Impact of 3D Var GSI-ENKF hybrid data assimilation system

V S PRASAD^{1,*} , C J JOHNY¹ and JAGDEEP SINGH SODHI²

¹National Centre for Medium Range Weather Forecasting, A-50, Sector-62, Noida, India.

²Andhra University, Visakhapatnam, Andhra Pradesh, India.

*Corresponding author. e-mail: vsprasad@ncmrwf.gov.in

The hybrid two-way coupled 3DEnsVar assimilation system was tested with the NCMRWF global data assimilation forecasting system. At present, this system consists of T574L64 deterministic model and the grid-point statistical interpolation analysis scheme. In this experiment, the analysis system is modified with a two-way coupling with an 80 member Ensemble Kalman Filter of T254L64 resolution and runs are carried out in parallel to the operational system for the Indian summer monsoon season (June–September) for the year 2015 to study its impact. Both the assimilation systems are based on NCEP GFS system. It is found that hybrid assimilation marginally improved the quality of the forecasts of all variables over the deterministic 3D Var system, in terms of statistical skill scores and also in terms of circulation features. The impact of the hybrid system in prediction of extreme rainfall and cyclone track is discussed.

1. Introduction

Numerical weather prediction models depend on the background state provided by the short-range forecast from the previous cycle. There is some amount of uncertainty in the background state and the assimilation system takes account of this uncertainty in a covariance matrix and is referred as background error. Accurate knowledge of these error statistics is extremely important for the success of the assimilation process. The assimilation system applies optimal corrections to the background state by using observations and taking account of the information on uncertainty in observations, as well as in the background state represented in respective error covariance matrices. The influence of each observation on analysis and its distribution among different model variables is determined by these error covariance matrices (Buehner 2010). Accurate specification of these uncertainties is important for the success of assimilation system. The

NMC (National Meteorological Center) method used in variational schemes in estimating uncertainty in the background state is static and does not show flow-dependent error characteristics (Parrish and Derber 1992). This static background error covariance used in these methods does not represent errors associated with the flow of the day, and background forecast error covariance are assumed to be static, isotropic and nearly homogeneous (Parrish and Derber 1992; Courtier *et al.* 1998). But in reality, background-error covariance depend on flow-of-day and may vary considerably from the static background error covariance (Wang *et al.* 2007). Anisotropy and spatial inhomogeneity are introduced in background error covariance used in 3D Var models by recursive filters (Wu *et al.* 2002; Purser *et al.* 2003). Ensemble data analysis schemes based on different forms of ensemble Kalman filters (EnKF) provided an alternative to variational methods. In these methods, the flow-dependent errors, which are ignored in variational

Keywords. Assimilation; GSI; ensemble; EnKF.

schemes, are estimated by considering few ensemble forecasts. In 3D Var schemes, analysis increments are centred on observation locations and not dependent on background flow and possess an isotropic structure. The computational cost of EnKF increases with the volume of data, since it employs sequential assimilation techniques (Zhang *et al.* 2013). But some EnKF schemes, like local ensemble transform Kalman filter (LETKF), reduced the computational cost considerably by updating the analysis at each grid point and assimilating observations in that local region centred around that point independently (Liu *et al.* 2008). In NCEP GFS ensemble data assimilation at a lower resolution outperformed reduced resolution version of the operational 3D Var system with the biggest improvement in data-sparse regions (Szunyogh *et al.* 2008; Whitaker *et al.* 2008). At the same time, the variational method can assimilate a large number of observations, efficiently, with less computational expense. Hybrid methods take advantage of the relative strengths of both ensemble and variational methods by combining them into a variational framework (Barker 1998; Hamill and Snyder 2000).

