GSI Three-Dimensional Ensemble–Variational Hybrid Data Assimilation Using a Global Ensemble for the Regional Rapid Refresh Model

MING HU

NOAA/OAR/Earth System Research Laboratory, and Cooperative Institute for Research in Environmental Sciences,
University of Colorado Boulder, Boulder, Colorado

STANLEY G. BENJAMIN

NOAA/OAR/Earth System Research Laboratory, Boulder, Colorado

THERESE T. LADWIG

NOAA/OAR/Earth System Research Laboratory, and Cooperative Institute for Research in Environmental Sciences,
University of Colorado Boulder, Boulder, Colorado

DAVID C. DOWELL, STEPHEN S. WEYGANDT, CURTIS R. ALEXANDER, AND
JEFFREY S. WHITAKER

NOAA/OAR/Earth System Research Laboratory, Boulder, Colorado

ABSTRACT

The Rapid Refresh (RAP) is an hourly updated regional meteorological data assimilation/short-range model forecast system running operationally at NOAA/National Centers for Environmental Prediction (NCEP) using the community Gridpoint Statistical Interpolation analysis system (GSI). This paper documents the application of the GSI three-dimensional hybrid ensemble–variational assimilation option to the RAP high-resolution, hourly cycling system and shows the skill improvements of 1–12-h forecasts of upper-air wind, moisture, and temperature over the purely three-dimensional variational analysis system. Use of perturbation data from an independent global ensemble, the Global Data Assimilation System (GDAS), is demonstrated to be very effective for the regional RAP hybrid assimilation. In this paper, application of the GSI-hybrid assimilation for the RAP is explained. Results from sensitivity experiments are shown to define configurations for the operational RAP version 2, the ratio of static and ensemble background error covariance, and vertical and horizontal localization scales for the operational RAP version 3. Finally, a 1-week RAP experiment from a summer period was performed using a global ensemble from a winter period, suggesting that a significant component of its multivariate covariance structure from the ensemble is independent of time matching between analysis time and ensemble valid time.

1. Introduction

The Rapid Refresh (RAP; Benjamin et al. 2016, hereafter B16) was developed as an hourly updated data assimilation–model forecast cycling system to meet the growing requirements for increased accuracy in short-range weather guidance for aviation, energy, severe weather, hydrology, agriculture, and other sectors. The RAP replaced the Rapid Update Cycle (RUC; Benjamin et al. 2004a,b) within the operational model suite at

NOAA's National Centers for Environmental Prediction (NCEP) in May 2012. A second version of the RAP with further advances in data assimilation and model design was implemented at NOAA/NCEP in February 2014, and a third version was implemented in August 2016. B16 summarizes the RAP and its key techniques in both data assimilation and model forecasting for hourly cycling. This paper focuses on the RAP forecast improvements gained by applying three-dimensional ensemble–variational hybrid analysis in the Gridpoint Statistical Interpolation analysis system (GSI) with the ensemble defined by Global Data Assimilation System (GDAS)

ensemble Kalman filter (EnKF) data assimilation system. For numerical weather prediction (NWP) systems with rapid update cycling like RAP and RUC with emphasis on short-range forecasts, the data assimilation component is especially important. One of the main improvements of the RAP system compared to the RUC is the use of GSI for the data assimilation component. GSI was developed at NCEP as a unified global–regional variational data analysis system (Wu et al. 2002). GSI was made available as a community data assimilation system by the Developmental Testbed Center (Shao et al. 2016). It serves both operational and research community users with extended options for variational or ensemble–variational hybrid analysis or as an observation operator for an EnKF analysis (Wu et al. 2002; Whitaker et al. 2008; Kleist et al. 2009). The use of GSI allows RAP development to take advantage of broad GSI development efforts from both the operational and research community. In return, new data assimilation enhancements developed for the RAP are shared back via the GSI repository as options for other operational and research applications.

The background error covariance (BEC) plays an important role in three-dimensional variational data assimilation. It is an important factor to define the analysis increment from observation innovations and the balance among different analysis variables, especially between mass and wind fields. The fixed (or static) BEC used by GSI is constructed based on perturbations generated with the National Meteorological Center (NMC) method (Parrish and Derber 1992) via a collection of model forecast error fields over a long period. In recent years, a major development in the GSI is to use ensemble-based perturbations to obtain flow-dependent BECs (Wang et al. 2007) for the GSI variational analysis using global ensemble forecasts (Wang 2010; Whitaker et al. 2008) or regional ensemble forecasts. Application of the variational-based solver using augmented control vector to incorporate the ensemble information is referred to as ensemble–variational (EnVar) analysis in Lorenc (2013) and as the GSI-hybrid EnVar or hybrid analysis for the specific version in this paper.

Several hybrid data assimilation methods have been proposed to combine the advantages of the ensemble and the variational methods (Hamill and Snyder 2000; Lorenc 2003; Buehner 2005; Wang et al. 2007). The GSI-hybrid analysis uses a combined covariance through a variational-based control variable method (Lorenc 2003). Both three-dimensional (3D) and four-dimensional (4D) versions of hybrid EnVAR have been successfully developed and tested for global and regional GSI applications by NCEP and its partner developers (Wang et al. 2013; Kleist and Ide 2015a,b). Wang et al. (2013) showed that the NCEP GFS using 3D EnVAR hybrid assimilation produces more skillful deterministic

forecasts out to day 5 relative to those using 3DVAR. Kleist and Ide (2015a) further demonstrated that using the GSI 3D EnVar hybrid method made improvements in the analysis and subsequent forecast for the NCEP/GFS system in the context of observing system simulation experiments (OSSEs). The test and evaluation of the same 3D EnVar hybrid system applied for the North American Mesoscale Forecast System (NAM) yielded similar improvement as in the GFS system (Wu et al. 2017). The successful development and application of the 3D EnVar in GFS and NAM encouraged the Earth System Research Laboratory (ESRL) RAP development group to apply this method in the RAP system. The RAP is unique in that it is a 1-h assimilation cycle with the WRF-ARW Model, while the GFS uses a 6-hourly cycle and the NAM has used a 3-hourly cycle.

