# Impact of TAMDAR Data on Hurricane Paula with an Eigen-Structure Dependent Inflation Scheme in WRFDA/ETKF Hybrid

Dongmei Xu[1,2], Arthur P. Mizzi[2], Xiang-Yu Huang[2*]

[1] University of Information Science and Technology, Nanjing, China

[2] National Center for Atmospheric Research, Boulder, Colorado, USA

# Abstract

This paper documents the application of the WRFDA/Hybrid ETKF data assimilation system to a hurricane case on the AirDat, LLC (http://www.airdat.com/) operational grid with the Tropospheric Airborne Meteorological Data Reporting (TAMDAR) observations.

This study proposed an advanced inflation scheme by projecting the estimation of the forecast error onto each particular eigenvector of the estimated ensemble covariance matrix. As a result, it restores lesser amounts of covariance in the eigen-directions associated with stronger projection of the forecast error onto the ensemble subspace, offsetting the under/over inflation in the previous inflation schemes. The preliminary results show that the revised inflation factor scheme ameliorates the ETKF ensemble spread instability by maintaining greater inter-cycle continuity for the eigen-structure of the posterior spread. As one of the first studies to apply TAMDAR data to hurricane forecasting using Hybrid system, the results show the TAMDAR data in Hybrid system generally improved the environment compared to 3DVAR for wind speed and specific humidity. TAMDAR observations improve the forecast skill for hurricane parameters such as track location and central pressure. Using TAMDAR data within Hybrid can improve maximum wind forecast.

Keywords: WRFDA/Hybrid, ETKF, TAMDAR, covariance inflation, hurricane forecasting

# 1. Introduction

The most common data assimilation schemes are Variational (e.g., Parrish and Derber 1992; Courtier et al. 1994, 1998; Gauthier et al. 1998; Rabier et al. 1998, 2000), which are widely used in many operational centers. A different approach is ensemble kalman filter (e.g., Evensen 1994; Houtekamer and Mitchell 1998, 2001, 2005; Anderson 2001; Whitaker and Hamill 2002; Snyder and Zhang 2003; Hamill and Whitaker 2005), which can provide a practical way of representing the probability distribution function of forecast errors. However, many of the current ensemble-based data assimilation techniques serially process the observations and the computational cost is large with large amount of observations. Hybrid techniques are efficient to improve analysis by introducing flow-dependent background error information from ensemble forecasts into a variational data assimilation system and easy to implement. It has been proposed (eg., Hamill and Snyder 2000; Lorenc 2003; Etherton and Bishop 2004; Wang et al. 2007, 2008a), implemented and tested with real numerical weather prediction (NWP) model for real data (e.g., Buehner 2005; Wang et al., 2008b; Buehner et al., 2010 a, b; Wang 2011; Wang et al., 2011; Whitaker et al., 2011).

Tropospheric Airborne Meteorological Data Reporting (TAMDAR) is a mulit-function (observing moisture, icing, turbulence, Temperature, wind, and pressure) in-situ atmospheric sensor for aircraft. TAMDAR sensors are installed on commercial aircraft and continuously transmit atmospheric observations. While preliminary tests show that TAMDAR improved regional numerical weather prediction (e.g., Jacobs et al., 2006; Liu et al., 2007; Moninger et al., 2007, 2010). TAMDAR assimilation in the Weather Research and Forecasting (WRF, Skamarock et al. 2008) model and its Data Assimilation

\

3

(WRFDA, Barker et al. 2012) system is also proved to be promising (Wang and Huang, 2012; Gao et al., 2012), to date there is no published study applying Hybrid method to the assimilation of TAMDAR data for Tropical cyclones (TCs) predictions. This study serves as a pilot study applying the hybrid ensemble-3DVAR system-WRFDA/Hybrid ETKF developed for the WRF model (Wang et al. 2008a) to explore its potential for assimilating TAMDAR observation for hurricane forecasts.

WRFDA/ Hybrid ETKF system was enhanced to assess the impact of TAMDAR data on analysis and forecasts of hurricane Paula (2010) on AirDat operational domain. The ETKF scheme does not have covariance localization to avoid the spurious covariance between distant grid points and a single inflation is applied for the whole domain. A hurricane or tropical cyclone may therefore be sensitive to large perturbations, causing forecast model unstable. A new adaptive inflation scheme is introduced for WRFDA/ Hybrid ETKF to offset the under-sampling and inappropriate inflation issues.

This paper is organized as follows: Section 2 introduces the WRFDA/Hybrid ETKF system and advanced improvements. Experimental design is described in section 3. Ensemble posterior spread, inflation and forecast verification are presented in section 4. The summary and future perspectives are presented in last section.

## 2. The WRFDA/Hybrid and ETKF system

### 2.1 WRFDA/Hybrid System

The detailed description of the WRFDA/Hybrid data assimilation system developed for WRF was documented in Wang et al. (2008a). For state vector $\mathbf{x}$, the analysis increment of the Hybrid scheme, $\mathbf{x}'$, is the sum of two terms,

\

$$\mathbf{x}^{'} = \mathbf{x}_{1}^{'} + \sum_{k=1}^{K} \alpha_{k} \circ \mathbf{x}_{k}^{e}. \tag{1}$$

, where the first term $\mathbf{x}_{1}^{'}$ is the increment associated with WRFDA/3DVAR static background covariance and the second term is the increment associated with flow-dependent covariance. Here, $K$ is the ensemble number and $\alpha_{k}$ is the extended control variable as defined by Lorenc (2003). $\mathbf{x}_{k}^{e}$ is the $k^{th}$ ensemble perturbation state vector. The symbol '∘' denotes the Schur product (element by element product) of the vectors $\alpha_{k}$ and $\mathbf{x}_{k}$. The cost function for WRFDA/Hybrid is

$$J(\mathbf{x}_{1}^{'}, \mathbf{a}) = \beta_{1} \frac{1}{2} \mathbf{x}_{1}^{'T} \mathbf{B}^{-1} \mathbf{x}_{1}^{'} + \beta_{2} \frac{1}{2} \mathbf{a}^{T} \mathbf{C}^{-1} \mathbf{a} + \frac{1}{2} (\mathbf{d} - \mathbf{H} \mathbf{x}^{'})^{T} \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H} \mathbf{x}^{'}) \tag{2}$$

, where $\beta_{1} \frac{1}{2} \mathbf{x}_{1}^{T} \mathbf{B}^{-1} \mathbf{x}_{1}^{'}$ is the traditional WRFDA/3DVAR background term associated with the static covariance $\mathbf{B}$ and $\beta_{2} \frac{1}{2} \mathbf{a}^{T} \mathbf{C}^{-1} \mathbf{a}$ is the Hybrid term associated with flow-dependent covariance. Here, $\mathbf{a}$ defined as the extra control variable is a vector formed by concatenating $K$ vectors $\alpha_{k}$, $k = 1, \cdots, K$, as $\mathbf{a}^{T} = (\alpha_{1}^{T}, \alpha_{2}^{T}, \alpha_{3}^{T} .... \alpha_{k}^{T})$. The innovation vector $\mathbf{d}$ is defined as, $\mathbf{d} = \mathbf{y}_{o} - H(\mathbf{x}_{b})$, where $\mathbf{y}_{o}$ is the observation vector, $\mathbf{x}_{b}$ is the background forecast state vector, and $H$ is observation operator ($\mathbf{H}$ the linearized observation operator). $\frac{1}{2} (\mathbf{d} - \mathbf{H} \mathbf{x}^{'})^{T} \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H} \mathbf{x}^{'})$ is the observation term associated with observation error covariance $\mathbf{R}$. The weights of the static covariance and flow-dependent covariance are determined by factors $\beta_{1}$ and $\beta_{2}$, with the constraint $\frac{1}{\beta_{1}} + \frac{1}{\beta_{2}} = 1$. $\mathbf{C}$ is the

\

ensemble covariance localization, containing the horizontal localization in this study, which is modeled by a recursive filter transform.