In hybrid method, data assimilation is done in the variational framework by effective incorporation of ensemble covariances in estimating background error covariance (Wang 2010). Hybrid schemes are adapted from an existing variational framework, and hence it becomes possible to include ensemble information relatively easily, while pure ensemble assimilation systems are built on a completely different framework. A hybrid data assimilation scheme introduced by Hamill and Snyder (2000), combining EnKF with 3D Var, significantly improved analysis with the use of flow-dependent information from ensembles in estimating background error covariance. They showed that on using a sufficiently large number of ensembles, background error covariance can be completely replaced with the ensemble-based background error covariance. In their hybrid formulation, perturbed observations are assimilated for each ensemble member at every analysis step. So, the computational cost will be very high on using a large number of ensembles. Lorenc (2003) introduced another hybrid scheme with a simpler way to incorporate information from ensemble into the 3D VAR scheme by augmenting state vector with another set of control variables, preconditioned upon the square-root of ensemble covariance. Both the hybrid schemes employ different methods to include ensemble information in the cost function, but they are theoretically equal (Wang *et al.* 2007).

Hybrid option included in the grid-point statistical interpolation (GSI) system by employing extended control variable as detailed in Wang (2010) is used for the GFS model. This hybrid GSI system has

a dual-resolution capability, where the ensemble members are of lower resolution than the subsequent analysis and deterministic forecast, to reduce the computational expense in operational applications (Kleist 2012). Kleist (2012) tested the hybrid algorithm in the GSI system using single observation experiment with the GFS model. Kleist and Ide (2015a) performed an observing system simulation experiment (OSSE) to find the impact of hybrid configuration in NCEP GFS system using deterministic models of T382 and 80 member ensemble of T190 resolution. It is found that the quality of analysis is considerably improved with hybrid configuration. Hamill *et al.* (2011) found hybrid algorithms in NCEP GFS reduced track errors in the prediction of cyclone track. Wang *et al.* (2013) found hybrid 3D EnVar forecasts are more skilful than deterministic with a reduced single resolution experiment with a reduced resolution of T190 for both deterministic and ensembles in NCEP GFS. Wang and Lei (2014) tested hybrid 4D EnVar in NCEP GFS with similar single resolution experiment at T190 resolution with real observations. They found that the 4D variant did not produce much improvement over the 3D counterpart and it showed less impact over tropics. OSSE study done by Kleist and Ide (2015b) with hybrid 4D EnVar in NCEP GFS also showed improvements in moving from 3D to 4D covariances in the hybrid EnVar formulation are not large, while more impact is seen in the tropics. NCEP operationalized hybrid assimilation system with the deterministic system of resolution of T574 and 80 member ensemble of T254 resolution in 2012. In their operational configuration, 25% weight is given to static covariance and 75% weight is given to ensemble covariance. In our centre (NCMRWF), we have tested hybrid 3D EnVar assimilation system with one-way coupling using 20 member ETR (ensemble transform with rescaling) ensembles and deterministic forecasts of T574 resolution and found good improvements over our deterministic system (Prasad and Johnny 2016). But ensembles used in that formulation have only 28 vertical levels. Reduced number of vertical levels and a small number of ensemble size may limit the efficiency of the system. In this work, we are using full ensemble data assimilation system with a sufficiently large number of ensembles. An intercomparison of present hybrid implementation with the previous one is discussed in section 3.1 to find the relative efficiency of both the systems. The Indian monsoon is a highly complex process with global scale and regional scale components, which make the prediction more difficult. In this study, we are investigating the impact of the two-way coupled hybrid 3D EnVar assimilation system based on NCEP GFS with our operational deterministic GSI system of T574 resolution

and 80-member EnKF ensembles during the 2015 Indian monsoon.

