except initialized at 0000 UTC July 27 for Typhoon Namtheun.

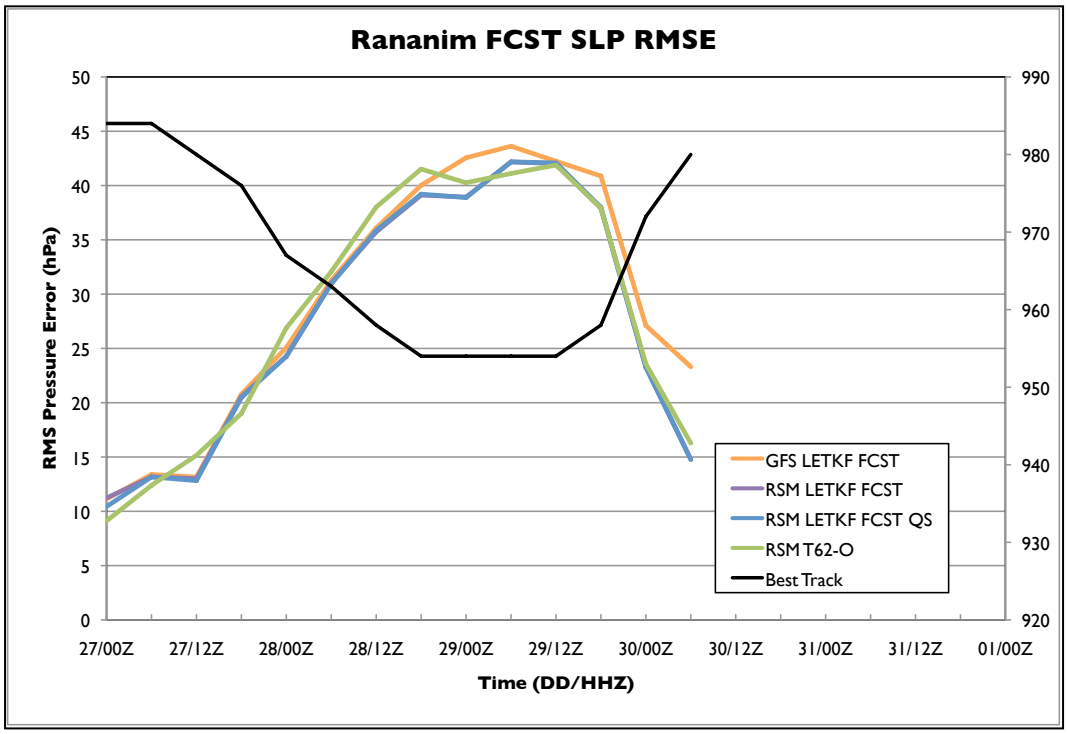

Fig. 33. Same as Fig. 30 except initialized at 1800 UTC August 9 for Typhoon Rananim.

Root mean square error for track position is shown in Fig. 32 for Mindulle. None of the cases displays a noticeable trend in time, which is unexpected since forecast error should grow with time. Fig. 33 shows the forecast track error for Tingting. Each of the track forecasts show an increasing trend in error over the entire 120 h, however the slope increases upward rapidly in the last 24 h. Namtheun track forecast error (Fig. 34) shows that the GFS LETKF FCST consistently performs best out of all FCST cases. Again, the forecast error increases with time for all forecasts with a steeper slope than occurred in the Tingting forecast. . Finally, Rananim track errors (Fig. 35) show that past 48 h, the GFS LETKF FCST provides an enhanced

track forecast over the RSM forecasts. The errors of all the RSM forecasts increase rapidly after approximately 24 h.

