GSI-based ensemble-variational hybrid data assimilation for HWRF for hurricane initialization and prediction: impact of various error covariances for airborne radar observation assimilation

Xu Lu
Nanjing University of Information Science and Technology, China and School of Meteorology, University of Oklahoma, Norman, OK

Xuguang Wang
School of Meteorology, University of Oklahoma, Norman, OK

Yongzuo Li
School of Meteorology, University of Oklahoma, Norman, OK

Mingjing Tong
Environmental Modeling Center, National Center for Environmental Prediction, College Park, Maryland

Xulin Ma
Nanjing University of Information Science and Technology, China

Oct.31, 2015

Submitted to Quarterly Journal of the Royal Meteorological Society

Modified and resubmitted on Apr. 16, 2016
Further modified and submitted on July 7, 2016
Minor modified and submitted on Aug 21, 2016

Corresponding author address:
Xuguang Wang
School of Meteorology
University of Oklahoma
120 David Boren Blvd.
Norman, OK, 73072
xuguang.wang@ou.edu

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.2914

Abstract

The hybrid EnKF (Ensemble Kalman Filter) -Var (Variational) data assimilation (DA) system based on GSI (Gridpoint Statistical Interpolation) is extended for the Hurricane WRF model (HWRF). Background ensemble forecasts initialized by the EnKF are used to provide the flow-dependent error covariance to be ingested by GSI using the extended control variable method. The hybrid system is then applied to assimilate airborne radar data.

In this paper, the newly developed HWRF hybrid system capable of assimilating airborne radar observations is introduced. The impact of using variously estimated background error covariances on TC (Tropical Cyclone) core analyses and subsequent forecasts is explored by a detailed study of hurricane Sandy (2012) and by systematic comparison of various sensitivity experiments for multiple cases during 2012-2013 seasons. The hybrid system using the HWRF EnKF ensemble covariance (Hybrid-HENS) is able to correct both the wind and mass fields in a dynamically and thermodynamically coherent fashion. In contrast, the wind and pressure adjustment by GSI3DVar using the static covariance are inconsistent. The wind and pressure relation in the covariances

derived from the GFS ensemble (Hybrid-GENS) improves upon the static covariance, but is still inconsistent compared to that of HWRF. Verifications against independent flight level and SFMR wind data, and HRD radar wind composite reveal that the Hybrid-HENS system improves the analyzed TC structure upon both GSI3DVar and Hybrid-GENS. Hybrid-HENS and Hybrid-GENS improve the track, MSLP and Vmax forecast relative to GSI3DVar. Hybrid-HENS further improves track forecasts compared to Hybrid-GENS. Hybrid-HENS provides the largest positive impact of the airborne radar data. In comparison, GSI3DVar shows consistently negative impact of the data when analyzing the structure and verifying track forecasts. Blending the static background error covariance in the hybrid system improves the maximum wind forecast while little benefit is found in the analyzed structures and the MSLP and track forecasts.

Key words: GSI, hybrid EnKF-Var data assimilation, Tail Doppler Radar, Hurricane

1. Introduction

While steady progress has been made in improving TC (Tropical Cyclone) track forecasts, difficulties still remain. Making progress for the intensity forecast is even more difficult (Rogers et al., 2013). This can be attributed to deficiencies in both models and current operational data assimilation systems. With respect to data assimilation methods, the current algorithms used in operations often do not permit effective utilization of existing observations.

The US operational Hurricane WRF (Weather Research and Forecasting Model; henceforth, HWRF) modeling and prediction system had adopted a three-dimensional variational (3DVar) data assimilation system using the National Centers for Environmental Prediction (NCEP) Gridpoint Statistical Interpolation algorithm (GSI, Wu et al., 2002). It used a quasi-static background error covariance (Parrish and Derber, 1992; Wu et al., 2002), which when applied to TCs, cannot represent the model forecast uncertainties accurately. Specifically, adjustments which are made when assimilating observations using GSI are not consistent with the TC in the background forecast and the cross-variable error statistics are either lacking or not suitable for TCs, which can result in an analyzed TC that lacks dynamic and thermodynamic coherency.

Recently, an ensemble-based data assimilation method (ENSDA) has been implemented and tested for TC forecasting (Zhang et al., 2009; Torn, 2010; Li et al., 2012; Aksoy et al., 2013). As shown in these early studies, ENSDA uses ensemble covariance to infer error statistics within the same variables and across different variables flow-dependently. Therefore, state variables updated by ENSDA are dynamically and thermodynamically consistent.

A hybrid ensemble-variational data assimilation (DA) method has been proposed (e.g., Hamill and Snyder, 2000; Lorenc, 2003; Etherton and Bishop, 2004; Wang et al., 2007b; Wang, 2010). In this method, the analysis increment is calculated using an ensemble-based, flow-dependent estimate of the background covariance within the variational framework (Wang et al., 2013). The hybrid method has been implemented for

both regional and global numerical weather prediction models (e.g., Wang et al., 2008a,b; Buehner et al., 2010a,b; Bishop and Hodyss 2011; Wang 2011; Li et al., 2012; Zhang and Zhang, 2012; Clayton et al., 2013; Wang et al., 2013; Wang and Lei, 2014; Kutty and Wang 2015). In particular, results from these studies indicate that the hybrid method may be able to take advantages of both the Var and EnKF methods (e.g., Wang et al., 2007a, 2009; Buehner et al., 2010b; Zhang and Zhang, 2012; Wang et al., 2013).

Recently, noted successes have been reported by using three-dimensional or four-dimensional hybrid DA for TC track forecasts in coarse-resolution NWP model settings (Wang, 2011; Hamill et al., 2011; Wang and Lei, 2014; Poterjoy and Zhang 2014) and for track and intensity forecasts at convection-allowing resolutions which assimilate ground-based radar observations (Li et al., 2012). A hybrid EnKF–variational data assimilation system including both three-dimensional (3DEnVar) and four-dimensional ensemble-variational (4DEnVar) frameworks was also developed based on the operational NCEP GSI system. Experiments with the Global Forecast System (GFS) showed that forecasts produced by 3DEnVar were more skillful than GSI 3DVar measured by verification metrics for overall global forecast (Wang et al., 2013) and for tropical cyclone forecast (Hamill et al., 2011). 4DEnVar further improved both the global and TC forecasts (Wang and Lei, 2014).

Encouraged by the improvement of global forecasts resulting from the hybrid DA method, focused efforts have been made to adapt, apply, and test the GSI-based hybrid DA strategy for the operational HWRF model. During the 2013 North Atlantic and

Eastern Pacific TC seasons, a hybrid DA method was adopted by the operational HWRF DA system where the background error covariance was derived from the GFS ensemble (Tong et al., 2016). While the GFS ensemble is able to provide flow-dependent ensemble covariance, the resolution of the GFS ensemble covariance is too coarse to properly resolve the error covariance of the TC core. Therefore, a self-consistent hybrid DA system for HWRF is developed where HWRF ensemble consistently initialized from HWRF EnKF analyses from previous DA cycle is ingested in the GSI hybrid, replacing the GFS ensemble.

An accurate TC forecast is also highly contingent upon effective utilization of all available observations. High-resolution airborne TC observations such as data from Tail Doppler radars (TDR) onboard reconnaissance aircrafts have been regularly collected since 1980s (Aberson et al., 2006). Because effectively assimilating cloud or rain contaminated radiances is still a challenge, these data remain to be the primary source in the operational data stream that provide valuable three-dimensional description of the TC core circulation. Recent studies using EnKF demonstrated positive impact of assimilating the airborne radar data (e.g., Weng and Zhang, 2012; Aksoy et al., 2012). Exploring the impact of airborne radar data using the hybrid DA method on TC core analysis and on subsequent forecast is, however, still limited. Therefore, as a first step of evaluating the potential of the newly extended hybrid DA system, this study focuses on the assimilation of airborne radar observations. Further development of the system to

assimilate all operational observations including TDR, conventional and satellite radiances are ongoing and results will be presented in future papers.

This manuscript first describes the newly extended self-consistent, GSI-based ensemble-variational hybrid data assimilation for HWRF. Using the newly developed system, the impact of using variously estimated background error covariances on TC core analyses and on subsequent forecasts assimilating the airborne radar observation is studied. Specifically, the following scientific questions are addressed: a) What is the impact of using flow-dependent ensemble covariances derived from coarse-resolution GFS ensembles versus the use of the static covariance for TC initialization and forecast? b) Does using HWRF's own EnKF ensemble provide further improvements relative to using the GFS ensemble? c) What are the impacts of blending the static covariance and the EnKF ensemble covariance for TC core analysis and forecast? And d) what are the differences of the impact of assimilating TDR data using the variously configured background error covariances? Compared to our earlier study of Li et al (2012), this study is among the first to explore the impact of assimilating airborne radar data using the hybrid DA method, addresses complementary scientific questions as outlined above, and uses many cases to draw systematic conclusions rather than a case study.

A detailed study of TC Sandy (2012) is documented first, followed by systematic results from multiple cases during 2012-2013 seasons. The rest of this paper is organized as follows: section 2 describes the hybrid system for HWRF. Section 3 discusses the experiment design using Sandy as an example. Detailed experiment results of Sandy

(2012) are discussed in section 4. Section 5 presents systematic results for various sensitivity experiments for multiple cases during 2012-2013 seasons. The final section concludes the study.

2. GSI-based hybrid EnKF-Var data assimilation system for HWRF

Following the GSI-based hybrid EnKF-Var DA system for GFS (e.g., Wang et al., 2013), a similar system is extended to HWRF. For consistency, the description of the system and methods therein for HWRF parallels that of Wang et al. (2013) and the texts in this section is derived from Wang et al. (2013) with adaptation to the HWRF modeling system.

As shown in Fig. 1, each cycle consists of the following three steps:

1) Use the augmented control vector (ACV) method in GSI variational minimization (GSI-ACV) to update the HWRF control background forecast, where the flow dependent ensemble based error covariance is incorporated.

2) Use an EnKF to update the HWRF forecast ensemble to produce the analysis ensemble.

3) Advance HWRF ensemble and control forecasts to the next analysis time.

The "GSI-ACV" component of the system was extended to HWRF (Fig. 1). The mathematical details of the implementation of ACV in GSI are described in Wang (2010). The formulas following the notation of Wang (2010) are briefly described below.

Lorenc (2003), Buehner (2005), Wang et al. (2007b) and Wang et al. (2008a) also adopted similar notations.

The term x', denoting analysis increment, is defined as

$$\mathrm{x}' = \mathrm{x}'_1 + \sum_{k=1}^{K} (\mathrm{a}_k \circ \mathrm{x}_k^e) . \quad (1)$$

The first term on the right, x'_1, is the increment corresponding to the static covariance. The second term on the right hand side is the increment corresponding to the flow-dependent ensemble covariance. x_k^e is the *k*th ensemble perturbation normalized by $\sqrt{K-1}$, where K is the ensemble size. The vector a_k is the augmented control vector (ACV) for the *k*th ensemble member. The symbol $\circ$ is used to define the Schur product. The solution of both x'_1 and a_k in eq. (1) is solved through minimizing eq. (2):

$$J(\mathrm{x}'_1, \mathrm{a}) = \beta_1 \frac{1}{2} (\mathrm{x}'_1)^T \mathrm{B}_1^{-1} (\mathrm{x}'_1) + \beta_2 \frac{1}{2} (\mathrm{a})^T \mathrm{A}^{-1} (\mathrm{a}) + \frac{1}{2} (\mathrm{y}^{o\prime} - \mathrm{Hx}')^T \mathrm{R}^{-1} (\mathrm{y}^{o\prime} - \mathrm{Hx}') \quad (2)$$

The first and last terms on the right hand side correspond to the traditional 3DVar background and observational terms where B_1, $\mathbf{R}$, $\mathrm{y}^{o\prime}$ and H are the static error covariance, observation error covariance, innovation vector and observation operator respectively. Different from traditional 3DVar, x' is defined by (1) in the observational term. In the second term of the right hand side, a is called ACV, formed by concatenating K vectors, a_k, $k = 1, \cdots K$. As proved by Wang et al. (2007a), the block-diagonal matrix A defines the localization applied to the ensemble covariance. Similar to the implementation to GFS hybrid DA system as described in Wang et al. (2013), the

spatial localization is applied both horizontally and vertically and no cross-variable covariance localization is applied for HWRF. For HWRF, a_k is applied for the variables such as surface pressure, wind, virtual temperature and relative humidity. At the end of the variational minimization, these variables were converted to the corresponding HWRF prognostic variables. For HWRF, both horizontal and vertical localizations are realized through a recursive filter transformation (Hayden and Purser, 1995). The distances are measured in kilometer and scale heights (i.e., natural log of the pressure) respectively for the horizontal and vertical localizations. For the vertical localization, the natural log of pressure is transformed into the HWRF Pressure-Sigma coordinate. As proved in Wang et al., 2007a and Wang et al. 2008, the inverses of the two factors β_1 and β_2 represent the weights assigned for the static and ensemble covariances respectively. These two factors satisfy $\frac{1}{\beta_1}+\frac{1}{\beta_2}=1$, which means the full background error covariance is a weighted sum of the static and ensemble covariances.

The EnKF component is also extended to HWRF. The same code has been interfaced with the GFS (Whitaker et al. 2008; Wang et al. 2013) and WRF ARW models (e.g., Johnson et al. 2015). This flavor of the EnKF adopts the ensemble square root filter algorithm (Whitaker and Hamill, 2002). Like the implementation with GFS and WRF ARW (Whitaker et al. 2008; Wang et al. 2013; Wang and Lei 2014; Johnson et al. 2015), the EnKF is interfaced with the GSI where the observation operators, observation pre-processing and quality control from GSI are used. Covariance localization with cut-off

distances similar to those used by hybrid is adopted in the EnKF to reduce the sampling errors. An adaptive, multiplicative inflation algorithm proposed by Whitaker and Hamill (2012) is adopted in the HWRF EnKF to alleviate the deficiency in the spread of the first guess ensemble.

3. Experiment design

In this section, Sandy (2012) is used as an example to illustrate the experiment design. Detailed results for this case will be discussed in section 4. As documented in the report from the National Hurricane Center (NHC) (http://www.nhc.noaa.gov/data/tcr/AL182012_Sandy.pdf) and as shown in Fig. 2, Sandy went through a complicated evolution during its life span as a late season TC. Figure 2 shows the track, Vmax and minimum sea level pressure (MSLP) evolution of Sandy and the official HWRF forecasts 5 days before Sandy made landfall at the Northeast US. The official HWRF forecasts initialized during October 27 and 28 in 2012 show eastward bias and forecasts initialized during October 26 and 27 overestimated Vmax and MSLP. As also discussed in the NHC report, Sandy caused $50 billion dollar damages over the US, and was the second costliest hurricane for the US since 1900.

a. The HWRF model configuration

The HWRF model, a nonhydrostatic primitive equation model on a rotated-E grid, (http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFv3.5a_Scient

ificDoc.pdf) is used in this study. As an initial study of the newly extended system, for simplicity, a single domain with a grid spacing of 0.06 degrees (approximately 9 km) is used. Future papers will document the hybrid system with moving nest capabilities where the inner nests are at higher resolution. The HWRF is configured with 230×450 horizontal grid points (Fig. 2a), and 61 vertical levels. The model top is at 2 hPa. Following the physics configurations in the scientific documents for HWRF, the HWRF Ferrier microphysics scheme based on Eta Grid-scale Cloud and Precipitation scheme (Ferrier, 2005), the Simplified Arakawa-Schubert (SAS) cumulus parameterization scheme (Han and Pan, 2011), the modified surface layer parameterization scheme, the GFDL slab scheme for land surface model (Tuleya, 1994), the non-local planetary boundary layer parameterization scheme (Hong and Pan, 1996), the Eta GFDL short wave (Lacis and Hanson, 1974) and long wave (Fels and Schwarzkopf, 1975; Schwartzkopf and Fels, 1991) radiation parameterization schemes are included in the HWRF simulation.

b. Airborne radar observations and processing

In this study, the airborne radial velocity data from the tail Doppler radar aboard the NOAA P-3 aircraft are assimilated in variously designed experiments. Fig.3 shows the flight tracks for seven NOAA N42RF flight missions that were carried out for hurricane Sandy between 25-31 October 2012. These flight missions were taken during 25 ~ 31 of October 2012 for Sandy. More details of flight missions and the data collected

can be found from http://www.aoml.noaa.gov/hrd/Storm_pages/sandy2012/radar.html.