2. Methodology

In this study, two numerical experiments are performed, one with GSI-based deterministic forecast in T574L64 configuration (GSI) and another with that of GSI-ENKF hybrid forecast. The GSI-ENKF hybrid system is a two-way coupled dual resolution system using 80-member T254L64 ensembles and a deterministic T574L64 member

(HGSI). Both the experiments were based on NCEP GFS model implemented at NCMRWF, and the number of ensemble members is chosen as 80 by considering the computational requirement of generating real-time forecasts. HGSI takes an additional computer resource of 252 processors and 1-hr computational time, when compared to operational GSI. Full details of this GSI system used for these experiments can be found in Hu *et al.* (2014) and the inclusion of ensemble covariance in GSI scheme and hybrid assimilation are detailed in Wang *et al.* (2013) and Kleist (2012). Figure 1 shows a flow chart of the hybrid assimilation system

Figure 1. Flow chart of hybrid system (HGSI) used in this study.

Table 1. *Types of observations used in the assimilation system (Indian satellite datasets used in our assimilation system are represented in bold).*

Observation category	Name of the observation
Surface	Surface, Mobile, Ship, Buoy (SYNOPS)
Upper air	TEMP (land and marine), PILOT (land and marine), Dropsonde, Wind profiler
Aircraft	AIREP, AMDAR, ACARS, KAMDAR
Atmospheric motion vectors from geo-stationary satellites	AMV from Meteosat-7, Meteosat-9, GOES-11, GOES-13, MTSAT-1R, MODIS (TERRA and AQUA), POLAR WINDS (NOAA, MODIS, METOP), INSAT 3D
Scatterometer winds and soil moisture	ASCAT winds from METOP-A and METOP-B satellites
NESDIS/POES ATOVS radiance data	1bamua, 1bmhs, 1bhirs3, 1bhirs4, CrIS and ATMS
Satellite derived ozone data	NESDIS/POES, METOP-2 and AURA orbital ozone data
Precipitation rates	GPM and SSM/I precip. rates
Bending angles from GPSRO	Atmospheric profiles from radio occultation data using GPS satellites (COSMIC series, GRACE, etc.)
Megha Tropique	SAPHIR radiance
AQUA AIRS & METOP/IASI brightness temperature	IASI, AIRS brightness temperatures
GPSIPW	GPSIPW (SUOMINET and EUMETNET)
Geostat. satellite radiances	GOES, Meteosat, INSAT-3D

used in this study. As shown in the figure, EnKF is a separate DA system that runs at each analysis time and updates every ensemble member using the information from observations. The error covariance used in the hybrid experiment is built by using 75% weight to ensemble covariance and 25% to static covariance. In every analysis cycle, ensemble members are re-centered around hybrid analysis to arrest any possible drift in the total system. In a hybrid system, deterministic and ensemble models can be coupled either in one-way or in two-way. In both methods, ensemble perturbations are used in generating hybrid control analysis, but in a two-way coupling, an additional step to re-center

ensemble analyses around hybrid analysis is included. These re-centering techniques help in arresting any divergence that can develop between hybrid and ensemble mean analyses. This, in turn, enables ensemble perturbations to sample deterministic forecast uncertainty well (Pan *et al.* 2014).

All the observations used in the operational deterministic forecast (GSI) are also used in hybrid experiment. The experiment was carried out during the Indian summer monsoon season, i.e., 1 June–30 September, 2015 with 10 days forecast integration at 00UTC and assimilation four times (00UTC, 06UTC, 12UTC and 18UTC) a day. The hybrid system uses horizontal ensemble localisation of

Figure 2. RMSE in model wind (m/s) forecasts at 850 hPa (bottom panel), 500 hPa (middle panel) and 200 hPa (top panel) pressure levels with respect to respective analysis in the domains global (left), tropics (middle) and RSMC region (right). Bottom panel shows the difference with respect to GSI and their statistical significance. The line above bar is significant at 95% confidence level.

