A GSI-Based Coupled EnKF-En3DVar Hybrid Data Assimilation System for the Operational Rapid Refresh Model: Tests at a Reduced Resolution

Yujie Pan[1,3], Kefeng Zhu[1], Ming Xue[1,2*], Xuguang Wang[1,2], Ming Hu[4,5], Stanley G. Benjamin[4], Stephen S. Weygandt[4], and Jeffrey S. Whitaker[4]

[1]Center for Analysis and Prediction of Storms and [2]School of Meteorology
University of Oklahoma, Norman Oklahoma 73072

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

[4]NOAA Earth System Research Laboratory, Boulder, Colorado

[5]Cooperative Institute for Research in Environmental Sciences
Colorado University at Boulder, Boulder, Colorado

Submitted to Monthly Weather Review
July 25, 2013
Revised November 2013

*Corresponding author address:
Dr. Ming Xue
Center for Analysis and Prediction of Storms
University of Oklahoma
120 David L. Boren Blvd, Norman OK 73072
mxue@ou.edu

Abstract

A coupled EnKF-En3DVar hybrid data assimilation system is developed for the operational Rapid Refresh (RAP) forecasting system. The three-dimensional ensemble-variational (En3DVar) hybrid system employs the extended control variable method, and is built on the NCEP operational Grid-point Statistical Interpolation (GSI) 3DVar framework. It is coupled with a GSI-based ensemble Kalman filter (EnKF) system for RAP, which provides ensemble perturbations. Recursive filters (RFs) are used to localize ensemble covariance in both horizontal and vertical within the En3DVar.

The coupled En3DVar hybrid system is evaluated with 3-hourly cycles over a 9-day period with active convection. All conventional observations used by operational RAP are included. The En3DVar hybrid system is run at 1/3 of the operational RAP horizontal resolution about 40-km grid spacing, and its performance is compared to parallel GSI and EnKF runs using the same data sets and resolution. Short-term forecasts initialized from the 3-hourly analyses are verified against sounding and surface observations.

When using equally weighted static and ensemble background error covariances and 40 ensemble members, the En3DVar hybrid system outperforms corresponding GSI and EnKF. When the RF coefficients are tuned to achieve a similar height dependency of localization as in the EnKF, the En3DVar results with pure ensemble covariance are close to EnKF. With 20 ensemble members, EnKF, GSI and En3DVar perform in ascending order, showing the advantage of the En3DVar hybrid for small ensembles. Two-way coupling between EnKF and En3DVar did not produce noticeable improvement over one-way coupling. Downscaled precipitation forecast skill on the 13-km RAP grid from the En3DVar hybrid is better than those from GSI analyses.

1. Introduction

Three-dimensional variational (3DVAR, Lorenc 1986) and four-dimensional variational (4DVAR, Le Dimet and Talagrand 1986; Talagrand and Courtier 1987) data assimilation (DA) methods have been used successfully at operational numerical weather prediction (NWP) centers for more than two decades (e.g., Parrish and Derber 1992; Courtier et al. 1998; Rabier et al. 2000). Typically, static, flow-independent background error covariance (BEC) is used in the background term of the variational cost function. Neglecting the flow dependent nature of the background error is a key deficiency, especially within a 3DVar framework where the NWP model is not directly used to incorporate model dynamics into the DA system (e.g., Parrish and Derber 1992; Purser et al. 2003b). This deficiency becomes more severe for mesoscale and convective-scale DA where even fewer state variables (compared to the full set) are directly observed and large-scale balance relationships, which are often built into 3DVar systems, become invalid (e.g., Gao et al. 2004; Ge et al. 2012). While some efforts had been made to build spatially inhomogeneous, anisotropic BEC into 3DVar frameworks (e.g., Wu et al. 2002; Purser et al. 2003a), major issues exist on how to determine the flow-dependent covariances and how to efficiently introduce them into a variational DA framework.

The ensemble Kalman filter (EnKF) algorithm, as initially developed by Evensen (1994) and Burgers et al. (1998), offers an alternative to the variational formulation. The EnKF employs the Monto Carlo sampling approach, where an ensemble of model forecasts is used to provide and evolve flow-dependent covariances, while the filter updates the ensemble states using an optimal estimation algorithm. Many subsequent studies have refined the filter algorithm by addressing a number of issues that are often related to the sampling error associated with the use of relatively small ensembles that is necessitated by practical computational constraints (e.g., Burgers et al.

1998; Houtekamer and Mitchell 1998; Anderson 2001; Hamill et al. 2001; Whitaker and Hamill 2002; Evensen 2003). Because of their ability to estimate flow-dependent BECs and to evolve them through assimilation cycles, and their relative ease of implementation, the ensemble DA methods (Evensen 1994; Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; Hunt et al. 2007) have gained much popularity within both the research and operational communities in recent years. The ensemble filters have been used in operational global forecast systems to provide ensemble-based BEC (e.g., Hamill et al. 2011b; Raynaud et al. 2011; Bonavita et al. 2012; Wang et al. 2013) as well as initial conditions for ensemble forecasts (e.g., Houtekamer et al. 2005; Whitaker et al. 2008; Hamill et al. 2011a). The application of EnKF to mesoscale models has also enjoyed encouraging successes (e.g., Fujita et al. 2007; Meng and Zhang 2007; Bonavita et al. 2008) while for the convective scale, EnKF has shown great ability in dealing with complex, nonlinear physical processes (e.g., Tong and Xue 2005) that may even involve two-moment microphysics parameterization (e.g., Xue et al. 2010; Jung et al. 2012; Putnam et al. 2013). Accurate representation of microphysical processes is especially important at the convective scale.

While EnKF provides a way of estimating flow-dependent BEC, the estimated covariance matrix is severely rank deficient due to the much smaller ensemble sizes typically used compared to the degrees of freedom of typical NWP model state (Houtekamer and Mitchell 1998; Hamill and Snyder 2000). The use of much larger ensembles is often computationally impractical while covariance localization that alleviates the rank deficiency problem has its own issues (Anderson 2007, 2012). An alternative for alleviating this problem is to combine the full-rank static BEC with the rank-deficient ensemble BEC, creating a so-called hybrid[1] algorithm.

Hamill and Snyder (2000) were the first to propose a 3DVar-based hybrid scheme in which

[1] In this study, we use the word ‘hybrid’ to mainly refer to the use of a combination of the static and ensemble-derived flow-dependent covariances, i.e., the hybrid covariance. Sometimes in this paper, hybrid is also used to refer to the En3DVar algorithm to be discussed later.

the static BEC in a 3DVar system was replaced by a linear combination of the static and ensemble-derived BEC. The system was tested with a low-resolution quasi-geostrophic model and simulated data in a perfect model setting. By running the hybrid analysis system multiple times with perturbed observations, the system is able to provide an ensemble of analyses. It was found that the analysis performs the best when BEC is estimated almost fully from the ensemble, especially when the ensemble size was large (100 in their case). When the ensemble is smaller, the system benefits from a lesser weighting given to the ensemble-based covariances. Wang et al. (2009) also found that a hybrid system based on an ETKF is more robust than EnKF for a two-layer primitive equation model when the ensemble size is small and when the model error is large. The hybrid formulation in these studies requires explicit evaluation and storage of the ensemble covarianes which is very expensive for full NWP models.

Lorenc (2003) proposed an elegant, alternative hybrid formulation, in which the control variables of the regular variational cost function are augmented by extended control variables (hereafter, ECV), which are preconditioned upon the square root of ensemble covariance. The ECV formulation involves adding an additional term to the variational cost function for the ECVs which has a similar form as the original background term, and is therefore relatively easy to implement based on an existing variational DA framework. Wang et al. (2007) proved that the ECV formulation is mathematically equivalent to that of Hamill and Snyder (2000). The potential for the hybrid system to perform better than a pure EnKF when the ensemble size is relatively small makes it attractive for operational implementation where computational constraint is often a significant issue. A variational framework used by the hybrid scheme also makes it easier to include equation constraints (Kleist et al. 2009b; Ge et al. 2012). Furthermore, for observations whose forward operators are non-local, such as those of satellite radiance data, the state-space-based covariance localization used in the hybrid formulation is potentially advantageous (Campbell et al. 2010). As suggested by Lorenc (2003), Buehner et al. (2010b, a) and further discussed by Liu and Xue (2013), both (traditional) 3DVar and 4DVar can be formulated to use the ensemble covariance with the extended control variable method, and we call such ensemble-variational formulations En3DVar and En4DVar[2], respectively, or EnVar in general.

Buehner (2005) implemented the ECV hybrid approach within the Canadian operational global 3DVar framework, and found that the hybrid scheme produced comparable or better forecasts that those initialized using 3DVar. Buehner et al. (2010b, a) further compared the performances of the coupled EnKF-En3DVar and EnKF-En4DVar with the pure 3DVar and 4DVar for global forecasts. Based on the variational DA framework of the Advanced Research WRF (WRF-ARW, Skamarock et al. 2005) model, Wang et al. (2008b, a) implemented the ECV-based hybrid, coupling it with an ensemble transform Kalman filter (ETKF, Bishop et al. 2001) that is used to update the ensemble perturbations (which we call ETKF-En3DVar hybrid). This WRF hybrid DA system was further applied for tropical cyclone DA (Wang 2011; Li et al. 2012) Most recently, Zhang and Zhang (2011) coupled a mesoscale EnKF system with WRF 4DVar through the WRF hybrid DA framework (hence EnKF-En4DVar hybrid but they called it E4DVar), and Zhang et al. (2013) further compared the performances of EnKF-En3DVar (they called it E3DVar) and EnKF-En4DVar hybrid for mesoscale applications. Mizzi (2012) reported results testing the GSI-based En3DVar hybrid, using ETKF, local ensemble transform Kalman filter (LEKF), and the regular EnKF for ensemble perturbation updating, respectively, and WRF-ARW as the prediction

[2] Here, En4DVar is an extension of the traditional 4DVar scheme to include the use of ensemble-derived background error covariance through the extended control variable method. The scheme still involves the use of an adjoint model. Liu et al. Liu, C., Q. Xiao, and B. Wang, 2008: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.*, **136,** 3363-3373. proposed an alternative algorithm that does not involve the use of a model adjoint, and En4DVar was used to refer to their algorithm. In Liu and Xiao Liu, C. and Q. Xiao, 2013: An ensemble-based four-dimensional variational data assimilation scheme. Part III: Antarctic applications with Advanced Research WRF (ARW) using real dataibid., **141,** 2721-2739. and Liu and Xue Liu, C. and M. Xue, 2013: A unified framework for four-dimensional ensemble-variational hybrid data assimilation. *Mon. Wea Rev.*, To be submitted., their algorithm is renamed 4DEnVar, to better differentiate the algorithm from traditional 4DVar.

model, for a hurricane case. In general, the introduction of flow-dependent ensemble covariance into 3DVar or 4DVar improves the forecast results. In fact, for the NCEP operational Global Forecasting System (GFS), an EnKF-En3DVar hybrid DA system (Hamill et al. 2011b; Whitaker et al. 2011) based on an EnKF (Hamill et al. 2011b) and the operational Grid-Point Statistical Interpolation (GSI) 3DVar (Kleist et al. 2009a) was developed and operationally implemented in 2012, replacing GSI 3DVar. Wang et al. (2013) reported the testing results from the GSI-based En3DVar hybrid system for GFS at a reduced resolution.

It has been a general decision at NCEP that the hybrid DA approach will be applied to its regional models as well, including the North America Mesoscale (NAM) model and the recently implemented (on 1 May 2012) Rapid Refresh (RAP) system, the replacement to the Rapid Update Cycle (RUC, Benjamin et al. 2004). Towards this end, an EnKF system was recently established for the RAP and tested at a reduced resolution by Zhu et al. (2013) using the operational observation data stream of RAP. The ensemble square-root filter (EnSRF) algorithm of Whitaker and Hamill (2002) was used in our study. As one of the ensemble-based Kalman filter algorithms, we will use EnKF as a general name to refer to this algorithm. Short-range (up to 18 hours) forecasts from 3-hourly EnKF analyses over a 9-day period were found to be consistently better than forecasts from corresponding GSI analyses, in terms of both model state forecasts and precipitation forecast skill scores. The primary goal of this current work is to extend the work of Zhu et al. (2013) by establishing and testing a coupled EnKF-En3DVar hybrid DA system for RAP that can potentially be implemented operationally. As the first step, we test and evaluate the hybrid DA system running at 1/3 of the native resolution of operational RAP; running the EnKF DA system at this reduced resolution is dictated by the expected availability of operational computing resources in the near future, and while running the En3DVAR hybrid analyses at the same resolution facilitates easy and direct comparisons with the EnKF results, and provides us with a benchmark against which a future dual-resolution implementation can be compared against. With the dual-resolution implementation, the En3DVar analyses will be run on the higher, native, grid resolution, using the reduced-resolution ensemble perturbations. In this paper, we focus on results obtained from all three systems, i.e., the GSI 3DVAR, EnKF and En3DVar hybrid, at the reduced, 40-km grid spacing resolution. Their performances are inter-compared.

The rest of the paper is organized as follows. The coupled EnKF-En3DVar hybrid system for RAP is first described in section 2. Experimental setup and testing results are discussed in sections 3 and 4, respectively. Downscaled precipitation forecasts on the 13 km RAP grid, starting from interpolated 40-km En3DVar hybrid, EnKF and GSI analyses, are compared in section 5. Finally, section 6 provides conclusions and additional discussions.

2. GSI-based EnKF-En3DVar hybrid system for Rapid Refresh

a) The Rapid Refresh System

The operational hourly-updated RUC system was designed to improve short-range weather forecasting through frequent updating of initial conditions with the latest observations (Benjamin et al. 2004). The RAP is a replacement of the RUC system and is based on the non-hydrostatic WRF-ARW dynamic core (Skamarock et al. 2005). RAP has been operational at NCEP since May 1, 2012 and currently uses the GSI 3DVar for hourly data assimilation cycles. Recently, an upgraded experimental version of the RAP (planned for operational implementation at NCEP) has employed an ensemble 3DVar hybrid analysis, using covariance information obtained from the 80-member GFS EnKF system and shown improved error statistics relative to the NCEP operational RAP. The GSI is an unified DA framework for both global and regional models (Kleist et al. 2009a). The horizontal grid spacing of RAP is ~13 km and has 50 vertical levels extending up to 10 hPa at the model top. Compared to RUC, the RAP system is capable of assimilating more observations, including satellite radiance data, and has a larger domain which covers the entire North America. The physics options used by the operational RAP include the Grell-G3 cumulus parameterization, Thompson microphysics, RRTM longwave radiation, Goddard shortwave radiation, MYJ turbulent mixing, RUC-Smirnova land-surface model. Details on these schemes can be found in Benjamin et al. (2009).

As with the RUC, the RAP employs a digital filter initialization (DFI) to reduce high-frequency noise during the initial period of model integration. In the operational RAP system, twice DFI (TDFI) (Lynch and Huang 1992), which applies the DFI twice, once on the adiabatic backward time integration and once on the full-physics forward time integration, is used. Considering that for high-resolution applications where diabatic processes are more important, adiabatic integration can introduce significant errors, Zhu et al. (2013) chose to employ the digital filter launching (DFL) procedure (Lynch and Huang 1994) instead in their EnKF system for RAP; DFL applies the DFI only once, on the forward integration time series. In this study, the same procedure is followed by the EnKF and En3DVar hybrid experiments. In our tests with 3-hourly cycles reported in this paper, the DFL employs a 40-minute filter window centered at 20 minutes of forecast time, and used Dolph filter (Lynch 1997) with a cutoff half width of 20 minutes.

b) The coupled EnKF-En3DVar hybrid system for RAP

As mentioned earlier, our En3DVar hybrid system is based on the operational GSI 3DVar system for RAP and it uses the operational data stream of RAP. To facilitate direct comparisons with the RAP EnKF and GSI 3DVar systems as reported in Zhu et al. (2013), we run our hybrid tests also at the reduced resolution of ~40 km grid spacing with 3-hourly assimilation cycles instead of the ~13 km grid spacing and hourly cycles of the operational RAP. The use of the reduced-resolution EnKF system is due to the expected constraint in available operational computational resources in the near future while the choice of 3-hourly cycles is to enable us to run a larger number of experiments and for more rapid prototyping of the system. The running of the continuously cycled experiments over a 9-day period is computationally expensive in terms of both CPU and storage requirements. Extensive experimentations and tuning were required to arrive at quasi-optimal configurations of the RAP EnKF system, including configurations of covariance inflation and localization. For future operational implementation, it is desirable to run the En3DVar at the native RAP resolution while using lower-resolution EnKF perturbations in a dual-resolution model to save computational cost; the implementation and testing of the dual-resolution coupled hybrid system for RAP will be done in the future.