As described in Marks and Houze (1984, 1987), Marks et al. (1992), Weng and Zhang (2012), and Li (2015), the tail Doppler radar onboard NOAA P-3 uses a fore/aft scanning technique (FAST). The angle between each radial beam and flight track is 70°. A cone surface is formed by the radial beams of each sweep. The antennas alternate between forward and afterward scanning. As the aircraft moves, FAST forms a three dimensional radial velocity volume. The two closest forward (afterward) sweeps are separated by about 1.4 km. Fig. 4 shows an example of a Tail Doppler Radar sweep. The negative values in blue means winds towards the radar, and the positive values in red means winds away from the radar. In this example, the radar is scanning a southerly wind with the fore scan pointing to the east.

As described in Li (2015), further quality control and data thinning are conducted for the airborne radar observations collected by the Hurricane Research Division (HRD). As discussed in Gamache (2005), several passes of quality control were made by HRD. These procedures include spurious observations removing, aircraft motion subtracting, and radar velocity de-aliasing. The data were then sent to the Environmental Modeling Center (EMC) in the BUFR (Binary Universal Form for the Representation) format for operational applications.

An enhanced thinning procedure is subsequently implemented in GSI to reduce the density of the data following Li (2015), given that in the radial velocity data sent off to EMC, the spacing along a radial beam is about 1.2km, which is much higher than the

model resolution adopted in the current study. The thinning procedure includes: 1) Separate fore/after scans based on the highest earth-relative elevation angle; 2) Divide the model space into many cubic boxes. 3) Different from the original thinning method of GSI, two radial velocities from forward and afterward scans instead of one radial velocity were maintained for each box. The primary goal of this modification is to keep wind components from different directions measured by the dual Doppler strategy. These two radial velocities are nearest to the center of the box. This thinning method is different from the "superobbing" technique applied by Zhang et al. (2009) which takes average within each box instead.

Fig. 3 shows horizontal distributions of the radial velocity observations of Sandy from the Doppler radar onboard the NOAA P3 aircraft after thinning for all legs and all missions. The plot is generated by aggregating from all levels. The swath of TDR data is about 1 latitude/longitude degree in width due to the maximum unambiguous range (Aksoy et al. 2012). Fig. 5 shows the vertical distributions of the total number of TDR observations collected during the first P3 mission of Sandy. The greatest number of observation is generally around 880 hPa where strongest precipitation occurs. The typical aircraft altitude is between 2500 m to 3800 m (e.g. http://www.aoml.noaa.gov/hrd/Storm_pages/sandy2012/20121027H2_wind.jpg).

Following Dowell et al. (2004) and Weng and Zhang (2012), the standard deviation of the radial velocity observation errors is assigned to be 3.0 ms^{-1}. Experiments where other values of observation errors such as 5.0 ms^{-1} are conducted

(Not shown). Initial examination of these experiments using other values of observation errors indicate that the comparison among experiments using different background error covariances does not vary qualitatively.

c. *Data assimilation experiments*

In the remainder of the section, nine experiments denoted as NoDA, GSI3DVar, Hybrid-HENS, Hybrid-HENS.5, Hybrid-GENS, Hybrid-HENS-Hrly, Hybrid-HENS.5-Hrly, GSI3DVar-Hrly and Hybrid-GENS-Hrly (see definitions in Table 1) are presented to address the series of scientific questions proposed in section 1.

As discussed in section 1, these experiments are different from our early hybrid DA study assimilating ground-based radar data such as Li et al. (2012) in both methodologies and scientific questions to be addressed. While Li et al. (2012) used the ensemble produced by perturbed observation method and focused its comparison with the static covariance, the hybrid experiments here use HWRF's own EnKF ensemble or the GFS ensemble and focus on the impact of various sources of ensembles on hybrid DA. Li et al. (2012) experimented with the ground-based radar whereas this study focuses on the airborne radar data. The current study also draws conclusion based on many cases rather than a single case. Details of each experiment in Table 1 are described below.

In the NoDA experiment, the initial condition and lateral boundary conditions are directly obtained from the ensemble mean analysis and forecast of the GFS hybrid Var-EnKF data assimilation system (Wang et al., 2013) (Fig 6a.). In other words, no

observations are assimilated in NoDA.

In the GSI3DVar experiment, the airborne radar data are assimilated by the GSI 3DVAR method. Fig. (6b) describes the procedures of the GSI3DVar experiment using the first TDR mission as an example. First, 40-member ensemble forecasts are carried out for 4 hours initialized from GFS ensemble analyses at 1800 UTC on October 2012. Average of the 40 ensemble members become the background at 2200 UTC on October 25 for the first DA cycle. Second, single deterministic analysis and forecast cycles follow. Each assimilation cycle ingests tail Doppler radial velocity data contained in each penetration leg. The leg-based TDR data assimilation is adopted for Sandy for the purpose of better illustrating the increment differences from various experiments. The assimilation time is 2200, 2330 UTC October 25, 0030, 0200 UTC October 26, which are close to the centers of the penetration legs. Each data assimilation cycle is done in the geographically fixed coordinate rather than in the storm-centered coordinate. Third, the analysis produced at the end of the final assimilation cycle is used to initialize a single deterministic forecast. As an initial study, the deterministic forecast is only made to 48 hours due to the computational limitation. Like NoDA, the HWRF boundary conditions for GSI3DVar are also provided by the GFS control forecast. The static background error covariance used in this study is from the operational HWRF, which is calculated through the NMC method (Parrish and Derber, 1992; Wu et al., 2012). As detailed in Wu et al. (2012), the static error covariance derived from this method is suitable for mid-latitude synoptic scale systems. Following Li et al. (2012) and Li (2015), the horizontal and

vertical correlation length scales of the default static covariance for GSI 3DVar are tuned and reduced by factors of 0.2 and 0.6 respectively to produce the best results for the GSI3DVar experiments.

The procedures in Hybrid-HENS are similar to GSI3DVar. The primary difference is in step 2 where the background covariance is calculated from HWRF's EnKF first guess ensemble rather than a static covariance (Fig.6c). Unlike GSI3DVar, a variational minimization of the hybrid cost function (section 2) is conducted to produce the control analysis, which initializes the subsequent control forecast. For all the Hybrid related experiments in this study, horizontal and vertical localization cut-off distances (Gaspari and Cohn 1999) are 450 km and 1 scale height respectively. The details about converting between the localization length scale in the e-folding distance used in the recursive filter and that in the cut-off distance used in the method by Gaspari and Cohen (1999) can be found in equation (4) of Pan et al. (2014). These values are selected after tests with a range of values. In this study, EnKF which provides ensemble covariance to Hybrid-HENS had 40 ensemble members. The inflation factor is implemented to relax the posterior ensemble variance to 90% of the prior ensemble variance. The lateral boundary conditions are obtained from the GFS hybrid Var-EnKF ensemble forecast (Whitaker et al., 2008; Wang et al., 2013). As in the GSI3DVar experiment, the ensemble background for the first data assimilation is initialized by the GFS ensemble analyses 6 hours prior to the synoptic times when the TDR data is available. For example, for the 00 UTC TDR mission, the background ensemble for the first assimilation is

initialized at 18 UTC. This 6-hour spin up time period is consistent with early studies that focused on the assimilation of inner core observations. In early studies, the first ensemble background is initialized less than 12 hours before the first assimilation (e.g. Zhang et al., 2009; Li et al., 2012; Aksoy et al., 2012; Zhang et al., 2015). Our subjective analysis also shows a 6-hour spin up is able to produce flow-dependent TC covariances for the first assimilation. For each subsequent HWRF EnKF DA and forecast cycle, the 40-member background ensemble forecasts are updated by assimilating the tail Doppler radial velocity data. 40-member ensemble background forecasts are subsequently initialized by the HWRF EnKF analyses.

As discussed earlier, Hybrid-HENS.5 is designed to reveal the impact of blending static covariance and flow-dependent ensemble covariances. Different from Hybrid-HENS where a 100% weight is used on the ensemble covariance, Hybrid-HENS.5 blends the static covariance and the flow-dependent ensemble covariance by assigning a 50% weight on each. Similar blending was used in early hybrid data assimilation studies (e.g., Wang et al. 2008ab; Li et al. 2012, etc.)

Hybrid-GENS follows the same steps as Hybrid-HENS except that in step 2, the background covariance is estimated from GFS ensemble perturbations (Fig. 6d).

Hybrid-HENS-Hrly, Hybrid-HENS.5-Hrly, GSI3DVar-Hrly and Hybrid-GENS-Hrly share the same steps as Hybrid-HENS, Hybrid-HENS.5, GSI3DVar and Hybrid-GENS respectively, except assimilating TDR data at hourly frequency and including data from both penetration and downwind legs (Fig. 6e, f and g).

4. Results for Sandy (2012)

a. Flow dependent and static increment

As discussed in section 1, ensemble-based DA can use ensemble covariance to realistically infer the flow-dependent error statistics among the same variables and cross different variables. Therefore, state variables including those directly and indirectly observed can be updated in both dynamically and thermodynamically consistent manners. Fig. 7 shows the wind field at 1000 hPa and the sea level pressure (SLP) field before and after assimilating the first penetration leg of TDR data during the first P-3 mission of Sandy. The background forecast (Fig. 7a) has a position error with the predicted position located to the west of the best track location. Using the static covariances, GSI3DVar is able to relocate the center of the minimum wind speed eastward after assimilating the TDR data. However, the adjustment of the location of the MSLP is not coherent with that of the wind field. Instead, the analyzed MSLP is located to the west of the minimum wind speed and still to the west of the best track location. For Hybrid-HENS, Hybrid-GENS, which use the ensemble derived covariances, both the centers of MSLP and minimum wind speed are consistently adjusted where the two centers are collocated and moved toward the best track location as a result of the DA. Compared to Hybrid-GENS, the analyzed TC center by Hybrid-HENS using HWRF's own EnKF ensemble is closer to the best track and the analyzed MSLP is deeper. The spatial pattern of the analyzed

TC by Hybrid-HENS is also less stretched along the North West – South East axis than Hybrid-GENS. With the use of the flow-dependent ensemble covariance, Hybrid-HENS.5 improves over GSI3DVar as the centers are now more consistently adjusted. Specifically, the centers of MSLP and wind speed are more collocated and the analyzed MSLP is also deeper than GSI3DVar. However, due to the use of static error covariances, the analyzed center locations remain inconsistent. These results suggest that the use of static error covariances is unable to coherently adjust the mass fields appropriately for TCs when only the momentum fields are directly observed. This result is consistent with what is found in Li et al. (2012), which adopted a different hybrid DA system and assimilated different types of observations

A single-observation experiment is conducted to reveal the fundamental difference among the static covariance, HWRF EnKF ensemble covariance and GFS ensemble covariance. In this single observation experiment, a 700-mb meridional wind observation to the east of the background storm center is assimilated. For consistency, the background field is the same as Fig. 7a. The observation increment is 5.0 ms^{-1}. Fig. 8 shows the resulting analysis increment of wind and geopotential height at 700hPa by using different background error covariances. The wind analysis increment obtained from using the static covariance in GSI3DVar (Fig. 8a) shows symmetric features centered around the meridional wind observation. The geopotential height increment of GSI3DVar (Fig. 8e) shows an east-west symmetric response to the wind increment. This symmetric geopotential height increment is largely consistent with the geostrophic and hydrostatic

balances, with the geopotential height field decreasing on the left and increasing on the right of the northward pointing meridional wind innovation. However, this geostrophic adjustment is not consistent with the existence of the TC vortex and its own dynamical balance conditions.

The wind analysis increments obtained from using the ensemble derived covariances in Hybrid-HENS (Fig. 8b) show more flow-dependent features appropriate for a TC vortex. Specifically, in response to the meridional wind innovation in the east of the storm center, a strengthened vortex circulation is produced around the storm center with a more realistic cyclonic wind increment pattern. The inherent scale of the wind increment, smaller than the GSI3DVar is also consistent with the background TC. For the geopotential height increments (Fig. 8f), the reduced central geopotential height is consistent with the enhanced cyclonic wind circulation in respond to the meridional wind innovation as indicated from the wind increments of Fig. 8b.

It is noted that while the maximum value of wind increment is similar between GSI3DVar and Hybrid-HENS, the geopotential height increment by GSI3DVar is much weaker. Further diagnostics using geostrophic and gradient wind balance equations show that the geopotential height gradient increment in response to a 5.0 ms^{-1} wind increment following the geostrophic balance is more than three times less than that following the gradient wind balance. This weakened geopotential height gradient increment implies that the cross-variable correlations between mass and wind variables in the static background error covariance are not suitable for estimating the error covariance of the

TC vortex. This result further explains why the GSI3DVar analysis cannot achieve the dynamic and thermodynamic coherency when only the momentum fields are directly observed.

Compared to GSI3DVar, increments by Hybrid-GENS showed much improved flow-dependent structures suitable for a TC (Fig. 8c and 8g). Fig. 8 however further reveals the difference of the error covariances derived from HWRF EnKF and GFS ensembles. The wind analysis from Hybrid-HENS using HWRF EnKF ensemble shows a location correction while Hybrid-GENS does not. For Hybrid-HENS using the self-consistent HWRF EnKF ensemble, the geopotential height increment suggests adjusting the storm position to the east, consistent with that of the 700-mb wind adjustment. For Hybrid-GENS, the geopotential height increment suggests correction of the storm position to the east whereas the 700-mb wind increment suggests no position correction. This inconsistent pressure and wind increment patterns in terms of storm location adjustment in Hybrid-GENS suggest the inconsistent wind and pressure relationship described by the GFS ensemble for the TC vortex. In addition, while the magnitude of wind increments is similar between Hybrid-HENS and Hybrid-GENS, the geopotential height increment by Hybrid-GENS is weaker. Fig. 8b and 8c also shows the correlation length scale of Hybrid-HENS is smaller than that of Hybrid-GENS. This result suggests that the wind and pressure relationship captured by the GFS ensemble is inconsistent and is inferior to the wind and pressure relationship described in the HWRF ensemble. Such inconsistency likely results in the dipole features in the final pressure analysis of Hybrid-

GENS as shown in section 4c and therefore explains the inferior performance of Hybrid-GENS than Hybrid-HENS.

By incorporating the static covariance in the Hybrid-HENS.5 (Fig. 8d and 8h), the wind increment patterns have more flow-dependent features in comparison to the GSI3DVar. For example, the wind increment is more cyclonic. Also, the geopotential height increment is larger and shows more axisymmetric features (Fig. 8h). In addition, the correlation length scale in the Hybrid-HENS.5 is smaller than that value in the GSI3DVar. However, the reflected wind and pressure relationship is not as good as using the full ensemble.

Early studies by e.g. Wang et al., 2007a and Wang et al 2009 using a primitive equation two layer model have found that blending the static covariance with the ensemble covariance can relax the covariance localization applied for the ensemble covariance and therefore the resultant analysis is more balanced. However, in those studies, the static covariance is well tuned and is consistent with the large scale dynamics in the model. The benefit of relaxing the localization scale by incorporating the static covariance in the hybrid system can be seen for certain fields for hurricanes (not shown). However, as discussed in this sub section and shown in the rest of the paper, the static covariance reflecting the large scale dynamic and thermodynamic relationship is not appropriate for hurricane data assimilation, which explains why including the static background error covariance in Hybrid-HENS shows little benefit.

Fig. 8 reveals the fundamental differences in the same variable and cross variable correlations between the static covariance and ensemble covariances. Further diagnostics based on the spatial distribution of the variance of different error covariances show that the static covariance based on the NMC method and used in the operational HWRF has larger variances than the ensemble covariance and its spatial distribution is more reflective of mid-latitude large scale system rather than a TC (Not shown). The diagnostics suggest that further development of the static variance may be needed to optimize its usage in the hybrid DA.

b. *Verification against independent flight level and SFMR wind observations*

The in-situ measurements made by the NOAA P-3 aircraft and the Stepped Frequency Microwave Radiometer (SFMR) retrieved surface wind speeds provides independent observations to verify the simulated TC inner-core structures. Several early studies used similar types of data for verification (e.g., Weng and Zhang 2012; Aksoy et al. 2013). The results for the first and last data assimilation corresponding to the first and last penetration flight legs of the first TDR mission are discussed for Sandy in this section. Systematic verification against these observations is documented in section 5.