700 km at a lower level, 1300 km at model top and vertical localisation of 0.5 log P units and tangent linear normal mode constraint is applied on the total increment for a single time. Ensemble perturbations are incorporated into the variational system through the extended control variable method (Kleist 2012):

$$J(x'_f, \alpha) = \beta_f \frac{1}{2} (x'_f)^T B^{-1} (x'_f) + \beta_e \frac{1}{2} (\alpha)^T L^{-1} (\alpha) + \frac{1}{2} (y'_o - Hx'_t)^T \times R^{-1} (y'_o - Hx'_t) \quad (1)$$

$$x'_t = C \left[x'_f + \sum_{n=1}^N (\alpha^n x_e^n) \right], 2 \times 4 \quad (2)$$

where β_f and β_e are weighing coefficients for fixed and ensemble variances, respectively, R is observation error covariance, H is observation operator, α^n is ensemble control variable for each member with L defined as its covariance matrix, n is index of ensemble member, x_t is total increment, y'_o is observation residuals, x'_f is static error covariance, x_e^n is ensemble covariance and C represents linear normal mode constraint operator. Assimilation system

in the centre uses conventional and satellite-based observations, including radiance observations from Indian satellites INSAT-3D and Megha-Tropiques. Table 1 shows the types of observations that are assimilated in the analysis system and the Indian satellite datasets used in our assimilation system are represented in bold font.

3. Results

Model forecasts fields of temperature, wind vector and specific humidity are verified against respective radiosonde observations and analysis over different regions. The comparison is performed over the

Table 2. *Percentage improvement in wind RMSE (computed against radiosonde observations) over 3D Var in 24 hr forecasts of hybrid schemes at 850, 500 and 200 hPa pressure levels.*

Pressure (hPa)	% improvement in RMSE	
	HGSI	One way hybrid
850	6.5	1.03
500	8.3	1.36
200	10.85	1.90

Figure 3. Vertical profile of RMSE difference between the experiments (hybrid-GSI) in analysis and first guess wind (m/s) against radiosonde observations in global domain (left) and tropical region (right) (mean value computed for the entire study period). Black line indicates zero difference position. Blue coloured error bars represent the differences significant at 95% confidence level. Mean RMSE difference values beyond the error bars are significant at 95% confidence level. Negative RMSE differences indicate the region where hybrid experiment performed better than 3D Var and vice versa.

domains like global, tropics (20°N–20°S) and regional specialized meteorological centre (RSMC) region (30°–120°E; 20°S–45°N) to identify regional impacts. Rainfall forecasts over Indian region (8°–38°N, 68°–98°E) are verified against gridded IMD-NCRWF merged rainfall data using rain gauge observations and TRMM rain estimates. The details of this IMD_NCMRWF rainfall data can be found in Mitra *et al.* (2009). The characteristics of monsoon

propagation over Indian region are investigated using different indices derived from model forecasts.

3.1 Wind

Errors in initial conditions can limit the atmospheric flow stability in models (Ferranti *et al.* 2002). Hybrid technique modifies atmospheric flow errors

Figure 4. RMSE in model specific humidity (g/kg) forecasts at 850 hPa (bottom panel), 500 hPa (middle panel) and 250 hPa (top panel) pressure levels with respect to respective analysis in the domains global (left), tropics (middle) and RSMC region (right). Bottom panel shows the difference with respect to GSI and their statistical significance. The line above bar is significant at 95% confidence level.

by the inclusion of flow-dependent error covariance. Global observing system with a large number of mass-type observations can limit errors in temperature analysis by assimilating satellite radiances (Kleist and Ide 2015a). So, hybrid assimilation is expected to provide more improvement in wind than other variables. There are some previous studies which found improvement in wind with hybrid assimilation is more than that on other variables (Wang *et al.* 2008; Wang and Lei 2014). RMSE in model wind forecasts with respect to respective analysis at different pressure levels over different regions is shown in figure 2. It can be seen that hybrid assimilation marginally improved wind forecasts at all levels and all the regions. The improvement of wind field forecasts in the case of HGSI system is very large in the upper levels. In tropics at 200 hPa level, improvement in wind forecasts with a hybrid system is not significant at 95% confidence level at longer (beyond one week) lead times. RMSE difference between the experiments (Hybrid-GSI) in model analysis and first guess (6-hr forecast) with that of radiosonde observations over tropics and over the globe are shown in figure 3. The negative values of RMSE difference indicates the improvement with hybrid experiment and differences beyond blue coloured bars are significant at 95% confidence level. It can be seen that