This paper discusses the results from experiments on application of the GSI 3D EnVar hybrid method to the hourly updating RAP. Section 2 describes RAP cycling workflow and configurations of GSI for the RAP. Section 3 introduces the configuration of GSI-hybrid assimilation for RAP operation and the baseline test of the 3D hybrid EnVar assimilation compared with that from the 3DVAR. In section 4, the experiments on hybrid assimilation parameters are discussed. Section 5 investigates the sensitivity of the assimilation skill to matching the valid time of the ensemble BEC and the RAP analysis time. Last, a summary of current work with GSI-hybrid assimilation and the future development for RAP are discussed in section 6.

2. RAP data assimilation

This section gives a brief summary on the GSI configuration for RAP to provide context for the GSI-hybrid tests. The details of the RAP model and analysis configuration can be found in B16.

The RAP uses the WRF-ARW release version with enhancements to the land surface model, planetary boundary layer scheme, and convection and cloud-microphysics parameterizations as described in B16. The RAP uses a rotated latitude–longitude grid with approximately 13-km horizontal grid spacing that covers Alaska, the Caribbean Sea, and all of North America. The model top is set at 10 hPa with 50 vertical levels. The lowest model computational level is located at about 8 m above the ground (sigma = 0.999). The RAP uses boundary conditions from GFS and runs hourly with new 18-h forecasts each hour. The GSI uses the RAP model 1-h forecast as the background and conducts the analysis on a rotated latitude–longitude A grid and on model vertical levels. GSI has the functionality to update different sets of model variables for use in different modeling systems. The variables updated in the RAP GSI, as
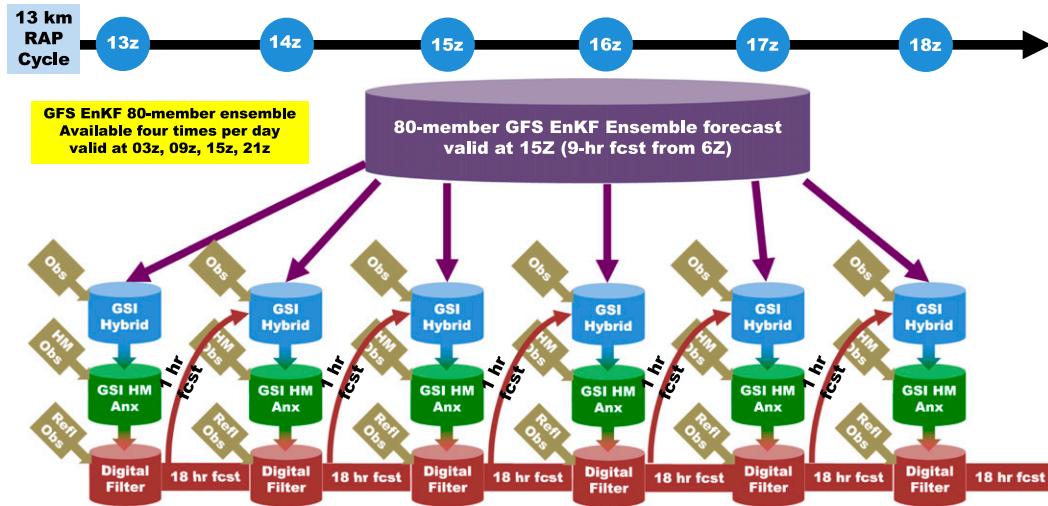
FIG. 1. RAP data assimilation flowchart. HM Anx stands for hydrometeor analysis, which is the GSI cloud/hydrometeor assimilation option used in RAP.

well as the static BEC and observation errors, are listed and explained in B16.

The use of GSI for the RAP data assimilation allows application of hybrid ensemble–variational assimilation using an ensemble-based covariance component from the independent GDAS/EnKF ensemble data assimilation. GSI also allows inclusion of new observations not previously incorporated into the RUC (details in B16). All observations used in the GSI are in WMO standard BUFR/PrepBUFR format. NCEP generates all hourly BUFR/PrepBUFR files for the RAP with narrow data cutoff time windows appropriate for a 1-h cycle, generally from 45 min before analysis time to 15 min after analysis.

Each RAP data assimilation cycle has three analysis components (Fig. 1). First, the GSI-hybrid EnVAR analyzes the temperature, wind, moisture, and surface pressure fields using conventional and radiance observations with radiance bias correction coefficients cycled with each RAP cycle. Then, the GSI hydrometeor analysis is run to improve the analysis of hydrometeors (precipitating and nonprecipitating) and water vapor consistency using surface cloud observations, satellite cloud-top products, and radar reflectivity. Finally, when the WRF Model is initialized, a diabatic digital filter is applied to filter noise and during its forward integration, 3D temperature tendencies are specified from radar reflectivity (and lightning) where available, resulting in modification to 3D divergence to retain radar information (Peckham et al. 2016; Weygandt and Benjamin 2007). A partial cycling is applied to the RAP as shown in Fig. 2 and described in B16. It brings in GFS atmospheric data every 12 h and combines with fully cycled

previous RAP land surface data to start a parallel 6-h spinup period for full RAP forecast cycles. The data assimilation behavior within the RAP studied in this paper is constrained by this partial cycling.