## 2.2 WRFDA/ETKF analysis scheme

ETKF is a form of Kalman Filter approach which estimates forecast error from the ensemble covariance matrix of the ensemble forecast perturbations (Bishop et al. 2001). The solution of ETKF is designed as:

$$\mathbf{X}^a = \mathbf{X}^b \mathbf{T} \tag{3}$$

, by transforming the matrix of forecast perturbations $\mathbf{X}^b$ into a matrix of analysis perturbations $\mathbf{X}^a$, whose columns contain $K$ analysis perturbations, $\mathbf{x}_k^{'a}, k = 1,...,K$ by a transformation matrix $\mathbf{T}$. The transformation matrix is chosen to ensure the matrix of analysis perturbations will be expected to stand for the true analysis error covariance, which is solved based on the distribution and the quality of the observation and forecast errors, as

$$\mathbf{T} = \mathbf{C}(\Gamma + \mathbf{I})^{-1/2} \mathbf{C}^T. \tag{4}$$

Here, $\mathbf{C}$ contains eigenvectors and $\Gamma$ contains the eigenvalues of the $K \times K$ matrix $(\mathbf{X}^b)^{\mathrm{T}} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{X}^b$ and $\mathbf{I}$ is the identity matrix. The computation cost of Eq. (4) is relatively cheap with an ensemble size $K$ of 100 or less. The ETKF transform vector depends on the eigen-structure of the ensemble covariance matrix in the observational space normalized by the observation errors. The magnitude of the ETKF spread contraction in any particular eigen-direction is inversely related to the associated eigenvalue. ETKF suffers from problems with under-sampling in that the number of ensemble members is much smaller than the number of degrees of freedom in a NWP

\

6

model. Thereby, forecast error is partly explained by the ensemble. Wang et al. (2007) introduced an adaptive fraction factor $\rho_c$ to account for this issue by denoting the usage of the scalar factor $\rho_c$ as the fraction of the forecast error projected onto the ensemble subspace,

$$\mathbf{T} = \mathbf{C}(\rho_c \Gamma + \mathbf{I})^{-1/2} \mathbf{C}^T \tag{5}$$

,where $\rho_c$ is estimated by,

$$\rho_c = \frac{\overline{\mathbf{R}^{-1}(y - \mathbf{H}\overline{\mathbf{X}^b})^T \mathbf{E}\mathbf{E}^T(y - \mathbf{H}\overline{\mathbf{X}^b})} - (K-1)}{\overline{\mathbf{R}^{-1}(y - \mathbf{H}\overline{\mathbf{X}^b})^T(y - \mathbf{H}\overline{\mathbf{X}^b})} - p}. \tag{6}$$

The overbar of Eq. (6) means the average of some independent samples, which covers the previous cycles just before the current one in this study. $p$ is the observation number and $\mathbf{E}$ contains the eigenvectors of the ensemble covariance in normalized observation space denoted as,

$$\mathbf{E} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{X}_b \mathbf{C}\Gamma^{-1/2} / \sqrt{K-1}. \tag{7}$$

To be specific, $proj_k$ is defined as the forecast error projected on the $k^{th}$ eigenvector space, thus,

$$\overline{\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\overline{\mathbf{X}^b})^T \mathbf{E}\mathbf{E}^T(\mathbf{y} - \mathbf{H}\overline{\mathbf{X}^b})} = \sum_i^{K-1}(proj_k)^2 \tag{8}$$

Fraction factor can only partly ameliorate the underestimate of analysis-error variance in Eq. (5). Another inflation factor is applied to increase the ensemble covariance by Wang and Bishop (2003). The idea is to multiply the initial perturbations obtained at time $t_i$ by an estimated inflation factor $\Pi_i$, that is,

$$\mathbf{X}_i^a = \mathbf{X}_i^f \mathbf{T}_i \Pi_i. \tag{9}$$

\

7

The inflation factor is to ensure at time $t_{i+1}$ the background ensemble forecast variance is consistent with the ensemble-mean background-error variance over global observation sites. The normalized innovation vector $\tilde{\mathbf{d}}$ is defined as, $\tilde{\mathbf{d}}_i = \mathbf{R}^{-1/2}\mathbf{y}_i - \mathbf{H}\bar{\mathbf{x}}_i^b$, where $\mathbf{y}_i$ is the observation vector, $\mathbf{H}\bar{\mathbf{x}}_i^b$ is the ensemble mean background forecast state vector at time $t_i$. The inflation factor is obtained by checking if $\tilde{\mathbf{d}}_i^T \tilde{\mathbf{d}}_i$ equals to $Tr(\mathbf{R}^{-1}\mathbf{H}c_i\mathbf{P}_i^e\mathbf{H}^T + \mathbf{I})$, where $Tr$ denotes the trace. If not, parameter $c_i$ is introduced to rescale, by

$$c_i = \frac{\tilde{\mathbf{d}}_i^T \tilde{\mathbf{d}}_i - p}{Tr(\mathbf{R}^{-1}\mathbf{H}\mathbf{P}_i^e\mathbf{H}^T)}. \tag{10}$$

Given $\Pi_{i-1}$ is the inflation factor at $t_{i-1}$, the inflation factor at $t_i$ is defined as

$$\Pi_i = \Pi_{i-1}\sqrt{c_i}. \tag{11}$$

From Eq. (11), $\Pi_i$ is a product of these $c$ parameters from the first cycle time to $t_i$, that is,

$$\Pi_i = \sqrt{c_1 c_2 \cdots c_i} \tag{12}$$

Thus, the inflation factors and fraction factors are designed to account for the covariance deficiency and sampling error. However, based on some sensitive experiments with different observation number configurations, we find the inflation scheme can't behave well when there were large variations in the number of observations in ETKF from one cycle to the next. It is consistent with the results from previous research (Mizzi, A.P., 2012a) based on different observation number configurations in the ETKF part, which

\

shows oscillations in inflation factor and posterior ensemble spread are due to big variations in the number of ETKF observations which may easily cause forecast model crash. Bi-cycle oscillations that have been observed are attributed to over/under inflation (Bowler et al. 2008). The main explanation is that, with limited ensemble size, for cycles with large observation number ($p \gg K$), denominator in Eq. 10 is systematically over-estimated in that the order observations will thus be projected to a very low-dimensional space. The $c_i$ in Eq. 10 is under-estimated. Vice versa, for cycles with smaller observation number, the denominator and $c_i$ is less overestimated. Over/under inflation will easily arouse increasing forecast error. Here's a simple illustration in Fig. 1. The solid line (Fig. 1a) denotes the time series of observation number and the dashed line denote the adaptive inflation factor. The solid line in Fig. 1b shows the posterior ensemble spread after inflation and the dashed line is estimated forecast error ($\dfrac{\tilde{\mathbf{d}}_i^T \tilde{\mathbf{d}}_i}{p}$).