A one-way coupled EnKF-En3DVar hybrid system is made up of four key steps: 1) GSI-based observation processing that includes both quality control and calculation of a full set of observation innovations; 2) EnKF analyses using the innovations calculated by the GSI and the background ensemble forecasts to yield an ensemble of analyses; 3) An En3DVar analysis using the background ensemble forecasts from the EnKF cycle for flow-dependent covariance estimation; and 4) carrying out ensemble forecasts from the EnKF ensemble analyses and a single control forecast from the En3DVar hybrid analysis to the next analysis time.

Fig. 1 shows a flowchart for both one-way and two-way coupled EnKF-En3DVar analysis-forecast cycle as employed in this paper. For 1-way coupling between the EnKF and En3DVar, the EnKF system provides the background ensemble forecast perturbations to the ECV-based En3DVar hybrid variational analysis , but does not feed back to the EnKF system. Two-way coupling includes an additional step that re-centers the EnKF analysis ensemble on the En3DVar control analysis (the thick black arrows and bold black box in Fig. 1). The two-way coupling implicitly assumes that the En3DVar control analysis is better than the EnKF ensemble mean analysis, and the re-centering should help prevent the divergence between the EnKF and En3DVar analyses so that the ensemble perturbations can sample the control forecast uncertainty well; divergence between the two systems may occur when continuous cycles are run for a long period of time.

As pointed out earlier, the GSI-based En3DVar hybrid analysis is achieved using the ECV method (Wang 2010). Within this framework, the analysis increment $\delta\mathbf{x}$ is a sum of two terms, defined as

$$\delta\mathbf{x} = \delta\mathbf{x}_1 + \sum_{k=1}^{K} (\mathbf{a}_k \circ \mathbf{x}_k^{'}) \quad , \qquad (1)$$

where $\delta\mathbf{x}_1$ is the analysis increment associated with static BEC $\mathbf{B}$ and the second term on the right hand side is the increment associated with the ensemble covariance. $\mathbf{x}_k^{'}$ is the k^{th} ensemble background perturbation normalized by $\sqrt{K-1}$, where K is ensemble size. Vectors $\mathbf{a}_k (k = 1, \cdots, K)$ in the second term are the extended control variables. Analysis increment $\delta\mathbf{x}$ is obtained by minimizing the following cost function:

$$\begin{aligned} J(\delta\mathbf{x}_1, \mathbf{a}) &= \beta_1 J_b + \beta_2 J_e + J_o \\ &= \frac{1}{2}\beta_1 \delta\mathbf{x}_1^T \mathbf{B}^{-1} \delta\mathbf{x}_1 + \frac{1}{2}\beta_2 \mathbf{a}^T \mathbf{A}^{-1}\mathbf{a} + \frac{1}{2}[\mathbf{y}_o - H(\mathbf{x}_b + \delta\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}_o - H(\mathbf{x}_b + \delta\mathbf{x})], \end{aligned} \qquad (2)$$

which gives the solutions of partial increment $\delta\mathbf{x}_1$ and ECV $\mathbf{a}$. Vector $\mathbf{a}$ is formed by concatenating K vectors $\mathbf{a}_k$. Compared to a traditional 3DVar cost function, a weighted sum of J_b and J_o is replaced by the sum of weighted J_b and J_e terms and J_o, where J_b is the traditional background term associated with static covariance $\mathbf{B}$, J_o is the observation term as in traditional 3DVar. J_e is the additional term associated with flow-dependent covariance for the ECV. Weighting factors β_1 and β_2 are placed in front of J_b and J_e terms, respectively, and they are constrained by

$$\frac{1}{\beta_1} + \frac{1}{\beta_2} = 1, \qquad (3)$$

to conserve the total variances.

The ECVs are constrained by a block-diagonal matrix $\mathbf{A}$, which defines the ensemble covariance localization (Lorenc 2003; Wang et al. 2007). In the GSI-based En3DVar hybrid implementation, the horizontal and vertical covariance localizations, or the effects of matrix **A** in Eq. (2), are achieved by applying recursive filter transforms (Hayden and Purser 1995), analogous to the treatment of **B** in Eq. (2). The parameters in the recursive filter will determine the correlation length scale in $\mathbf{A}$ as a precondition and therefore prescribe the covariance localization length scale for the ensemble covariance. The vertical covariance localization scale (CLS) is measured in either scaled height (the natural log of pressure) or the number of model levels while the horizontal CLS is measured either in kilometers or number of grid points in GSI. In this study, the natural log of pressure is used for the vertical, and kilometer is used for the horizontal localization.

Apart from the variational minimization of the En3DVar hybrid cost function given by Eq. (2), a major component of the overall coupled EnKF-En3DVar hybrid DA system is an ensemble DA system that provides the perturbations, which in our case the EnKF system described in Zhu et al. (2013). This EnKF system uses the serial ensemble square-root filter (EnSRF) algorithm of Whitaker and Hamill (2002) and its configuration settings follow the control experiment of Zhu et al. (2013). To facilitate fair comparisons between the En3DVar hybrid and EnKF experiments, the CLSs in the En3DVar hybrid system are specified to match the CLSs used by the EnKF as closely as possible in the control experiments, and the vertical and horizontal scales are measured in the natural log of pressure and kilometers, respectively. The *e-folding* distance from the Gaspari and Cohn (1999) localization function is $\sqrt{2}\sqrt{0.3}S_{GC}$ (where S_{GC} is cut-off radii in the EnKF), while *an e-folding* distance from the recursive filter is $2\sqrt{2}S_{RF}$ (Barker et al. 2004; Wang et al. 2008c) (where S_{RF} is recursive filter localization length scale). Thus, to keep the same *e-folding* distance for both EnKF and En3DVar, the cut-off radii in the EnKF S_{GC} can be converted to the recursive filter localization length scale S_{RF} in hybrid according to

$$S_{RF} = \sqrt{0.15} S_{GC} / \sqrt{2} \qquad (4)$$

3. Experiment designs

a. Model, observations, ensemble configuration and verification techniques

The test period, model domains and boundary conditions used in this study are the same as in Zhu et al. (2013). DA experiments at ~40 km grid spacing are run in continuous 3-hourly cycles throughout the 9-day retrospective testing period from May 8 to 16, 2010; the cycles start at 0000 UTC 8 May 2010 and end at 2100 UTC 16 May 2010. The 40 km model domain (as shown in Fig. 2) covers the entire North America with 207x207 grid points. A slightly smaller domain at ~13 km grid spacing, as indicated by the bold rectangle in Fig. 2a, is used for forecasts at the native RAP resolution and for precipitation verification. The domains have 50 vertical levels. Eighteen-hour deterministic forecasts (after applying DFL) are launched every three hours from the En3DVar hybrid control analyses as well as EnKF ensemble mean analyses on the 40 km domain. Three-hourly ensemble forecasts are produced within the assimilation cycles of EnKF, which are fed into the En3DVar control analysis (Fig. 1). Two outer loops and 50 iterations, the same as in operational RAP GSI 3DVar, were utilized for all the En3DVar and GSI experiments. The 13 km deterministic forecasts start from interpolated 40 km analyses at 0000 and 1200 UTC for precipitation forecast evaluation. The lateral boundary conditions for both grids come from operational GFS forecasts; perturbations created using the random-CV method in the WRF 3DVar (Barker et al. 2004) are added to GFS forecast boundary conditions for the ensemble forecasts and to the GFS analysis initial condition at 0000 UTC May 8, 2010 to start the initial ensemble of EnKF.

The observations used in this study are the same as those used in the operational RAP except for the exclusion of satellite radiance data. The realtime RAP system collects data from 1.5 h before and 0.5 h after the time of analysis. However, for 0000 and 1200 UTC it waits half an hour longer for more data (such as sounding data) to arrive. In our tests, the data sets assimilated at 3 hourly intervals are the data sets collected and used by the operational hourly RAP system; as a result, observations that arrived in realtime outside the 2 hour (2.5 hours for 0000 and 1200 UTC) windows are not used. They include surface observations (land reporting stations, mesonets, ships, and buoys, etc.), upper air observations (radiosondes, aircrafts, wind profilers, VAD data and satellite retrieval winds) and GPS precipitable water (PW), the same as in Zhu et al. (2013) except for the exclusion of GPS perceptible water (PW) data there. The exclusion of the PW data in Zhu et al. (2012) was due to an initial problem with the EnKF code, which has since been fixed. The distributions of most major observation types are shown in Fig. 2. The satellite radiance data are not included in the experiments reported here because our preliminary tests suggested that bias correction remains an important issue within the system that would require careful treatment for positive impact. Our more recently tests with the radiance data using the EnKF show small although generally positive impacts and the results will be reported separately in the future. Initial studies of EnKF for NCEP GFS global model also excluded satellite radiance data (Whitaker et al. 2008).

The short-range deterministic forecasts from the En3DVar, EnKF ensemble mean and GSI analyses are verified against surface and sounding observations. The Model Evaluation Tools (MET) developed by the Development Tested Center (DTC) (Brown et al. 2009) are employed here. MET contains comprehensive verification metrics for both deterministic and probabilistic forecasts. Root-mean square error (RMSE) is used as the primary verification metric for the 40 km deterministic forecasts here. The RMSEs for temperature (T), relative humidity (RH), and wind components U and V are calculated against upper air soundings, and those for surface pressure P, 2-

m *RH*, 2-m *T* and 10-m *U* and *V* are calculated against surface observations.

The statistical significance of RMSEs is determined by using bootstrap resampling (Candille et al. 2007; Buehner and Mahidjiba 2010; Schwartz and Liu 2013). The RMSEs from all cycles are randomly selected 3000 times, and for these samples, the mean is calculated, along with a two-tailed 90% confidence interval from 5% to 95%. To determine whether the improvements from En3DVar on GSI 3DVAR is statistically significant, the mean RMSE differences between En3DVar and GSI 3DVar together with a 90% confidence interval is computed and plotted in each figure. The RMSE differences from all cycles are also randomly selected 3000 times, and for these samples, a two-tailed 90% confidence interval from 5% to 95% is calculated. The same technique is also applied to the differences between En3DVar experiments and EnKF_Ctl to determine whether the improvement of En3DVar over EnKF is statistically significant. That the bounds of a 90% confidence interval between the forecast pair are all lower than zero means RMSEs from the first experiment are always lower than the second one at the 90% confidence level, therefore the improvement from the first experiment over the second one is statistically significant at the 90% confidence level. Conversely, that zero is included within the bounds of the 90% confidence level denotes statistically insignificant situations (Schwartz and Liu 2013; Xue et al. 2013).

For the 12-hourly forecasts on the 13 km grid, the Gilbert skill score (GSS) (Gandin and Murphy 1992), also known as the equitable threat score (ETS), and frequency bias (BIAS) are used to verify precipitation forecasts against NCEP Stage IV precipitation data (Lin and Mitchell 2005). The error and skill scores are aggregated over all forecasts within the 9-day test period. The same evaluation procedure was used in Zhu et al. (2013) although they only presented the GSSs.

b. Assimilation experiments

Experiments performed in this study are listed in Table 1. First, well-tuned En3DVar hybrid

1-way (Hybrid1W_Ctl) and 2-way coupled (Hybrid2W_Ctl), EnKF (EnKF_Ctl) control and GSI experiments are compared. The EnKF control experiment, EnKF_Ctl, uses 40 ensemble members and corresponds to experiment EnKF_CtrHDL from Zhu et al. (2013) except additional GPS PW data in this study, and uses a single suite of physics parameterizations in the ensemble to keep the setup simple (so that the EnKF, GSI and the En3DVar experiments all use the same set of physics in the forecast model). The En3DVar hybrid control experiment assigns equal weights ($1/\beta_2 = 0.5$) to the static and ensemble BECs. The EnKF codes and configurations are the same as the EnKF control experiment in Zhu et al. (2013), except for the exclusion of GPS PW data there. A combination of static and adaptive covariance inflation is applied in EnKF as in Zhu et al. (2013).

There are mainly two sets of tunable parameters in the En3DVar hybrid scheme. One set is the covariance weighting factors, which define the weights placed on the BECs. Four sensitivity experiments test the relative weights given to the static and ensemble BECs, with $1/\beta_2$=0.1, 0.5, 0.9, 1.0 (Hybrid01, Hybrid05, Hybrid09, Hybrid10) corresponding to 1/10, 1/2, 9/10, 100% weight given to the ensemble BEC, respectively.

The other set of tunable parameters includes the horizontal and vertical CLSs applied to the covariances. For weighting factor $1/\beta_2$=0.5 with 1-way coupling, we test three horizontal CLSs S_h =192, 300 and 356 km in Hybrid_HS, Hybrid1W_Ctl, and Hybrid_HL, respectively (corresponding to cut-off radii of 700, 1095, 1300 km according to Eq. (4)); three vertical CLSs S_v= -0.1, -0.3 and -0.5 (corresponding to cut-off radii of 0.36, 1.1 and 1.8 according to Eq. (4)) are tested in Hybrid_VS, Hybrid1W_Ctl and Hybrid_VL, respectively. The minus sign is due to the use of ln(p) as the length measure. To facilitate the comparison with control experiment Hybrid1W_Ctl, the mean domain-average RMSE difference, defined as

$$D = \frac{1}{N}\sum_{k=1}^{N}(RMSE^{k}{}_{Hybrid*} - RMSE^{k}{}_{Benchmark}) \ , \qquad (5)$$

where N is the total number of cycles and k refers to the k^{th} cycle, is calculated between experiment $Hybrid*$; the benchmark experiment is Hybrid1W_Ctl here and $Hybrid*$ refers to one of hybrid sensitivity experiments.

All CLSs used in the En3DVar hybrid experiments described above are constant with height. However, the cut-off radii used in the well-tuned EnKF control experiment of Zhu et al. (2013) (EnKF_CtrHDL in their paper) are height- and observation-type dependent based on the vertical position of the observations. These localization settings are shown in Fig. 3. The horizontal cut-off radius r_{cut} at the model top is 1.5 times the value at the surface for all state variables; as shown in Fig. 3a, r_{cut} increases from 700 km at the surface to 1050 km at the model top. The vertical cut-off radius $\ln(p_{cut})$ is not only height dependent, but also observation-type dependent. For *RH* and *T* observations (solid line in Fig. 3b), the vertical cut-off radii at the model top and surface are set to a quarter of 1.1 and half of 1.1, respectively. For wind observations (dash line in Fig. 3b), $\ln(p_{cut})$ is twice as large as that for *RH* and *T* observations. For surface pressure observations and GPS PW data (which are most strongly linked to low-level moisture), their vertical localization radii are set to a constant value of 1.6. These settings were used in the control experiment of Zhu et al. (2013), and their choices were guided by the correlation scales found in the NMC-method-derived error statistics used by GSI and were further tuned based on sensitivity experiments.

In the En3DVar system, height-dependent localization is straightforward to implement, but not observation-type-dependent localization, because unlike the serial EnKF scheme, En3DVar analyzes all observations simultaneously and the localization is performed in the state instead of the observation space (Campbell et al. 2010). Theoretically, if the localization treatment were the same for the EnKF ensemble mean analysis as for the En3DVar analysis and when the ensemble-derived covariance is used at 100%, the results from the two algorithms should be very close. We observed differences between such EnKF and En3DVar analyses in our experiments, and want to see if localization is the main cause for these differences. We are interested in finding out if the height- and observation-dependent covariance localization treatments would have similar effects in En3DVar as in EnKF. These are examined in the next three experiments (Hybrid_Con, Hybrid_HD and Hybrid3G), all performed with 100% ensemble covariance and all used one-way coupling.

Hybrid_Con uses constant CLSs corresponding to the cut-off radii of EnKF_Ctl at the surface. In Hybrid_HD, the height-dependent horizontal CLSs are chosen to match the height-dependent cut-off radii of EnKF_Ctl closely, while the vertical CLSs for all variables are chosen to be the same as that for wind observations in EnKF_Ctl (Table 1).

The only way to apply different localization to different observations is to break the En3DVar analysis into multiple steps of analysis, with each step analyzing a sub-set or a sub-group of observations. To do this, the corresponding EnKF analysis that provides the ensemble perturbations also needs to be broken up into multiple steps and the EnKF and En3DVar need to be run in alternating order. The disk I/O costs reading and writing the ensembles will be much increased so will be the costs of En3DVar minimizations. Doing this significantly increases the overall computational costs for operational implementation (the costs of associated gridded data IO are also significant, apart from CPU costs) but is doable in a research mode. Towards this end, experiments EnKF3G and Hybrid3G are run, where each analysis is broken into 3 steps, with each step analyzing one of the three groups of observations consisting of 1) *RH* and *T*, *2) U* and *V*, 3) and *PS* and GPS PW data, respectively. Within each step, the EnKF ensemble analysis is followed by an En3DVar hybrid analysis step using the latest EnKF-updated ensemble perturbations.