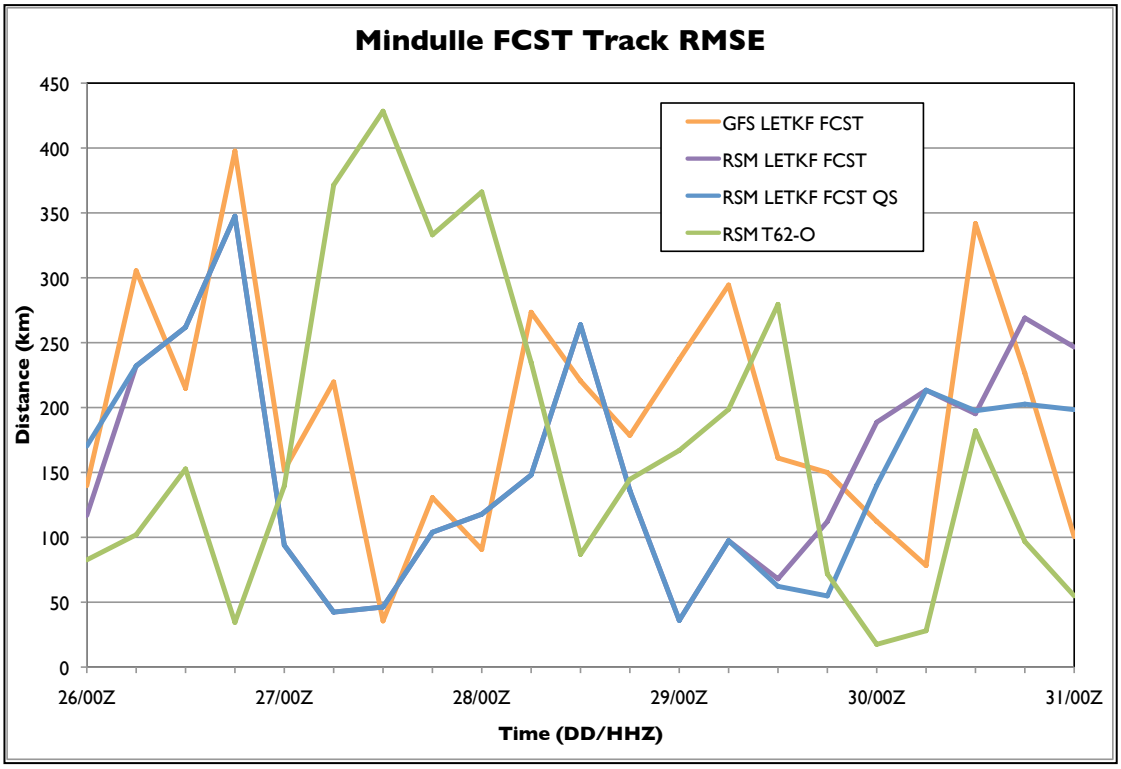

Fig. 34. RMSE for track forecast initialized on June 26 0000 UTC for Typhoon Mindulle.