Note that the inflation factors and posterior ensemble spreads are approximately parallel with an increasing trend. For the second cycle with a larger observation number, we underestimate the inflation factor in a large scale and when we come to the third cycle with fewer observations, we underestimate the inflation, still, which is less evident for this cycle compared to the previous cycle. Thus, there will be large oscillation of inflation factor and the inflation factor can not converge to be stable, causing increasing forecast error. As the forecast is becoming larger and larger, the inflation factor aims to approximate the forecast error, thereby yielding huge inflation and posterior spread, which will shock the forecast model.

\
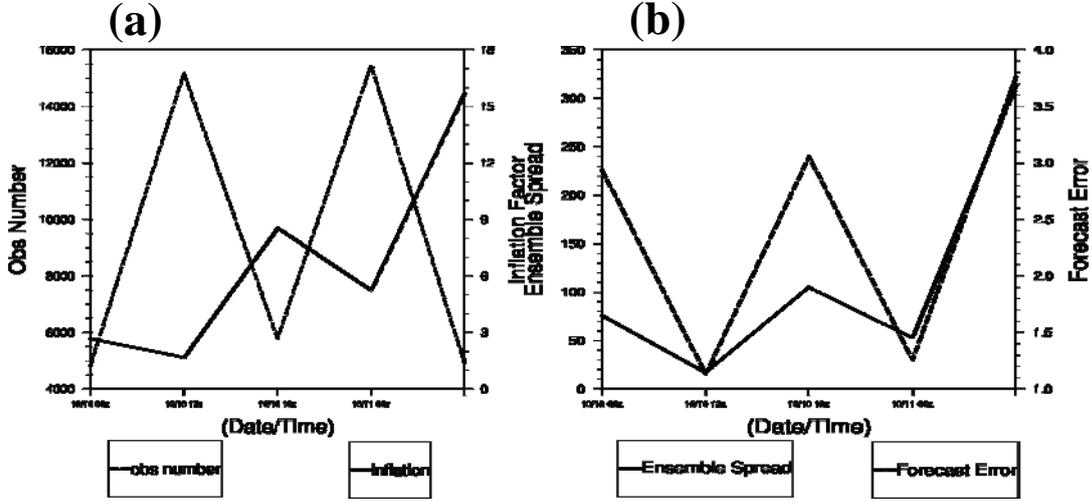
*Fig. 1. (a). The times series of adaptive inflation factors (the solid line) and number of observations used in the WRFDA/ETKF (the dashed line); (b). The time series of the ensemble posterior spread (the solid line) and the estimation of forecast error (the dashed line)*

The solution to this under/over inflation and large spread problem is to find a more explicit method to make the ensemble forecast variance better account for the forecast error variance at the first place. In the modified inflation scheme, following Eq. 6 and Eq. 8, we made the $\rho$-factor dependent on the eigenvector as,

$$\rho(k) = \rho_c \frac{proj_k^2}{R^{-1}(\mathbf{y} - \mathbf{H}\bar{\mathbf{X}}^b)^T \mathbf{E}\mathbf{E}^T(\mathbf{y} - \mathbf{H}\bar{\mathbf{X}}^b)}(K-1). \tag{13}$$

So it corrects the inflation in accordance with proportional amount of the forecast error variance projected onto a particular $k^{th}$ eigenvector under the constraint that the average of the $\rho(k)$ equals to $\rho_c$. The adaptive inflation and fraction algorithm here aim to ameliorate the problem by distinguishing between large and small background forecast

\

10

errors explained by the different ensemble eigenvectors, instead of using a constant fraction factor to rescale all the eigenvaules. It restores lesser amounts of covariance in the eigen-directions associated with stronger projection of the forecast error onto the ensemble subspace and account for the angle difference between ensemble-based ensemble eigenvector and forecast error-based eigenvector. Then the Eq. 5 becomes,

$$\mathbf{T} = \mathbf{C}(\mathbf{D} \circ \mathbf{\Gamma} + \mathbf{I})^{-1/2} \mathbf{C}^{T}. \tag{14}$$

$\mathbf{D}$ is a vector containing $\rho(k), k = 1,2,\cdots,(K-1)$, as elements. The symbol '$\circ$' denotes the Schur product (element by element product) of the vectors $\mathbf{D}$ and $\Gamma$. Our experiments showed that this correction facilitate the use of that inflation scheme with large variations in the number of ETKF observations from one cycle to the next.

An illustration of the WRFDA/Hybrid ETKF analysis and ensemble generation cycle for a $K$-member ensemble is shown in Figure 2 with five stages. (1) Generate a short term of ensemble forecasts from the initial ensemble which is the (Global Forecast System) GFS analysis plus correlated random perturbations following Torn et al. (2006) and Wang et al. (2008a, b). (2) Calculate the ensemble mean and perturbations. (3) Update the ensemble mean and perturbations with the WRFDA/Hybrid and the WRFDA/ETKF, respectively. $ob.etkf.e001,\cdots ob.etkf.e00k$, $k = 1,\cdots,K$ are observations used in WRFDA/ETKF which are filtered and reordered observations by the procedures (QC-OBS and VERIFY) in WRFDA, respectively. (4) Obtain analysis ensemble by adding updated ensemble mean and perturbations. (5) Update lateral boundary conditions (LBC) and lower boundaries and conducting short term ensemble forecast to next assimilation time or deterministic forecast to diagnose outputs from analysis ensemble mean.

\
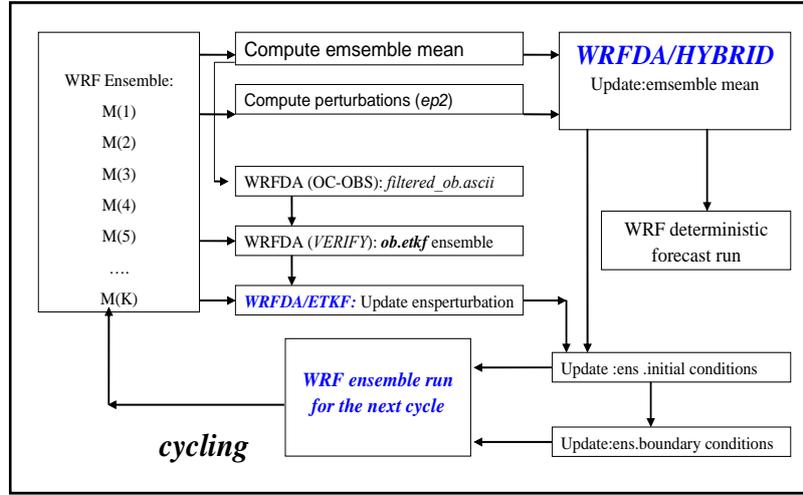
*Fig. 2. Illustration of the WRFDA/Hybrid ETKF analysis and ensemble generation cycle for a K -member ensemble.*

In the WRFDA/Hybrid system, the weighting between the variational increment and the ensemble increment is controlled by $\dfrac{1}{\beta_1}$. When $\dfrac{1}{\beta_1} = 1.0$, we get the variational increment. When $\dfrac{1}{\beta_1} = 0.0$, we get the ensemble increment. When $\dfrac{1}{\beta_1}$ is between 0.0 and 1.0, we get the Hybrid increment. Additional parameters in WRFDA/Hybrid are: (i) H – the horizontal localization length scale in km, (ii) N – the number of ensemble members. As the localization scales decrease, the radius of influence for a particular observation decreases, reducing the amount of noise in the increment. Based on the results of sensitive tests in previous research (Mizzi, A.P., 2012a), for the experiments discussed in this report, we generally set $\dfrac{1}{\beta_1} = 0.75$, H=750 km, and N=20 as ensemble members.