3. RAP GSI-hybrid configuration and baseline test

The first operational version of the RAP (version 1; B16) used three-dimensional variational data assimilation with static background error covariance (BEC). RAP version 2 introduced the use of the GSI 3D EnVAR hybrid analysis technique for hourly updating. The initial hybrid configuration was based on experience from other NCEP operational systems (GFS, NAM, and HWRF). RAP version 2 uses a 50% contribution each from both static and ensemble-based BEC for the overall BEC, and the configuration of the RAP GSI-hybrid application is listed in Table 1. The RAP system relies on the GDAS EnKF 80-member ensemble forecast [T254 resolution, ~55 km and increased to T574 (semi Lagrangian), ~30 km after January 2015] for its ensemble component of the BEC. This unique setup is possible because GSI has the functionality to map global ensemble perturbations to any regional grid allowing this dual-resolution hybrid analysis, which enables the RAP to apply hybrid analysis without need for expensive RAP-specific ensemble. The GDAS ensemble during this study only runs 4 times per day (0000, 0600, 1200, and 1800 UTC), comprising a 9-h forecast with 3-hourly output at 3-, 6-, and 9-h duration. In real time, the 9-h GDAS ensembles are available about 6 h after the valid synoptic time from which these ensembles were initialized. Because of the GDAS ensemble latency and
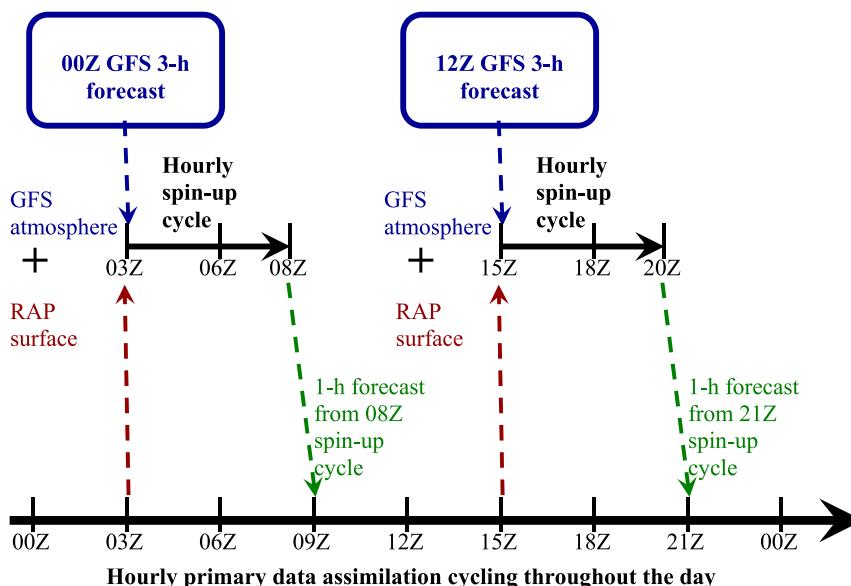
FIG. 2. RAP partial cycle structure. Atmospheric information from GFS is introduced twice daily at 0900 and 2100 UTC after 6-h parallel hourly spinup cycles in RAP during 0300–0800 and 1500–2000 UTC. The alternative RAP 1-h forecast between 0800 and 2000 UTC parallel cycle (green arrow) is introduced for the background field at 0900 and 2100 UTC. Land surface fields, by contrast, are fully cycled within the RAP through combining the GFS forecast and RAP land surface fields at 0300 and 1500 UTC (brown arrow) as the background to start the parallel cycles.

hourly cycle data assimilation, the RAP GSI 3D EnVAR hybrid assimilation must reuse the same set of GDAS ensemble information for its six consecutive hourly cycles (Fig. 1). In addition to the time availability, the grid spacing of the GDAS ensemble forecast used in this experiment is approximately 60 km, while the RAP analysis grid and ensemble grid is 13 km for the experiments in this section.

The GSI 3D EnVAR hybrid technique applied for RAP version 2 (Table 1) is compared in this paper to the use of a static BEC alone, for retrospective experiments during 28 May–4 June 2012. Figure 3 shows 6-h forecast root-mean-square error (RMSE) for upper-air winds, relative humidity, and temperature against sounding observations. The 3D EnVAR hybrid retrospective RAP experiment (blue line) outperforms the 3DVAR experiment (red line) by substantial (and statistically significant) margins (see one standard deviation boxes in experiment differences shown by the black line near zero in these graphics). The EnVAR hybrid experiment has consistently smaller errors as evident in time series and the vertical profiles of temperature, wind, and moisture. Wind forecasts show the largest improvement from the use of hybrid analysis, attributable to the inclusion of the ensemble-based BEC. The temperature improvement is relatively smaller; however, the improvements are still significant for most levels. These results indicate upper-air forecast error is reduced significantly by using the hybrid analysis, even though the same set of GFS ensembles are used for six continuous RAP cycles. The results are consistent with the results from other GSI-hybrid analysis experiments based on

TABLE 1. RAP GSI-hybrid analysis configurations for RAP version 2 and version 3 and options tested in experiments.