hybrid wind forecasts are closer to observations in both the regions, while GSI analysis is closer to the observations in lower levels. Fit of analysis fields to the observations cannot be considered as a representative of the characteristics of the analysis error, unless we have an independent set of observations (Wang *et al.* 2008). Here observations used for verifying analysis are also part of the assimilation system and hence analysis-observation fit may not be considered as a true representation of the quality of analysis (Zhang *et al.* 2011). The fit between variational analysis and the observations are likely to be closer than that of hybrid analysis as they employ a cost function that minimises, iteratively, the distance between the analysis and observations (the observational cost function J_o) with less confidence on the first guess (Zhang and Zhang 2012). Privé *et al.* (2013) clearly demonstrated that in a data analysis system, analysis fields can be drawn near to the observations than background either by changing error variances in either observations or background fields. Few other studies also clearly proved that variational analysis is closer to observations than hybrid analysis (Wang *et al.* 2008; Zhang and Zhang 2012). But these observations can act as independent observations for verifying forecasts (Wang *et al.* 2008).

Figure 5. Vertical profile RMSE difference between experiments (hybrid-GSI) in analysis and first guess specific humidity (g/kg) against radiosonde observations in global domain (left) and tropical region (right) (mean value computed for the entire study period). Black dashed line indicates zero difference position. Blue-coloured error bars represent the differences significant at 95% confidence level. Mean RMSE difference values beyond the error bars are significant at 95% confidence level. Negative RMSE differences indicate the region where hybrid experiment performed better than 3D Var and vice versa.

The performance of present hybrid implementation is compared to the previous one-way hybrid (Prasad and Johny 2016) implementation mentioned in section 1. Wind is chosen for comparison as the variable largely benefited in both the hybrid implementations. Table 2 shows percentage improvement in RMSE (computed against radiosonde observations) over the 3D Var in 24-hr wind forecasts in both the hybrid implementations.

It can be seen that the present hybrid implementation improved wind forecasts six times than with the previous one.

3.2 Humidity

Figure 4 shows RMS errors in humidity forecasts with respect to analysis at various pressure levels and over different regions. It can be observed that

Figure 6. RMSE in model temperature (K) forecasts at 850 hPa (bottom panel), 500 hPa (middle panel) and 200 hPa (top panel) pressure levels with respect to respective analysis in the domains global (left), tropics (middle) and RSMC region (right). Bottom panel shows the difference with respect to GSI and their statistical significance. The line above bar is significant at 95% confidence level.

in the case of HGSI experiment forecast fields of specific humidity over all the regions are marginally improved when compared with its own analyses. In figure 5, the differences in the specific humidity fields between model analysis and respective first guess are compared against radiosonde observations. It can be seen that model first guess and forecasts are improved by the HGSI system in the most levels except a few in the upper levels. These differences are not that significant, in the case of first guess fields at the upper tropospheric levels (300 hPa) at the 95% confidence level.

3.3 Temperature

The verification of temperature forecasts with respect to own analysis over different regions is shown in figure 6. Temperature forecasts are marginally improved by hybrid assimilation in all the regions, except at 850 hPa in the tropics. Comparison of first guess and analysis with respect to radiosonde observations is shown in figure 7. In comparison with the radiosonde observations first guess is improved by the hybrid system over tropics and also on a global scale. The hybrid analysis is closer to the observations in tropics at all heights, while in the global domain, GSI analysis is closer to observations, except at upper levels. In