\

12

## 3. Experimental design

The WRF model is described by Skamrock et al. (2008). We ran WRF on the AirDat tropical domain, which has a horizontal resolution of 15 km on a 718 x 373 horizontal grid with 43 vertical levels. The model top was located at 30 hPa.

WRF was initialized with GFS analyses and GFS forecast-based boundary conditions. We started the assimilation experiments at 0600UTC, 10 October 2010 and cycled for six days until 0006UTC, 15 October 2010 cycling every six hours. During that period, Hurricane Paula formed within the domain over the western Caribbean at 81.8° W, 13.9° N on 11 October 2010 and moved northward over the Yucatan Channel. On 14 October 2010, Hurricane Paula made landfall in Cuba at 97° W, 21° N as a tropical storm.

For any particular cycle, the WRFDA/Hybrid and WRFDA/ETKF used all conventional observations (including radiosondes, aircraft, satellite-derived winds, and land and oceanic surface stations etc.), obtained from the NCEP operational datasets. In this study, we ran three assimilation experiments: (i) one experiment without TAMDAR observations using WRFDA/Hybrid (hereafter Hybrid), and (ii) one experiment with TAMDAR data using Hybrid, and (iii) the experiment using WRFDA/3DVAR (hereafter 3DVAR) with TAMDAR data. Figure 3 shows the TAMDAR observation locations with the best-fit track for Hurricane Paula.

\
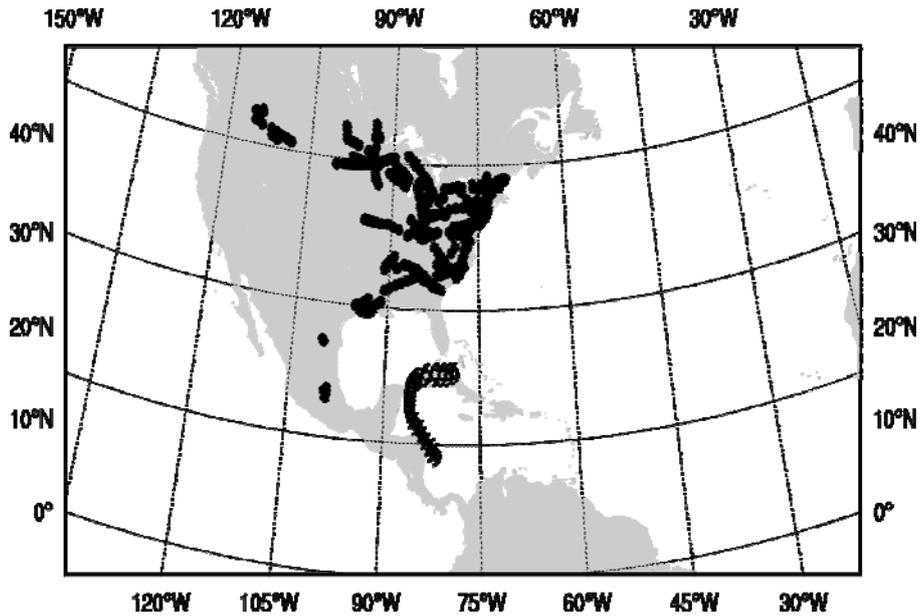
*Fig. 3. TAMDAR observation locations (dots) and Hurricane Paula's best track (the hurricane symbols).*

For cycling, the initial and boundary conditions for each ensemble was generated by adding random perturbations drawn from a normal distribution based on the WRFDA background error covariance to the analysis and boundary conditions valid at 0000UTC, 10 October 2010. That initial ensemble was used to obtain an ensemble of WRF 6-hr forecasts. That forecast ensemble became the background forecasts for the first assimilation cycle on 0600UTC, 10 October 2010.

## 4. Results

4. 1 Inflation and fraction factors

This section shows the differences between the inflation and ensemble spread for the assimilation experiments with and without TAMDAR data in Hybrid experiments. Figure 4a displays numbers of observations used in WRFDA/ETKF. For both observation

\

configurations, the oscillations of number of observations are observed between synoptic hours (0000UTC and 1200UTC) and other hours (0600UTC and 1800UTC), even though TAMDAR observations fill in some the gap of the current data source. Based on some sensitive experiments (not shown), instability issues are always observed during the first 3 or 4 cycle times, when using default inflation schemes introduced in section 2.2. Figure 4b shows time series of the inflation and corresponding $\rho$-factors. For display purposes, the $\rho$-factors are averaged in eigenvector space. The magnitudes of the inflation are larger than inflation factors from traditional Ensemble Kalman Filter without localization. With the new inflation scheme, the inflation factor and fraction factors stay relatively stable (compared to Fig 1a.). When there are a relatively large number of observations from experiments with TAMDAR data in ETKF, the inflation is large, and when there are a relatively small number of observations in ETKF, the inflation is small. That behavior is consistent with Ensemble Kalman Filter theory: the greater the number of observations, the greater the spread reduction. Slightly smaller fraction factors for experiments with large observation number are obtained as we expect in that smaller percentage of forecast error is explained by ensemble space when observation number is much larger than ensemble number.
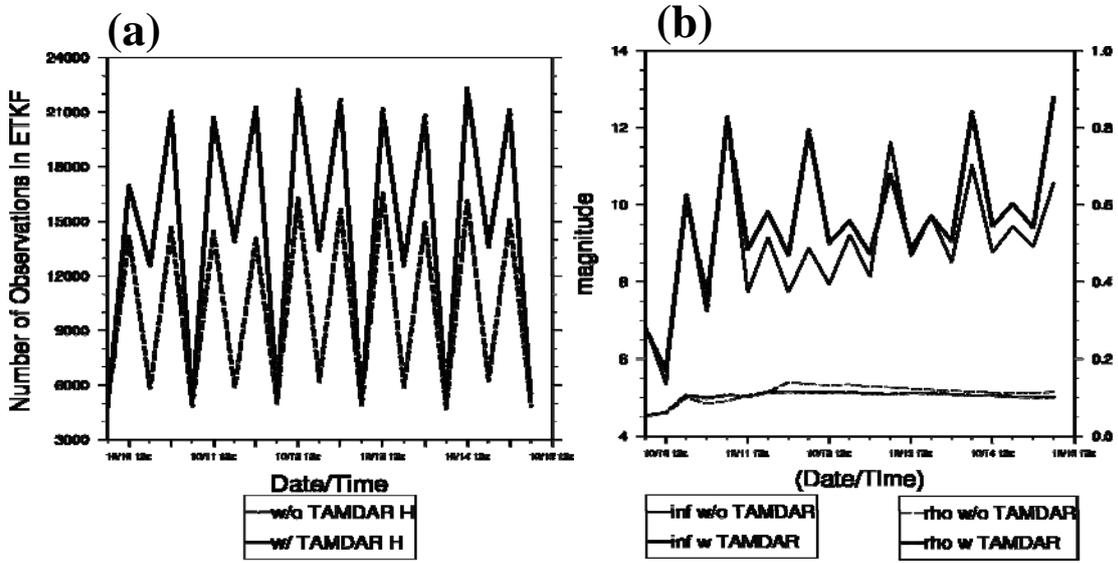
*Fig. 4. (a). Times series of number of observations used in the WRFDA/ETKF with and without TAMDAR data. (b). Time series of the inflation and $\rho$ -factors for the WRFDA/ETKF with and without TAMDAR data. The inflation factor magnitude is displayed on the left ordinate axis, and the $\rho$-factor is display on the right axis.*