| Options	Values for RAP version 2	Values for RAP version 3	Options tested in experiments
Static/ensemble BEC ratio	0.5/0.5	0.25/0.75	0.25/0.75, 0.0/1.0
Ensemble grid resolution	3 times coarser than analysis grid	3 times coarser than analysis grid	Same as analysis grid
GFS ensemble update frequency	Available four times per day	Available four times per day	Available hourly
Horizontal localization scale (km)	110	110	160, 220, 330
Vertical localization scale	Three levels	Three levels	Nine levels −0.15 (~100 hPa)
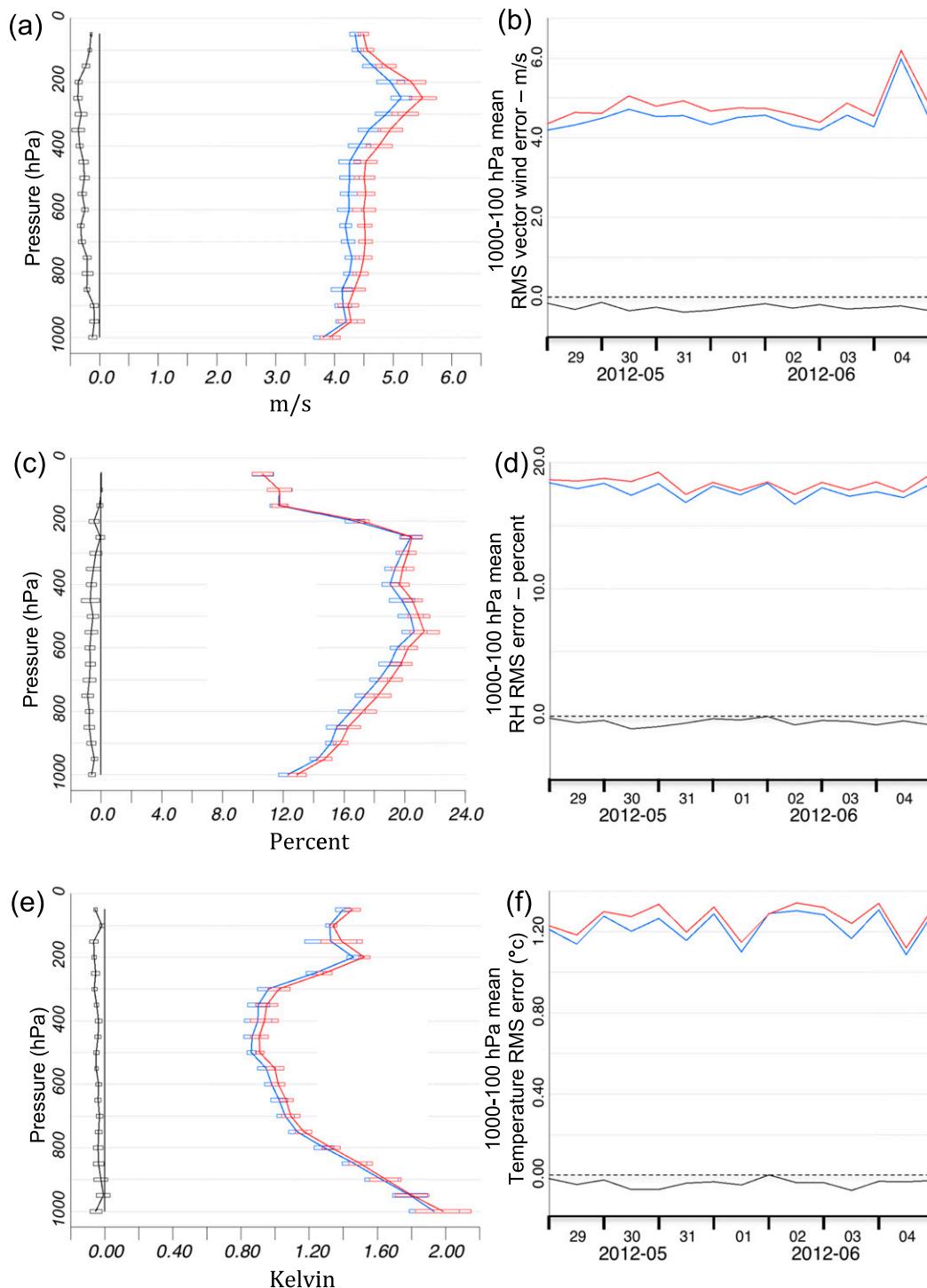
FIG. 3. (left) Upper-air RMSE profile and (right) time series for (a),(b) 6-h forecast wind (m s^{-1}); (c),(d) relative humidity (%); and (e),(f) temperature [K in (e) and °C in (f)] errors against sounding observations in 1000–100 hPa for 28 May–5 Jun 2012 RAP experiments using RAP GSI ensemble–variational hybrid analysis (EnVar, blue line) and the RAP GSI 3D-variational analysis (3DVAR, red line). Difference (EnVar − 3DVAR) is plotted in black. Boxes show 95% confidence (Weatherhead et al. 1998).

the GFS (Wang et al. 2013; Kleist and Ide 2015a,b) and NAM system (Wu et al. 2017). The positive results on hybrid analysis drawn from those shorter-range retrospective experiments here were found to be consistent with day-to-day time series results (Figs. 3b,d,f and similar to James and Benjamin 2017), suggesting that those results from shorter-range experiments are likely representative of longer-period results.

The 3- and 6-h forecast RMSE for surface fields (10-m wind, 2-m dewpoint, and 2-m temperature) are shown in Fig. 4. The impact of EnVAR hybrid analysis on low-level temperature, wind, and moisture is mostly neutral, which may indicate small spread and lack of the local weather details in the GDAS ensemble forecasts in the low levels. The surface wind and dewpoint forecasts show some minor improvement from the use of the hybrid analysis (Figs. 4a–d). Forecasts for 2-m temperature show neutral impact from the hybrid analysis (Figs. 4e,f). Similar neutral impact from the hybrid analysis is found in ceiling and precipitation verification from these two experiments (not shown). Overall, improved forecast skill from use of the EnVAR hybrid assimilation is more evident for upper-air fields and less so for near-surface fields in this study.