comparison with radiosonde observations in the tropics, temperature behave differently from wind and humidity. The general characteristics of wind and humidity analysis are same in both global and tropic domains with GSI analysis closer to observations at lower levels and hybrids closer to observations at upper levels. It can be attributed to multivariate aspects of analysis increments in hybrid systems noted by Kleist (2012) in which he obtained large improvements in wind and humidity and not in temperature. It has to be noted that relative to analysis, hybrid temperature forecasts of all lead times degraded in tropics at 850 hPa, while relative to radiosonde observations such a degradation is not seen. Hamrud *et al.* (2015) showed that EnKF analyses tendency to generate spurious inertia-gravity waves in short range forecasts and it is due to their inherent problems in representing the constraints in the model dynamics. The corrections are applied to the initial condition variables separately, thus, this may create a dynamical imbalance with one another and thereby may produce unphysical inertial-gravity waves. This may be attributed to the effect disruption of mass-wind field geostrophic balance due to the localization of error covariance. For improving the balance between analysis increments, Tangent linear normal mode constraint (TLNMC) is

Figure 7. Vertical profile of RMSE difference between experiments (hybrid-GSI) in analysis and first guess temperature (K) against radiosonde observations in global domain (left) and tropical region (right) (mean value computed for the entire study period). The black dashed line indicates zero difference position and the blue-coloured error bars represent the differences significant at 95% confidence level. Mean RMSE difference values beyond the error bars are significant at 95% confidence level. Negative RMSE differences indicate the region where hybrid experiment performed better than 3D Var and vice versa.

employed, but it is slightly detrimental in the tropics (Thomas *et al.* 2014). In the tropics at lower level humidity, its variations are large. The imbalances in correction made to individual variables at lower levels in the tropics might be causing degradation in hybrid temperature forecasts in the tropics.

3.4 Rainfall

The mean statistics of rainfall forecasts by model and of observations over the Indian region (8° – 38° N and 68° – 98° E) is computed for the entire

period of study. Bar plots of RMSE and correlation in 24 hr accumulated rainfall forecasts against that of IMD-NCMRWF gridded rainfall data and distribution of rainfall and mean characteristics in the observations and model forecasts in the season are given in figure 8. Hybrid experiment reduced errors in rainfall prediction in the forecasts at all lead times. Hybrid forecasts show high correlation with rainfall observations in comparison with GSI. Both the experiments are able to predict the mean character of rainfall and low rainfall events. But they are not able to pick the distribution of

Figure 8. Bar plots of (a) RMSE and (b) correlation in model forecasts of 24-hr accumulated rainfall against IMD-NCMRWF observations and (c) mean of rainfall in model forecasts and observations in mm computed for season.

Figure 9. 24-hr accumulated rainfall (mm) around Mumbai on 19th June. (a) Observations and (Hybrid – b) model forecasts and (c) GSI.

Figure 10. Track of Ashobaa cyclone predicted in numerical experiments based on June 11th 00 UTC initial condition hybrid (green), GSI (blue) and JTWC best track (red).

high rainfall events. During this study period, a heavy rainfall event occurred in Mumbai on 19th June. Both the experiments are able to simulate rainfall events with 24 hr accumulated rainfall, as shown in figure 9. The amount of rainfall predicted in the case of the hybrid experiment is 110 mm and in the case of GSI, it is 60 mm, whereas the actual amount is 200 mm. Thus, the hybrid GSI system is able to improve the prediction of mean statistics and extreme rainfall event.

3.5 Cyclone track

In the investigation period, a depression formed on 7th June in the Arabian Sea, developed into a cyclonic storm (Ashobaa), moved north-west and dissipated as a low-pressure system on June 12 near the Oman coast. Figure 10 shows the cyclone track predicted by the models correspond to initial conditions at 00UTC on the 11th along with the Joint Typhoon Warning Center (JTWC) best track. Both the experiments predicted the movement of the system in a different direction with the hybrid showing movement towards the west, while GSI is showing the movement of the system towards south-west. The hybrid experiment is

able to simulate the system and its direction of movement somewhat correctly, while there was some overprediction in its intensity forecast, which led to the prediction of system movement towards land in subsequent days. Tropical cyclone tracker used in this study is the modified version of the original NCEP cyclone tracker (Marchok 2002) computes track position based on mean sea level pressure, relative vorticity, vector wind and geopotential height.