Figure 5 shows time series of the averaged $\rho$ factors and the $\rho$ factors for the first three eigenvectors. The rho factors vary in the orthogonal basis eigen-structure space around the constant factor. The largest projection wouldn't eventually fall on the leading vector all the time, which implies the expected angle difference between ensemble-based eigenvectors and forecast error based eigenvectors. For orthogonal directions with smaller $\rho$ factors, the inflation scheme contracts more eigenvalues, eventually giving fewer weighting for those eigenvectors in ETKF transformation procedure. The differences between the orthogonal directions are relatively large for the first 10 cycles and then the projections start to converge to be flat between those directions.

\

16
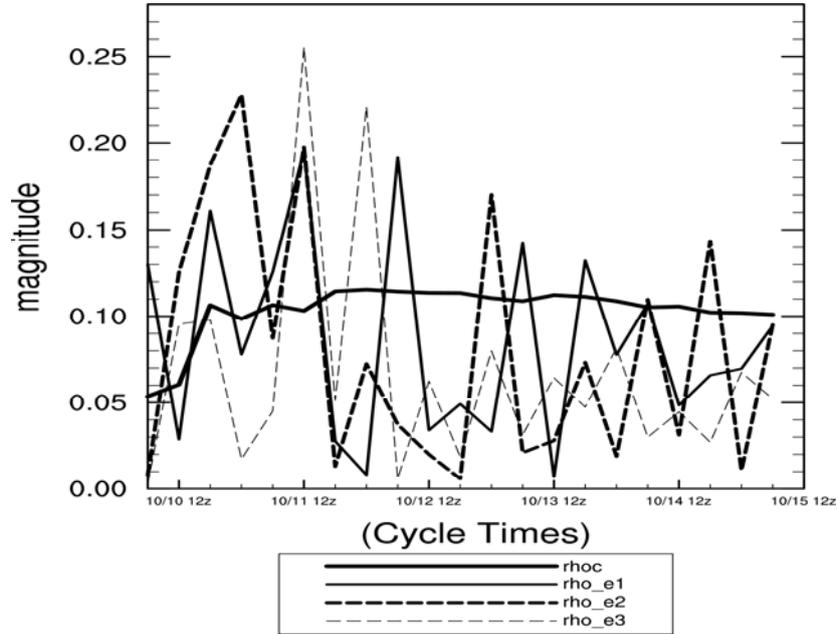
*Fig. 5. Time series of averaged $\rho$ factors (the bold solid line) and the $\rho$ factors for first three leading eigenvectors (the thin solid line for the first leading eigenvector, the bold and thin dashed lines for the second and third eigenvector, respectively) for the experiments Hybrid w/ TAMDAR.*

Figure 6 illustrated the time series of the posterior ensemble spread (the ensemble spread after applying inflation to the ensemble) for Temperature corresponding to the inflation factors displayed in Figure 4b. Note that it takes more cycles for the experiment without TAMDAR data to converge, in that including TAMDAR data fills the gap of observation source and stabilizes the ETKF (the same phenomenon was observed for the inflation and $\rho$ factors in Figure 4b). Figure 6 shows that including TAMDAR data generally reduced the posterior ensemble spread compared to not including TAMDAR data, which indicates that TAMDAR data improve forecast accuracy because the inflation target is the dependent on the sum of the squared innovations. Thus, as the forecast

\

17

accuracy increases, the sum of the squared innovations decreases, and the spread decreases.
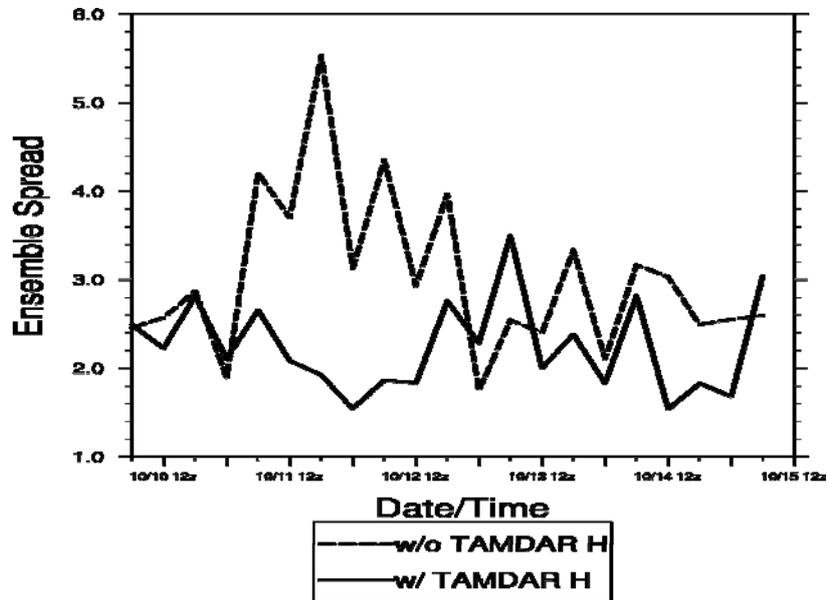


*Fig. 6. Time series of the posterior ensemble spread for temperature for w/o TAMDAR and w/TAMADAR experiments.*

4.2 Forecast error verification

In this section, we examine the forecast root mean square error (RMSE) from the three cycling experiments. For verification, we used all radiosonde and surface synoptic observations that passed the WRFDA quality control procedures based on the background forecasts from conventional WRFDA cycling. We began the verification at 0600UTC, 10 October 2010, and continued through 0600UTC, 15 October 2010. We verified our results on the entire horizontal and vertical domain. Figure 7 displays vertical profiles of the 12-hr forecast RMSE for wind speed (V), temperature (T) and specific humidity (Q) averaged over all cycle times. Figure 8 shows similar results except for the 24-hr the