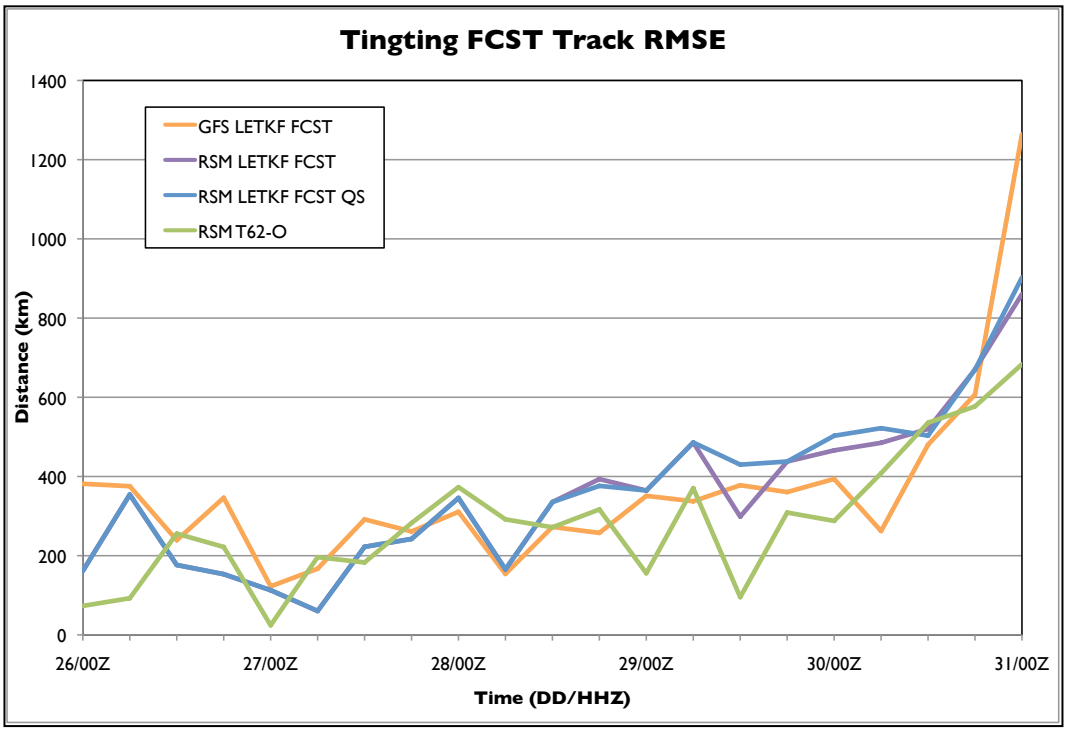

Fig. 35. Same as Fig. 34 except for Typhoon Tingting.

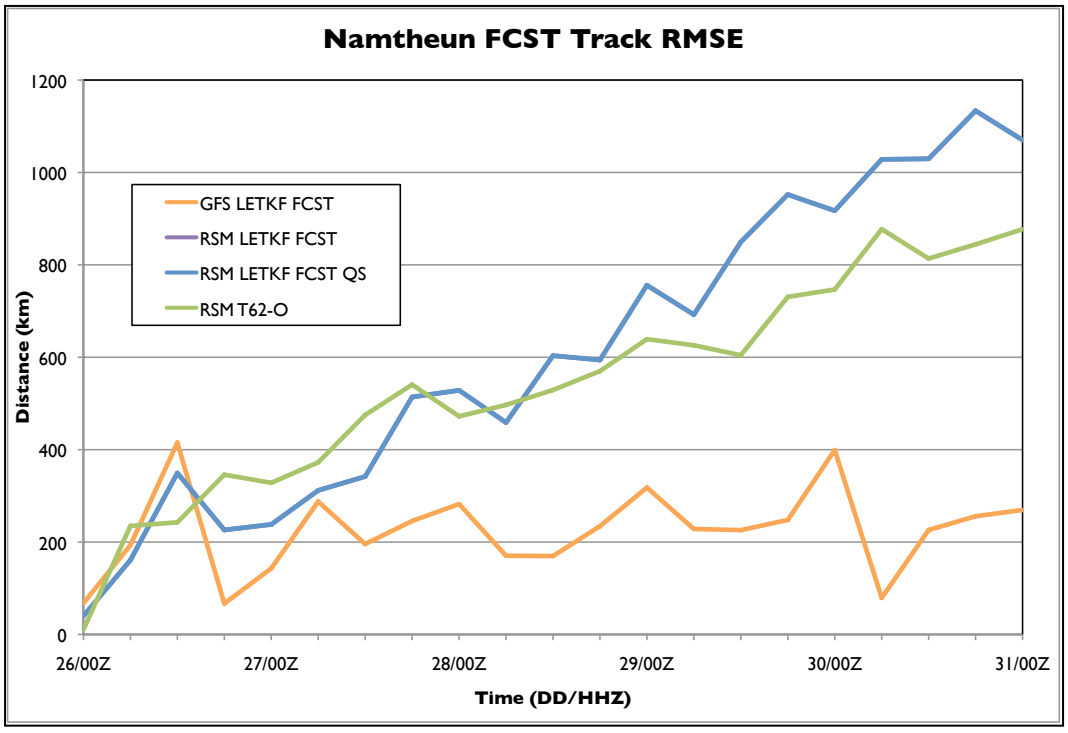

Fig. 36. Same as Fig. 34 except initialized at 0000 UTC July 27 for Typhoon Namtheun.