Because the EnKF includes both static and adaptive covariance inflation (Zhu et al. 2013), it is difficult to maintain the same amount and effects of inflation when each EnKF analysis in broken into three steps. Applying the static inflation every EnKF sub-step can over-inflate the covariance, while applying it only in the last step would change the overall behavior of the filter. Because our primary goal here is to determine if the difference between the EnKF and En3DVar analyses (with 100% ensemble covariance) is primarily caused by the observation-based localization, to avoid the above issue, we run EnKF3G without any covariance inflation and examine the RMSE differences between the EnKF and En3DVar analyses. We just need to find out if the En3DVar hybrid analyses are closer to the EnKF analyses when observation-type dependent localization is similarly used in the En3DVar through the split-step procedure.

The mean domain average absolute RMSE difference, defined as

$$DB = \frac{1}{N}\sum_{k=1}^{N}\left| RMSE^{k}{}_{Hybrid*} - RMSE^{k}{}_{Benchmark} \right| \ , \qquad (6)$$

is used to measure how close the En3DVar and EnKF analyses are. The differences between Hybrid_Con and EnKF_Ctl, Hybrid_HD and EnKF_Ctl, Hybrid3G and EnKF3G (Table 2) will be calculated to examine the impacts of constant localization, height-dependent localization, and observation-dependent localization, respectively. The statistical significance of DB is also determined by using bootstrap resampling. The DBs at cycles are randomly selected 3000 times, for this sample; a mean is calculated, along with a two-tailed 90% confidence interval from 5% to 95%. If the error bars from the experiments pair do not overlap, the differences between En3DVar and EnKF are significantly reduced at the 90% confidence level.

Finally, to see how the En3DVar hybrid scheme compares with the EnKF and GSI for smaller ensemble sizes, we run the EnKF and 1-way and 2-way coupled En3DVar hybrid with 20 instead of 40 ensemble members, and the experiments are called EnKF20, Hybrid1W20 and Hybrid2W20 (Table 1). The ensemble covariance is used at 50% in the En3DVar hybrid analyses. The relative percentage improvement (RPI) comparing to experiment GSI

$$RPI = \left(\frac{1}{N}\sum_{k=1}^{N} RMSE^{k}_{\exp *} - \frac{1}{N}\sum_{k=1}^{N} RMSE^{k}_{GSI}\right) / \frac{1}{N}\sum_{k=1}^{N} RMSE^{k}_{GSI} \qquad (7)$$

are used to discuss the results of these experiments in the following sections. The *exp** refer to one of ensemble size sensitivity experiments.

4. Results of experiments

a. Single observation tests

Single-observation tests are often performed to verify DA code correctness and evaluate algorithm behaviors. EnKF differs from 3DVar in its use of flow-dependent BEC derived from the forecast ensemble. For the En3DVar scheme that uses a combination of static and flow-dependent covariances, the analysis increment from a single observation should be somewhere between those of EnKF and 3DVar, which represent the extreme ends of the En3DVar hybrid analysis, corresponding to 100% and 0% use of the ensemble covariance, respectively. Single observation tests also reveal clearly how spatial covariance localization works, or if it works as expected. Here, we place a temperature observation at 500 hPa over Norman, Oklahoma, with a 1 K innovation over the background and an observation error standard deviation of 0.8 K. The background ensemble for the single observation test is taken from the EnKF system after 5 days of 3-hourly analysis cycles employing the full set of observations; the GSI and En3DVar analyses use the mean of the 3-hour ensemble forecasts as the background, therefore the background used by the En3DVar, EnKF and GSI are the same. The key parameter settings used in these tests are the same as the corresponding control experiments with full data sets.

The resulting GSI analysis increment has a circular shape in T (Fig. 4a), reflecting its static, isotropic spatial covariance structure while the T increment of EnKF is stretched along the direction of geopotential height contours (Fig. 4b) reflecting the flow-dependent covariance structures. With $1/\beta_2$ = 0.5, the En3DVar hybrid T increment is also stretched along the direction of geopotential height contours (Fig. 4c) but not as much as in the case of EnKF and the increment is broader.

Overall, it lies in-between the increments of EnKF and GSI. In all cases, the maximum increment is a little over 0.4 K, about half of the observation innovation, which is consistent with expectations.

In the vertical cross section, the T increment of GSI is not exactly elliptic but has a tendency to follow the terrain-following coordinate surfaces (Fig. 4d). This is because the (isotropic) recursive filters used to model the static BEC are applied along coordinate lines (Purser et al. 2003b). As a consequence of the balance operators and background error statistics used in the GSI (Kleist et al. 2009a), the wind increment is close to zero at the observation level, cyclonic below and anticyclone above the observation; these structures are consistent with the thermal wind balance. The locations of the T increment maxima of EnKF (Fig. 4e) and En3DVar (Fig. 4f) are shifted slightly above the observation location, and the increments are wider and deeper for the En3DVar than for EnKF. The latter is because for an observation at 500 hPa, the constant CLS for En3DVar is about 100 km wider in the horizontal, and about 0.6 deeper in the vertical than the corresponding height-dependent cut-off radii of EnKF at the same height level as indicated by Fig. 3. The wind increments for the EnKF are more complicated (Fig. 4e); they do not show the simple thermal wind balance, indicating significant unbalanced components in the analysis. Their magnitudes are about twice as large as the GSI wind increments. For the En3DVar hybrid, the wind increments appear to be a combination of the GSI and EnKF wind increments, containing a larger-scale balanced component also (Fig. 4f).

Overall, we see that the En3DVar hybrid analysis increments appear to be a combination of the GSI and EnKF analysis increments, reflecting the combined use of static and flow-dependent background covariances (c.f., Eq. 1). Other single-observation experiments with different CLSs and different covariance weights show that the En3DVar hybrid system responds as expected to the changes in these parameters (results not shown). These results suggest that the En3DVar hybrid system works correctly.

b. GSI, EnKF and En3DVar hybrid control experiments

The RAP system had been run experimentally in real-time for several years at the NOAA Earth System Research Laboratory (ESRL) before being officially implemented at NCEP in May 2012. In this study, we borrow from a recent configuration of the experimental 13-km RAP for our 40-km grid spacing tests.

In this section, we present and compare the results from the En3DVar hybrid 1-way (Hybrid1W_Ctl) and 2-way coupled (Hybrid2W_Ctl), EnKF (EnKF_Ctl) control experiments, and those of the GSI experiment.

The RMSE profiles of the 3-hour forecasts verified against sounding data are shown in Fig. 5. These forecasts were launched from the GSI, EnKF ensemble mean, and En3DVar hybrid analyses. The RMSE for each pressure level was calculated by averaging values obtained from all cycles within a layer 50 hPa above and below that pressure, except for the lowest and topmost levels. The RMSEs of EnKF_Ctl are overall lower than those of GSI except for the temperature at the upper levels where the error can be ~0.1 K greater. The performances of one-way and two-way coupled En3DVar hybrid schemes are very close. With half static and half flow-dependent covariances in these experiments, Hybrid1W_Ctl and Hybrid2W_Ctl outperform GSI, and are also generally better than EnKF_Ctl except for *RH* above 500 hPa, *V* at 100 hPa, and *T* below 900 hPa.

The average RMSEs for all levels over the entire domain are shown in Fig. 6 for forecast hours 3 through 18. Generally, both EnKF and En3DVar hybrid significantly outperform GSI for all the variables throughout the forecast period at the 90% confidence level (the intervals of error differences do not include zero). For *RH*, the average RMSEs of En3DVar hybrid are slightly higher than those of EnKF_Ctl by 9 hours, which appears to be related to the larger errors at 3 hours at the upper levels (Fig. 5a); they become slightly smaller after 9 hours. However, the improvement of En3DVar hybrid over EnKF for *RH* is not statistically significant. For *T* and *U*, the domain-averaged RMSEs of En3DVar hybrid are significantly and consistently smaller than those of GSI and EnKF throughout the forecast period (Fig. 6b,c). For *V*, the errors of the En3DVar and EnKF are very similar and are all clearly lower than those of GSI. The reason that En3DVar performs better than EnKF for *U* may be related to the dominance of the east-west flows that may increase the validity of the static covariance. Overall, the En3DVar hybrid out-performs GSI and EnKF for *T* and *V* for the 18 hours of the forecast.

Fig. 7 shows the average RMSEs for 3-18 hour forecasts against surface observations. For 2 m *T* and 10 m *U*, EnKF and En3DVar outperform GSI at all forecast hours significantly, with the EnKF significantly outperforming the En3DVar hybrid at most forecast hours. For 2 m *RH* and 10 m *V*, EnKF occasionally underperforms GSI slightly but at most forecast hours it is better. The En3DVar hybrid schemes improve over EnKF further, enough to ensure better or equal performance than GSI for all hours, and more clearly so for *RH*. For surface pressure, EnKF underperforms GSI initially but becomes better after 9 hours; throughout the forecast period, the En3DVar hybrid outperforms both GSI and EnKF significantly. In general, there is little difference between the 1-way and 2-way En3DVar hybrid schemes; this result may be due to the relative short 9-day testing period; if the cycles were run for a much longer time period, a larger divergence between the EnKF and En3DVar hybrid may develop in a 1-way coupling mode, then 2-way coupling would show a bigger advantage. When the En3DVar hybrid runs at a higher resolution than the EnKF in a dual-resolution mode, there may also be more beneficial with the 2-way coupling.

Overall, the En3DVar hybrid schemes significantly outperform GSI 3DVar for all the variables at all forecast hours for sounding and surface observations; and are comparable, even better than EnKF for some variables. The results indicate the benefit of combining the static and flow-dependent covariances. In the next section, the sensitivity to the covariance weighting factors is examined.

Finally, one may have concern that the 9-day cycled assimilation period is not long enough for the ensemble DA system to spin up (over the course of evaluating and testing our EnKF and En3DVar hybrid systems, we had run over 100 cycled experiments so extending the experiment period would be expensive). To answer this question, we examine how the short-range forecast errors evolve through the 9-day period. Fig. 8 shows the domain-averaged 3-hour forecast RMSEs verified against sounding data at 0000 and 1200 UTC through the test period. We can see that the relative performances of GSI 3DVar, EnKF and En3DVar hybrid do not change much throughout the 9-day period, even in the earlier days. These results indicate that the ensemble system had spin-up rather quickly.

c. Sensitivity to covariance weighting factors

We perform a set of four 1-way-coupled En3DVar hybrid experiments with $1/\beta_2$ =0.1, 0.5, 0.9 and 1.0, which are the weights given to the ensemble covariance. The one with $1/\beta_2$ =0.5, called Hybrid05 here, is the same as experiment Hybrid1W_Ctl discussed earlier (Table 1). The vertical profiles of 3 hour forecast RMSEs verified against sounding data are shown in Fig. 9 for these four experiments, GSI, and EnKF_Ctl. It can be seen that the En3DVar hybrid and EnKF schemes generally outperform GSI, except for *RH* at 700 to 400 hPa levels for Hybrid09 and Hybrid10, i.e., the hybrid scheme with 90 or 100% ensemble covariances. Introducing 10% static covariance into the En3DVar hybrid framework reduces the error slightly (comparing Hybrid09 to Hybrid10 for *RH*), and further increasing it to 50% brings the *RH* errors below those of GSI at all levels (Fig. 9a).

The average RMSEs for all levels over the entire domain are shown in Fig. 10 for forecast hours 3 through 18. All En3DVar hybrid experiments significantly outperform GSI 3DVar for all variables throughout the entire forecast period at the 90% confidence level, except for the *RH* of Hybrid10 after 9 hours. The errors of Hybrid05 are about the lowest among all En3DVar hybrid experiments, while errors of Hybrid10 are the greatest and significantly worse than those of EnKF_Ctl. RMSE differences between Hybrid01 and EnKF_Ctl are generally smaller than those between Hybrid09 and EnKF_Ctl for *T*, *U* and *V*.

Overall, introducing 10% ensemble covariance into the variational framework in Hybrid01 has a much larger impact (compare Hybrid01 to GSI) than adding 10% static covariance into the En3DVar hybrid framework (compare Hybrid09 to Hybrid10), and the errors of Hybrid01 are generally between those of Hybrid05 and GSI and are closer to those of Hybrid05, especially for wind fields. Hybrid05 gives the smallest errors on average.

It can also be noticed from Fig. 9 that EnKF_Ctl outperforms Hybrid10, except for wind between 500 and 200 hPa levels. As pointed out earlier, if covariance localization treatments were the same in EnKF and Hybrid10, their results should be very close given that the ensemble covariance is used at 100% in both cases. The use of height- and observation-type-dependent covariance localization in the EnKF but not in the En3DVar hybrid is suspected to be the main cause of the differences; it had been found to help improve the EnKF results in Zhu et al. (2013) but is not used within the En3DVar hybrid. In the following section, we want to see if doing something similar within the En3DVar hybrid framework can improve the En3DVar hybrid results too.

d. Sensitivity to ensemble covariance localization

In Zhu et al. (2013), several tests with the horizontal and vertical covariance localizations were performed. In this paper, the EnKF experiment uses the same configuration as experiment

EnKF_CtrlHDL of Zhu et al (2013), with height- and observation-type-dependent localization radii. For the En3DVar analysis, covariance localization also requires tuning. Because En3DVar realizes covariance localization in the state or grid point space, it is difficult if at all possible to use observation-type-dependent localization.

In this section, we first look at the experiments that use smaller or larger horizontal and vertical CLSs than those used in Hybrid1W_Ctl. For weighting factor $1/\beta_2$=0.5, we test three horizontal CLSs, S_h=700, 1095 and 1300 km (in Hybrid_HS, Hybrid1W_Ctl, and Hybrid_HL, respectively) and three vertical CLSs S_v= 0.36, 1.1 and 1.8 (for Hybrid_VS, Hybrid1W_Ctl and Hybrid_VL, respectively). The domain-average forecast RMSE differences between 3 hour forecasts and those of Hybrid1W_Ctl are shown in Fig. 11. When the CLSs increase or decrease from those of control experiment, the En3DVar hybrid performs worse for almost all variables, except for *T* when the horizontal CLS is increased (Fig. 11). However, even though reduced CLSs are not preferred according to Fig. 11, the *RH* errors are reduced at levels above 800 hPa when using reduced CLSs (not shown), suggesting that we may be able to benefit from the use of observation-type and/or height-dependent CLSs, as in the case of EnKF (Zhu et al. 2013). Doing so may also help further improve the En3DVar hybrid performance.

For a better comparison with EnKF_Ctl, height-dependent localization is introduced into the En3DVar hybrid framework when the ensemble covariance is used at 100% (Hybrid_HD). As shown in Fig. 12, Hybrid_HD outperforms Hybrid_Con (with constant localization scale and 100% ensemble covariance) and is much closer to EnKF_Ctl for *RH*, *U* and V. For wind, Hybrid_HD is even slightly better than EnKF_Ctl at the middle levels (Fig. 12c,d) while Hybrid_Con is poorer than EnKF_Ctl at all levels. For *RH*, EnKF_Ctl still has smaller RMSEs than hybrid_HD above 700 hPa, and this appears to be because of the smaller cut-off radii in the EnKF when observation-type-dependent localization is used.

The cut-off radii used in the EnKF control experiment are height- and observation-type dependent. Within the serial EnKF algorithm where observations are assimilated one by one, and localization is applied to the covariances between individual observation and the state variables, observation-type-dependent localization can be easily implemented. However, the En3DVar hybrid algorithm analyzes all observations simultaneously by variational minimization in the state space in which covariance localization is applied (Campbell et al. 2010), it is impossible to apply the observation-type-dependent localization used by EnKF within En3DVar using a single analysis step.

To study the differences caused by observation-type-dependent localization scales, we break each coupled EnKF-En3DVar analysis into three sub-steps of coupled EnKF (EnKF3G) - En3DVar (Hybrid3G) analyses, with each step analyzing a sub-group of observations that share the same height-dependent localization scales. Here, we use the absolute RMSE differences between a pair of hybrid and EnKF experiments (Table 2) together with the 90% confidence interval as determined by the bootstrap resampling procedure to determine the statistical significance of the differences. When the error bars do not overlap, we consider the RMSE differences between En3DVar and EnKF statistically significant. As shown in Fig. 13, the RMSE differences are reduced, by about 1/4 to 1/3 for *RH*, *U* and V when height-dependent localization is used in Hybrid_HD compared to Hybrid_Con, and the reduction is smaller and statistically significant except for *T*. When observation-type-dependent localization is used, the differences between Hybrid3G and EnKF3G are further reduced significantly for *RH*, *U* and *V*. For *RH* (which has the largest difference between the En3DVar hybrid and EnKF according to Fig. 12a), the RMSE difference is about 0.5% versus the 1.25% of the constant localization case. The reductions for T, *U* and *V* are smaller but still evident.