Fig. 9 and 10 compare the surface wind speeds, derived from GSI3DVar, Hybrid-HENS, Hybrid-GENS and that observed by the SFMR on board of the P3. The wind speeds from the model are interpolated to the observation location at 2200 UTC 25 October for the first penetration leg (Fig. 9) and 0200 UTC 26 October 2012 for the 4th

penetration leg (Fig. 10). In Fig. 9, the lowest wind speed indicates the along-track wind minimum captured by the observation. The flight track is overlaid on the corresponding analyses in Fig. 7. For the first DA cycle ingesting the first penetration leg of TDR data, the first guess misplaces Sandy by 80 km and the corresponding peak wind is also weaker and broader. After assimilating the TDR data, the along-track wind minimum is corrected to the location nearly consistently with that identified by the SFMR wind speed observations for Hybrid-HENS and Hybrid-GENS. The location of the peak wind, especially along the west side of the penetration leg, is also nearly consistent with the SFMR wind. Hybrid-HENS fits the SFMR wind more closely than Hybrid-GENS. Specifically, the analysis from Hybrid-GENS has a relatively broader peak on the west side of the penetration leg than Hybrid-HENS. The minimum wind speed in Hybrid-HENS fits the SFMR more closely than Hybrid-GENS. GSI3DVar over corrects the location and therefore fit the SFMR the worst. The fit of Hybrid-HENS.5 to SFMR is worse than to that of Hybrid-HENS.

Fig. 10 shows the comparison at the final data assimilation cycle ingesting the last penetration leg of TDR data. The flight track is overlaid to the corresponding analyses in Fig. 11. Without assimilating the TDR data, wind speed from NoDA is out of phase compared to the SFMR observations. First guess surface wind fields from all experiments assimilating TDR data show improved fit to SFMR wind observations after assimilating TDR data in previous cycles except GSI3DVar, suggesting the positive impact of assimilating TDR data if using ensemble covariances to estimate background

error covariances. First guess surface wind from Hybrid-HENS fits SFMR wind the best. The vortex in Hybrid-GENS first guess field is weaker than the SFMR wind and that of Hybrid-HENS, with relatively weak peak wind. Using static covariance, the first guess from Hybrid-HENS.5 is comparable or slightly worse than Hybrid-HENS. After assimilating the last penetration leg of the TDR data, analysis from Hybrid-HENS and Hybrid-GENS both improve their fits to the SFMR wind. In comparison, the analysis from GSI3DVar fits SFMR wind observation worse compared to not assimilating the last penetration leg of TDR data. Verification against flight level wind observations show similar results (not shown). These results in Fig. 9-10 together with results in Fig. 7-8 suggest the impact of assimilating the TDR data can be highly dependent on the type of background error covariance. Positive impacts of assimilating TDR are consistently found when using ensemble-estimated error covariances in data assimilation. In comparison, the use of the full static covariance can lead to weak positive and some negative impacts from data. Within the methods of using ensembles in data assimilation, using self-consistent HWRF EnKF ensemble is able to best extract the information to produce most accurate storm structure and therefore produces the largest amount of positive impact of the TDR data.

c. Analyzed tropical cyclone structure

The structure of the TC simulated by different data assimilation methods is further examined by comparing the analyses at the end of the data assimilation cycles with the

radar wind analyses produced by the NOAA Hurricane Research Division (HRD). The HRD radar wind analyses are derived by compositing quality controlled TDR radial velocity data (Gamache, 2005; HRD radar wind data can be found: http://www.aoml.noaa.gov/hrd/Storm_pages/sandy2012/radar.html). Fig. 11 shows the HRD radar wind analysis and the model-derived wind at 1km height. Radar wind analysis showed two (inner and outer) wind maximum bands in the northwest and north quadrants with a wind maximum about 45-50 ms^{-1}. The storm center for NoDA is about 80 km away from the best track position. After assimilating the TDR data, in the final analysis, the storm center for all experiments assimilating the TDR data is much closer to that of the best track, resulting in a 75% reduction of position error.

The major difference among the experiments assimilating TDR is the depiction of the storm structure. Rather than concentrating wind maxima in the north and northwest quadrants, GSI3DVar has multiple spurious large-wind regions. For example, a wind speed greater than 60 ms^{-1} is in the southeast and south quadrants, which is absent in the HRD wind analyses. The structure of the storm including both max wind speed and spatial distribution of the wind speed in Hybrid-GENS is much improved compared to GSI3DVar. Compared to the radar wind analysis, wind speed of Hybrid-GENS is still too strong especially over the northwest quadrant. Hybrid-HENS compares most closely with the HRD radar wind analysis. Compared to Hybrid-GENS, Hybrid-HENS shows more detailed structures such as the northeast-southwest elongated wind speed minimum center and details of two banded wind maxima. In addition, in Hybrid-GENS, there is a

secondary minimum in the pressure analysis to the west of the storm center. Compared to Hybrid-HENS, the ensemble in Hybrid-GENS is at a coarser resolution and is produced independently by another model. To isolate the impacts of these two differences, another hybrid experiment is conducted. This experiment is configured the same as Hybrid-HENS except that the HWRF EnKF ensemble is at coarser resolution similar to the GFS. It is found that with the coarse-resolution, Hybrid-HENS still does not show such dipole features (Not shown). This result implies that the coarse-resolution is not the main issue that results in the unrealistic pattern in Hybrid-GENS. Therefore, such unrealistic pattern is likely due to the inconsistency of using GFS ensemble to estimate the HWRF background forecast error covariance. For example, based on the discussions from section 4a, it is likely that the wind and pressure relationship described by the GFS ensemble is inconsistent with wind and pressure relationship from the HWRF first guess. Involving ensemble in Hybrid-HENS.5 improves the wind speed pattern as compared to the GSI3DVar. However, the analyzed wind field by Hybrid-HENS.5 still has spuriously large wind maximum centers resembling that of GSI3DVar. Among all data assimilation experiments, Hybrid-HENS shows the largest positive impact of the TDR data in analyzing the structure of Sandy.

Fig. 12 shows the south-north vertical cross section of horizontal wind speed. Radar wind analysis indicates a northward tilted eye and a maximum wind speed of 40-45 ms^{-1} to the north of the eye between 1 km and 2 km. The simulated eye in NoDA is upright instead of being tilted. As in Fig. 11, GSI3DVar has multiple spurious maxima

with values larger than that in radar wind analysis. Although there is an indication of tilt, the simulated eye in GSI3DVar is wider than the HRD radar wind analyses. Using GFS ensemble in the assimilation, Hybrid-GENS improves the depiction of the vertical structure of the storm compared to GSI3DVar. Specifically, the magnitude of wind maxima at low levels matches the radar wind analysis much closely and the upper level wind maxima become much smaller or non-existent. Compared to Hybrid-GENS, Hybrid-HENS simulated storms show much more detailed structure that better matches the radar wind analysis. For example, the separation of wind maxima to the north of the storm is closer to the HRD wind analyses whereas such separation is larger in Hybrid-GENS. The northward tilted eye is narrower than Hybrid-GENS, better fitting the HRD wind analysis. Hybrid-HENS.5 improves the analyzed vertical structure compared to GSI3DVar. However, spurious wind maxima still exists at upper levels on the north side of the storm. In summary, Fig. 11 and 12 suggest that using flow dependent ensemble covariance even from the independently generated GFS ensemble improves the analysis of the storm structure. Using self-consistent HWRF ensemble provides further improvements. Assimilating TDR data using ensemble-based covariance improves both the analyzed location and storm structure. The information in TDR data is best used by Hybrid-HENS compared to other data assimilation methods.

To quantitatively compare the analyzed 3-dimensional TC structure and the HRD radar wind analysis, spatial correlation of the analyzed 3D wind speed field is computed and shown in Fig. 12. The analyzed field is relocated so that the center of the analyzed

field is consistent with that of the radar wind analysis. Such reposition is to fulfill the goal of only comparing the TC structure isolating the impact of location. Among Hybrid-HENS, Hybrid-HENS.5, Hybrid-GENS and GSI3DVar, Hybrid-HENS shows the highest correlation coefficient value of 0.71 averaged over 7 missions. GSI3DVar has the worst coefficient of 0.58.

d. *Verification of track, MSLP and maximum wind forecasts*

The track, MSLP and Vmax forecasts are verified against the best track data for all seven TDR missions for Sandy. The blue lines in figure 2 show the forecasts initialized by Hybrid-HENS. The performance of track, MSLP and Vmax forecasts by Hybrid-HENS are compared with GSI3DVar, Hybrid-GENS and Hybrid-HENS.5 for Sandy. The results (not shown) are briefly described below. Systematic comparison among different methods using all TDR cases during 2012-2013 including Sandy will be shown and discussed in details in section 5. Briefly, for Sandy, compared to GSI3DVar, Hybrid-HENS shows smaller errors in track forecasts. Hybrid-HENS however does not show consistent improvement over GSI3DVar in MSLP and Vmax forecast except at early forecast lead times. The use of consistent HWRF ensemble in Hybrid-HENS produces better track forecast compared to the use of GFS ensemble in hybrid (Hybrid-GENS) and does not show consistent improvement for MSLP and Vmax forecast. Compared to Hybrid-HENS, the use of static covariance in hybrid (Hybrid-HENS.5) improves the Vmax forecast and does not show consistent improvement for track and

MSLP forecasts. The experiments are therefore extended to other cases during 2012-2013 seasons.

5. Results for hurricanes surveyed by NOAA P3 Tail Doppler Radar during 2012-2013 seasons

During the 2012-2013 Atlantic hurricane seasons, NOAA surveyed the following hurricanes with the TDR on board of the P3 aircrafts: 9 missions for Isaac (00 UTC August 23-12 UTC August 24 in 2012; 00 UTC August 27-00 UTC August 29 in 2012), 2 missions for Leslie (12 UTC September 7 - 00 UTC September 8 in 2012), 7 missions for Sandy (00 UTC October 26 - 00 UTC October 29 in 2012), 5 missions for Ingrid (00 UTC September 14 - 00 UTC September 16 in 2013) and 4 missions for Karen (00 UTC October 4 - 12 UTC October 5 in 2013). The case details and the number of TDR data assimilated can be found in Table 2. Experiments are carried out for all these cases to further address the scientific questions proposed in section 1. The student t-test (Klotzbach and Gray, 2009) is performed to examine the statistical significance of the differences among the experiments. Significance test using the Bootstrap resampling (Wang and Bishop, 2005) is also conducted and the results (not shown) are consistent with the t-test results.

In the experiments for Sandy, for the purpose of illustrating the differences of increments from various background error covariances, each DA window is approximately centered around the middle of the penetrating leg and only the TDR data

corresponding to the penetrating leg is assimilated. Similar method was used by early studies such as Zhang et al (2011) and Weng and Zhang (2012). This approach neglects the data corresponding to the downwind legs and requires pre-determining the beginning and ending times of each flight leg which are often irregular. For systematic studies involving multiple cases or for operational implementation, using DA windows with a regular and fixed length is more convenient and ingests TDR data from all flight legs. This approach was also used by early studies such as Aksoy et al. (2012) and Aksoy (2013). Fig. 13 shows there are no statistically significant differences between the leg-by-leg assimilation and the hourly assimilation in track, Vmax and MSLP forecasts for Hybrid-HENS. Consistently, structure for Sandy (Fig. 11g and Fig. 12g) simulated in the final analyses by Hybrid-HENS-Hrly and Hybrid-HENS both compare favorably with the HRD radar wind analysis. The correlation coefficient between the analyzed structure and the HRD radar wind analysis by Hybrid-HENS-Hrly is only slightly higher than Hybrid-HENS. Therefore, in this section where systematic studies with multiple cases are performed, the TDR data is assimilated with a fixed, one-hour window, in which both penetration and downwind observations are ingested. Aksoy (2013) suggested that the storm-relative rather than earth-relative observation framework can lead to more homogenized impact of the penetrating and downwind legs and therefore potentially improve extracting the information from the TDR observations. Applying this approach may therefore further improve the analysis and forecast for Hyrid-HENS-Hrly. The

impact of the storm relative observation framework in a hybrid DA system and the optimal assimilation window length will be left for future studies.

a. Impact of inclusion of static covariance on track, MSLP and Vmax forecasts

The impact of the static covariance in the hybrid data assimilation for hurricane analysis and forecast assimilating airborne radar data is addressed. Fig. 13 shows the mean errors of forecasts of all the TDR missions during 2012-2013 seasons comparing the hybrid without static covariance (Hybrid-HENS-Hrly) and the hybrid where 50% weight is given to the static covariance (Hybrid-HENS.5-Hrly). Compared to Hybrid-HENS-Hrly, Hybrid-HENS.5-Hrly degrades the track forecasts (Fig. 13a) and MSLP forecast (Fig. 13b). For example, Hybrid-HENS-Hrly demonstrates statistically significant better track forecasts at almost all lead times except the 24 hour and better MSLP forecasts at 6-18 and 42-48 hour lead times than Hybrid-HENS.5-Hrly. However, there are more instances that Hybrid-HENS.5-Hrly produces more accurate Vmax (Fig.13c) forecasts than Hybrid-HENS-Hrly. Early studies have shown and indicated that when ensemble size was small, or model error was big or the ensemble was run at a reduced resolution, the inclusion of static covariance would improve the performance of the hybrid. These studies mostly focused on coarse-resolution or large-scale applications (e.g., Wang et al. 2007a, 2009, 2013). As discussed in section 4a, although the static covariance is tuned to improve the GSI3DVar experiments, inherent deficiencies of the static error covariances such as inappropriate wind and mass relationship for hurricane

applications can overwhelm its potential advantages in the current study. This result is consistent with Li et al. (2012) which applied a different hybrid DA method assimilating ground-based radar data for a single hurricane case.

b. Impact of flow dependent ensemble covariance and dependence of TDR data impact on DA methods

This sub-section focuses on addressing the impact of using different sources of ensemble perturbations and the dependence of the TDR data impact on different DA methods using cases during 2012-2013 seasons. Fig. 14a shows the mean error of track forecasts for NoDA, GSI3DVar-Hrly, Hybrid-HENS-Hrly and Hybrid-GENS-Hrly. Among all the experiments, hybrid using HWRF's own EnKF ensemble (Hybrid-HENS-Hrly) produces the smallest track errors. T-test shows that the improvement of Hybrid-HENS-Hrly relative to NODA, GSI3DVar-Hrly and Hybrid-GENS-Hrly is statistically significant for most of the lead times up to the 48-hour. The improvement of Hybrid-GENS-Hrly relative to GSI3DVar-Hrly is statistically significant at 0-12 and 30-36 hour lead times. Compared to GSI3DVar-Hrly, Hybrid-HENS-Hrly ingesting both the GFS ensemble and the HWRF EnKF ensemble improves the track forecasts, suggesting the positive impacts of using flow-depend ensemble covariance in data assimilation. Compared to Hybrid-GENS-Hrly, Hybrid-HENS-Hrly produces more accurate track forecasts, suggesting further improvement by ingesting the self-consistent, flow-dependent HWRF EnKF ensemble. Compared to NoDA, both Hybrid-GENS-Hrly and

Hybrid-HENS-Hrly demonstrate positive impact of assimilating the TDR data in general. In comparison, assimilating the TDR data using GSI3DVar demonstrates negative impacts after the first couple of hours of forecasts.