\

forecast RMSE. Note that the Hybrid generally yields lower RMSE compared to 3DVAR for V at middle troposphere and Q and the RMSE reduction is more evident for 24-hr forecast. Also, the Hybrid with TAMDAR data generally yields lower RMSE compared to the Hybrid without TAMDAR data for V and Q for 12-hr and for V, T and Q for 24-hr forecast.
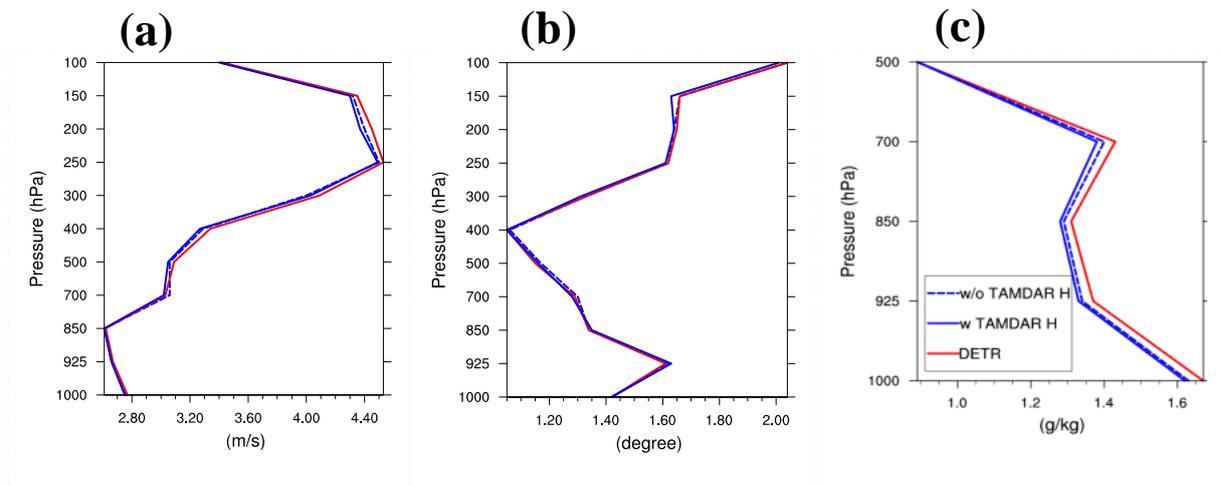


*Fig.7. Vertical profiles of 12-hr forecast RMSE for (a) Wind speed; (b) Temperature; (c) specific humidity. Blue dashed lines are for Hybrid without TAMDAR data, blue solid lines are for Hybrid with TAMDAR data and red solid line for 3DVAR.*
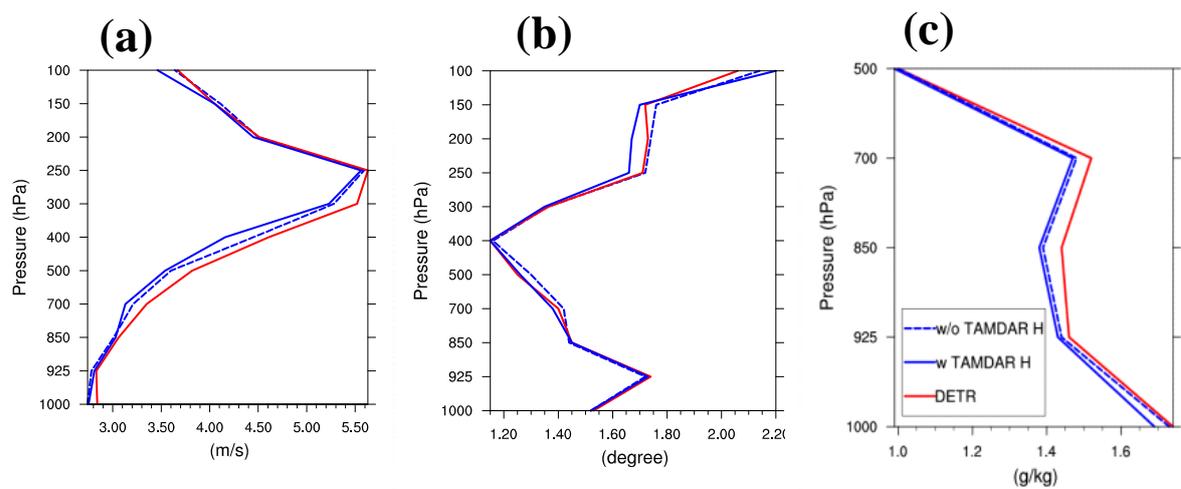


*Fig.8. Same as Figure 7 except for the 24-hr forecast RMSE.*

\

The differences in RMSE for the different profiles in Figures 7 and 8 are often small. To assess the significance of those differences, we applied the Student T-Test to the 24-hr forecast RMSE. We tested two null hypotheses that: (i) there was no difference between the mean RMSE for the Hybrid with TAMDAR data and deterministic 3DVAR experiments; and (ii) there was no difference between the mean RMSE for the Hybrid with TAMDAR data and the Hybrid without TAMDAR data experiments. A 0.05 significance level and a pooled estimator for the sample variance are applied. Table 1 lists significance testing results for 500 hPa, and Table 2 displays those for 850 hPa. For 20 degrees of freedom (RMSE results from 21 cycles times) and a 0.05 significance level, the critical t-statistic ($t_{crit}$) is $t_{crit} = 2.086$. Therefore, we reject the null hypothesis when the sample t-statistic ($t_{sample}$) exceeds $t_{crit}$.

*Table 1. Sample t-statistics for the 24-hr forecast RMSE at 500hPa with a 0.05 confidence level and 20 degrees of freedom.*

| 500hPa (df=20, $t_{0.05}$=2.086) | Hybrid w/ TAMDAR vs. Deterministic WRFDA | Hybrid w/ TAMDAR vs. Hybrid w/o TAMDAR |
|---|---|---|
| V | 2.37 | 2.25 |
| T | 1.82 | 2.11 |
| Q | 2.56 | 2.24 |

*Table 2. Same as Table 1 except for 850hPa.*

| 850hPa (df=20, $t_{0.05}$=2.086) | Hybrid w/ TAMDAR vs. Deterministic WRFDA | Hybrid w/ TAMDAR vs. Hybrid w/o TAMDAR |
|---|---|---|
| V | 2.33 | 2.14 |
| T | 1.42 | 1.15 |
| Q | 2.45 | 2.23 |

Tables 1 and 2 show that we can reject the null hypothesis for every case except when comparing the Hybrid with TAMDAR data and deterministic 3DVAR experiments

\

for T. That result suggests that: (i) the Hybrid with TAMDAR data significantly improves the 24-hr forecast skill for V and Q when compared to deterministic WRFDA; and (ii) the Hybrid with TAMDAR data significantly improves the 24-hr forecast skill for V, T, and Q when compared to the Hybrid without TAMDAR data.