Fig. 14 shows the profiles of the RMSE differences together with the 90% confidence interval. In reference to Fig. 12, those levels where domain average absolute RMSE differences of 1GHD are greater than 1GC correspond to the levels where the Hybrid_HD outperforms Hybrid_HC, given that EnKF_Ctl is generally the best among the three experiments. For 3GHD, the average absolute RMSE differences are the smallest for *RH* at all levels, for *T* above 800 hPa and for *U* and *V* above 600 hPa. For *U* and *V*, the 3GHD differences are slightly larger below 700 hPa than 1GHD and clearly smaller than 1GC. These results show that when similar height- and observation-type-dependent covariance localizations are used in the En3DVar framework using 100% ensemble covariance, some of the differences between the EnKF and En3DVar analyses are significantly reduced, and such localization treatment generally brings the En3DVar results closer to the better EnKF results. The reduction in the RMSE differences for *RH* is greater than those for *T*, *U* and *V* when height- and observation-type-dependent localization are used. Because the relative humidity tends to contain smaller scale structures than temperature and wind fields and can benefit from tighter localization more when using height- and observation-type-dependent localization. However, because there are still differences between the EnKF and En3DVar algorithm, some differences still exist between their results, as indicated by the red bars in Fig. 13. When the ensemble covariance is used at 50%, height-dependent localization did not improve the En3DVar hybrid results as much as in the 100% case (not shown).

In summary, the use of height-dependent localization in the En3DVar hybrid framework when using full ensemble covariance improves the resulting model forecasts at almost all levels and forecast hours. Height- and observation-type-dependent localizations used in EnKF are responsible for about half of the differences between the EnKF and the En3DVar with full ensemble covariance.

e. Sensitivity to ensemble size

Previous studies (e.g., Hamill and Snyder 2000; Wang et al. 2008a) had found that the En3DVar hybrid system is more robust than EnKF when the ensemble size is small. As mentioned earlier, EnKF20, Hybrid1W20 and Hybrid2W20 use 20 instead of 40 ensemble members for control experiments (EnKF40, Hybrid1W40 and Hybrid2W40 corresponding to EnKF_Ctl, Hybrid1W_Ctl and Hybrid2W_Ctl respectively in this section). Fig. 15 shows the relative percentage improvement (RPI) of EnKF and 1-way coupled En3DVar hybrid compared to GSI for different forecast hours. The RPIs of experiments with 20 members are compared with the corresponding control experiments using 40 members. A negative RPI indicates an improvement (error reduction) over GSI. With an ensemble size of only 20, EnKF20 performs the worst for the variables and almost all forecast hours, and even worse than GSI for *RH* at all forecast hours, *T* at 12 hours and *U* at 3 hours. The EnKF and En3DVar hybrid with 40 ensemble members almost always improve more than the corresponding ones with 20 members. The improvement of the En3DVar hybrid over EnKF for 20 members is consistent for all variables and all forecast hours, and actually reverses the direction of improvement with 40 members (i.e., better rather worse) for *RH* at 3 and 6 hours and *V* at 18 hours. These results show even greater value of the static covariance utilized through the En3DVar hybrid framework when the ensemble size is small. The 2-way interactive En3DVar hybrid experiment Hybrid2W20 compared with Hybrid2W40 has very similar relative performance as the corresponding En3DVar hybrid 1-way coupling experiments with 20 and 40 members; the results are therefore not shown.

5. Precipitation forecast skills on 13-km grid

In this section, precipitation forecasts on the 13 km grid initialized from the 40 km GSI, EnKF_Ctl ensemble mean, Hybrid1WCtl and Hybrid2WCtl analyses (labeled as GSI13, EnKF13, Hybrid1W13 and Hybrid2W13, respectively) are compared. Considering extensive CPU and storage requirements, we launched the forecasts only twice a day at 00 and 12 UTC. The precipitation forecasts are verified against the NCEP Stage IV precipitation data. GSSs calculated for the 0.1, 1.25 and 2.5 mm h^{-1} thresholds are calculated as in Zhu et al. (2013).

The GSSs and BIASs are shown in Fig. 16. Both EnKF and En3DVar hybrid outperforms GSI on average for all forecast hours and thresholds shown. EnKF13 has higher GSSs than Hybrid for 0.1 mm h^{-1} after 3 hours. For greater thresholds of 1.25 and 2.5 mm h^{-1}, forecasts of Hybrid1W13 are comparable to EnKF13 by 7 hours, and are better than Hybrid2W13 during the first four hours, which is consistent with the *RH* domain-averaged RMSEs verified against sounding (Fig. 6a). Fig. 6d,e show that EnKF generally has the highest positive BIASs. The En3DVar hybrid scheme has the lowest BIASs in the first 5 hours of the forecast, and values between those of GSI and EnKF after 5 hours. The differences among the biases are relatively small and all of them have positive biases for both thresholds examined. Overall, in terms of GSS, the En3DVar hybrid outperform GSI at almost all forecast hours, and are comparable to EnKF13 for the larger thresholds at the early forecast hours but become slightly worse at the later hours. The somewhat the worse performance of precipitation forecasts with the En3DVar hybrid appears to be consistent with the deterioration of humidity forecasts of the En3DVar hybrid scheme compared to EnKF even though there is a general improvement with other variables (Fig. 15).

6. Summary and discussions

In this paper, a coupled EnKF-En3DVar hybrid data assimilation system based on the NCEP operational GSI variational framework is established and tested for the Rapid Refresh (RAP) forecasting system. It uses a recently developed, well-tuned, 40-member EnKF system (Zhu et al. 2013) to update and provide the ensemble perturbations. A 9-day spring period starting from May 8, 2010 that contains active convection is used to examine the performance of the system through comparisons with parallel experiments using the EnKF and GSI 3DVar. The En3DVar hybrid, EnKF and GSI experiments use the same observational data sets as the operational RAP system except for the exclusion of satellite radiance data. The experiments are performed at a reduced resolution of ~40 km grid spacing with 3-hourly assimilation cycles rather than at the native 13 km grid spacing with hourly cycles of the operational RAP. The systems are evaluated based on forecast RMSEs verified against surface observations and upper air sounding data for 3 to 18 hour forecasts. The effects of static and ensemble covariance weighting factors, covariance localization configurations, and ensemble size are also examined through sensitivity experiments.

With equal weighting for the ensemble and static covariances, the En3DVar hybrid scheme outperforms GSI for all variables at all levels with statistical significance, and are slightly better than EnKF, especially for later forecast hours. The En3DVar hybrid scheme benefits from the combined use of static and ensemble covariances. Introducing 10% flow-dependent covariance into the standard 3DVar framework has a much bigger positive impact than including 10% static covariance in the En3DVar framework. The forecasts from En3DVar analyses with 100% ensemble covariance and constant CLSs are worse than those from pure EnKF analyses using height- and observation-type-dependent covariance localization, especially for relative humidity. The height-dependent localization scheme in which the horizontal localization cut-off radii increase with height, and the observation-type-dependent localization scheme in which the cut-off radii for relative humidity and temperature observations are set to be smaller than those for winds led to smaller forecast RMSEs for the pure EnKF, especially at the high and low levels. Using similar height-dependent localization, the RMSEs of En3DVar with 100% ensemble covariance are significantly reduced and become close to those of pure EnKF. When using similar observation-type-dependent covariance localization by running the coupled EnKF-En3DVar analyses in three steps with each of the steps analyzing a subset of observation variables (in a similar way as in

EnKF), the results of the En3DVar with 100% ensemble covariance become even closer to those of EnKF. The benefit of height- and observation-type-dependent localization is negligible when the ensemble covariance is used at 50%. The multi-step EnKF-En3DVAR analysis procedure is, unfortunately, not very practical due to much increased computational costs. It is straightforward for pure EnKF because the algorithm is serial, where observations are assimilated sequentially.

Deterministic forecasts were launched on a 13 km grid from interpolated 40-km En3DVar hybrid control (with equal weighting for static and ensemble covariances), EnKF ensemble mean and GSI analyses at 0000 and 1200 UTC of each day. Hourly accumulated precipitation is better predicted in the En3DVar hybrid and EnKF experiments than GSI. But for a threshold of 0.1 mm h^{-1}, En3DVar hybrid does not improve the precipitation forecast as much as the EnKF does. This appears to be consistent with the lack of improvement in the humidity forecast when using constant localization. Apart from the differences in the localization in the two schemes, a thorough understanding of the cause of the poorer precipitation forecasts for this threshold than EnKF will require much investigation.

Despite the encouraging results, the En3DVar hybrid system still has much room for further improvement. Adding satellite and radar data and examining their impacts are among the desired tasks, as is a dual-resolution implementation where the En3DVar is performed at a higher resolution than the EnKF cycles. These aspects will be pursued in future studies.

Acknowledgments. This work was primarily supported by the FAA Aviation Weather Research Program through grant DOC-NOAA NA080AR4320904, and by MOST of China grant 2013CB430100. Additional support was provided by NSF grants AGS-0802888, OCI-0905040, AGS-0941491, AGS-1046171, and AGS-1046081, and by the NOAA Warn-on-Forecast program. The 4th author is also supported by NOAA THORPEX NA08OAR4320904, NASA NIP NNX10AQ78G and NOAA HFIP Grant NA12NWS4680012.

References

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129,** 2884-2903.

Anderson, J. L., 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, **59,** 210-224.

Anderson, J. L., 2012: Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation. *Mon. Wea. Rev.*, **140,** 2359-2371.

Barker, D. M., W. Huang, Y. R. Guo, A. J. Bourgeois, and Q. N. Xiao, 2004: A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results. *Mon. Wea. Rev.*, **132,** 897-914.

Benjamin, S. G., D. Dévényi, S. S. Weygandt, K. J. Brundage, J. M. Brown, G. A. Grell, D. Kim, B. E. Schwartz, T. G. Smirnova, T. L. Smith, and G. S. Manikin, 2004: An hourly assimilation–forecast cycle: The RUC. *Mon. Wea. Rev.*, **132,** 495-518.

Benjamin, S. G., W. R. Moninger, S. S. Weygandt, M. Hu, D. Devenyi, J. M. Brown, T. Smirnova, J. Olson, C. Alexander, K. Brundage, G. Grell, S. Peckham, T. L. Smith, S. R. Sahm, and B. Jamison, 2009: Technical review of rapid refresh/RUC project, NOAA/ESRL/GSD internal review, 3 Nov 2009.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129,** 420.

Bonavita, M., L. Torrisi, and F. Marcucci, 2008: The ensemble Kalman filter in an operational regional NWP system: Preliminary results with real observations. *Quart. J. Roy. Meteor. Soc.*, **134,** 1733-1744.

Bonavita, M., L. Isaksen, and E. Hólm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **138,** 1540-1559.

Brown, B. G., J. H. Gotway, R. Bullock, E. Gilleland, and D. Ahijevych, 2009: The Model Evaluation Tools (MET): Community tools for forecast evaluation. *25th Conf. Int. Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Paper 9A.6.

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. and A. Mahidjiba, 2010: Sensitivity of Global Ensemble Forecasts to the Initial Ensemble Mean and Perturbations: Comparison of EnKF, Singular Vector, and 4D-Var Approaches. *Monthly Weather Review*, **138,** 3886-3904.

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

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

Burgers, G., P. J. v. Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126,** 1719-1724.

Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Wea. Rev.*, **138,** 282-290.

Candille, G., C. Côté, P. L. Houtekamer, and G. Pellerin, 2007: Verification of an Ensemble Prediction System against Observations. *Monthly Weather Review*, **135,** 2688-2699.

Courtier, P., E. Andersson, E. Keckley, J. Pailleux, D. Vasiljevic, M. Hamurd, A. Hollingsworth, F. Rabier, and M. Fisher, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: formulation. *Quart. J. Roy. Met. Soc.*, **124,** 1783-1808.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99,** 10143-10162.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53,** 343-367.

Fujita, T., D. J. Stensrud, and D. C. Dowell, 2007: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev.*, **135,** 1846-1868.

Gandin, L. S. and A. H. Murphy, 1992: Equitable skill scores for categorical forecasts. *Mon. Wea. Rev.*, **120,** 361-370.

Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Ocean. Tech.*, **21,** 457-469.

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

Ge, G., J. Gao, M. Xue, and K. K. Droegemeier, 2012: Diagnostic pressure equation as a weak constraint in a storm-scale three dimensional variational radar data assimilation system. *J. Atmos. Ocean. Tech.*, **29,** 1075-1092.

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

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129,** 2776-2790.

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

Hamill, T. M., J. S. Whitaker, D. T. Kleist, M. Fiorino, and S. G. Benjamin, 2011b: Predictions of 2010’s tropical cyclones using the GFS and ensemble-based data assimilation methods. *Mon. Wea. Rev.*, **139,** 3243-3247.

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

Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126,** 796-811.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133,** 604-620.

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

Jung, Y., M. Xue, and M. Tong, 2012: Ensemble Kalman filter analyses of the 29-30 May 2004 Oklahoma tornadic thunderstorm using one- and two-moment bulk microphysics schemes, with verification against polarimetric data. *Mon. Wea. Rev.*, **140,** 1457-1475.

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

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, R. M. Errico, and R. Yang, 2009b: Improving incremental balance in the GSI 3DVAR analysis system. *Mon. Wea. Rev.*, **137,** 1046-1060.

Le Dimet, F. X. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A,** 97-110.

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

Lin, Y. and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. Hydrology*, San Diego, CA, Amer. Meteor.Soc., Paper 1.2.

Liu, C. and M. Xue, 2013: A unified framework for four-dimensional ensemble-variational hybrid data assimilation. *Mon. Wea Rev.***,** To be submitted.

Liu, C. and Q. Xiao, 2013: An ensemble-based four-dimensional variational data assimilation scheme. Part III: Antarctic applications with Advanced Research WRF (ARW) using real data. *Mon. Wea. Rev.*, **141,** 2721-2739.

Liu, C., Q. Xiao, and B. Wang, 2008: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.*, **136,** 3363-3373.

Lorenc, A., 2003: The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129,** 3183-3204.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112,** 1177-1194.

Lynch, P., 1997: The Dolph-Chebyshev window: A simple optimal filter. *Mon. Wea. Rev.*, **125,** 655-660.

Lynch, P. and X.-Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.*, **120,** 1019-1034.

Lynch, P. and X.-Y. Huang, 1994: Diabatic initialization using recursive filters. *Tellus*, **46A,** 583-597.

Meng, Z. and F. Zhang, 2007: Tests of an Ensemble Kalman Filter for Mesoscale and Regional-Scale Data Assimilation. Part II: Imperfect Model Experiments *Mon. Wea. Rev.*, **135,** 1403–1423

Mizzi, A. P., 2012: Comparison of a GSI/ETKF Regional Hybrid Using WRF/ARW with GSI/LETKF and GSI/EnKF Regional Hybrids. *16th Conference on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface*, New Orleans, LA, Amer. Met. Soc.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120,** 1747-1763.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003a: 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.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003b: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131,** 1524-1535.

Putnam, B. J., M. Xue, Y. Jung, N. A. Snook, and G. Zhang, 2013: The analysis and prediction of microphysical states and polarimetric variables in a mesoscale convective system using double-moment microphysics, multi-network radar data, and the ensemble Kalman filter. *Mon. Wea. Rev.***,** Accepted.

Rabier, F., H. Jarvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Met. Soc.*, **126,** 1143-1170.

Raynaud, L., L. Berre, and G. Desroziers, 2011: An extended specification of flow-dependent background error variances in the Meteo-France global 4D-Var system. *Quart. J. Roy. Meteor. Soc.*, **137,** 607-619.

Schwartz, C. S. and Z. Liu, 2013: Convection-permitting forecasts initialized with continuously cycling limited-area 3DVAR, ensemble Kalman filter, and "hybrid" variational-ensemble data assimilation systems. *Monthly Weather Review*.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. D. Powers, 2005: A Description of the Advanced Research WRF Version 2, 88 pp.

Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoins vorticity equation. Part I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113,** 1311-1328.

Tong, M. and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS Experiments. *Mon. Wea. Rev.*, **133,** 1789-1807.

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., C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble/VAR hybrid analysis schemes. *Mon. Wea. Rev.*, **135,** 222–227.

Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiment. *Mon. Wea. Rev.*, **136,** 5132-5147.

Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008b: 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., D. M. Barker, C. Snyder, and T. M. Hamill, 2008c: A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part II: Real Observation Experiments. *Monthly Weather Review*, **136,** 5132-5147.