To further understand the GSI-hybrid analysis experiment results, the horizontally varying ensemble spreads of zonal wind and temperature from the 6-h GDAS ensemble forecast valid at 1200 UTC 1 June 2016 is shown in Fig. 5. Spread is shown for both a middle level (RAP level 19, near 500 hPa) and a low level (RAP level 3, near 80 m above the surface) with NCEP surface analysis map. The general pattern of the ensemble spreads are associated with the three low pressure systems located, respectively, near the northeastern United States, Hudson Bay, and west of Vancouver Island. The temperature spread at the lower level is smaller than that for the middle level (Figs. 5b and 5c) but the U-component spread at the lower level is larger (Figs. 5a and 5c). The spread fields are smooth and reflect mainly large-scale weather patterns. Further investigation of the GDAS ensemble with a rank histogram using the surface observations shows the classic U-shape distribution indicating the underdispersion of the ensemble forecast (figure not shown).

A special set of the GDAS ensemble forecasts covering this RAP retrospective period were rerun with hourly (instead of 3-hourly) output and forecasts out to 12 h. The retrospective RAP EnVAR experiment was repeated with 7- to 12-h GDAS ensemble forecast perturbations valid at each hourly RAP analysis time. The new EnVAR-hourly GFS experiment results are shown in Fig. 6 (blue line). The EnVAR-hourly GFS experiment has RMSE for upper air that is similar to the first EnVAR experiment. Thus, the GSI EnVar hybrid experiment

using the 6-hourly ensemble forecasts (currently used in operational RAP) is not further improved upon by more frequent availability of the GDAS ensemble forecast. The similar performance indicates that the GDAS ensemble BEC likely represents forecast error structure from larger-scale weather patterns (Fig. 5). Also, these results indicate that variance and multivariable correlations could be more important for RAP GSI-hybrid applications than the real-time flow dependence aspect of the perturbations, which will be discussed in more detail in section 5.

Applying GSI-hybrid assimilation in RAP greatly improved the RAP upper-air forecast accuracy (e.g., Fig. 4), but initially increased the wall-clock time used to run GSI about 300%. To obtain better efficiency, RAP GSI EnVar hybrid uses ensemble perturbations distributed on a coarser grid that is an integer times the analysis grid space. A retrospective experiment is conducted to investigate the impact of the coarser ensemble grid. The previous EnVar hybrid experiment (in Fig. 4) used an ensemble grid equal to the analysis grid, approximately 13-km horizontal grid spacing. The new comparison experiment set the ensemble grid to 3 times the analysis grid, resulting in an ensemble horizontal grid spacing of approximately 40 km. The upper-air RMSE profiles of 6-h forecasts for wind, relative humidity, and temperature against sounding observations are shown in Fig. 7. The EnVar hybrid experiment and the comparison experiment with coarser ensemble data have similar error statistics. Thus, the GSI-hybrid using the coarser ensemble grid data can produce the same quality forecast as one using the ensemble perturbations on the analysis grid. This result is not surprising because the grid spacing of the GDAS ensemble forecast used in this experiment is approximately 60 km. Therefore, interpolation of the GDAS data directly to the 13-km RAP analysis grid and to 3 times (40 km) the RAP analysis grid, produces similar results in GSI.

To summarize this section, using the GSI EnVAR hybrid data assimilation helps the RAP system by leading to significant improvements in the RAP upper-air forecast over the 3DVAR method. The use of the 6-hourly available GDAS/EnKF global ensemble on a 3-times coarser grid works equally well for the RAP forecast. These experiments enabled the application of the GSI EnVar hybrid for RAP version 2 to obtain significant positive impact with essentially no additional wall-time increase.