4. Conclusions

Here, the performance of the hybrid assimilation system based on NCEP GFS implemented in our centre compared with deterministic GSI during the 2015 monsoon season. Hybrid assimilation improved forecasts of wind, temperature, humidity and rainfall fields except for lower level tropical temperatures. Hybrid assimilation produced more accurate background (short range forecast) than 3D Var. At lower levels, 3D Var analysis is close to the observations except for temperature in the tropics, while in the case of forecasts, hybrid fits observations closely. The hybrid system

demonstrated higher prediction skill in the case of extreme rainfall event that occurred in Mumbai. The hybrid experiment is able to predict the direction of movement of cyclonic storm formed in the Arabian Sea.

Acknowledgements

Authors are thankful to the Head, NCMRWF for the support and encouragement. We thank NOAA Center for Weather and Climate Prediction, USA and Daryl Kleist, University of Maryland, USA for their scientific support and discussion in designing the experiment. We also thank the Monsoon Desk at NCEP, NOAA, for providing codes and technical support.

References

- Barker D M 1998 Var scientific development paper 25: The use of synoptic-dependent error structure in 3D Var; UK MetOffice Tech. Rep. (Available from the Met Office, Saughton House, Broomhouse Dr., Edinburgh EH11 3XQ, United Kingdom).
- Buehner M 2010 *Error Statistics in Data Assimilation: Estimation and Modelling*; Springer, Berlin, Heidelberg, pp. 93–112.
- Courtier P et al. 1998 The ECMWF implementation of three-dimensional variational assimilation (3DVAR) I: Formulation; *Quart. J. Roy. Meteor. Soc.* **124** 1783–1807.
- Ferranti L, Klinker E, Hollingsworth A and Hoskins B J 2002 Diagnosis of systematic forecast errors dependent on flow pattern; *Quart. J. Roy. Meteorol. Soc.* **128** 1623–1640.
- Hamill T M and Snyder C 2000 A hybrid ensemble Kalman filter-3D variational analysis scheme; *Mon. Wea. Rev.* **128** 2905–2919.
- Hamill T M, Whitaker J S, Kleist D T, Fiorino M and Benjamin S 2011 Prediction of 2010's tropical cyclones using the GFS and ensemble based data assimilation methods; *Mon. Wea. Rev.* **139** 3243–3247.
- Hamrud M, Bonavita M and Isaksen L 2015 EnKF and hybrid gain ensemble data assimilation. Part I: EnKF implementation; *Mon. Wea. Rev.* **143** 4847–4864.
- Hu M, Shao H, Stark D and Newman K 2014 Gridpoint Statistical Interpolation (GSI; Version 3.3 User's Guide), Developmental Testbed Centre, NCAR, NOAA, USA, <http://www.dtcenter.org/com-GSI/users/index.php>.
- Kleist D T 2012 An evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS; PhD Thesis, Dept. of Atmospheric and Oceanic Science, University of Maryland-College Park, 149p, <http://drum.lib.umd.edu/handle/1903/13135>.
- Kleist D T and Ide K 2015a An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results; *Mon. Wea. Rev.* **143** 443–451.
- Kleist D T and Ide K 2015b An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4D-EnVar and hybrid variants; *Mon. Wea. Rev.* **143** 452–469.
- Liu J, Fertig E J, Li H, Kalnay E, Hunt B R, Kostelich E J, Szunyogh I and Todling R 2008 Comparison between local ensemble transform Kalman filter and PSAS in the NASA finite volume GCM perfect model experiments; *Nonlin. Process. Geophys.* **15** 645–659.
- Lorenc A C 2003 The potential of the ensemble Kalman filter for NWP – a comparison with 4D-VAR; *Quart. J. Roy. Meteor. Soc.* **129** 3183–3203.