Wang, X., T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 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., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP Global Forecast System: Single resolution experiments. *Mon. Wea Rev.***,** Accepted.

Whitaker, J., D. T. Kleist, X. Wang, and T. Hamill, 2011: Tests of a hybrid variational-ensemble global data assimilation system for hurricane prediction. *24th Conf. Wea. Forecasting/20th Conf. Num. Wea. Pred.*, Amer. Meteor. Soc., Paper J16.2.

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

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136,** 463-482.

Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130,** 2905-2916.

Xue, M., Y. S. Jung, and G. F. Zhang, 2010: State estimation of convective storms with a two-moment microphysics scheme and an ensemble Kalman filter: Experiments with simulated radar data. *Quart. J. Roy. Meteor. Soc.*, **136,** 685-700.

Xue, M., J. Schleif, F. Kong, K. W. Thomas, Y. Wang, and K. Zhu, 2013: Track and Intensity Forecasting of Hurricanes: Impact of Convection-Permitting Resolution and Global

Ensemble Kalman Filter Analysis on 2010 Atlantic Season Forecasts. *Wea. Forecasting*. In press.

Zhang, F., M. Zhang, and J. Poterjoy, 2013: E3DVar: Coupling an ensemble Kalman filter with three-dimensional variational data assimilation in a limited-area weather prediction model and comparison to E4DVar. *Mon. Wea. Rev.*, **141,** 900-917.

Zhang, M. and F. Zhang, 2011: 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.

Zhu, K., Y. Pan, M. Xue, X. Wang, J. S. Whitaker, S. G. Benjamin, S. S. Weygandt, and M. Hu, 2013: A regional GSI-based ensemble Kalman filter data assimilation system for the Rapid Refresh configuration: Testing at reduced resolution. *Mon. Wea Rev.*, **141,** 4118-4139.

List of figures

Fig. 1. Flowchart of a full GSI-based EnKF-En3DVar hybrid data assimilation cycle, with one-way or two-way coupling between the EnKF (upper portion) and En3DVar hybrid control analysis (lower portion denoted En3DVar). The thick upward pointing arrow indicates the feedback of the En3DVar hybrid analysis to the EnKF in the two-way coupling procedure, when the En3DVar hybrid control analysis is used to replace the ensemble mean of the EnKF analyses**.**

Fig. 2. Example of the horizontal distributions of observation at 0000 UTC May 14: (a) sounding (circles) and profile (pluses), (b) surface stations over land and for ships, buoys, etc., (c) aircraft observations, (d) satellite retrieval winds, and (e) GPS precipitable water (PW) data. The small box in (a) is the domain used by 13 km forecasts. (a)-(d) are adopted from Zhu et al. (2013).

Fig. 3. Profiles of (a) horizontal and (b) vertical cut-off radii for the EnKF control experiment. The horizontal axis is the cut-off radius of a given observation at a particular vertical position given in pressure. The vertical axis is the vertical position of observations given by the pressure

Fig. 4. Analysis increments at 500 hPa resulting from a single 500 hPa temperature observation over Norman Oklahoma (shown by the black dot) that is 1 K above the background for (a) GSI, (b) EnKF, and (c) En3DVar Hybrid schemes, valid at 0300 UTC 13 May 2010. The contours and shading are for the background geopotential height (gpm) and temperature increments, respectively. Lower panels are analysis increments in an east-west vertical cross section through the observation point, for (d) GSI, (e) EnKF, and (f) En3DVar hybrid. Shaded is the temperature increment. Thick contours (solid for positive and dash for negative) are for the north-south wind increment; thin contours are for potential temperature from 294 to 338 K at 4 K intervals.

Fig. 5. Vertical profiles of mean 3-hour forecast RMSEs verified against sounding data for (a) *RH*, (b) T, (c) *U*, and (d) V for the labeled experiments. Error bars represent the two-tailed 90% confidence interval (5% on the left and 95% on the right) using the bootstrap distribution method.

Fig. 6. The 9-day and domain-averaged forecast RMSEs verified against sounding data for (a) RH, (b) T, (c) U, and (d) V, for different forecast hours.

Fig. 7. The 9-day and domain-averaged forecast RMSEs verified against surface observations and the 90% confidence interval of the RMSE differences between En3DVar hybrid experiments and GSI/EnKF_Ctl for (a) surface pressure, (b) 2-m RH, (c) 2-m temperature, (*d*) 10-m U, and (e) *10-m* V for different forecast hours. The horizontal axis is forecast hour. The error bars in domain-averaged forecast RMSEs represent the two-tailed 90% confidence interval**.**

Fig. 8. Domain-averaged 3-hour forecast RMSEs (upper panels in each frame) verified against sounding data at 0000 and 1200 UTC through test period and the 90% confidence interval of RMSE differences (lower panel of each frame) between the En3DVar hybrid and EnKF experiments and GSI for (a) RH, (b) T, (c) U, and (d) *V***.**

Fig. 9. The same as Fig. 5 but for experiments GSI, Hybrid01, Hybrid05, Hybrid09, Hybrid10 and EnKF_Ctl.

Fig. 10. The same as Fig. 6, except for for experiments GSI, Hybrid01, Hybrid05, Hybrid09, Hybrid10 and EnKF_Ctl.

Fig. 11. Mean forecast RMSE differences between different experiments and Hybrid1W_Ctl, verified against sounding data, for 3-hour forecast averaged over the 9-day forecast period over the entire model domain.

Fig. 12. The same as Fig. 5 but for experiments EnKF_Ctl, Hybrid_HD and Hybrid_Con.

Fig. 13. Nine-day and domain-averaged absolute 3-hour forecast RMSE differences verified against sounding data, where 1G means difference between Hybrid_Con and EnKF_Ctl, 1GHD means difference between Hybrid_HD and EnKF_Ctl, and 3GHD means difference between Hybrid3G and EnKF3G. The error bars represent the two-tailed 90% confidence interval.

Fig. 14. Nine-day and domain-averaged profiles of absolute RMSE differences between Hybrid_Con and EnKF_Ctl (labeled 1GC), Hybrid_HD and EnKF_Ctl (labeled 1GHD), Hybrid3G and EnKF3G (labeled 3GHD) for (a) *RH*, (b) T, (c) *U*, and (d) *V*. The error bars represent the two-tailed 90% confidence interval.

Fig. 15. The relative percentage improvement (RPI, negative represents an improvement or error reduction) of Hybrid1W20, Hybrid1W40, EnKF20 and EnKF40 comparing to experiment GSI for (a) RH, (b) T, (*c) U, and* V . The horizontal axis is forecast hour.

Fig. 16. Averaged precipitation GSSs and BIASs of 13-km forecasts as a function of forecast length for thresholds (a) (b) 0.1 mm h-1, (c) (d) 1.25 mm h-1, and (e) (f) 2.5 mm h-1.

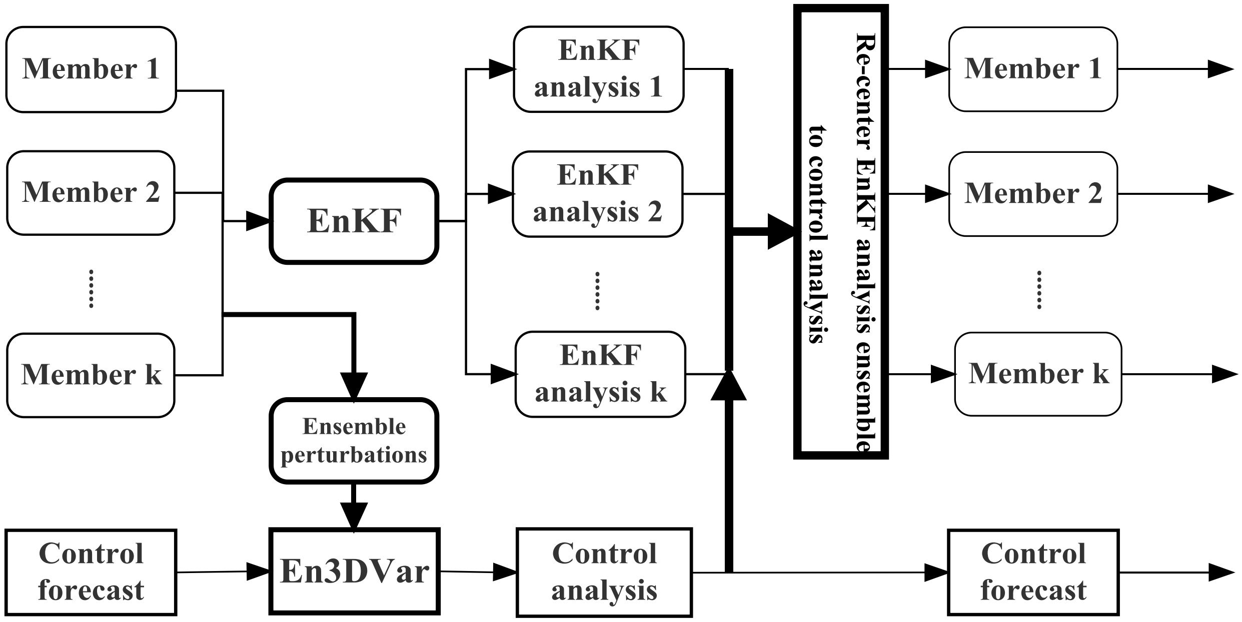


Fig. 1. Flowchart of a full GSI-based EnKF-En3DVar hybrid data assimilation cycle, with one-way or two-way coupling between the EnKF (upper portion) and En3DVar hybrid control analysis (lower portion denoted En3DVar). The thick upward pointing arrow indicates the feedback of the En3DVar hybrid analysis to the EnKF in the two-way coupling procedure, when the En3DVar hybrid control analysis is used to replace the ensemble mean of the EnKF analyses.

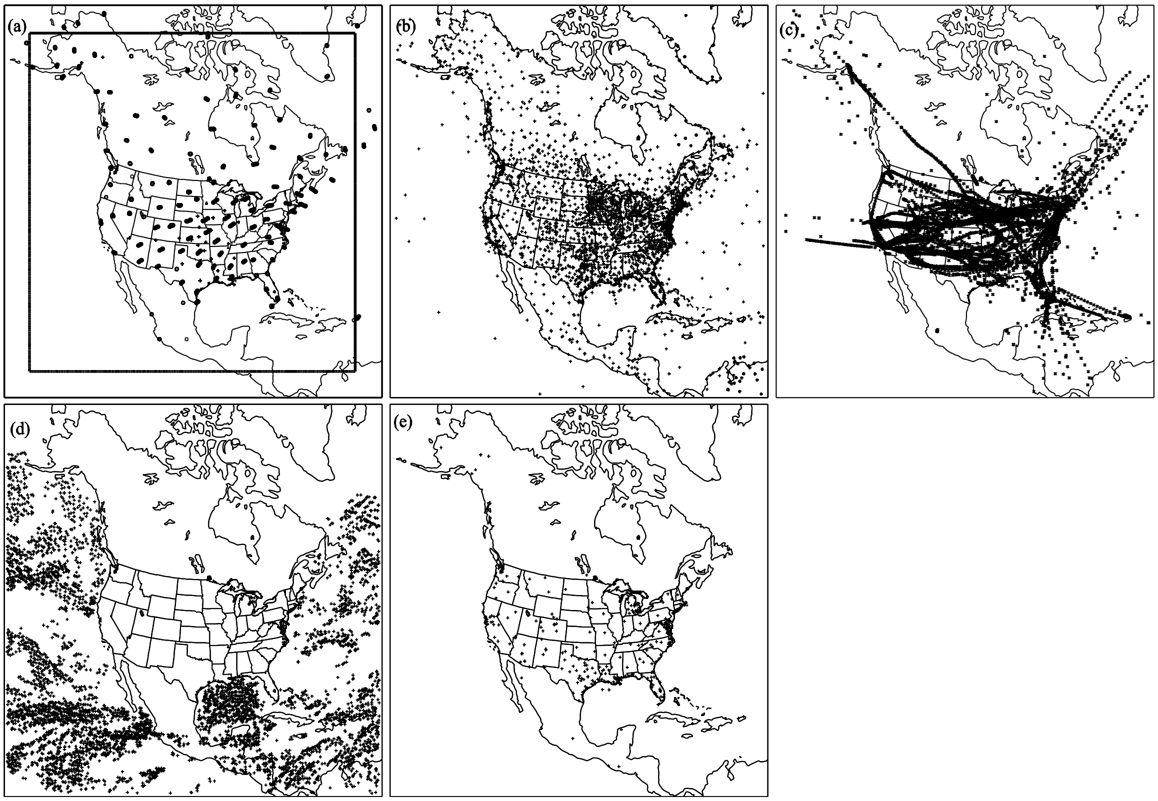


Fig. 2. Example of the horizontal distributions of observation at 0000 UTC May 14: (a) sounding (circles) and profile (pluses), (b) surface stations over land and for ships, buoys, etc., (c) aircraft observations, (d) satellite retrieval winds, and (e) GPS precipitable water (PW) data. The small box in (a) is the domain used by 13 km forecasts. (a)-(d) are adopted from Zhu et al. (2013).

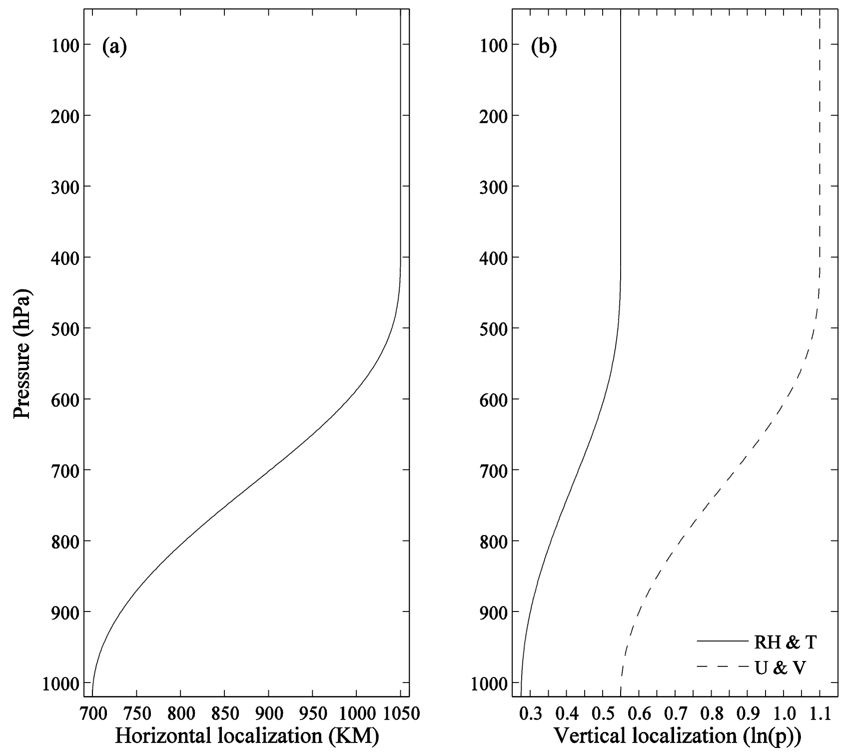


Fig. 3. Profiles of (a) horizontal and (b) vertical cut-off radii for the EnKF control experiment. The horizontal axis is the cut-off radius of a given observation at a particular vertical position given in pressure. The vertical axis is the vertical position of observations given by the pressure.