For the MSLP forecasts in Fig. 14b, both Hybrid-HENS-Hrly and Hybrid-GENS-Hrly produce more accurate forecasts compared to GSI3DVar-Hrly for the first 24-36 hours suggesting the need of using flow dependent covariance to assimilate inner core data to improve MSLP forecasts. The improvements by Hybrid-HENS-Hrly and Hybrid-GENS-Hrly are statistically significant at 0-18 hour and 12-24 hour lead times respectively. Ingesting HWRF EnKF ensemble demonstrates statistically significant improvement relative to ingesting GFS ensemble for the MSLP forecasts at 0-12 and 48 hour lead times and degrades the performance at 24-30 hour lead times. Like the track forecasts, TDR data assimilated by Hybrid-GENS-Hrly and Hybrid-HENS-Hrly demonstrates positive impact compared to the experiments without assimilating the TDR data. For example, Hybrid-HENS-Hrly demonstrates statistically significant better MSLP forecasts than NoDA for all lead times up to 48 hours except the 18 hour. For the maximum wind forecasts (Vmax, Fig. 14c), Hybrid-HENS-Hrly and Hybrid-GENS-Hrly produce consistently more accurate forecasts compared to GSI3DVar-Hrly after the first 18 hours, although the improvement is not as apparent or statistically significant as the track and MSLP forecasts. Assimilating the TDR data using Hybrid-HENS-Hrly and Hybrid-GENS-Hrly still in general demonstrate positive impact of the data for the Vmax

forecasts compared to not assimilating the TDR data (NoDA) with four lead times, 0, 12, 36 and 48 hour, showing statistical significance.

Following the same method of quantitatively verifying the analyzed TC structures for Sandy (2012) (section 4d), the correlation coefficients between the analyses and the HRD radar composites are calculated (Fig.15) for Hybrid-HENS-Hrly, Hybrid-GENS-Hrly and GSI3DVar-Hrly. A higher correlation coefficient indicates that the analyzed structures are more consistent with the HRD radar wind composites. For all the TDR missions during 2012-2013 seasons[1], the hybrid system ingesting the GFS ensemble shows a higher correlation (0.610) than GSI3DVar-Hrly (0.431) on average. Ingesting self-consistent HWRF EnKF ensemble, the analyzed TC structure by Hybrid-HENS-Hrly is even more consistent with the HRD radar wind composite with the Hybrid-HENS-Hrly shows a higher correlation (0.665) than the Hybrid-GENS-Hrly (0.610) on average. The statistical t-test shows that both improvements are significant at a confidence level of 99%.These results suggest the importance of using flow-dependent covariances, in particular, the consistent HWRF ensemble in assimilating inner core observations for TC analysis.

The impact of using various sources of flow dependent covariances is further explored by quantitative verifications of the first guess and analysis against independent observations. Fig. 16 shows the root mean square fit of the first guess and analysis to

[1] HRD composites for the 3rd mission of Karen 2013 and the 3rd mission of Ingrid 2013 were not available.

SFMR and flight level wind for all TDR missions during 2012-2013 seasons. The RMS fit is calculated by removing the bias following Aksoy et al. (2009). For both the fits of first guess and analysis to SFMR and flight level wind, Hybrid-GENS-Hrly improved upon GSI3DVar-Hrly. The statistical t-test shows that Hybrid-GENS-Hrly is significantly better than GSI3DVar-Hrly at a confidence level of 99% for both fist guess and analysis in both the SFMR and flight level wind verifications. Using the HWRF EnKF ensemble, Hybrid-HENS-Hrly further improves the fit to these independent observations relative to using the GFS ensemble. The statistical t-test shows that Hybrid-HENS-Hrly is significantly better than Hybrid-GENS-Hrly at a confidence level of 99% for the analysis in both the SFMR and flight level wind verifications. The confidence level for the first guess of Hybrid-HENS-Hrly is significantly more accurate than Hybrid-GENS-Hrly at 95% confidence level for the flight level wind verification and is not significantly better than Hybrid-GENS-Hrly for the SFMR verification. The inferior performance of Hybrid-GENS-Hrly relative to Hybrid-HENS-Hrly is likely because the GFS ensemble does not best reflect the error covariance in the HWRF control forecast. By design, GFS ensemble is to sample the first guess errors in the GFS data assimilation system, which used different model, different resolution and assimilated different observations compared to the HWRF system. Nevertheless, Hybrid-GENS-Hrly is still able to improve upon GSI3DVar-Hrly using the static covariance, which suggests that although there are deficiencies in using the GFS ensemble for the HWRF hybrid data assimilation system, there is still useful information within the GFS ensemble as indicated when comparing

the results from the static covariance experiments. For instances when computing resources are limited and the HWRF's own EnKF ensemble are not generated, the use of the GFS ensemble provides a quick and inexpensive fix to the static covariance.

6. Summary and Conclusion

The GSI-based hybrid EnKF-Var data assimilation system is extended for HWRF for the initialization and prediction of TCs. The hybrid method is then applied for the assimilation of airborne radar data for the first time. In this paper, methods describing the newly developed GSI-based ensemble-variational hybrid data assimilation for HWRF assimilating airborne radar observations are first described using a detailed study of Sandy (2012), followed by systematic comparison of various sensitivity experiments for multiple cases during 2012-2013 seasons. Using the newly developed system, the impact of using variously estimated background error covariances on TC core analyses and on subsequent forecasts assimilating the airborne radar observation is studied.

In the GSI-based hybrid DA system for HWRF, both the EnKF and the extended control variable components which are interfaced with GSI (Whitaker et al. 2008; Wang 2010; Wang et al. 2013) are expanded to HWRF. The EnKF ensemble covariance is incorporated into the GSI variational minimization through the use of extended control variables (Wang, 2010). Further, the data thinning method in GSI is further enhanced for the tail Doppler radar where the fore and aft sweeps are separated when thinning is performed to maintain the dual Doppler information.

The tail Doppler radar radial velocity data onboard of NOAA P-3 are assimilated. Experiments addressing the impact of various error covariances as outlined in section 1 are designed. Experiment "NoDA" is first conducted without assimilating any radar data. Experiment "GSI3DVar" adopts the GSI 3DVar where the full static covariance is used. Experiment "Hybrid-GENS" ingests flow-dependent background covariance from independently generated GFS ensemble. In comparison, in experiment "Hybrid-HENS", the hybrid DA system ingests fully flow-dependent background error covariance provided by HWRF EnKF. Experiment "Hybrid-HENS.5" incorporates half static covariance and half HWRF ensemble covariance. Finally, systematic experiments are made assimilating TDR data at hourly frequency (Hybrid-HENS-Hrly, Hybrid-HENS.5-Hrly, Hybrid-GENS-Hrly and GSI3DVar-Hrly).

Although only radar wind is assimilated, detailed study for Sandy shows that Hybrid-HENS, Hybrid-GENS, using flow-dependent ensemble covariance, is able to adjust both the wind and mass field to match the best track position in a more dynamically and thermodynamically coherent fashion. Using the static covariance, adjustment of mass and momentum field by GSI3DVar follows more in the line of large scale flow with approximate geostrophic balance, which is not suitable for TC inner core. The wind and pressure relation in the covariances derived from the GFS ensemble improves upon that of the static covariance, but is still inferior compared to that of HWRF.

Verification of first guesses and analyses against independent flight level and

SFMR observations is made for both Sandy and other TDR cases during 2012-2013 seasons. Using the GFS ensemble in data assimilation, first guesses and analyses from the hybrid system fit the independent observations much closer than 3DVar using static covariance. Using HWRF's own EnKF ensemble, fits of first guesses and analyses from the hybrid system to SFMR and flight level wind data are improved further compared to using the GFS ensemble.

The structure of the TC simulated by different data assimilation methods is further examined by comparing the analyses at the end of the data assimilation cycles with the radar wind analyses produced by the NOAA Hurricane Research Division (HRD) for both Sandy and other TDR cases during 2012-2013 seasons. Hybrid-HENS using both GFS ensemble and HWRF's own EnKF ensemble better correlates with the HRD radar wind analysis compared to 3DVar. Hybrid-HENS ingesting HWRF's own EnKF ensemble compares more closely with the HRD radar wind analysis than the hybrid ingesting GFS ensemble. Examining the results of Sandy in details shows that 3DVar depicts multiple spurious wind maxima, deviating from the radar wind analysis the most. For the hybrid ingesting GFS ensemble, there is a secondary minimum in the pressure analysis. Further experiments and diagnostics show that such unrealistic pattern is likely due to the inconsistency of using GFS ensemble to estimate the HWRF background forecast error covariance.

Track, MSLP and Vmax forecasts are verified against the best track data for TDR cases during 2012-2013 seasons. Compared to 3DVar, the hybrid ingesting both the GFS

ensemble and the HWRF EnKF ensemble improves the track forecasts, suggesting the positive impacts of using flow-depend ensemble covariance in data assimilation. Compared to hybrid using GFS ensemble, hybrid using HWRF EnKF ensemble produces more accurate track forecasts, suggesting further improvement by ingesting the self-consistent, flow-dependent HWRF EnKF ensemble. For the MSLP forecasts, hybrid using both GFS ensemble and HWRF EnKF ensemble produces more accurate forecasts compared to GSI3DVar for the first 24-36 hours suggesting the need of using flow dependent covariance to assimilate inner core data to improve MSLP forecasts. Compared to ingesting GFS ensemble, ingesting HWRF EnKF ensemble improves the MSLP forecasts at 0-12 and 42-48 hours but degrades the forecast at 24-36 hours. For the maximum wind forecasts, hybrid using both GFS ensemble and HWRF EnKF ensemble produces consistently more accurate forecasts compared to 3DVar after the first 18 hours, although the improvement is not as apparent as the track and MSLP forecasts.

Detailed study of Sandy reveals that compared to GSI3DVar, the hybrid system incorporating the flow dependent HWRF EnKF ensemble with a 50% weight depicts more realistic analysis increments, better fit to flight level and SFMR wind and compared more favorably with the HRD radar wind composite. The same configuration also produces better track, MSLP and Vmax forecasts compare to GSI3DVar for all TDR cases during 2012-2013 seasons. Blending the static covariance with the ensemble covariance does not show apparent improvement in these verifications over the experiment where only pure ensemble is used except that there are more instances that

the blending produces more accurate Vmax.

The impacts of assimilating the TDR data at hourly frequency (named Hybrid-HENS-Hrly) are compared with the leg by leg assimilation (Hybrid-HENS). For the hourly assimilation, the data falling into the hourly window are included for assimilation and therefore there is no need to determine the beginning and ending times of the flight leg. Although additional downwind-leg information is added in the hourly assimilation, no distinct differences between the leg-by-leg assimilation and the hourly assimilation are found in track, Vmax and MSLP forecasts for Hybrid-HENS. This could be attributed to the lack of efficient usage of the downwind-leg information. The "storm-relative observation strategy" proposed by HRD can be a potential tool to help further improve the usage of TDR data.

Diagnostics of analysis increment, verification against independent flight level and SFMR wind observations and HRD radar wind composite reveals that the impact of TDR observations can be highly dependent on the type of background error covariance. Using self-consistent HWRF EnKF ensemble is able to best extract the information to produce most accurate storm structure and therefore produce the largest amount of positive impact of the TDR data on the analysis and forecast. In comparison, traditional GSI 3DVar method shows negative impact of assimilating the TDR data in analyzing the structure of the TC and demonstrates negative impacts after the first couple of hours of forecasts.

Research and development are ongoing to include other observational data in

addition to the TDR data and to further extend the system with higher resolution moving nests and dual resolution DA capabilities, which will be documented in future papers. This study, using the new hybrid DA system, focuses on exploring the impact of various error covariances on TC core analysis and on subsequent forecasts. Comparison of different ensemble data assimilation algorithms such as the hybrid and the pure EnKF for hurricane prediction requires carefully designed experiments, which is an interesting topic for future studies.

Acknowledgement

The research documented in this paper was supported by grants NA12NWS4680012, NA14NWS4830008 and NA14NWS4680021 from the National Ocean and Atmosphere Administration (NOAA) Hurricane Forecast Improvement Project (HFIP) project and the ONR grant N00014-14-1-0125. The experiments were performed on the NOAA supercomputer Jet. The first and fifth authors were also partially supported by China National Science Foundation 41275111. Henry Winterbottom and Jeff Whitaker are acknowledged for initial discussion at early stage of the experiment.

References

Aberson SD, Black ML, Black RA, Cione JJ, Landsea CW, Marks Jr FD, Burpee RW. 2006. Thirty years of tropical cyclone research with the NOAA P-3 aircraft. *Bull. Amer. Meteor. Soc.,* **87,** 1039-1055.

Aksoy A, Aberson SD, Vukicevic T, Sellwood KJ, Lorsolo S, Zhang X. 2013. Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using NOAA/AOML/HRD's HEDAS: Evaluation of the 2008–11 vortex-scale analyses. *Mon. Wea. Rev.,* **141,** 1842-1865.

Aksoy A, Dowell DC, Snyder C. 2009. A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.,* **137(6),** 1805-1824.

Aksoy A, Lorsolo S, Vukicevic T, Sellwood KJ, Aberson SD, Zhang F. 2012. The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for high-resolution data: The impact of airborne Doppler radar observations in an OSSE. *Mon. Wea. Rev.,* **140,** 1843-1862.

Aksoy A. 2013. Storm-relative observations in tropical cyclone data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.,* **141**, 506-522.

Anderson JL, Collins N. 2007. Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.,* **24,** 1452-1463.

Bishop CH, Hodyss D. 2011. Adaptive ensemble covariance localization in ensemble 4d-var state estimation. *Mon. Wea. Rev.,* **139,** 1241-1255.

Buehner M. 2005. Ensemble-derived stationary and flow-dependent background-error covariances: evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.,* **131,** 1013-1043

Buehner M, Houtekamer PL, Charette C, Mitchell HL, He B. 2010a. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.,* **138,** 1550-1566.

Buehner M, Houtekamer PL, Charette C, Mitchell HL, He B. 2010b. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Wea. Rev.,* **138**, 1567-1586.

Clayton AM, Lorenc AC, Barker DM. 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q.J.R. Meteorol. Soc.,* **139,** 1445-1461.

Dowell DC, Zhang F, Wicker LJ, Snyder C, Crook NA. 2004. Wind and Temperature Retrievals in the 17 May 1981 Arcadia, Oklahoma, Supercell: Ensemble Kalman Filter Experiments. *Mon. Wea. Rev.,* **132(8),** 1982–2005.

Etherton BJ, Bishop CH. 2004. Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev.,* **132,** 1065-1080.

Ferrier BS. 2005. An efficient mixed-phase cloud and precipitation scheme for use in Operational NWP Models. *EOS Trans. AGU,* **86,** Jt. Assem. Suppl., Abstract A42A-02.

Fels SB, Schwarzkopf MD. 1975. The simplified exchange approximation: A new method for radiative transfer calculations. *J. Atmos. Sci.*, **32,** 1475-1488.

Gamache JF. 2005. `Final report on JHT project entitled: Real-Time Dissemination of Hurricane Wind Fields Determined from Airborne Doppler Radar`. Available online at http://www.nhc.noaa.gov/jht/2003-2005reports/DOPLRgamache_JHTfinalreport.pdf

Gaspari G, Cohn SE. 1999. Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,* **125,** 723-757.

Hamill TM, Snyder C. 2000. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.,* **128,** 2905-2919.

Hamill TM, Whitaker JS, Fiorino M, Benjamin SG. 2011. Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.,* **139,** 668-688.

Han J, Pan HL. 2011. Revision of convection and vertical diffusion schemes in the NCEP global forecast system. *Wea. Forecasting,* **26,** 520-533.

Hayden CM, Purser RJ. 1995. Recursive filter objective analysis of meteorological fields: Applications to NESDIS operational processing. *J. Appl. Meteor.,* **34,** 3-15.

Hong SY, Pan HL. 1996. Nonlocal Boundary Layer Vertical Diffusion in a Medium-Range Forecast Model. *Mon. Wea. Rev.,* **124,** 2322-2339.

Johnson A, Wang X, Carley JR, Wicker LJ, Karstens C. 2015. A Comparison of Multiscale GSI-Based EnKF and 3DVar Data Assimilation Using Radar and Conventional Observations for Midlatitude Convective-Scale Precipitation Forecasts. *Mon. Wea. Rev.,***143**, 3087–3108. http://doi.org/10.1175/MWR-D-14-00345.1

Klotzbach PJ, Gray WM. 2009. Twenty-five years of Atlantic basin seasonal hurricane forecasts (1984-2008). *Geophysical Research Letters*, **36**(9), 1–5.