- Marchok T 2002 How the NCEP Tropical Cyclone Tracker Works; Proc. 25th Conference on Hurricanes and Tropical Meteorology, San Diego, CA.
- Mitra A K, Bohra A K, Rajeevan M and Krishnamurti T N 2009 Daily Indian precipitation analyses formed from a merge of rain-gauge with TRMM TMPA satellite derived rainfall estimates; *J. Meteor. Soc. Japan* **87A** 265–279.
- Pan Y, Zhu K, Xue M, Wang X, Hu M, Benjamin S G, Weygandt S S and Whitaker J S 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.
- Parrish D F and Derber J C 1992 The National Meteorological Center's spectral statistical interpolation analysis system; *Mon. Wea. Rev.* **120** 1747–1763.
- Prasad V S and Johny C J 2016 Impact of hybrid GSI analysis using ETR ensembles; *J. Earth Syst. Sci.* **125(3)** 521–538.
- Privé N C, Errico R M and Tai K-S 2013 The influence of observation errors on analysis error and forecast skill investigated with an observing system simulation experiment; *J. Geophys. Res. Atmos.* **118** 5332–5346.
- Purser J R, Wu W S, Parrish D F and Roberts N M 2003 Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances; *Mon. Wea. Rev.* **131** 1536–1548.
- Szunyogh I, Kostelich E J, Gyarmati G, Kalnay E, Hunt B R, Ott E, Satterfield E and Yorke J A 2008 A local ensemble transform Kalman filter data assimilation system for the NCEP global model; *Tellus* **60A** 113–130.
- Thomas C, Kleist D and Mahajan R 2014 Improving balance in the NCEP Hybrid Ensemble-Var data assimilation system; World Weather Open Science Conference 2014, Montreal, Canada.
- Wang X 2010 Incorporating ensemble covariance in the grid-point statistical interpolation variational minimization: A mathematical framework; *Mon. Wea. Rev.* **138** 2990–2995.
- Wang X and Lei T 2014 GSI-based four-dimensional ensemble-variational (4DEnVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP global forecast system; *Mon. Wea. Rev.* **142** 3303–3325.
- Wang X, Snyder C and Hamill T M 2007 On theoretical equivalence of differently proposed ensemble/3D-Var hybrid analysis schemes; *Mon. Wea. Rev.* **135** 222–227.
- Wang X, Barker D M, Snyder C and Hamill T M 2008 A hybrid ETKF-3DVar data assimilation scheme for WRF model. Part 2: Real observation experiments; *Mon. Wea. Rev.* **136** 5132–5147.
- Wang X, David P, Daryl K and Whitaker J S 2013 GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP global forecast system: Single-resolution experiments; *Mon. Wea. Rev.* **141** 4098–4117.

- Whitaker J S, Hamill T M and Wei X 2008 Ensemble data assimilation with NCEP global forecast system; *Mon. Wea. Rev.* **136** 463–482.
- Wu W S, Purser R J and Parrish D F 2002 Three-dimensional variational analysis with spatially inhomogeneous covariances; *Mon. Wea. Rev.* **130** 2905–2916.
- Zhang M and 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.
- Zhang M, Zhang F, Huang X Y and Zhang X 2011 Inter-comparison of an ensemble Kalman filter with three- and four-dimensional variational data assimilation methods in a limited-area model over the month of June 2003; *Mon. Wea. Rev.* **139** 566–572.
- Zhang F, Zhang M and Poterjoy J 2013 E3DVar: Coupling an ensemble Kalman filter with three dimensional variational data assimilation in a limited area weather prediction model and comparison to 4DVar; *Mon. Wea. Rev.* **141** 900–917.

MS received 8 January 2016; revised 12 August 2016; accepted 24 August 2016

Corresponding editor: ASHOK KARUMURI