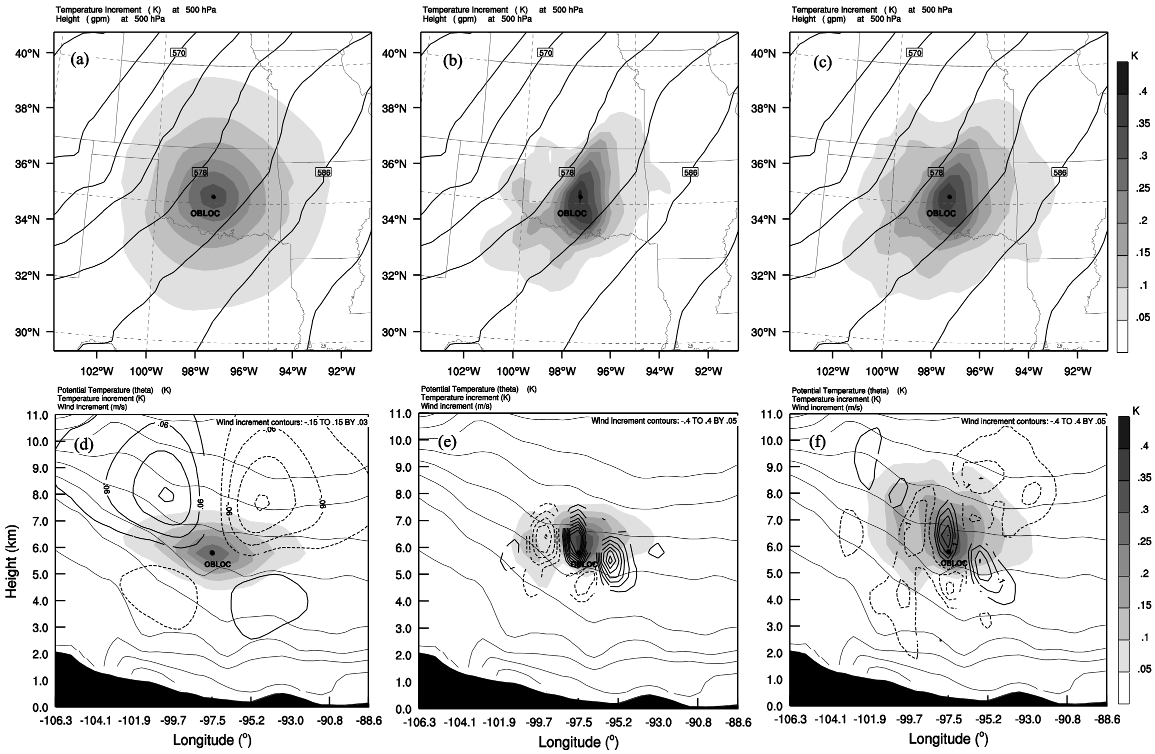
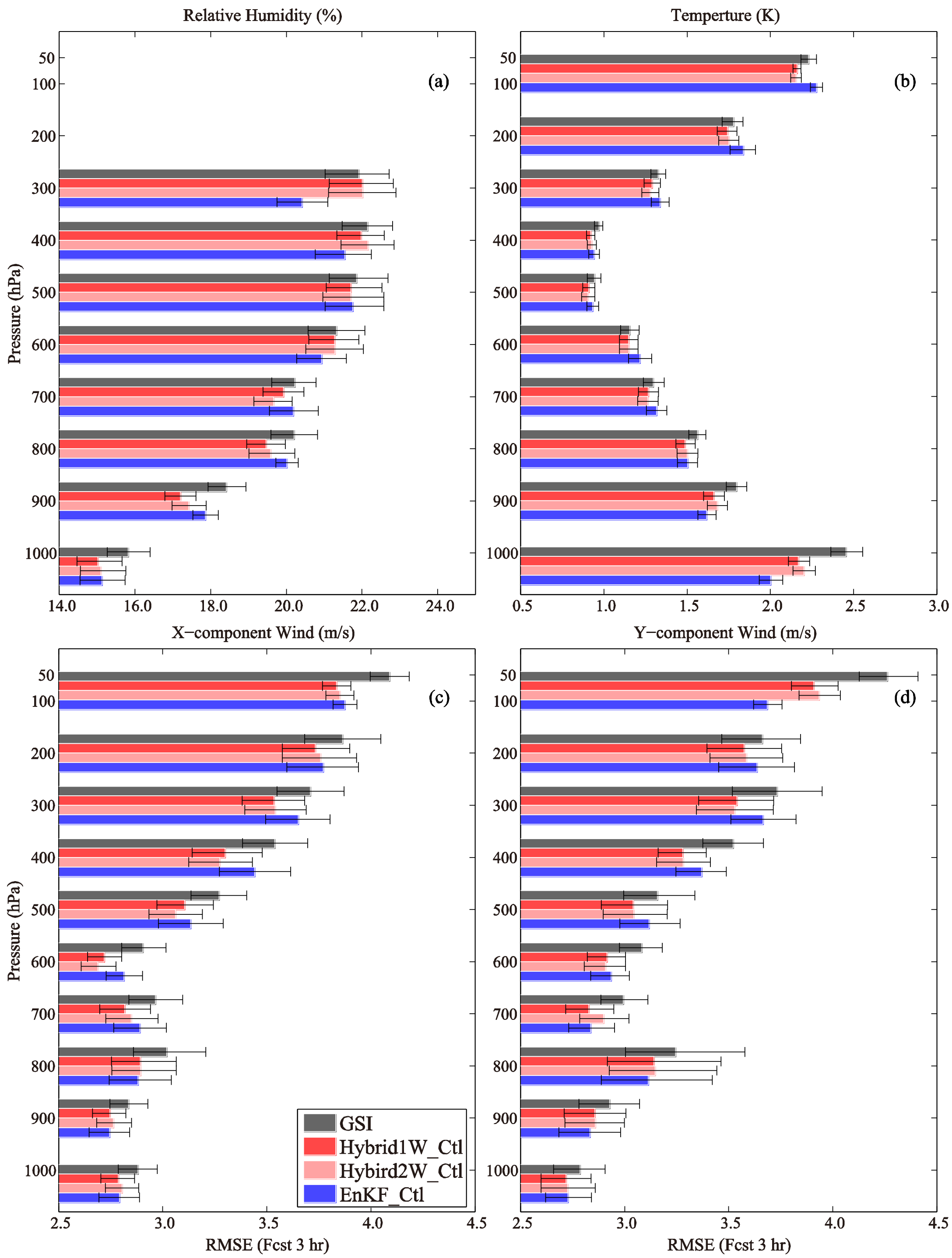


Fig. 4. Analysis increments at 500 hPa resulting from a single 500 hPa temperature observation over Norman Oklahoma (shown by the black dot) that is 1 K above the background for (a) GSI, (b) EnKF, and (c) En3DVar Hybrid schemes, valid at 0300 UTC 13 May 2010. The contours and shading are for the background geopotential height (gpm) and temperature increments, respectively. Lower panels are analysis increments in an east-west vertical cross section through the observation point, for (d) GSI, (e) EnKF, and (f) En3DVar hybrid. Shaded is the temperature increment. Thick contours (solid for positive and dash for negative) are for the north-south wind increment; thin contours are for potential temperature from 294 to 338 K at 4 K intervals.

985 

986 Fig. 5. Vertical profiles of mean 3-hour forecast RMSEs verified against sounding data for (a) *RH*,
987 (b) T, (c) *U*, and (d) V for the labeled experiments. Error bars represent the two-tailed 90%
988 confidence interval (5% on the left and 95% on the right) using the bootstrap distribution method.

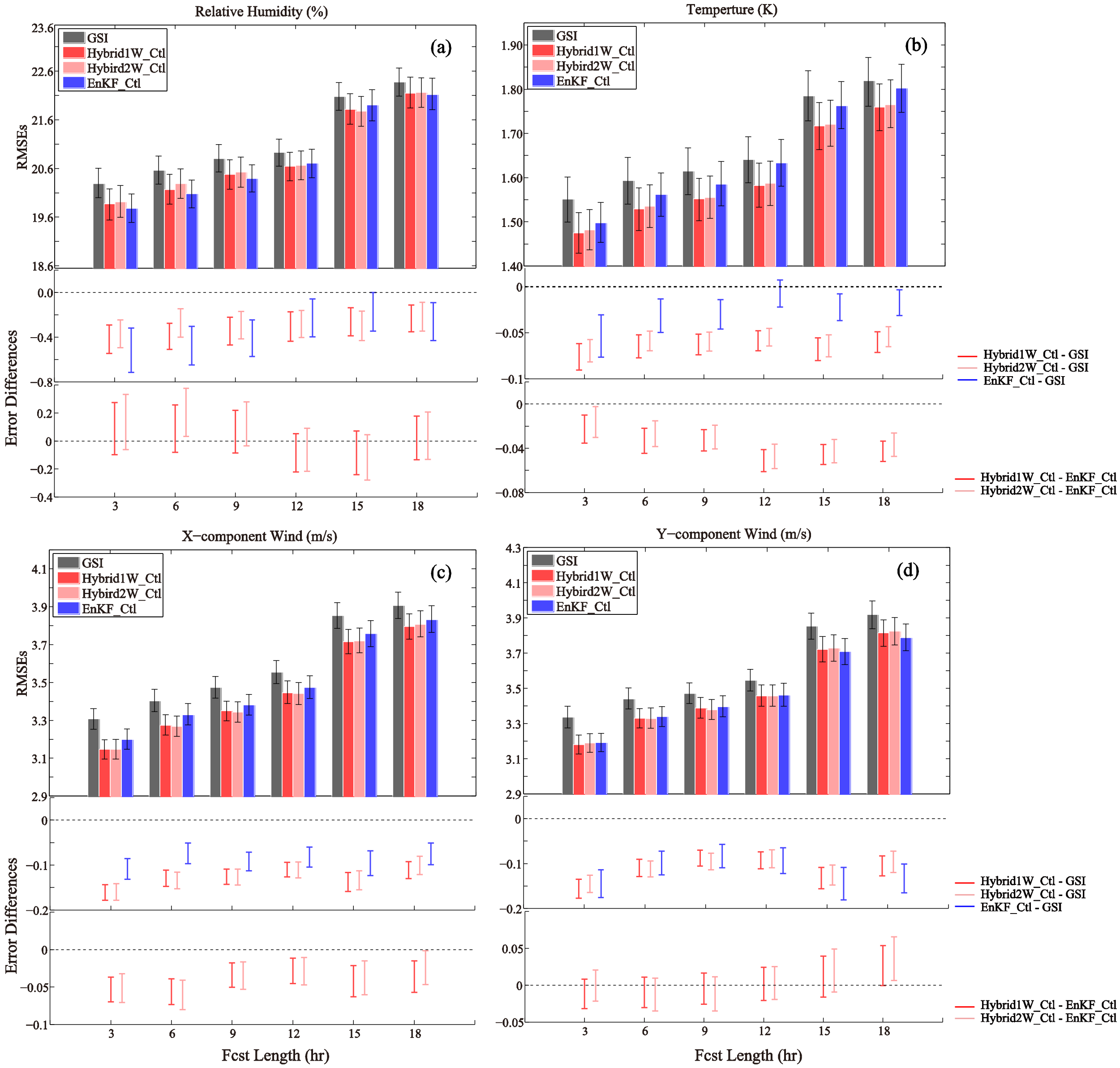


Fig. 6. The 9-day and domain-averaged forecast RMSEs verified against sounding data for (a) *RH*, (b) *T*, (c) *U*, *a*nd (d) V, for different forecast hours. Error bars represent the two-tailed 90% confidence interval (5% at the bottom and 95% on the top) using the bootstrap distribution method.

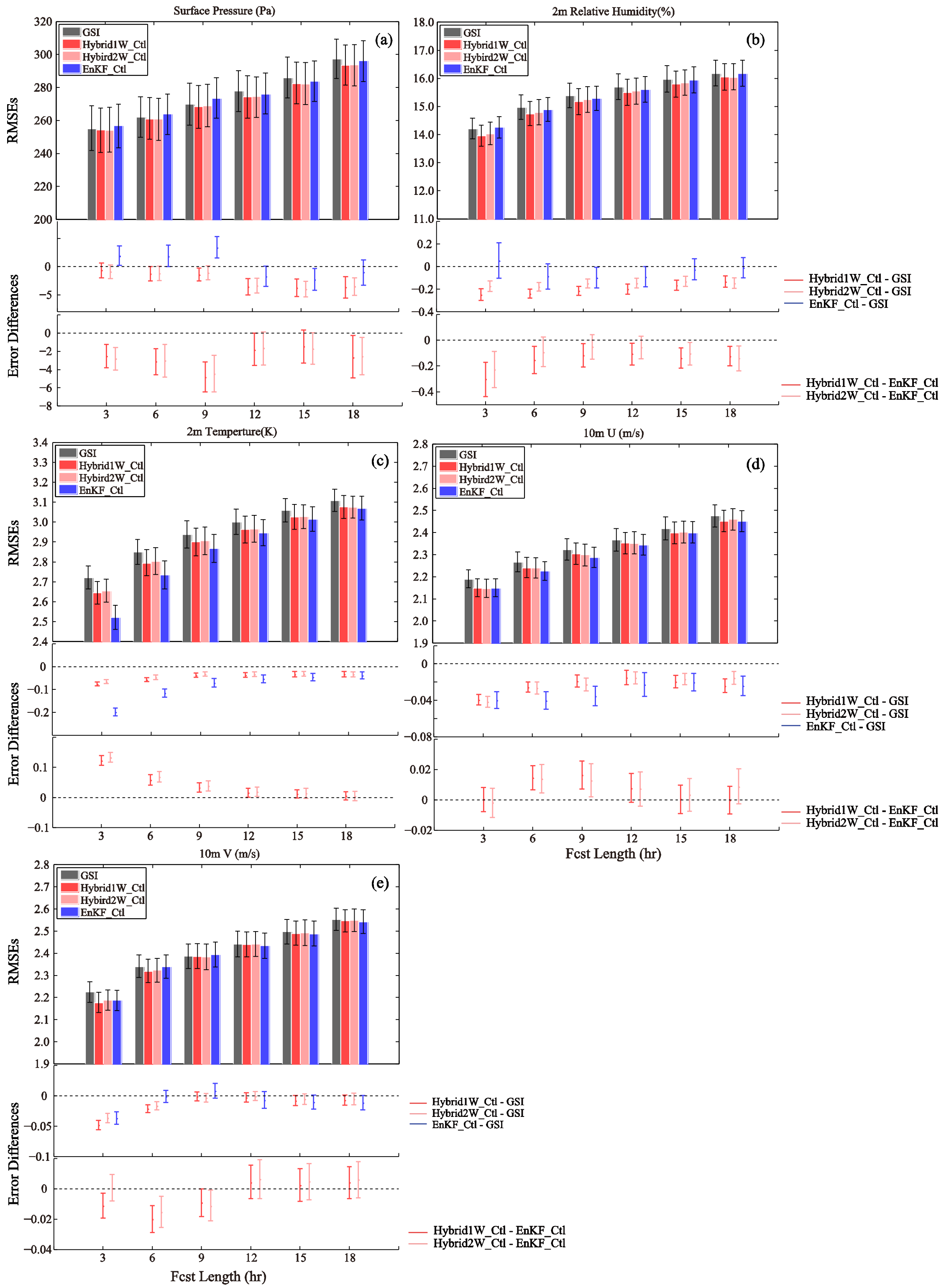


Fig. 7. The 9-day and domain-averaged forecast RMSEs verified against surface observations and the 90% confidence interval of the RMSE differences between En3DVar hybrid experiments and GSI/EnKF_Ctl for (a) surface pressure, (b) 2-m *RH*, (c) 2-m temperature, (d) 10-m *U*, and (e) 10-m *V* for different forecast hours. The horizontal axis is forecast hour. The error bars in domain-averaged forecast RMSEs represent the two-tailed 90% confidence interval.

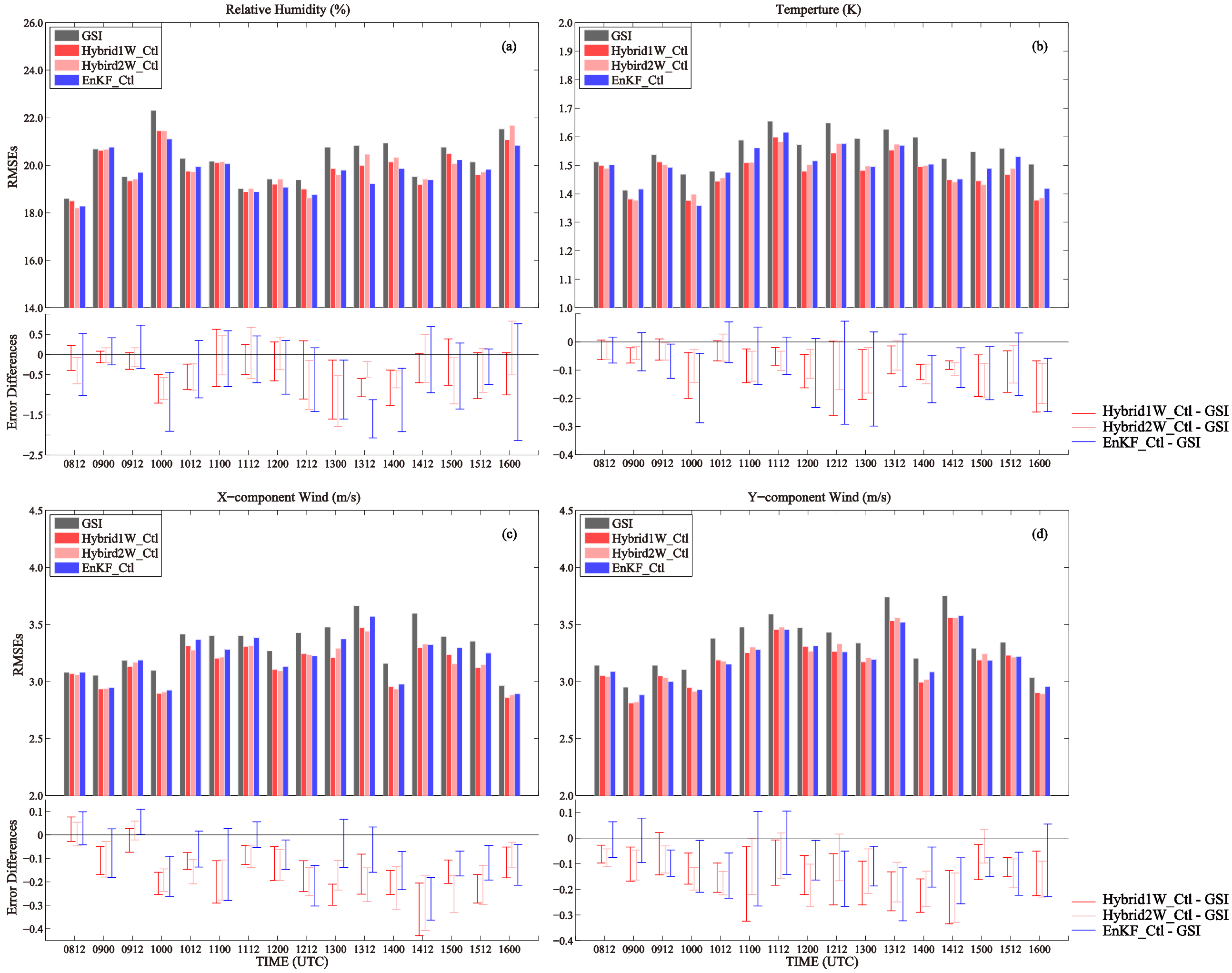


Fig. 8. Domain-averaged 3-hour forecast RMSEs (upper panels in each frame) verified against sounding data at 0000 and 1200 UTC through test period and the 90% confidence interval of RMSE differences (lower panel of each frame) between the En3DVar hybrid and EnKF experiments and GSI for (a) *RH*, (b) *T*, (c) *U*, and (d) *V*.