4. Sensitivity experiments for RAP data assimilation

After the operational implementation of the GSI-hybrid in RAP version 2, additional experiments were
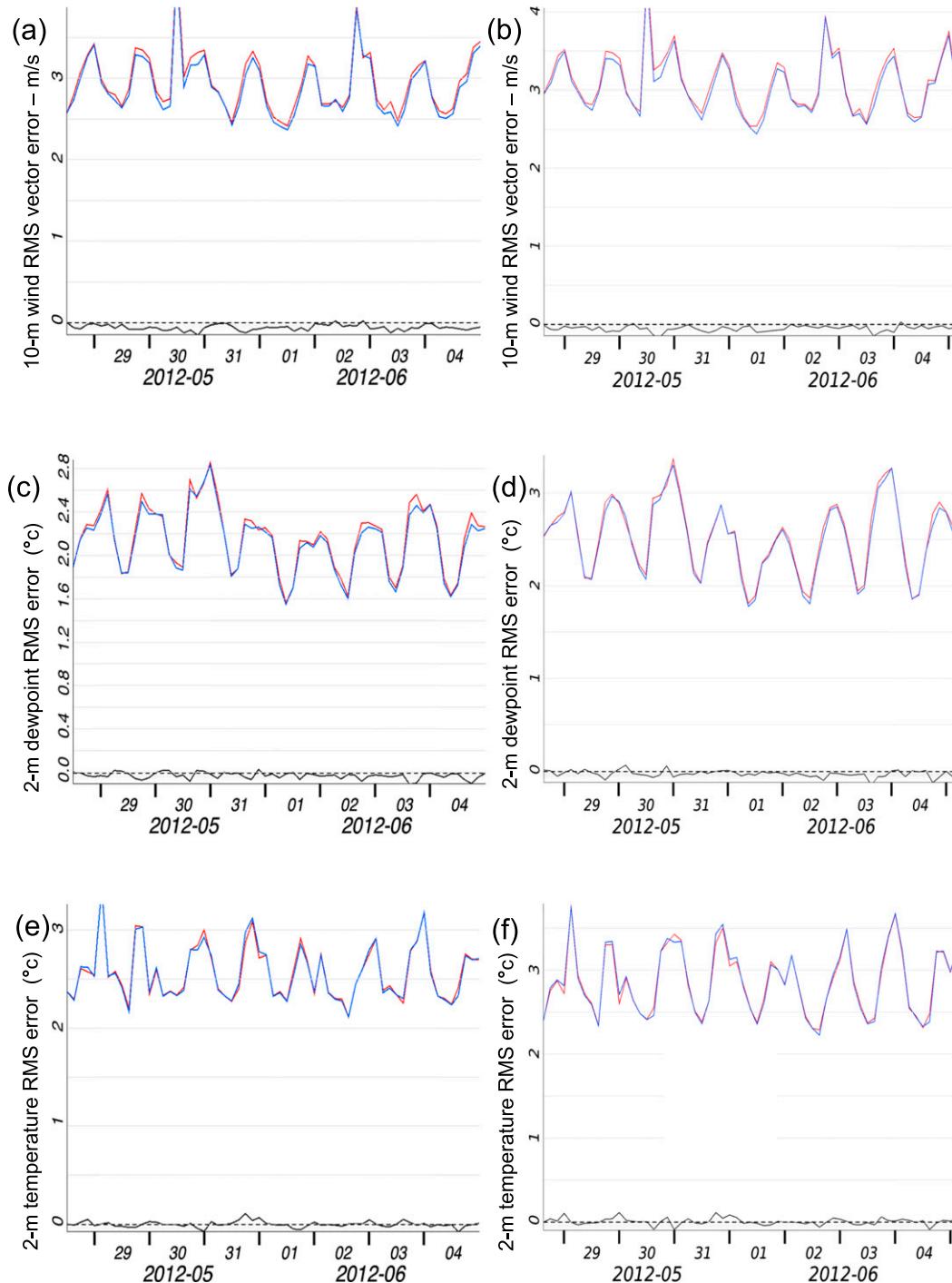
FIG. 4. For the same period as in Fig. 3, surface RMSE time series for (left) 3- and (right) 6-h forecast for (a), (b) 10-m wind (m s^{-1}); (c),(d) 2-m dewpoint (°C); and (e),(f) 2-m temperature (°C) errors against METAR observations between experiments using RAP GSI ensemble–variational hybrid analysis (EnVar, blue line) and the RAP GSI 3D-variational analysis (3Dvar, red line). Difference (EnVar − 3DVAR) is plotted in black.

conducted to further investigate sensitivity in RAP to hybrid assimilation parameters, such as the ratio between ensemble BEC and static BEC, horizontal localization, and vertical localization.

a. Ratio between ensemble BEC and static BEC

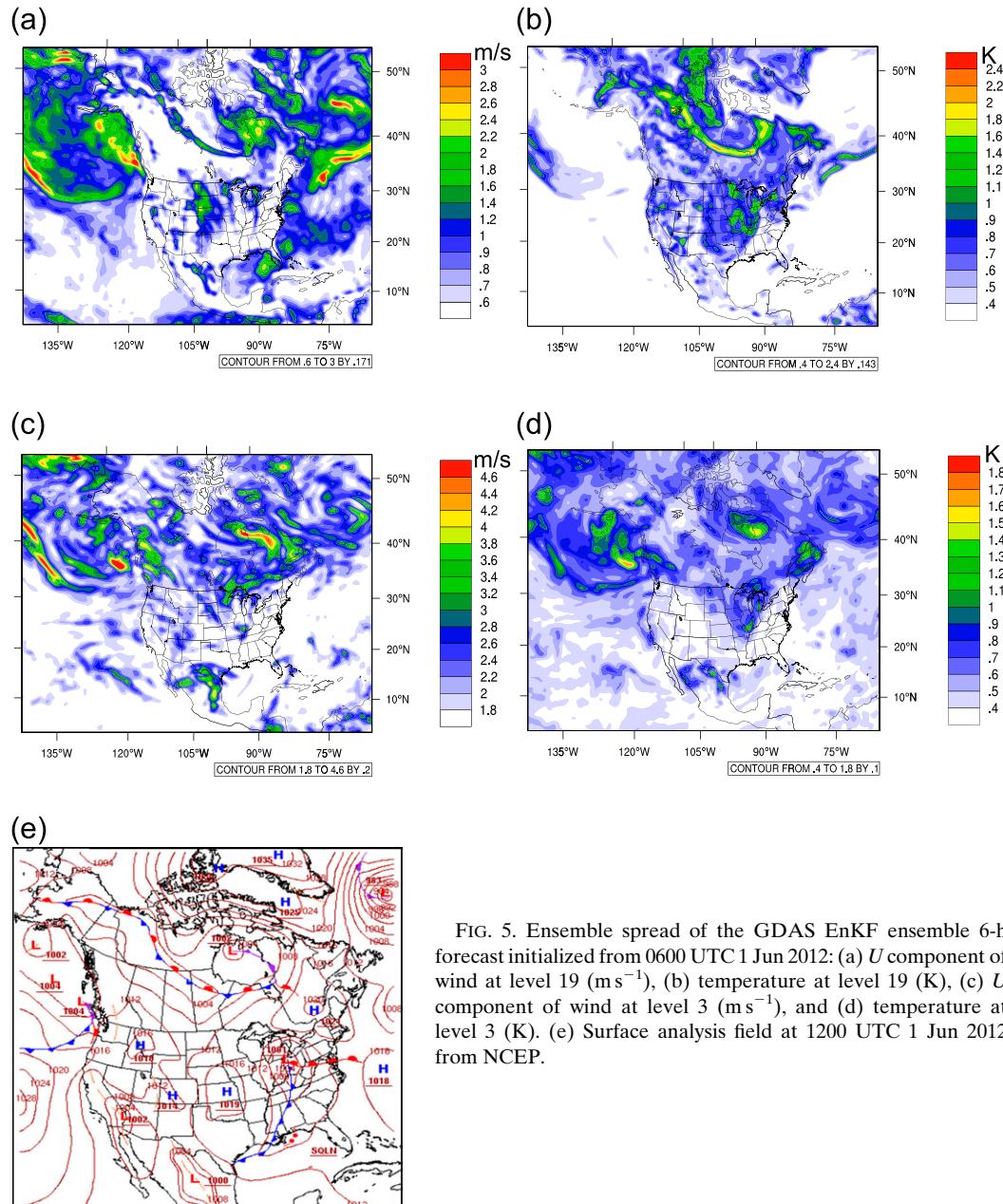Forecast skill improvement from applying hybrid analysis with the ensemble BEC is thought to be through