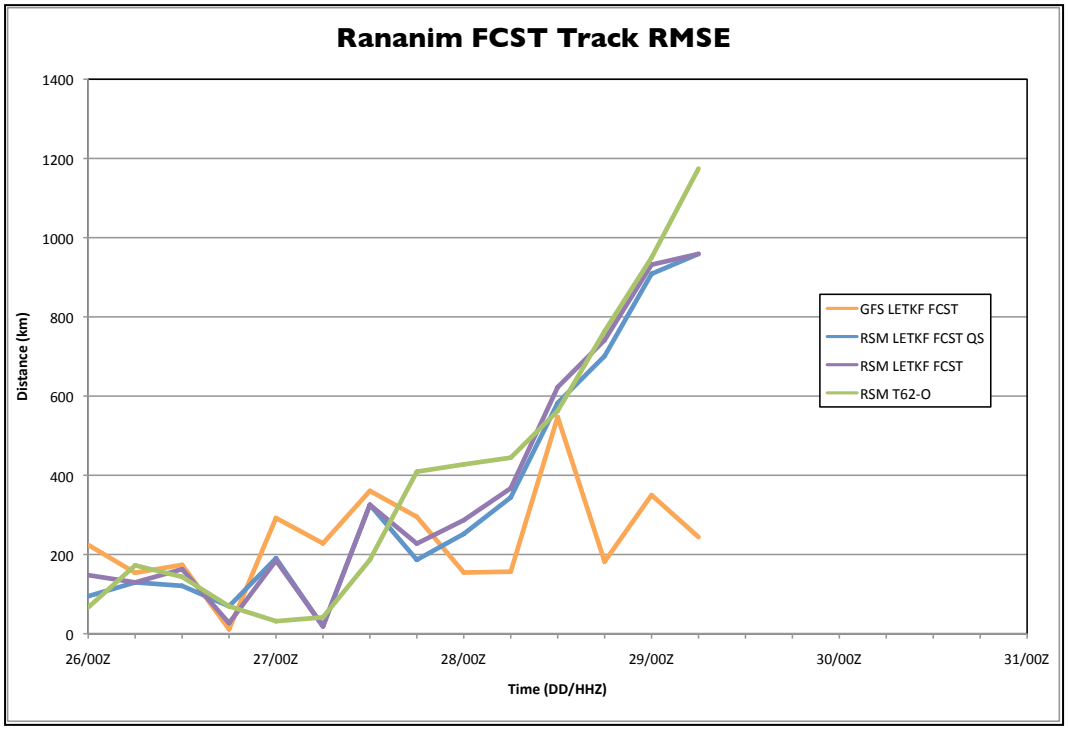

Fig. 37. Same as Fig. 34 except initialized at 1800 UTC August 9 for Typhoon Rananim.

7. CONCLUSIONS

We successfully implemented a coupled global-limited-area ensemble based data assimilation system with the aim of improving the analysis of tropical cyclones (TCs) in the Northwest Pacific: analysis/forecast experiments for a two-month period of the 2004 typhoon season showed the system to be adequate at analyzing tropical cyclones without TC relocation or bogusing techniques and the regional component of the data assimilation system enhanced the TC analyses and forecasts for many of the TCs. The inclusion of additional scatterometer observations and targeted dropsonde observations from the DOTSTAR program led to very modest changes in the quality of the analyses and forecasts.

The comparison of the global component of our system with the NCEP Reanalysis showed that the track analysis in our system was more accurate by an average of 25 – 180 km over the lifetime of the TCs. Another comparison with the operational, high-resolution, NCEP global analysis, which assimilated a large number of satellite radiance observations in addition to the observations assimilated in our system, demonstrated that the use of the LETKF analysis scheme alone was sufficient to achieve a similar performance in analyzing the track as the operational system with its advanced features. The comparison of track analysis errors between the ensemble-based global and regional analyses showed that the regional analyses were more accurate by 20 – 175 km over the lifetime of most TCs. The analysis track error distribution also depended on cyclone intensity: low intensity cyclones tended to produce more variable track errors with many outliers.

Comparisons of the minimum sea level pressure (SLP) analyses were made against the benchmarks. When comparing our analyses of SLP to the benchmarks we found that our global analysis was more accurate than the NCEP Reanalysis by about 2 hPa on average; the operational GFS and the global component of our system were of similar quality; and, the regional component improved the global ensemble-based analysis of the SLP by as much as a few hPa in some instances. We showed, in agreement with the published literature, that minimum SLP analysis errors increased with increasing cyclone intensity. This response is likely due to the grid spacing being insufficient in a coarse-resolution model to resolve the structure of the eye of the TC. The addition of extra data in the regional data assimilation step did not consistently enhance position or intensity analysis. The inclusion of QuikSCAT data led to statistically significant improvement for only one of the eight TCs. In the case of the one storm for which DOTSTAR data were available, the SLP error was reduced, but track error was not significantly changed.