Kutty G, Wang X. 2015. A Comparison of the Impacts of Radiosonde and AMSU Radiance Observations in GSI Based 3DEnsVar and 3DVar Data Assimilation Systems for NCEP GFS. *Advances in Meteorology*, vol. 2015, Article ID 280546, 17 pages, 2015. doi:10.1155/2015/280546

Lacis AA, Hansen J. 1974. A parameterization for the absorption of solar radiation in the earth's atmosphere. *J. Atmos. Sci.,* **31,** 118-133.

Li Y, Wang X, Xue M. 2012. Assimilation of radar radial velocity data with the WRF hybrid ensemble–3DVAR system for the prediction of Hurricane Ike (2008). *Mon. Wea. Rev.,* **140**, 3507-3524.

Li Y. 2015. Assimilation of radar observations with ensemble variational hybrid data assimilation method for the initialization and prediction of hurricanes`, Doctoral dissertation. Available from

https://shareok.org/bitstream/handle/11244/14050/2015_Li_Yongzuo_Dissertation.pdf?sequence=2

Lorenc AC. 2003. The potential of the ensemble Kalman filter for NWP—a comparison with 4D‐Var. *Quart. J. Roy. Meteor. Soc.,* **129,** 3183-3203.

Marks Jr FD, Houze Jr RA. 1984. Airborne doppler radar observations in Hurricane Debby. *Bull. Amer. Meteor. Soc.* **65,** 569-582.

Marks Jr FD, Houze Jr RA. 1987. Inner core structure of Hurricane Alicia from airborne Doppler radar observations. *J. Atmos. Sci.,* **44,** 1296-1317.

Marks Jr FD, Houze Jr RA, Gamache JF. 1992. Dual-aircraft investigation of the inner core of Hurricane Norbert. Part I: Kinematic structure. *J. Atmos. Sci.,* **49,** 919-942.

Pan Y, Zhu K, Xue M, Wang X, Hu M, Benjamin SG, Weygrant SS, Whitaker JS. 2014. A GSI-Based Coupled EnSRF–En3DVar Hybrid Data Assimilation System for the Operational Rapid Refresh Model: Tests at a Reduced Resolution. *Mon. Wea. Rev.*, **142**, 3756–3780.

Pielke Jr RA, Gratz J, Landsea CW, Collins D, Saunders MA, Musulin R. 2008. Normalized hurricane damage in the United States: 1900–2005. *Natural Hazards Rev.,* **9,** 29-42.

Poterjoy J, Zhang F. 2014. Intercomparison and Coupling of Ensemble and Four-Dimensional Variational Data Assimilation Methods for the Analysis and Forecasting of Hurricane Karl (2010). *Mon. Wea. Rev.*, **142**, 3347–3364.

Rogers R and coauthors. 2013. NOAA'S hurricane intensity forecasting experiment: A progress report. *Bull. Amer. Meteor. Soc.*, **94(6)**, 859–882. http://doi.org/10.1175/BAMS-D-12-00089.1

Schwarzkopf MD, Fels SB. 1991. The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes. *J. Geophys. Res,* **96,** 9075-9096.

Tong M, Tallapragada V, Liu E, Kieu C, Kwon I, Wang W, Liu Q, Zhang B. 2016. Impact of Assimilating Aircraft Reconnaissance Observations on Tropical Cyclone Initialization and Prediction using Operational HWRF and GSI Ensemble-Variation Hybrid Data Assimilation. *Mon. Wea. Rev*. submitted.

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

Tuleya RE. 1994. Tropical storm development and decay: Sensitivity to surface boundary conditions. *Mon. Wea. Rev.,* **122,** 291-304.

Weng Y, Zhang F. 2012. Assimilating airborne Doppler radar observations with an ensemble Kalman filter for convection-permitting hurricane initialization and prediction: Katrina (2005). *Mon. Wea. Rev.,* **140,** 841-859.

Wu WS, Purser RJ, Parrish DF. 2002. Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.,* **130,** 2905-2916.

Wang X. 2010. Incorporating ensemble covariance in the Gridpoint Statistical Interpolation variational minimization: A mathematical framework. *Mon. Wea. Rev.,* **138,** 2990-2995.

Wang X. 2011. Application of the WRF hybrid ETKF–3DVAR data assimilation system for hurricane track forecasts. *Wea. Forecasting,* **26,** 868–884.

Wang X, Bishop CH. 2005. Improvement of ensemble reliability with a new dressing kernel. *Quart. J. Roy. Meteor. Soc.*, **131**(607), 965–986.

Wang X, Barker D, Snyder C, Hamill TM. 2008a. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.,* **136,** 5116–5131.

Wang X, Barker D, Snyder C, Hamill TM. 2008b. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiments.*Mon. Wea. Rev.,* **136,** 5132–5147.

Wang X, Hamill TM, Whitaker JS, Bishop CH. 2007a. A comparison of hybrid ensemble transform Kalman filter-optimum interpolation and ensemble square root filter analysis schemes. *Mon. Wea. Rev.,* **135,** 1055-1076.

Wang X, Hamill TM, Whitaker JS, Bishop CH. 2009. A comparison of the hybrid and EnSRF analysis schemes in the presence of model errors due to unresolved scales. *Mon. Wea. Rev.,* **137,** 3219-3232.

Wang X, Lei T. 2014. GSI-based four dimensional ensemble-variational (4DEnsVar) data assimilation: formulation and single resolution experiments with real data for NCEP Global Forecast System. *Mon. Wea. Rev.*, **142,** 3303-3325.

Wang X, Parrish D, Kleist D, Whitaker J. 2013. GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.,* **141,** 4098-4117.

Wang X, Snyder C, Hamill TM. 2007b. On the theoretical equivalence of differently proposed ensemble-3DVAR hybrid analysis schemes. *Mon. Wea. Rev.,* **135,** 222-227.

Whitaker JS, Hamill TM. 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130,** 1913-1924.

Whitaker JS, Hamill TM. 2012. Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.,* **140,** 3078-3089.

Whitaker JS, Hamill TM, Wei X, Song Y, Toth Z. 2008. Ensemble Data Assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482.

Xiao Q, Zhang X, Davis C, Tuttle J, Holland G, Fitzpatrick PJ. 2009. Experiments of hurricane initialization with airborne Doppler radar data for the Advanced Research Hurricane WRF (AHW) model. *Mon. Wea. Rev.,* **137,** 2758-2777.

Xu Q, Gong J. 2003. Background error covariance functions for Doppler radial‐wind analysis. *Quart. J. Roy. Meteor. Soc.,* **129,** 1703-1720.

Zhang F, Weng Y, Sippel JA, Meng Z, Bishop CH. 2009. Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.,* **137,** 2105-2125.

Zhang M, Zhang F. 2012. E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. *Mon. Wea. Rev.,* **140,** 587-600.

Table captions

Table 1 List of experiments and their descriptions.

Table 2 List of cases and mission times and the TDR data assimilated at each times.

Figure captions

Fig. 1 Flow chart of GSI based EnKF-Variational hybrid data assimilation system for HWRF (adapted from Wang et al. 2013).

Fig. 2 a) Track, b) minimum sea level pressure and c) maximum wind obtained from best track data for Sandy (2012) during October 26 – October 31 in 2012 (Black). Red lines are operational HWRF forecasts initialized every 12 hours from 00UTC October 26 to 00UTC October 29 in 2012 corresponding to each TDR mission. Blue lines are corresponding Hybrid-HENS forecasts. Dash lines show the forecasts for each TDR mission and solid lines show the average of forecasts valid at the same time. The starting and ending point of the forecast from each mission is marked with the mission number.

Fig. 3 Flight tracks (blue line) and horizontal distribution of airborne radar data (grey dot) for 7 NOAA P3 tail Doppler radar missions. The black line is the best track from NHC.

Fig. 4 An example of a Tail Doppler Radar fore scan sweep. The radar is pointing to the east. U, N, D and S stand for the upward, northward, downward (ocean surface) and southward directions, respectively. The radar scans counterclockwise from U to N to D to

S and back to U. The negative values in blue mean wind towards the radar, and the positive values in red mean wind away from the radar. The radar is scanning a southerly wind.

Fig. 5 Vertical distribution of the number of Tail Doppler Radar data collected during the first NOAA P3 mission of Sandy around 00 UTC on Oct. 26th, 2012.

Fig. 6 The flow charts for a) NoDA, b) GSI3DVar, c) Hybrid-HENS, Hybrid-HENS.5, d) Hybrid-GENS, e) Hybrid-HENS-Hrly, Hybrid-HENS.5-Hrly f) Hybrid-GENS-Hrly and g) GSI3DVar-Hrly experiments.

Fig. 7 1000hPa wind speed (shaded) and SLP (contour) for a) background first guess, and analysis after assimilating the first penetration leg of TDR data for b) GSI3DVar, c) Hybrid-HENS, d) Hybrid-GENS and e) Hybrid-HENS.5 experiments. The blue dot is the best track storm location at 2200Z on Oct. 25th. The black line denotes the corresponding flight track associated with the first assimilated penetration leg.

Fig. 8 The analysis increment of a) ~ d) wind (purple vector and shaded, ms^{-1}) and e) ~ h) geopotential height (shaded, m) at 700hPa by assimilating a meridional wind observation at the black dot position. The observation is 5 ms^{-1} larger than the background state. The black contour indicates the background sea level pressure. The analysis increment is computed using a), e) GSI3DVar, b), f) Hybrid-HENS, c), g)

Hybrid-GENS, d), h) Hybrid-HENS.5.

Fig. 9 SFMR wind speed observation (black) along the flight track during the first penetration leg of TDR observation. The x axis is the distance from the starting point of the leg. The green line is the first guess and the red line is the analysis along the same flight track for a) GSI3DVar, b) Hybrid-HENS, c) Hybrid-GENS, d) Hybrid-HENS.5 experiments. The values in the parenthesis indicate the RMS fit of the analysis and first guess to the observation.

Fig. 10 Same as Fig 9 except for assimilating the last penetration leg of TDR data. Results for NoDA experiment is also included in e).

Fig. 11 Wind (shaded and vector) and pressure (contour) at 1km height for a) HRD radar wind composite, b) GSI3DVar, c) NoDA, d) Hybrid-HENS, e) Hybris-GENS, f) Hybrid-HENS.5 and g) Hybrid-HENS-Hrly. Black dot is the interpolated best track position. The blue line denotes the corresponding flight track associated with the last assimilated penetration leg. The red line denotes the corresponding cross-sections to be presented in Fig. 12.

Fig. 12 Wind (shaded), potential temperature (solid line) and relative humidity (dash line) of the south to north vertical cross section for a) HRD radar wind composite, b)

GSI3DVar, c) NoDA, d) Hybrid-HENS, e) Hybris-GENS, f) Hybrid-HENS.5 and g) Hybrid-HENS-Hrly.

Fig. 13 a) Track, b) SLP and c) Vmax Mean Error statistics for Hybrid-HENS (green solid), Hybrid-HENS.5-Hrly (light blue dash) and Hybrid-HENS-Hrly(dark blue solid) experiments during 2012-2013 seasons. The light blue dots above the x axis denote that the Hybrid-HENS-Hrly is significantly better than Hybrid-HENS.5-Hrly at a confidence level greater than 90%. The light blue crosses above the x axis denote that the Hybrid-HENS.5-Hrly is significantly better than Hybrid-HENS-Hrly at a confidence level greater than 90%.

Fig. 14 a) Track, b) SLP and c) Vmax Mean Error statistics for Hybrid-HENS-Hrly (blue solid), Hybrid-GENS-Hrly (green dash), NoDA (orange solid) and GSI3DVar-Hrly (light blue solid) experiments during 2012-2013 seasons. The light blue, orange, and green dots above the x axis denote that the Hybrid-HENS-Hrly is significantly better than GSI3DVar-Hrly, NoDA, Hybrid-GENS-Hrly at a confidence level greater than 90%, respectively. The green crosses right above the x axis denote that the Hybrid-GENS is significantly better than GSI3DVar at a confidence level greater than 90%.

Fig. 15 Correlation coefficient values for all missions during 2012-2013 seasons for Hybrid-HENS-Hrly (Blue), GSI3DVar-Hrly (Red) and Hybrid-GENS-Hrly (Purple)

against HRD composite. The values in the parenthesis denote the average values of all cases for each experiment.

Fig. 16 Independent verification of analyses (blue) and first guess (red) against a) SFMR and b) Flight Level data for experiments during 2012-2013 seasons for Hybrid-HENS-Hrly, GSI3DVar-Hrly and Hybrid-GENS-Hrly experiments.

Table 1 List of experiments and their descriptions.

Experiment name	Description
NoDA	No airborne radar data assimilation. HWRF forecast initialized from GFS analysis.
GSI3DVar	Assimilate airborne radar data using GSI 3DVar with static covariance.
Hybrid-HENS	Assimilate airborne radar data using the newly developed self-consistent EnKF-Var data assimilation system for HWRF. HWRF ensemble initialized from HWRF's EnKF was ingested in the hybrid. A full weight was given to the ensemble covariance.
Hybrid-HENS.5	Same as Hybrid-HENS except equal weights were given to the static and ensemble covariances.
Hybrid-GENS	Assimilate airborne radar data using the HWRF hybrid ingesting GFS ensemble.
Hybrid-HENS-Hrly	Same as Hybrid-HENS except assimilating airborne radar data at hourly frequency and including both penetration and downwind leg data.
Hybrid-HENS.5-Hrly	Same as Hybrid-HENS.5 except assimilating airborne radar data at hourly frequency.
GSI3DVar-Hrly	Same as GSI3DVar except assimilating airborne radar data at hourly frequency
Hybrid-GENS-Hrly	Same as Hybrid-GENS except assimilating airborne radar data at hourly frequency.

Table 2 List of cases and mission times and the TDR data assimilated at each times.

Storm name	Mission time	Number of TDR assimilated	Storm name	Mission time	Number of TDR assimilated	Storm name	Mission time	Number of TDR assimilated
Isaac	00 UTC 23 Aug 2012	45727	Leslie	12 UTC 07 Sep 2012	7996	Ingrid	00 UTC 14 Sep 2013	19618
	12 UTC 23 Aug 2012	45940		00 UTC 08 Sep 2012	14425		12 UTC 14 Sep 2013	26472
	00 UTC 24 Aug 2012	19730	Sandy	00 UTC 26 Oct 2012	48023		00 UTC 15 Sep 2013	14157
	12 UTC 24 Aug 2012	11717		12 UTC 26 Oct 2012	22700		12 UTC 15 Sep 2013	16003
	00 UTC 27 Aug 2012	28575		00 UTC 27 Oct 2012	4703		00 UTC 16 Sep 2013	19124
	12 UTC 27 Aug 2012	49341		12 UTC 27 Oct 2012	8429	Karen	00 UTC 04 Oct 2013	15552
	00 UTC 28 Aug 2012	42300		00 UTC 28 Oct 2012	34545		12 UTC 04 Oct 2013	12974
	12 UTC 28 Aug 2012	50582		12 UTC 28 Oct 2012	30596		00 UTC 05 Oct 2013	13163
	00 UTC 29 Aug 2012	20162		00 UTC 29 Oct 2012	30973		12 UTC 05 Oct 2013	113

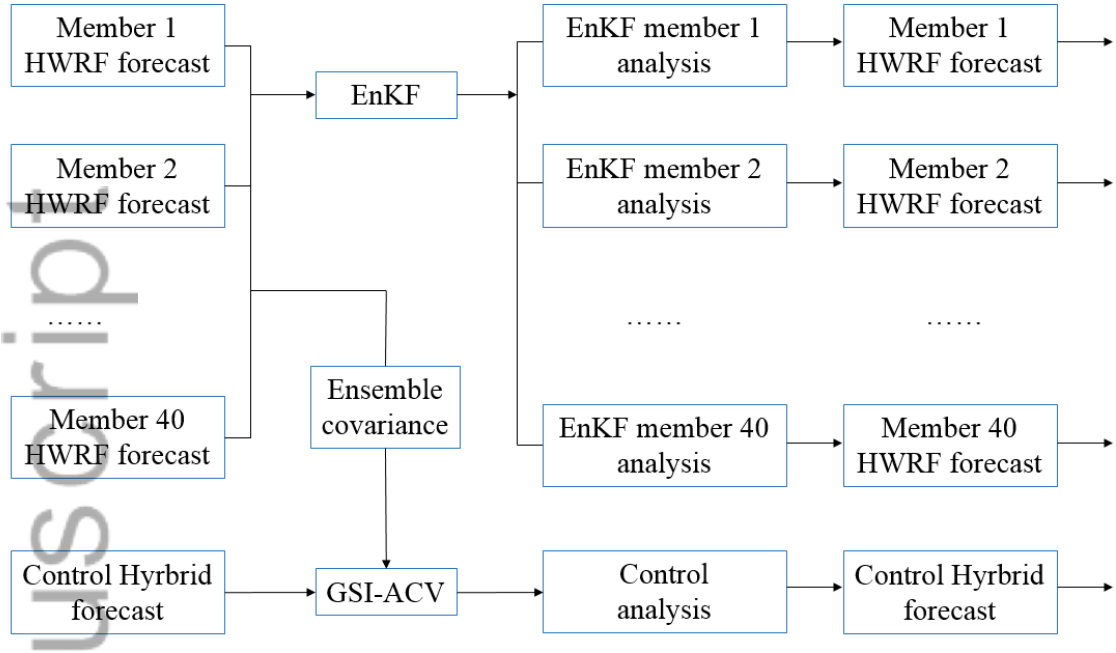

Fig. 1 Flow chart of GSI based EnKF-Variational hybrid data assimilation system for HWRF (adapted from Wang et al. 2013).