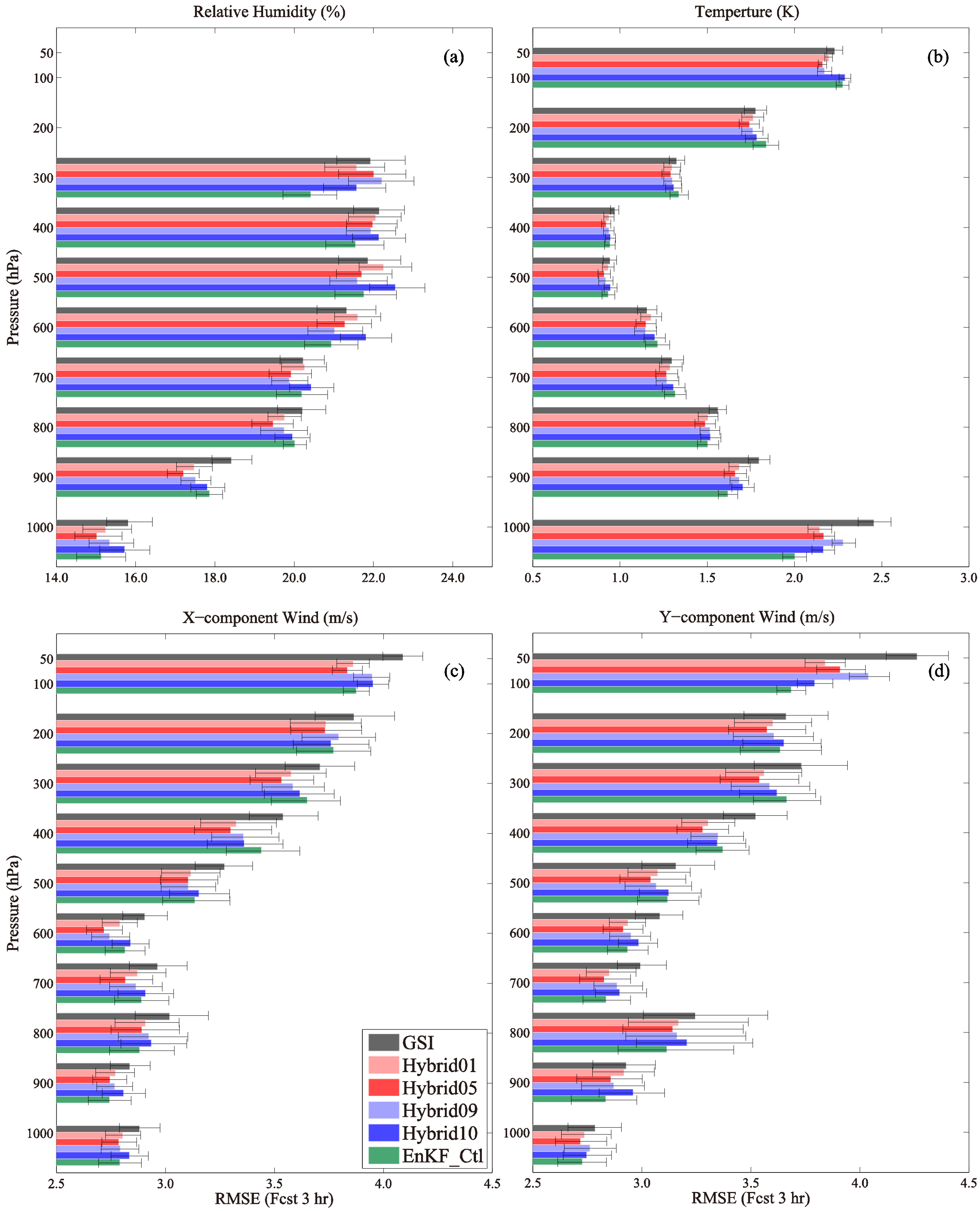


Fig. 9. The same as Fig. 5 but for experiments GSI, Hybrid01, Hybrid05, Hybrid09, Hybrid10 and EnKF_Ctl.

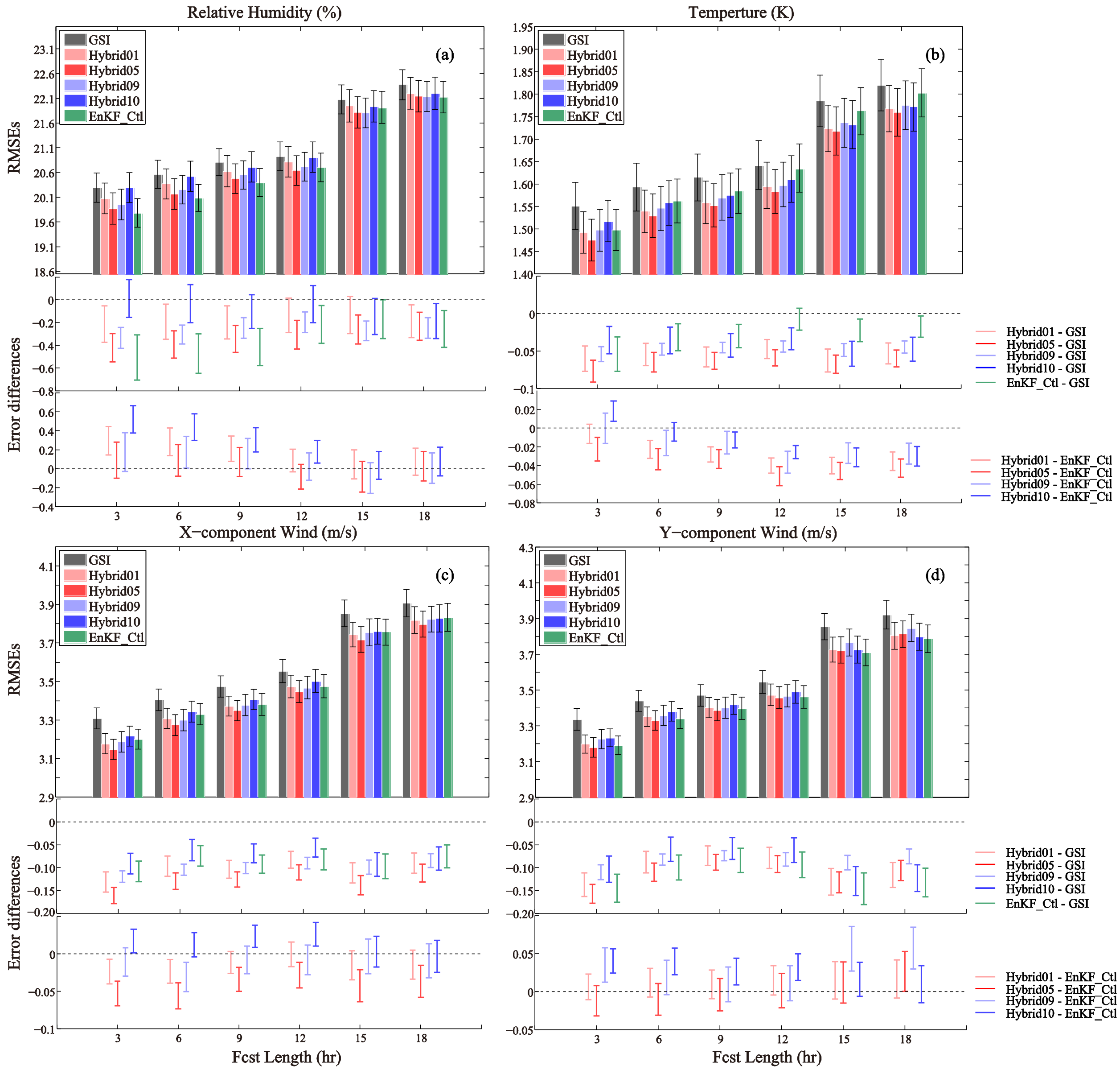


Fig. 10. The same as Fig. 6, except for for experiments GSI, Hybrid01, Hybrid05, Hybrid09, Hybrid10 and EnKF_Ctl.

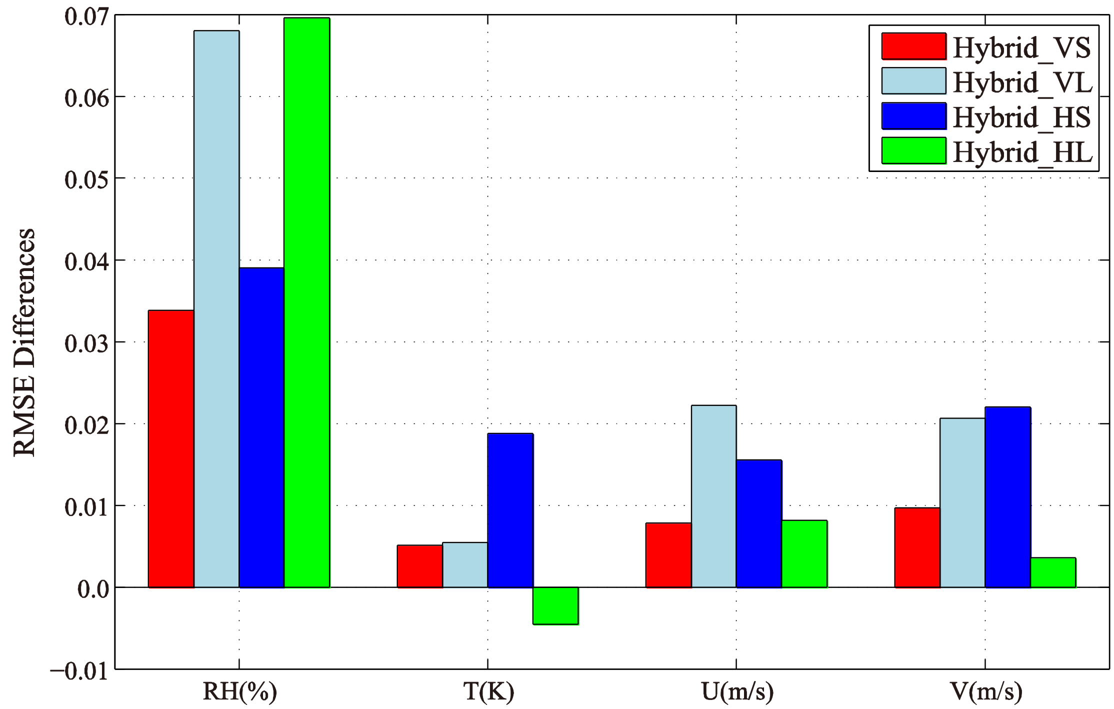


Fig. 11. Mean forecast RMSE differences between different experiments and Hybrid1W_Ctl, verified against sounding data, for 3-hour forecast averaged over the 9-day forecast period over the entire model domain.

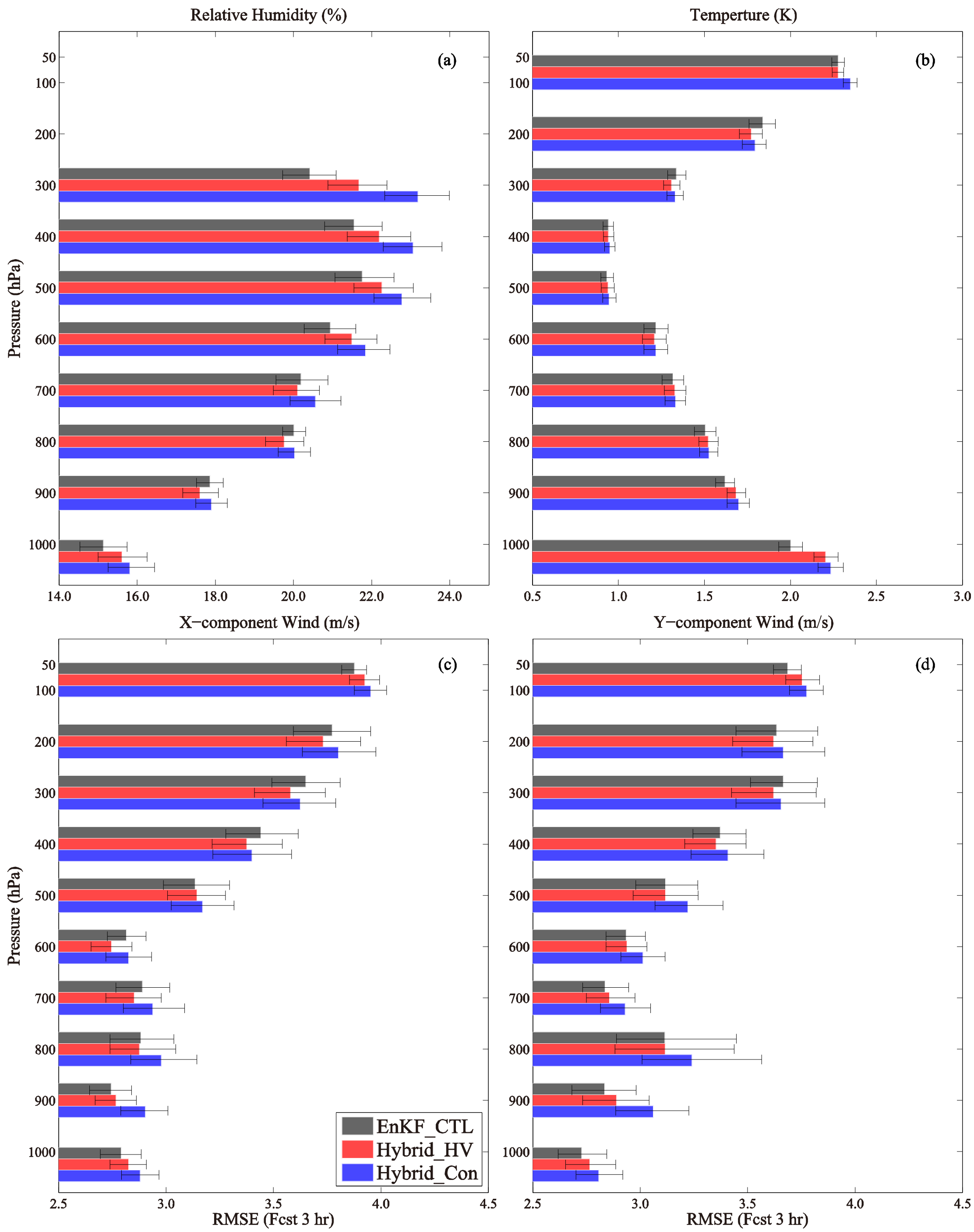


Fig. 12. The same as Fig. 5 but for experiments EnKF_Ctl, Hybrid_HD and Hybrid_Con.