Deterministic forecasts were started from the mean analyses of global and regional components of the ensemble-based system, and regional forecasts were also started from the operational NCEP analyses. The comparison of the track forecasts shows that the accuracy of the track forecasts from the regional ensemble-mean were not significantly different, on average, from those started from the NCEP operational analysis; and, differences between errors in the global and regional track forecasts were not statistically significant. The inclusion of QuikSCAT data led to a significant improvement for only one of the four storms. The quality of the SLP forecasts was

similar for all regional forecasts and the regional forecasts all proved more accurate SLP forecasts than the global forecast. On average, the regional experiments performed better for longer than 48 h SLP forecasts, while the global forecast performed better in predicting the track for longer than 48 h.

REFERENCES

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. *Mon. Wea. Rev.,* in press.

Atagan, J. F., A. Preble, A. C. Bryant, and Y. Pitts, 2004: 2004 Annual Tropical Cyclone Report. U.S. Naval Pacific Meteorology and Oceanography Center/ Joint Typhoon Warning Center, Pearl Harbor, Hawaii.

Chu, J., C. R. Sampson, A. S. Levine, and E. Fukada, 2002: The Joint Typhoon Warning Center Tropical Cyclone Best-tracks, 1945-2000. Joint Typhoon Warning Center Report; Reference Number: NRL/MR/75540-02-16

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

———, and ———, 1989b: Contributions to tropical cyclone motion by small, medium, and large scales in the initial vortex. *Mon. Wea. Rev*., **117**, 721-727.

Grasso, L. D., 2000: The difference between grid spacing and resolution and their application to numerical modeling. *Bull. Amer. Meteor. Soc*., **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. *Mon. Wea. Rev*., **139**, 668-688.

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

Juang, H.-M. H., and M. Kanamitsu, 1994: The NMC nested regional spectral model. *Mon. Wea. Rev.,* **122,** 3–26

Kalnay , E., M. Kanamitsu, R. Kistler, W. Collins., D. Deaven, et al., 1996: The NCEP/NCAR 40-year reanalysis project, *Bull. Amer. Meteor. Soc.*, **77**, 437−470

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

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP global data assimilation system. *Weather Forecast*. **24**, 1691-1705.

Merkova, D., I. Szunyogh, and E. Ott, 2011: Strategies for coupling global and limited-area ensemble Kalman filter assimilation. Nonlin. Processes Geophys., **18**, 415-430.

Montgomery, D. C., and G. C. Runger, 2007: *Applied statistics and probability for engineers*. 4th ed. John Wiley and Sons, 768 pp.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, 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. *Mon. Wea. Rev*., **120**, 1747-1763.

Rappaport, E. N., J. L. Franklin, L. A. Avila, S. R. Baig, J. L. Beven II, et al., 2009: Advances and challenges at the National Hurricane Center. *Weather Forecast*. **24**. 395-419.

Skamarock, W. C., 2004:Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev*., **132**, 3019-3032.

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.

———, ———, ———, 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.

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

———, 2010: Performance of a Mesoscale Ensemble Kalman Filter (EnKF) during the NOAA High-Resolution Hurricane Test. *Mon. Wea. Rev*., **138**, 4375-4392.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev*., **130**, 1913-1924

Wu, C., P. Lin, S. Aberson, T. Yeh, W. Huang, et al., 2005: Dropwindsonde observations for typhoon surveillance near the Taiwan region (DOTSTAR): An Overview. *Bull. Amer. Meteor. Soc.*, **86**, 787-790.

VITA

Name:	Christina R. Holt
Address:	Department of Atmospheric Sciences, 1204 Eller O&M 3150 TAMU, College Station, TX 77843
Email Address:	crh403@geos.tamu.edu
Education:	B.S., Meteorology, University of South Alabama, 2009 M.S., Atmospheric Sciences, Texas A&M University, 2011