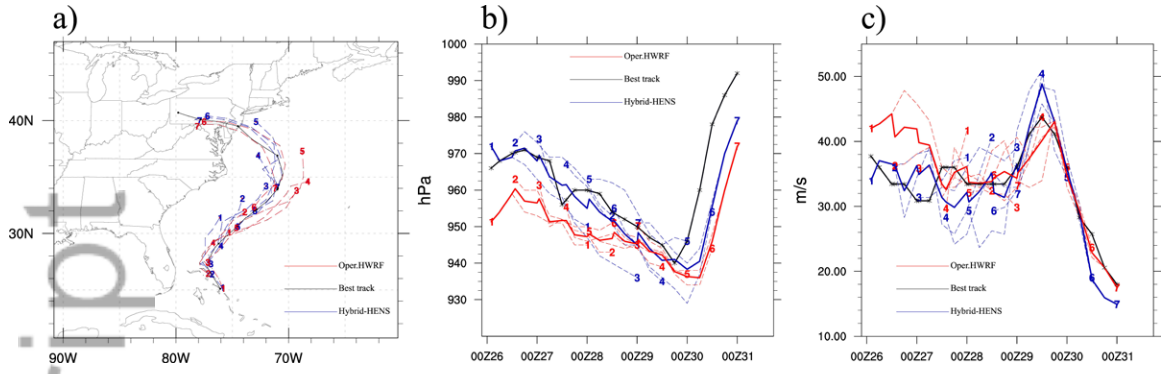

Fig. 2 a) Track, b) minimum sea level pressure and c) maximum wind obtained from best track data for Sandy (2012) during October 26 – October 31 in 2012 (Black). Red lines are operational HWRF forecasts initialized every 12 hours from 00UTC October 26 to 00UTC October 29 in 2012 corresponding to each TDR mission. Blue lines are corresponding Hybrid-HENS forecasts. Dash lines show the forecasts for each TDR mission and solid lines show the average of forecasts valid at the same time. The starting and ending point of the forecast from each mission is marked with the mission number.

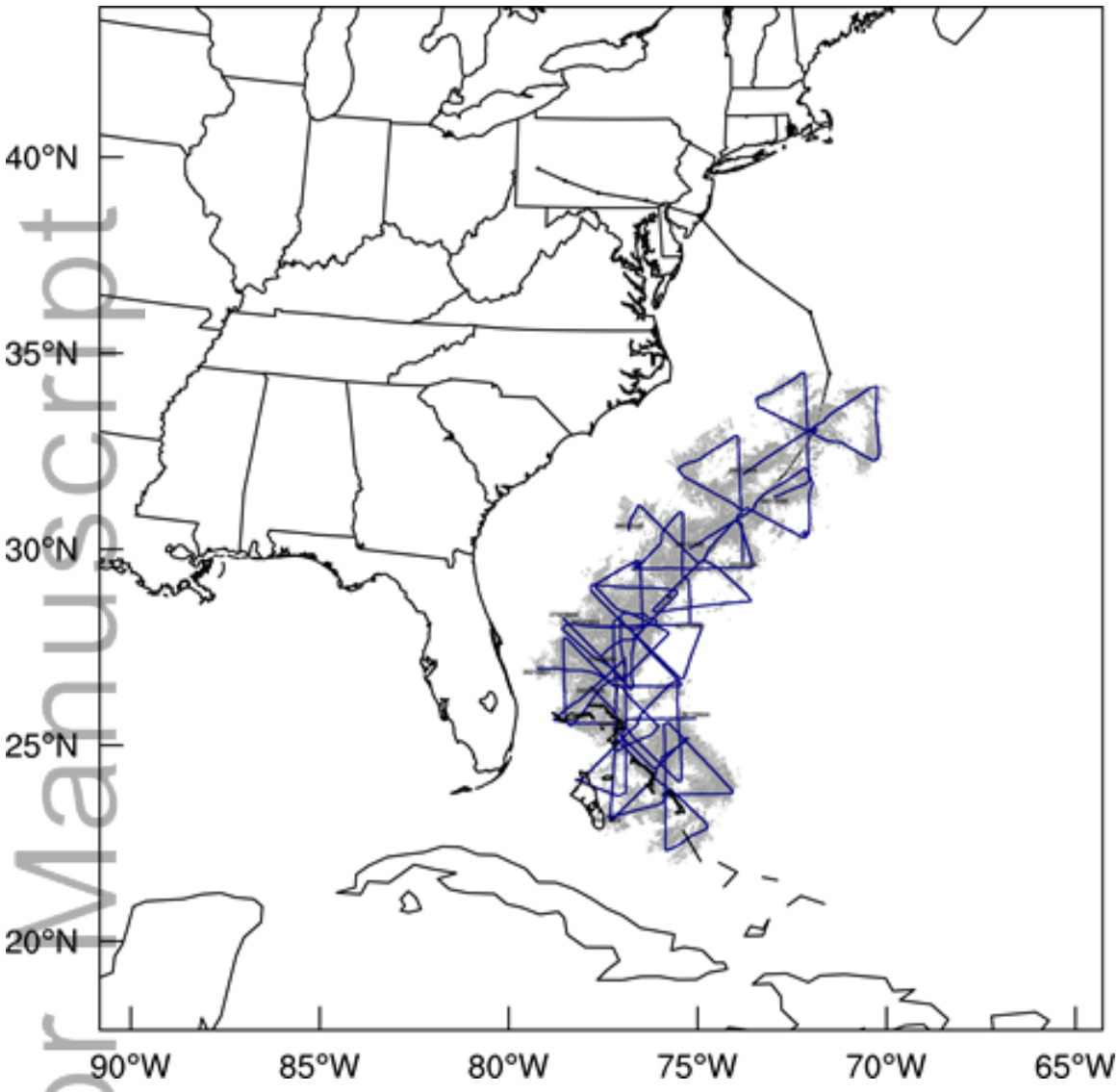

Fig. 3 Flight tracks (blue line) and horizontal distribution of airborne radar data (grey dot) for 7 NOAA P3 tail Doppler radar missions. The black line is the best track from NHC.

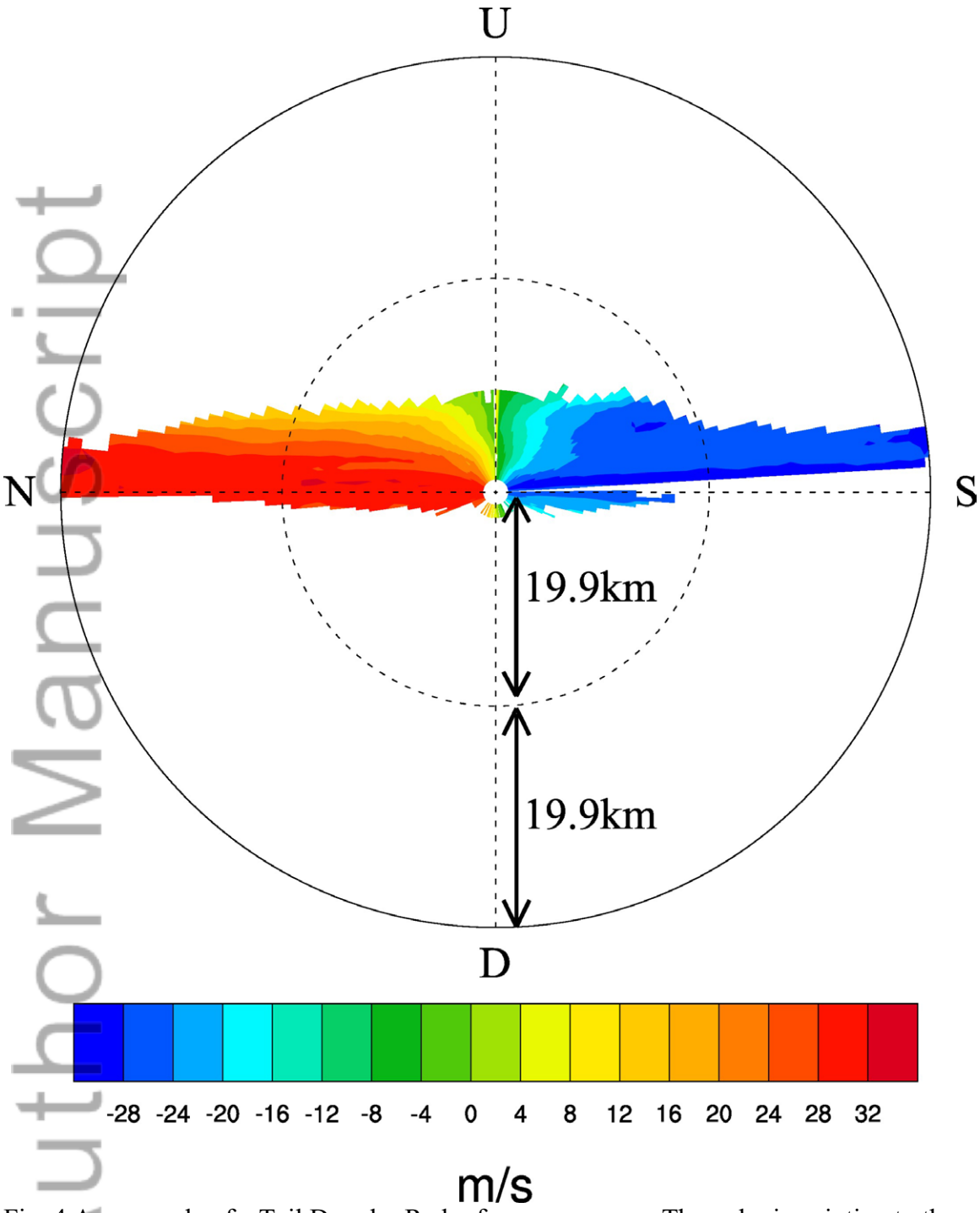

Fig. 4 An example of a Tail Doppler Radar fore scan sweep. The radar is pointing to the east. U, N, D and S stand for the upward, northward, downward (ocean surface) and southward directions, respectively. The radar scans counterclockwise from U to N to D to

S and back to U. The negative values in blue mean wind towards the radar, and the positive values in red mean the wind away from the radar. The radar is scanning a southerly wind.

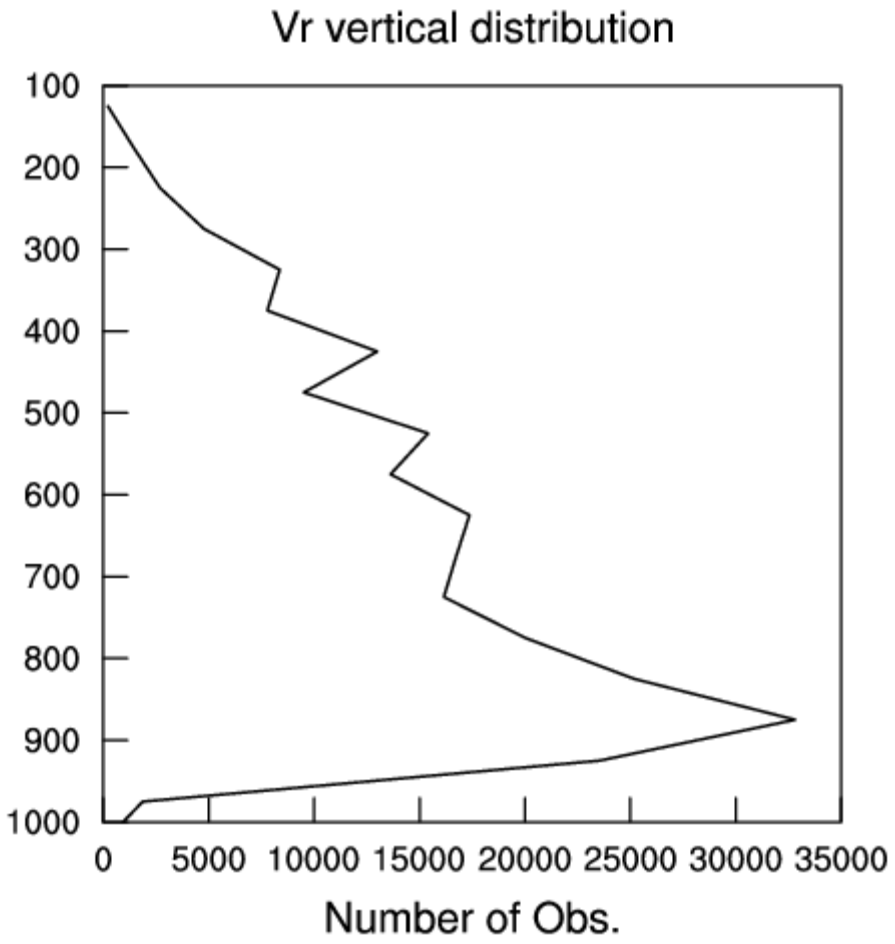

Fig. 5 Vertical distribution of the number of Tail Doppler Radar data collected during the first NOAA P3 mission of Sandy around 00 UTC on Oct. 26th, 2012.

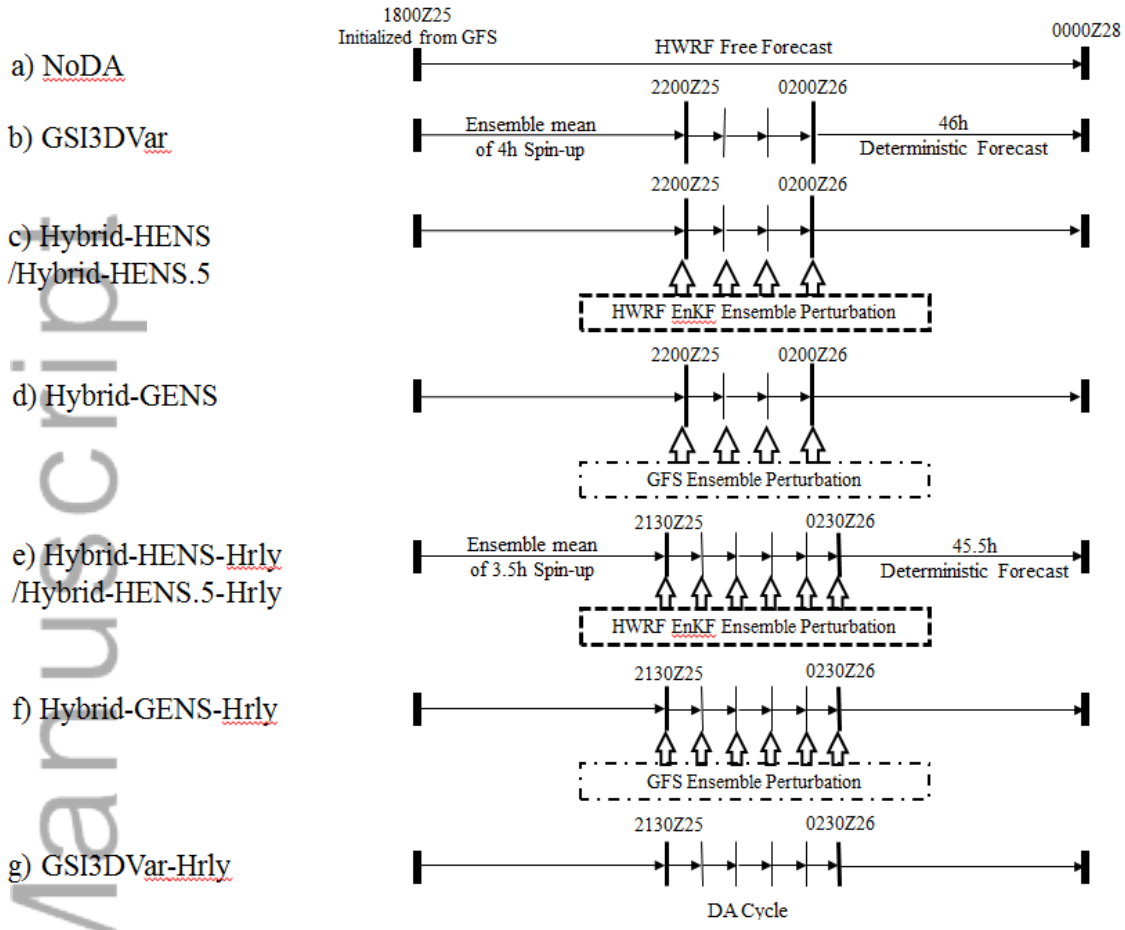

Fig. 6 The flow charts for a) NoDA, b) GSI3DVar, c) Hybrid-HENS, Hybrid-HENS.5, d) Hybrid-GENS, e) Hybrid-HENS-Hrly, Hybrid-HENS.5-Hrly, f) Hybrid-GENS-Hrly and g) GSI3DVar-Hrly experiments.