FIG. 5. Ensemble spread of the GDAS EnKF ensemble 6-h forecast initialized from 0600 UTC 1 Jun 2012: (a) U component of wind at level 19 (m s^{-1}), (b) temperature at level 19 (K), (c) U component of wind at level 3 (m s^{-1}), and (d) temperature at level 3 (K). (e) Surface analysis field at 1200 UTC 1 Jun 2012 from NCEP.

its provision of flow-dependent background error structure information through use of real-time ensemble forecasts in the analysis. The combined weight of the ensemble BEC and the static BEC in the GSI-hybrid analysis is set to add up to 1.0 in the GSI version used for this work. More weight given to the ensemble BEC means that more ensemble-based information will be included in the analysis. The RAP version 2 GSI-hybrid assimilation used 50% ensemble BEC and 50% static BEC. Two retrospective experiments (Table 1) were conducted with the ensemble BEC weight increased to

75% (25% static BEC) and to 100% (no static BEC). The impact of the ensemble BEC ratio is compared by examining the upper-air RMSE profiles of 3- and 6-h forecasts (Fig. 8).

When the fraction for ensemble BEC is increased from 50% (red line) to 75% (blue line), 3-h wind forecast skill is improved for all levels with most levels showing significant improvement (Fig. 8a). The 6-h forecast wind field also shows some improvement using increased ensemble BEC at most levels between 700 and 200 hPa (Fig. 8b). The larger improvement to the
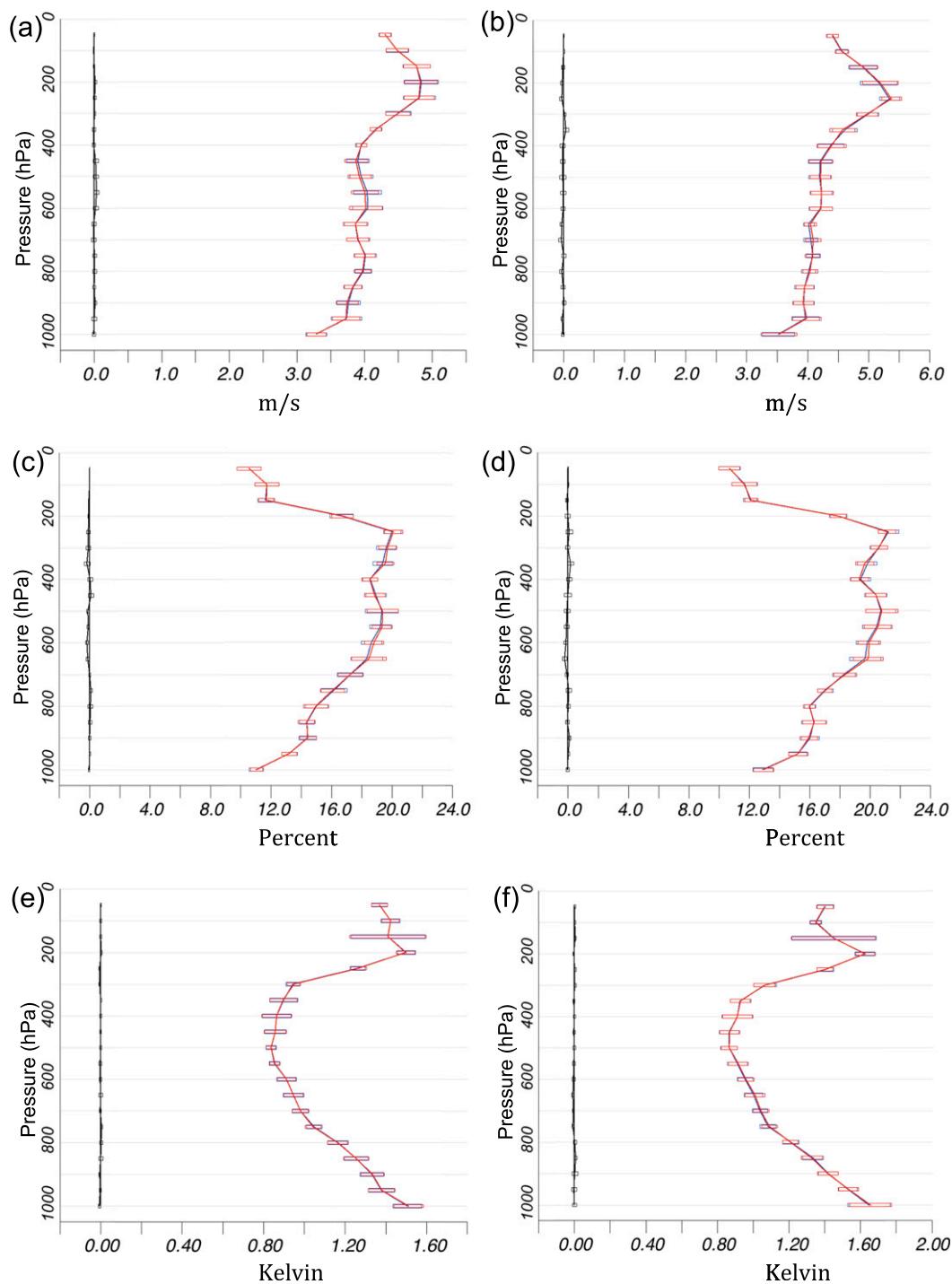