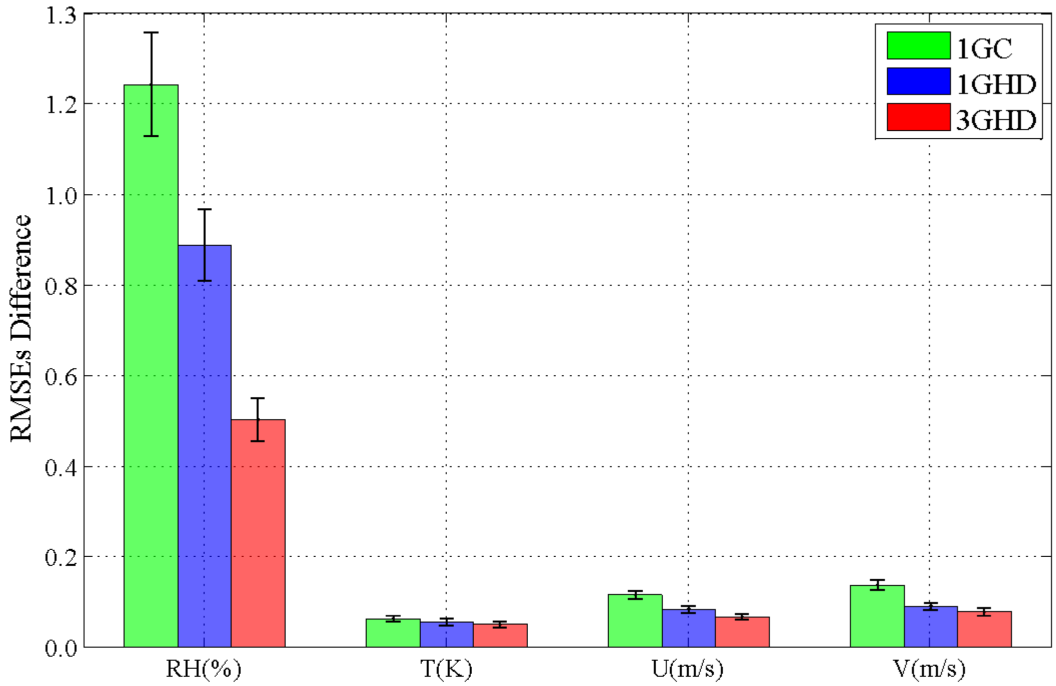


Fig. 13. Nine-day and domain-averaged absolute 3-hour forecast RMSE differences verified against sounding data, where 1G means difference between Hybrid_Con and EnKF_Ctl, 1GHD means difference between Hybrid_HD and EnKF_Ctl, and 3GHD means difference between Hybrid3G and EnKF3G. The error bars represent the two-tailed 90% confidence interval.

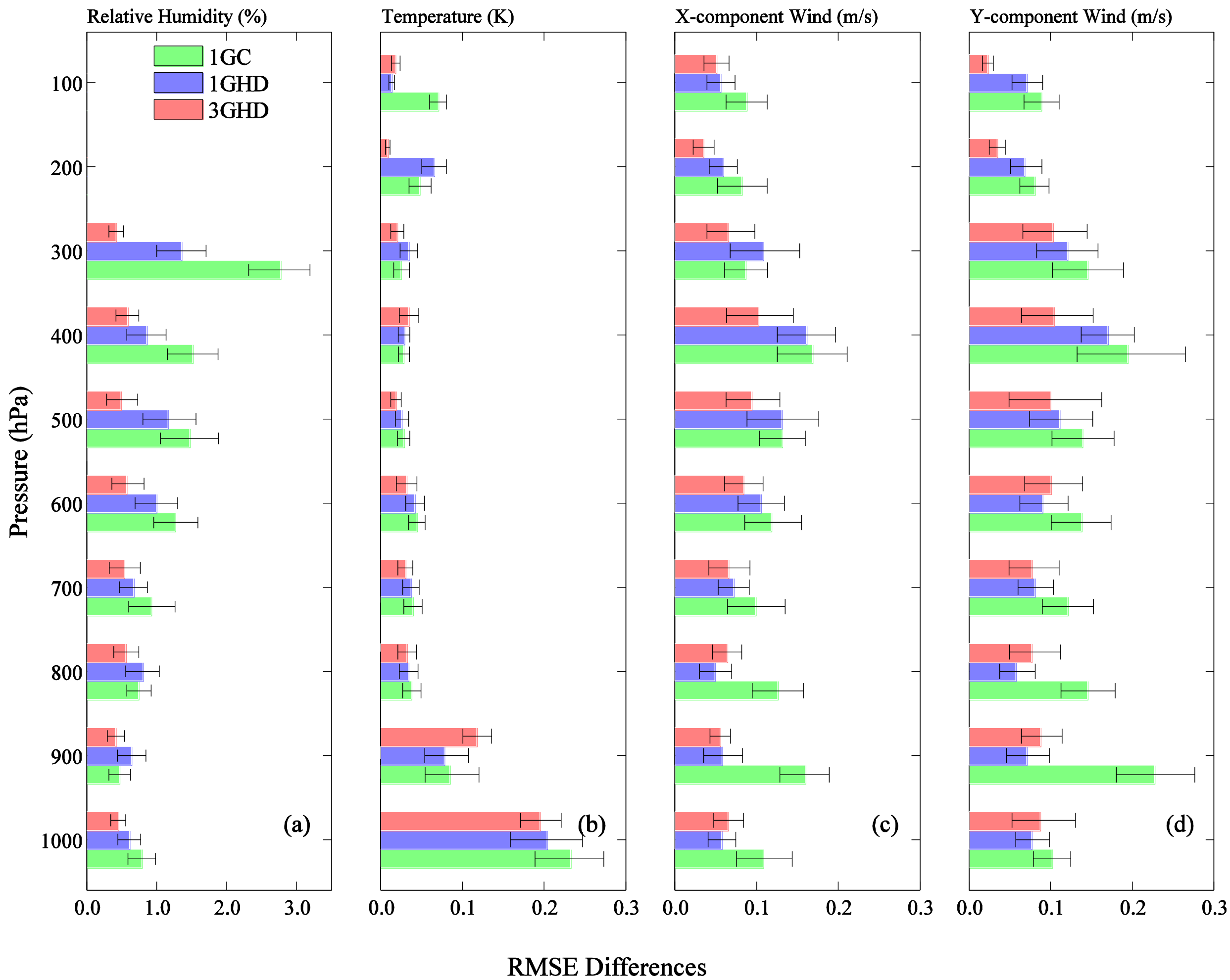


Fig. 14. Nine-day and domain-averaged profiles of absolute RMSE differences between Hybrid_Con and EnKF_Ctl (labeled 1GC), Hybrid_HD and EnKF_Ctl (labeled 1GHD), Hybrid3G and EnKF3G (labeled 3GHD) for (a) *RH*, (b) *T*, (c) *U*, and (d) *V*. The error bars represent the two-tailed 90% confidence interval.

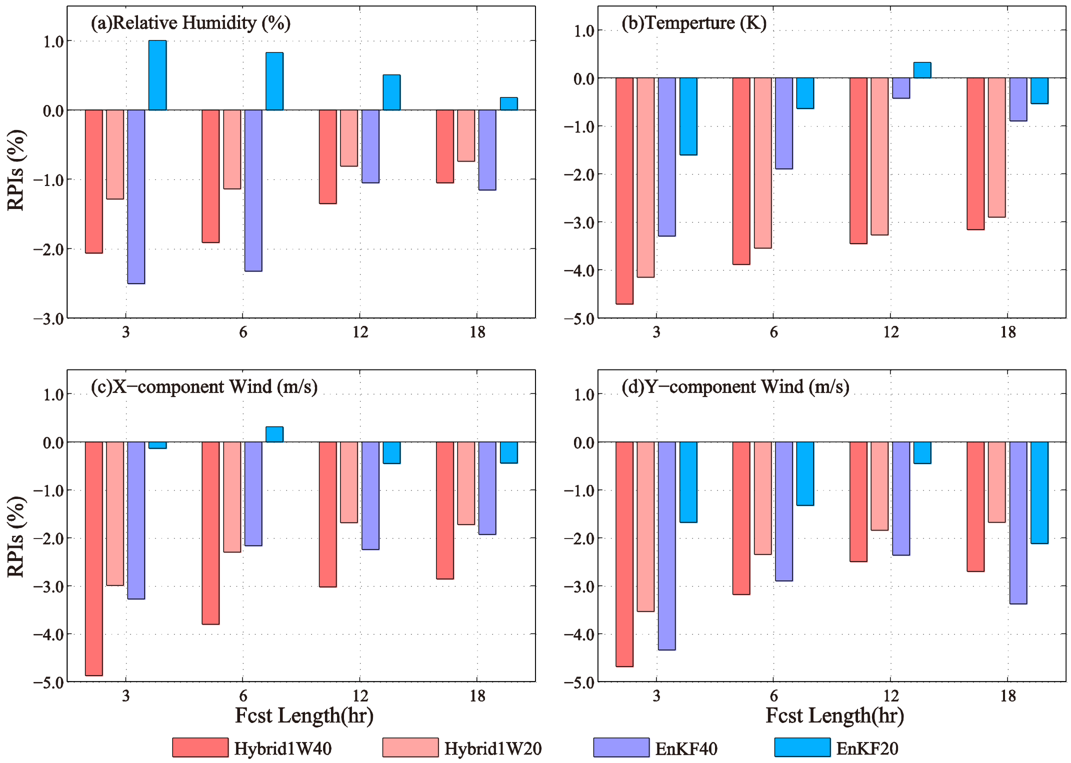


Fig. 15. The relative percentage improvement (RPI, negative represents an improvement or error reduction) of Hybrid1W20, Hybrid1W40, EnKF20 and EnKF40 comparing to experiment GSI for (a) *RH*, (b) *T*, (c) *U*, and *V* . The horizontal axis is forecast hour.

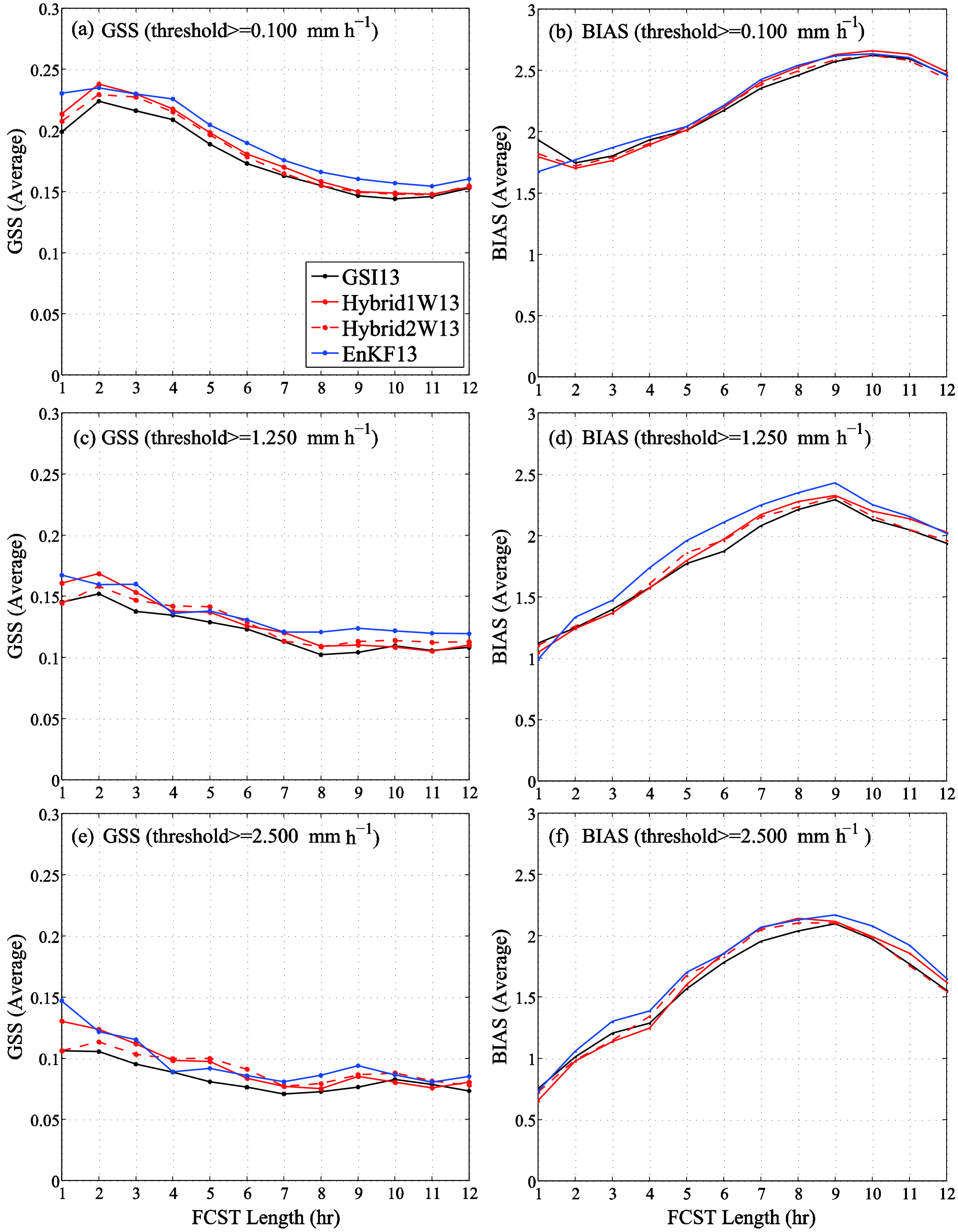


Fig. 16. Averaged precipitation GSSs and BIASs of 13-km forecasts as a function of forecast length for thresholds (a) (b) 0.1 mm h^{-1}, (c) (d) 1.25 mm h^{-1}, and (e) (f) 2.5 mm h^{-1}.

Table 1. List of data assimilation experiments. In the horizontal and vertical localization columns, $\nearrow$ means increasing with height.

Experiment group	Experiment (including alternative names)	Ensemble covariance weighting factor $(1/\beta_2)$	Horizontal cut-off radius for hybrid/EnKF (km)	Vertical cut-off radius for hybrid/EnKF in ln(*p*)	Ensemble size	EnKF-En3DVar Coupling
Control experiments	GSI	N.A.				
	EnKF_Ctl /EnKF40	-	700 $\nearrow$ 1050	*RH* and *T*: 1.1/4 $\nearrow$ 1.1/2 *U* and *V*: 1.1/2 $\nearrow$ 1.1 ps and pw: 1.6	40	-
	Hybrid1W_Ctl /Hybrid1W40 /Hybrid05	0.5	~1095	1.1	40	1-way
	Hybrid2W_Ctl /Hybrid2W40	0.5	~1095	1.1	40	2-way
Sensitivity experiments on covariance weighting factors	Hybrid01	0.1	~1095	1.1	40	1-way
	Hybrid09	0.9	~1095	1.1	40	1-way
	Hybrid10	1.0	~1095	1.1	40	1-way
Sensitivity experiments on localization scales	Hybrid_HS	0.5	~701	1.1	40	1-way
	Hybrid_HL	0.5	~1300	1.1	40	1-way
	Hybrid_VS	0.5	~1095	0.36	40	1-way
	Hybrid_VL	0.5	~1095	1.8	40	1-way
Sensitivity experiments on height- and observation-type-dependent localization scales	Hybrid_Con	0.0	~701	1.1	40	1-way
	Hybrid_HD	0.0	700 $\nearrow$ 1050	1.1/2 $\nearrow$ 1.1	40	1-way
	Hybrid3G	0.0	700 $\nearrow$ 1050	*RH* and *T*: 1.1/4 $\nearrow$ 1.1/2 *U* and *V*: 1.1/2 $\nearrow$ 1.1 p_s and PW: 1.6 (observations are assimilated in 3 groups)	40	1-way
	EnKF3G	-	700 $\nearrow$ 1050	*RH* and *T*: 1.1/4 $\nearrow$ 1.1/2 *U* and *V*: 1.1/2 $\nearrow$ 1.1 p_s and PW: 1.6 (observations are assimilated in 3 groups)	40	1-way

Sensitivity experiments on ensemble size	EnKF20	-	700 $\nearrow$ 1050	*RH* and *T*: 1.1/4 $\nearrow$ 1.1/2 *U* and *V*: 1.1/2 $\nearrow$ 1.1 p_s and PW: 1.6	20	-
	Hybrid1W20	0.5	~1095	0.3	20	1-way
	Hybrid2W20	0.5	~1095	0.3	20	2-way

Table 2 list of mean domain average absolute RMSE difference pair

Name	Hybrid*	Benchmark
1GC	Hybrid_Con	EnKF_Ctl
1GHD	Hybrid_HD	EnKF_Ctl
3GHD	Hybrid3G	EnKF3G