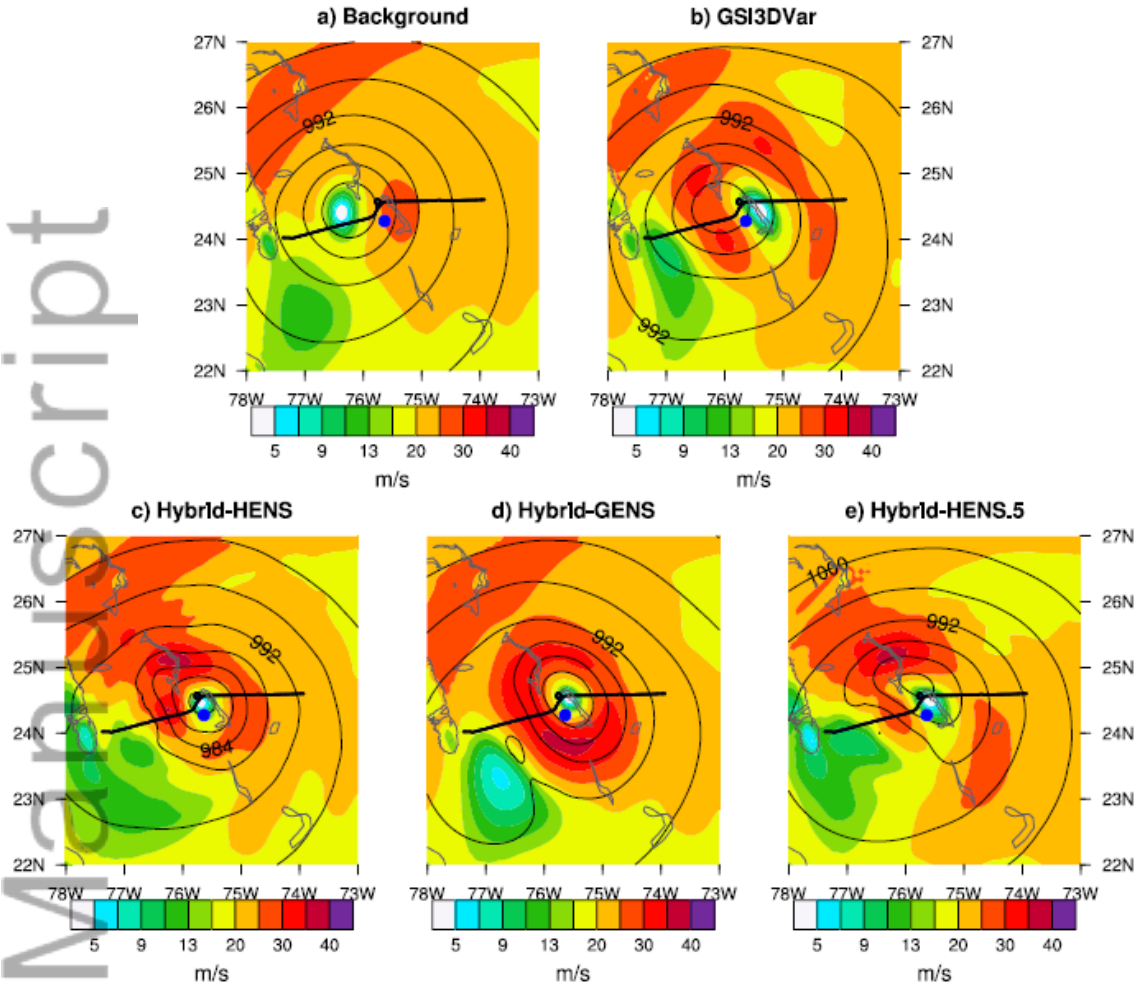

Fig. 7 1000hPa wind speed (shaded) and SLP (contour) for a) background first guess, and analysis after assimilating the first penetration leg of TDR data for b) GSI3DVar, c) Hybrid-HENS, d) Hybrid-GENS and e) Hybrid-HENS.5 experiments. The blue dot is the best track storm location at 2200Z on Oct. 25th. The black line denotes the corresponding flight track associated with the first assimilated penetration leg.

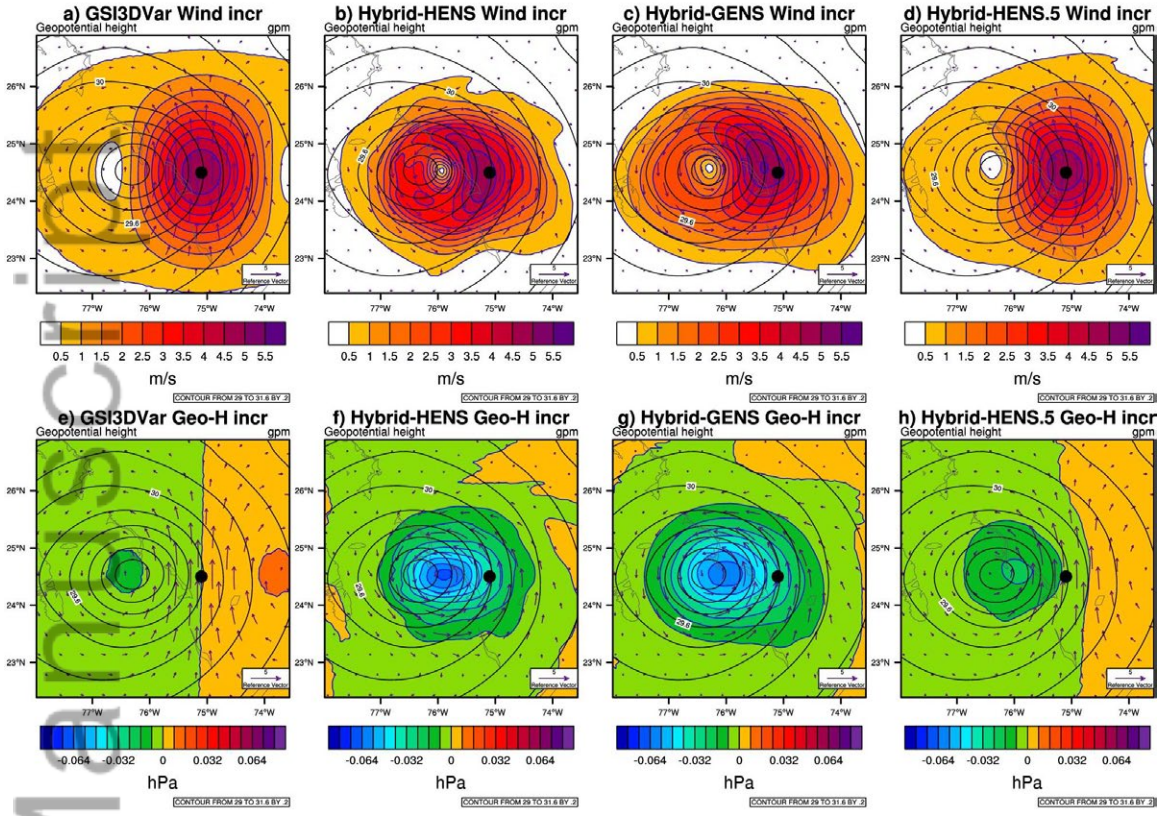

Fig. 8 The analysis increment of a) ~ d) wind (purple vector and shaded, ms^{-1}) and e) ~ h) geopotential height (shaded, m) at 700hPa by assimilating a meridional wind observation at the black dot position. The observation is 5 ms^{-1} larger than the background state. The black contour indicates the background sea level pressure. The analysis increment is computed using a), e) GSI3DVar, b), f) Hybrid-HENS, c), g) Hybrid-GENS, d), h) Hybrid-HENS.5.

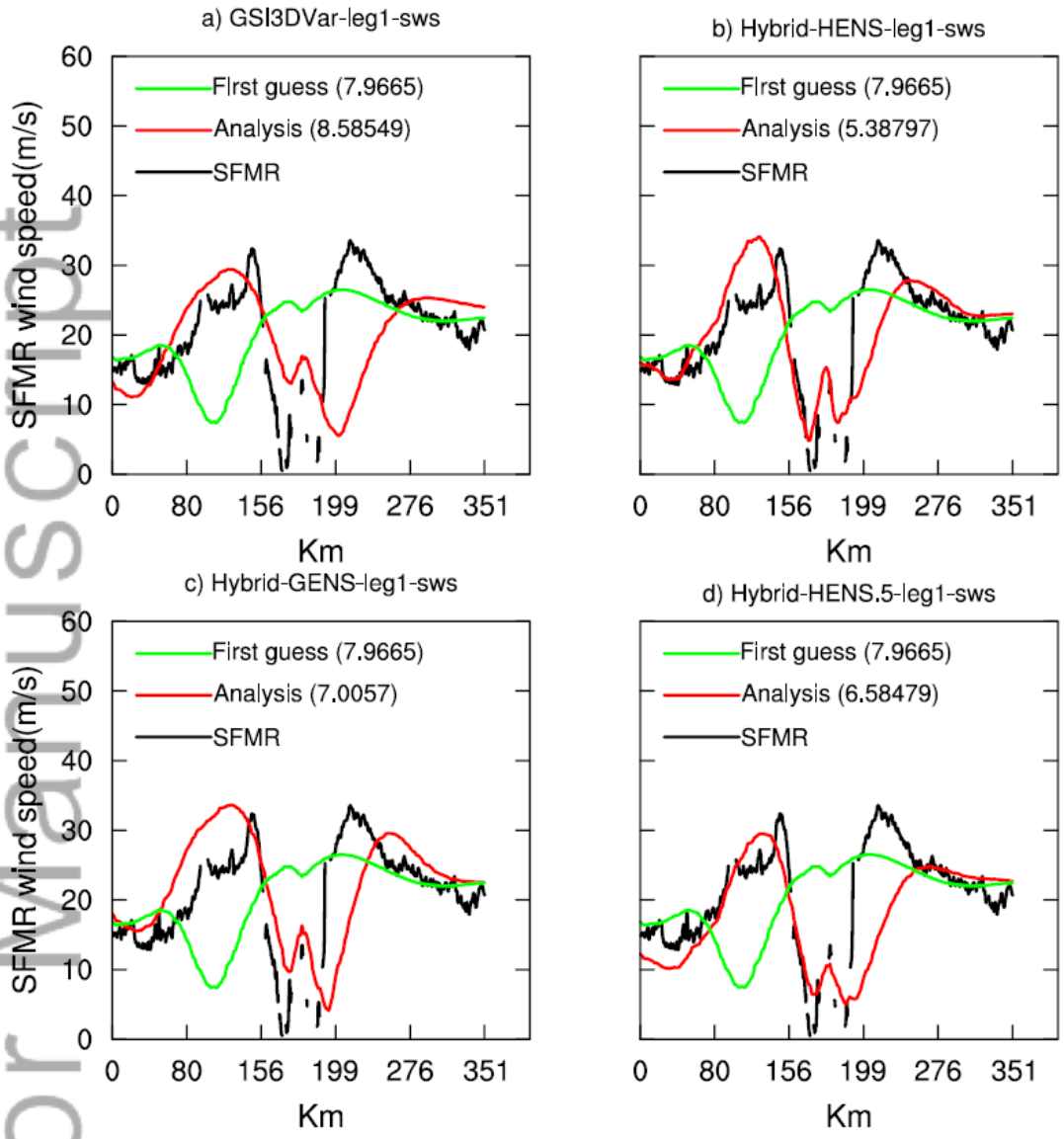

Fig. 9 SFMR wind speed observation (black) along the flight track during the first penetration leg of TDR observation. The x axis is the distance from the starting point of the leg. The green line is the first guess and the red line is the analysis along the same flight track for a) GSI3DVar, b) Hybrid-HENS, c) Hybrid-GENS, d) Hybrid-HENS.5 experiments. The values in the parenthesis indicate the RMS fit of the analysis and first guess to the observation.

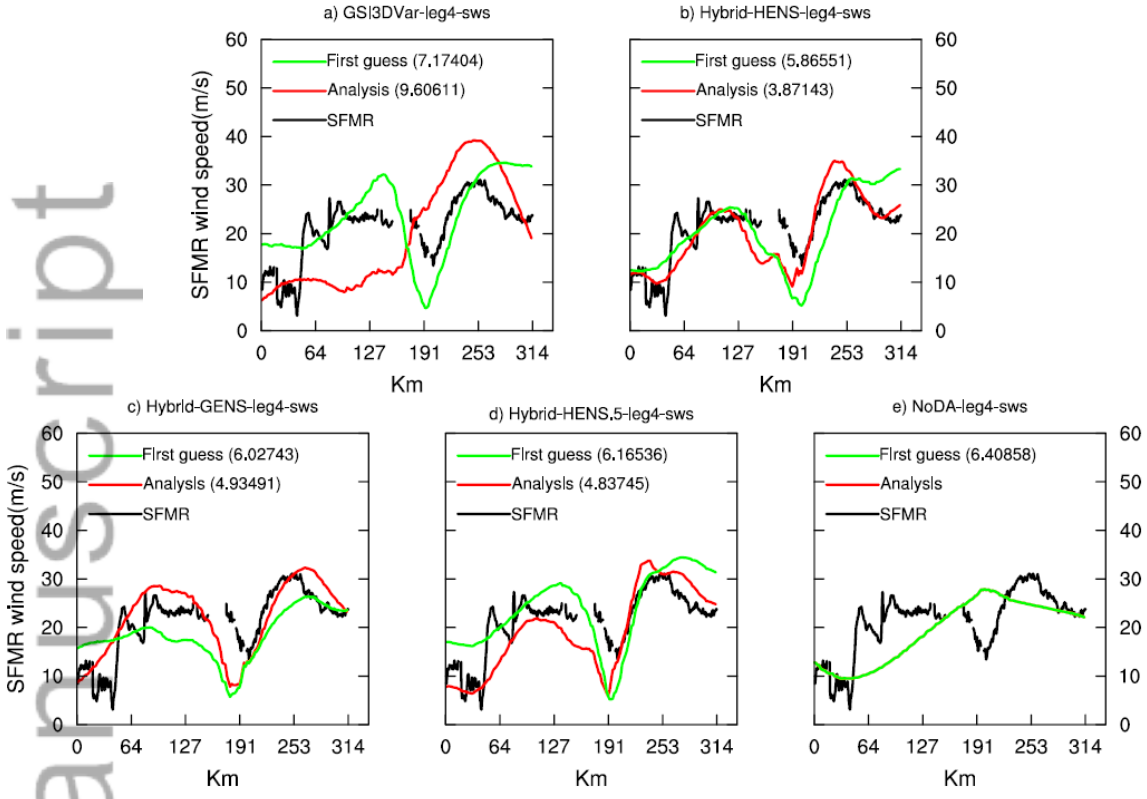

Fig. 10 Same as Fig 9 except for assimilating the last penetration leg of TDR data. Results for the NoDA experiment is also included in e).