4.3 Track and intensity forecast verification
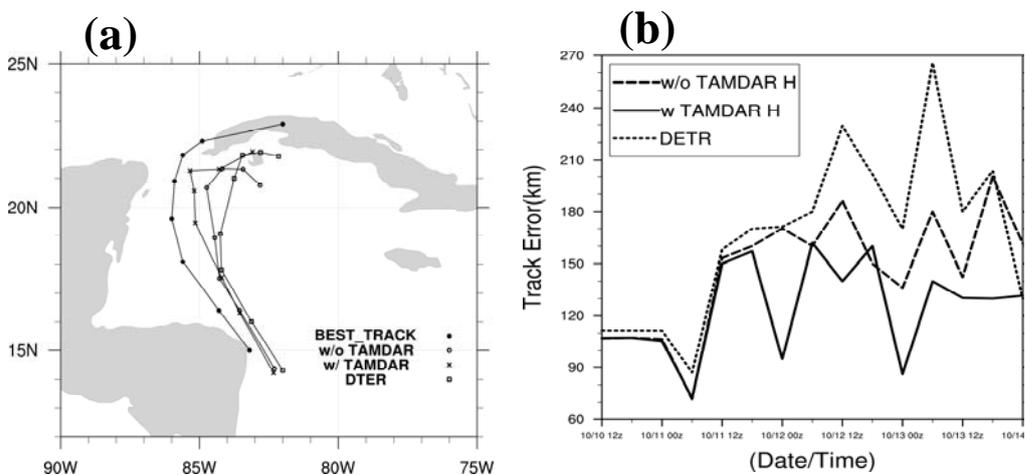


*Fig. 9. (a) 24-hr forecast track and best track positions every twelve hours and (b) absolute track error every six hours for all three cycling experiments from 1200UTC October 10, to 0000UTC, 14 October*
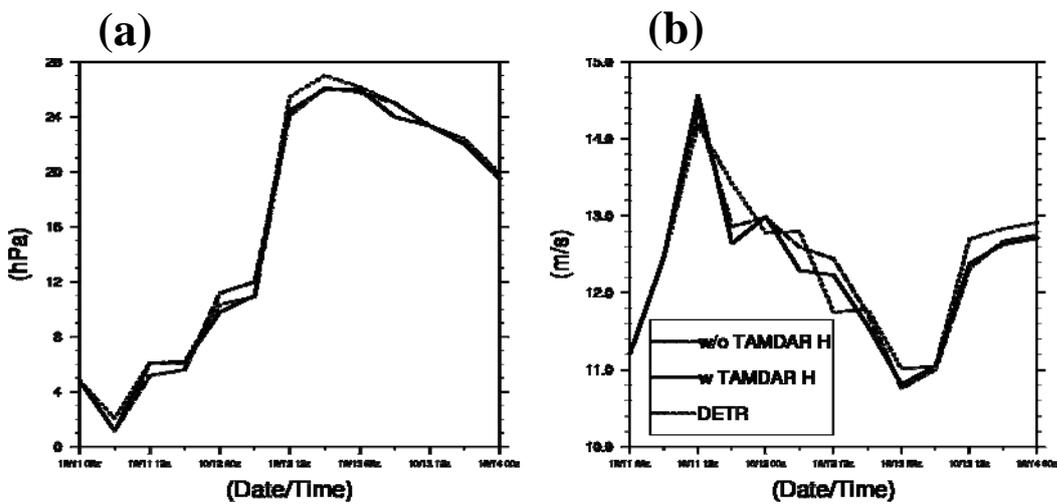
*Fig. 10. 24-hr forecast error from three cycling experiments for (a) central pressure and (b) maximum wind speed from 0000UTC, 11 October to 0000UTC, 14 October.*

Figure 9 shows the 24-hr forecast track for the three experiments together with the best-fit track and the associated 24-hr forecast track error (distance from best-fit track) as a function of time. Generally, all the experiments have similar track except that the deterministic WRFDA track began recurving earlier than the Hybrid without TAMDAR data, which began recurving earlier than the Hybrid with TAMDAR data, which began recurving earlier than the best track. Those recurvature errors are seen in the track error time series in the companion figure where the Hybrid with TAMDAR data consistently produces the smallest track error and deterministic WRFDA produces the largest track errors. Generally Hybrid reduces the confidence in the static background error by considering the ensemble variance during the assimilation process, and additional observations (the TAMDAR data) increase the confidence in the observations.

Figure 10 presents a comparison of the time series for the central pressure and maximum wind speed 24-hr forecast error for Hurricane Paula from 0000UTC, 11 October to 0000UTC, 14 October. Figure 10a shows that the WRFDA/Hybrid ETKF with TAMDAR data experiment gives reduced errors and the maximum error reduction occurs after approximate seven cycles. From Figure 10b, there's mixed results between WRFDA and WRFDA/Hybrid for maximum wind error and the TAMDAR data gave positive impact after approximate eight cycles.

# 5. Summary and future work

This paper explored the impact of TAMDAR data with Hybrid method on the hurricane forecasting skill. Previous work had found unstable issues with the six-hour cycling with an ETKF-based Hybrid using previous inflation prototypes due to large variations in the number of ETKF observations between cycles. Our results showed that the modified version of the Wang et al. (2007) inflation scheme (the eigenvector dependent $\rho$-factor) ameliorated that problem, by maintaining greater inter-cycle continuity for the eigen-structure of the posterior spread. We found that: (i) the response of the inflation factor and ensemble spread when adding of TAMDAR data were consistent with Ensemble Kalman filter theory; and (ii) cycles with larger observation numbers easily yield underestimation of inflation factors which can be rescaled with the eigenvector dependent $\rho$-factor fraction scheme; (iii) eigenvector dependent $\rho$-factor can distinguish between large and small background forecast errors explained by the different ensemble eigenvectors and stabilize the ensemble spread after ETKF.

Comparison between Hybrid and deterministic 3DVAR is conducted. The results show that the Hybrid system can outperform 3DVAR system for wind speed and specific humidity and the magnitude of the RMSE reductions for the Hybrid compared with deterministic 3DVAR increases with the forecast lead time. Hybrid has positive impact on track error and central pressure with neutral impact on maximum wind.

This paper also examined the impact of using TAMDAR observations in Hybrid cycling for hurricane forecasting. Our results show that TAMDAR data improved forecast skill at the 12 and 24-hr lead times and that the magnitude of the improvement increased as the forecast lead time grows. TAMDAR data improved the forecast skill for

\

23

hurricane parameters such of location, central pressure. Using TAMDAR data within Hybrid framework can improve maximum wind forecast.

Only one hurricane case is studied using in this study. Additional studies on more cases and over longer time periods are needed to fully assess the impact of TAMDAR data assimilation on tropical cyclones within the Hybrid framework. Further strategies of adaptive inflation schemes with observation density dependent and location dependent will be implemented and explored for ETKF. The LETKF component of WRFDA (Huang et al. 2009) will also be used to assess the impact of TAMDAR data.

\

1 **REFERENCE**

2 Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon.*

3     *Wea. Rev*., **129,** 2884–2903.

4 Barker, D. M., Huang X.-Y., Liu Z. Q., Auligné, T., Zhang, X., Rugg, S., Ajjajj R.,

5     Bourgeois, A., Bray, J., Chen, Y. S., Demirtas, M., Guo,Y.R., Henderson, T., Huang,

6     W., Lin H. C., Michalakes, J., Rizvi, S., and Zhang, X. Y., 2012: The Weather

7     Research and Forecasting (WRF) Model's Community Variational/Ensemble Data

8     Assimilation System: WRFDA. *Bull. Amer. Meteor. Soc.*, **93**, 832-843

9 Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the

10     ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev,* **129**,

11     420–436.

12 Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson and E. B. Sarah, 2008: The

13     MOGREPS short-range ensemble prediction system. *Quart. J. Royal Met. Soc*., **134**:

14     703-722

15 Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error

16     covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Royal. Met. Soc*.,

17     **131**, 1013-1043.