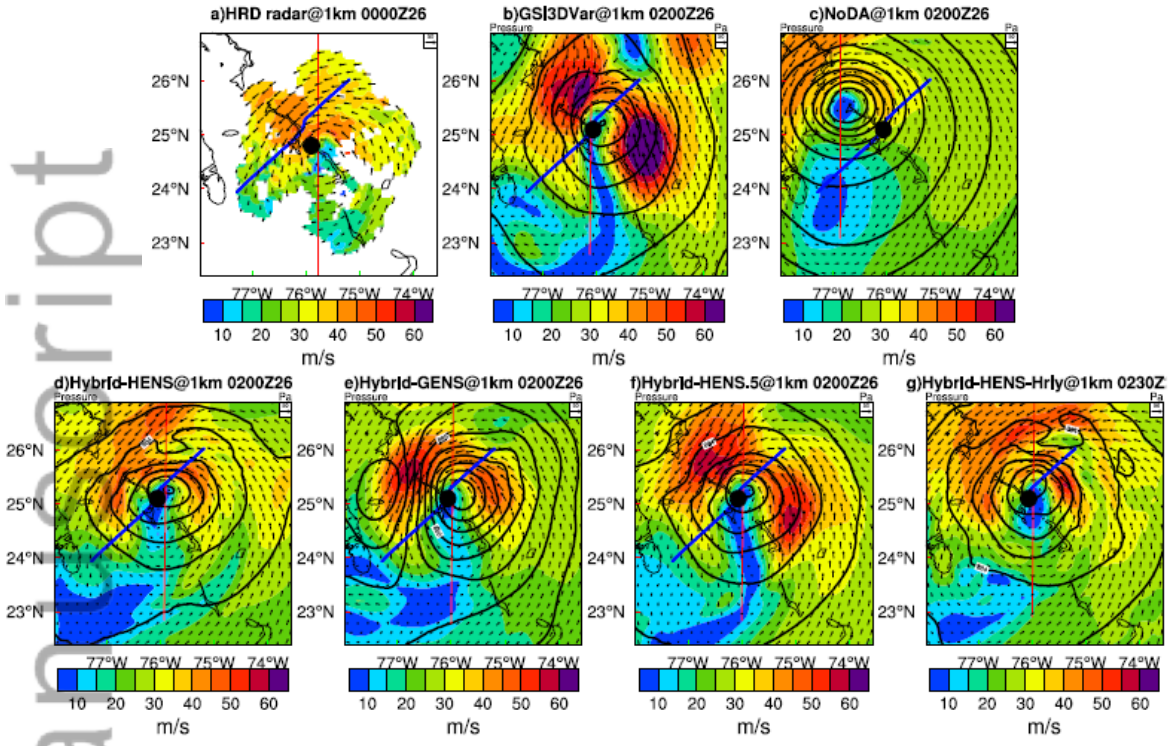

Fig. 11 Wind (shaded and vector) and pressure (contour) at 1km height for a) HRD radar wind composite, b) GSI3DVar, c) NoDA, d) Hybrid-HENS, e) Hybris-GENS, f) Hybrid-HENS.5 and g) Hybrid-HENS-Hrly. Black dot is the interpolated best track position. The blue line denotes the corresponding flight track associated with the last assimilated penetration leg. The red line denotes the corresponding cross-sections to be presented in Fig. 12.

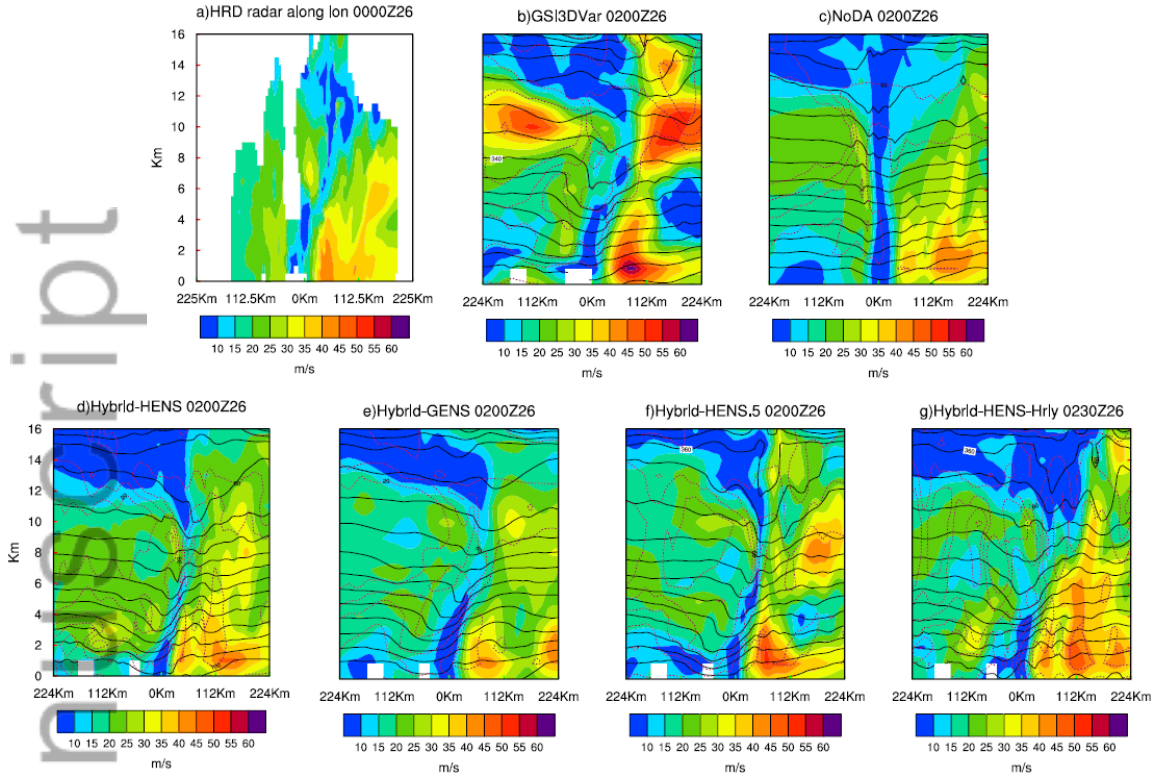

Fig. 12 Wind (shaded), potential temperature (solid line) and relative humidity (dash line) of the south to north vertical cross section for a) HRD radar wind composite, b) GSI3DVar, c) NoDA, d) Hybrid-HENS, e) Hybrid-GENS, f) Hybrid-HENS.5 and g) Hybrid-HENS-Hrly

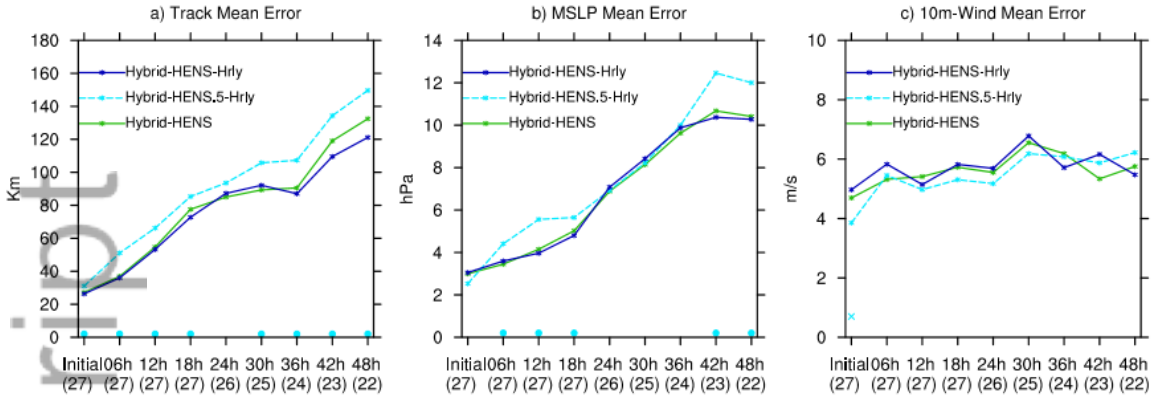

Fig. 13 a) Track , b) SLP and c) Vmax Mean Error statistics for Hybrid-HENS (green solid), Hybrid-HENS.5-Hrly (light blue dash) and Hybrid-HENS-Hrly(dark blue solid) experiments during 2012-2013 seasons. The light blue dots above the x axis denote that the Hybrid-HENS-Hrly is significantly better than Hybrid-HENS.5-Hrly at a confidence level greater than 90%. The light blue crosses above the x axis denote that the Hybrid-HENS.5-Hrly is significantly better than Hybrid-HENS-Hrly at a confidence level greater than 90%.

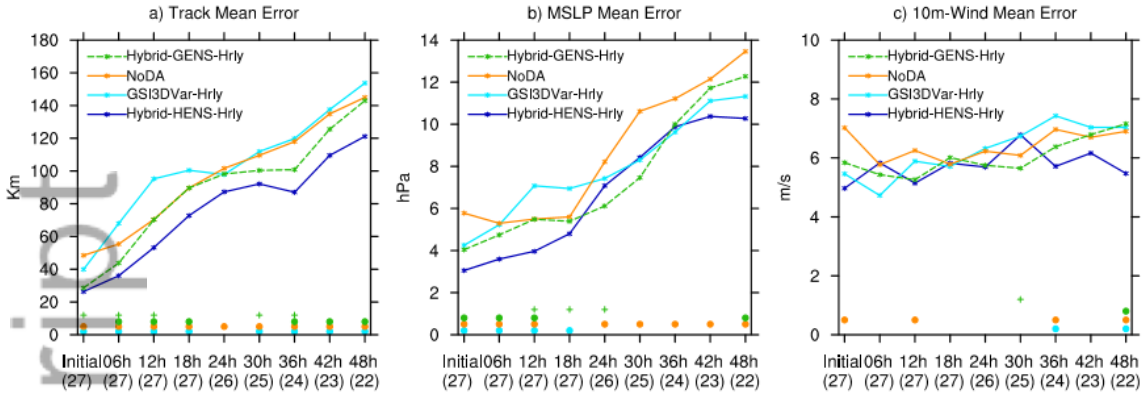

Fig. 14 a) Track , b) SLP and c) Vmax Mean Error statistics for Hybrid-HENS-Hrly (blue solid), Hybrid-GENS-Hrly (green dash), NoDA (orange solid) and GSI3DVar-Hrly (light blue solid) experiments during 2012-2013 seasons. The light blue, orange, and green dots above the x axis denote that the Hybrid-HENS-Hrly is significantly better than GSI3DVar-Hrly, NoDA, Hybrid-GENS-Hrly at a confidence level greater than 90%, respectively. The green crosses right above the x axis denote that the Hybrid-GENS is significantly better than GSI3DVar at a confidence level greater than 90%.

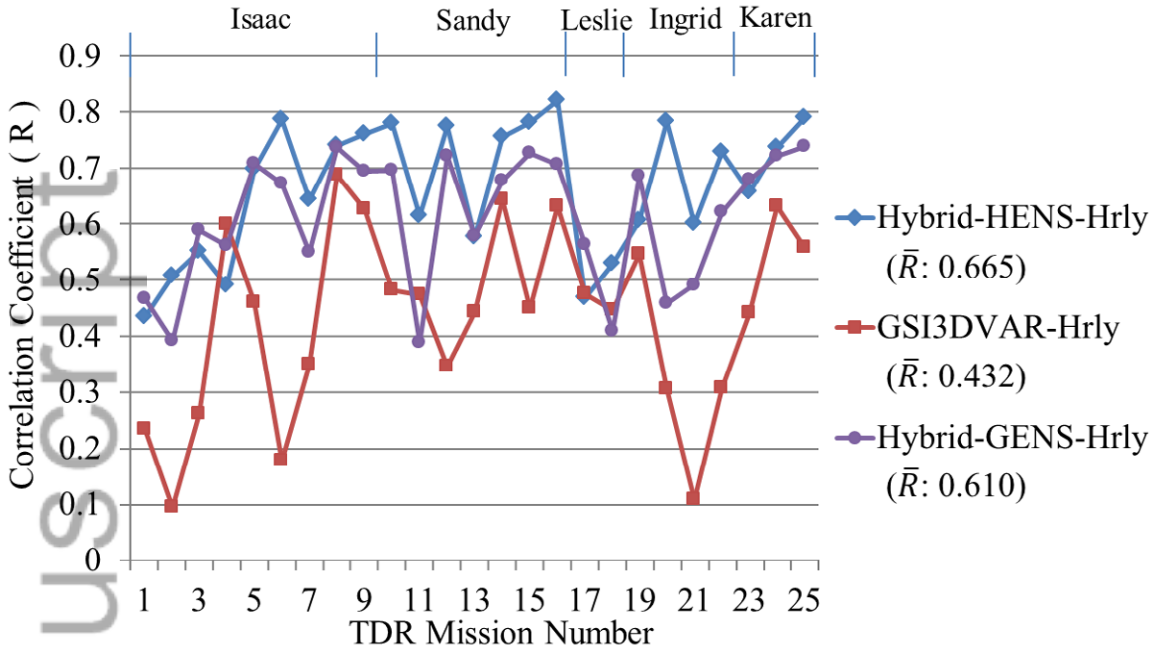

Fig. 15 Correlation coefficient values for all missions during 2012-2013 seasons between Hybrid-HENS-Hrly (Blue), GSI3DVar-Hrly (Red) and Hybrid-GENS-Hrly (Purple) against HRD composite. The values in the parenthesis denote the average values of all cases for each experiment.

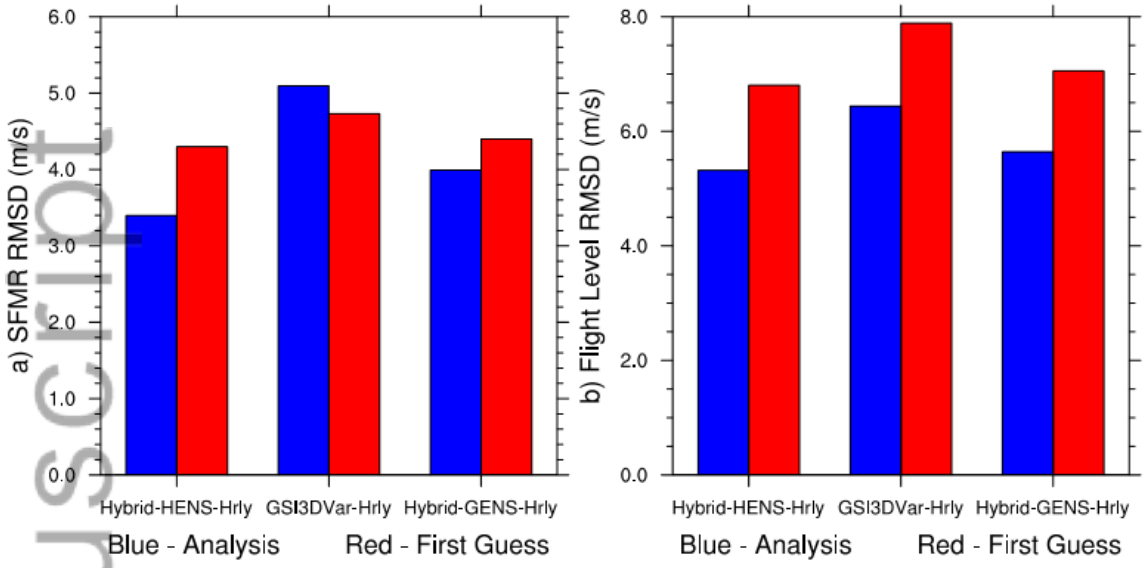

Fig. 16 Independent verification of analyses (blue) and first guess (red) against a) SFMR and b) Flight Level data for experiments during 2012-2013 seasons for Hybrid-HENS-Hrly, GSI3DVar-Hrly and Hybrid-GENS-Hrly experiments.