18 Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a:

19     Intercomparison of variational data assimilation and the ensemble Kalman filter for

20    global deterministic NWP. Part I: Description and single-observation experiments.

21    *Mon. Wea. Rev.*, **138**, 1550-1566.

22    Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010b:

23    Intercomparison of variational data assimilation and the ensemble Kalman filter for

24    global deterministic NWP. Part II: One-month experiments with real observations.

25    *Mon. Wea. Rev.*, **138**, 1567-1586.

26    Courtier, P., J. N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational

27    implementation of 4D-VAR, using an incremental approach. *Quart. J. Royal. Met. Soc.*,

28    **120**, 1367–1387.

29    Courtier, P., and Coauthors, 1998: The ECMWF implementation of three-dimensional

30    variational assimilation (3D-Var). I: Formulation. *Quart. J. Royal. Met. Soc.*, **124**,

31    1783–1807.

32    Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model

33    using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5),

34    10143–10162.

35    Etherton, B. J., and C. H. Bishop, 2004: The resilience of hybrid ensemble/3D-Var analysis

36    schemes to model error and ensemble covariance error. *Mon. Wea. Rev.*, **132**, 1065-

37    1080.

38    Gauthier, P. C., L. Cherette, L. Fillion, P. Koclas, and S. Laroche, 1998: Implementation of

39    a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I:

40    The global analysis. *Atmos.–Ocean,* **37,** 103–156.

41    Gao., F., X. Zhang., N. A. Jacobs, X.-Y Huang, and X. Zhang, 2012. Estimation of

42       TAMDAR Observational Error and Assimilation Experiments. *Weather and*

43       *Forecasting*, **27**, 856-877.

44    Hamill, T. M. and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational

45       analysisscheme. *Mon. Wea. Rev.*, **128**, 2905-2919.

46    Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman

47       filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

48    Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for

49       atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

50    Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Quart. J. Royal.*

51       *Met. Soc.*, **131,** 3269–3289.

52    Hamill T. M., and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales

53       in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.*,

54       **133**, 3132–3147.

55    Huang X. Y., Q. Xiao, D. M. Barker, X. Zhang, J. Michalakes, W. Huang, T. Henderson, J.,

56       Bray, Y. Chen, Z. Ma, Dudhia J., Y. R. Guo, Xiaoyan Zhang, D. J. Won, H. C. Lin

57       and Y. H. Kuo, 2009: Four-dimensional variational data assimilation for WRF:

58       Formulation and preliminary results. *Mon. Wea. Rev.,* **137**, 299–314.

59    Jacobs, N., Y. Liu, and C. Druse, 2006: Evaluation of temporal and spatial distribution of

60       TAMDAR data in short-range mesoscale forecasts, in: 10th Symposium on Integrated

61      Observing and As similation Systems for the Atmosphere, Oceans, and Land Surface

62      (IOAS-AOLS), Amer. Meteor. Soc., Atlanta, GA, **29** January–2 February 2006, J9.5.

63  Lorenc, A., 2003: The potential of the ensemble Kalman filter for NWP - a comparison

64      with 4D-Var. *Quart. J. Royal. Met. Soc.*, **129**, 3183-3204.

65  Liu, Y., N. A. Jacobs, W. Yu, et al., 2007: An OSSE study of TAMDAR data impact on

66      mesoscale data assimilation and prediction, in: 11th Symposium on Integrated

67      Observing and Assimilation and prediction, in: 11th Symposium on Integrated

68      Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface

69      (IOAS-AOLS), Amer. Meteor. Soc., San Antonio, TX, January 2007, J5.20.

70  Mizzi, A. P., 2012a:  A Hybrid GSI/ETKF data assimilation scheme for the WRF/ARW

71       model.

72  Moninger, W., S. G. Benjamin, B. D. Jamison, T. L. Smith, E. J. Sozoke, T. W. Schlatter,

73      2007: TAMDAR and its impact RAPID UPDATE CYCLE (RUC) forecasts, in: 22nd

74      Weather Analysis and Forecasting Conference, Amer. Meteor. Soc., Park City, UT,

75      June 2007, J4A. 2.

76  Moninger, W., S. Benjamin, B. Jamison, et al., 2010: Evaluation of regional aircraft

77      observations using TAMDAR, *Weather and  Forecasting*, **25**, 627–645.

78  Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral

79      statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.

80    Rabier, F., J. N. Thepaut, and P. Courtier, 1998: Extended assimilation and forecast

81        experiments with a four-dimensional variational assimilation system. *Quart. J. Royal.*

82        *Met. Soc.*, **124**, 1–39.

83    Rabier, F., H. Jarvinen, E. Klinker, J. F. Mahfouf, and A. Simmons, 2000: The ECMWF

84        operational implementation of fourdimensional variational assimilation. I:

85        Experimental results with simplified physics. *Quart. J. Royal. Met. Soc.*, **126**, 1143–

86        1170.

87    Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with

88        an ensemble Kalman filter. *Mon. Wea. Rev.*, **131,** 1663–1677.

89    Skamarock, W. C., J. B. Klemp, J. Dudhia, Gill, D. O., et al., 2008: A Description of the

90        Advanced Research WRF Version 3. *NCAR Tech.Note TN-475+STR*, 113pp.

91    Torn, Ryan D., G. J. Hakim, C. Snyder, 2006: Boundary conditions for limited-area

92        ensemble Kalman filters. *Mon.Wea. Rev.,* **134**, 2490–2502.

93    Whitaker, J., D. Kleist, X. Wang and T. Hamill, 2011: Tests of a hybrid variational-

94        ensemble global assimilation system for hurricane prediction. Paper J16.2. AMS

95        annual meeting, 2011, Seattle, WA.

96    Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed

97        observations. *Mon. Wea. Rev.*, **130,** 1913–1924.

98    Wang, H. and X.-Y. Huang, 2012: TAMDAR Observation Assimilation in WRF 3D-Var

99        and its Impact on Hurricane Ike (2008) Forecast. *Atmospheric and Oceanic Science*

100      *Letters.*, **5,** 206-211

101    Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform

102        Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.

103    Wang, X., T. M. Hamill, J.S. Whitaker, and C.H. Bishop, 2007: A comparison of Hybrid

104        ensemble transform Kalman filter-OI and ensemble square-root filter analysis schemes.

105        *Mon. Wea. Rev.*, **135**, 1055-1076.

106    Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A Hybrid ETKF-3DVAR

107        data assimilation scheme for the WRF model, Part I: Observing system simulation

108        experiment. *Mon. Wea. Rev.*, **136**, 5116-5131.

109    Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008b: A Hybrid ETKF-3DVAR

110        data assimilation scheme for the WRF model, Part II: Real observation experiment.

111        *Mon. Wea. Rev.*, **136**, 5132-5147

112    Wang, X., 2011: Application of the WRF hybrid ETKF-3DVAR data assimilation system

113        for hurricane track forecasts. *Wea. Forecasting*, **26**, 868-884.

114    Wang, X., T. Lei, J. Whitaker, D. Parrish, and D. Kleist, 2011: GSI-based hybrid ensemble-

115        variational data assimilation system for the Global Forecast system model: 3DVAR-

116        based hybrid and ensemble 4DVAR. Paper J16.5. AMS Annual meeting, Seattle, WA.

117    Jan. 23-27