FIG. 6. As in Figs. 3a,c,e, but for upper-air RMSE profile of (left) 3- and (right) 6-h forecast error for (a),(b) wind (m s^{-1}); (c),(d) relative humidity (%); and (e),(f) temperature (K) against sounding observations in 1000–100 hPa between RAP experiment using 6-hourly available GFS ensemble (EnVar, red line) and an experiment using hourly available GFS ensemble (EnVar-hourly GFS, blue line). Difference (EnVar-hourly GFS − EnVar) is plotted in black. Boxes show 95% confidence.
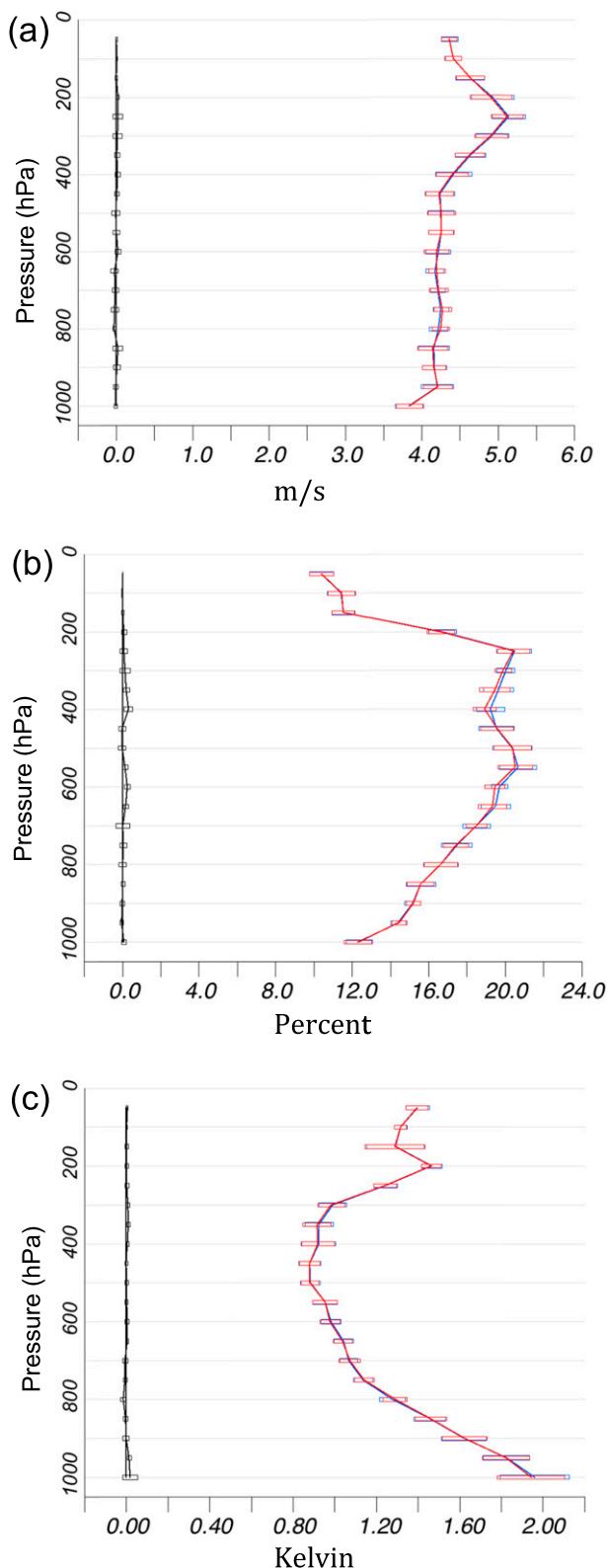
FIG. 7. As in Figs. 3a,c,e, but for RAP retrospective runs using the ensemble grid that is the same as the analysis grid (red line) and using the ensemble grid that is 3 times coarser than the analysis grid (blue line). Difference (3 times coarser grid − analysis grid) is plotted in black. Boxes show 95% confidence.

3-h forecast compared to the 6-h forecast indicates that the impact from using a larger BEC component from the ensemble BEC fades with longer forecast lead times, as the impact of the data assimilation also decreases with the longer forecast lead times. The 75% of ensemble BEC has a neutral impact on 3- and 6-h moisture and temperature forecasts (Figs. 8c,e). Increasing the ensemble BEC fraction to 100% (black lines) increases forecast errors compared to the 75% ensemble BEC experiment (blue lines) and the 50% ensemble BEC experiment (red lines). For example, the 3-h temperature forecasts from the 100% ensemble BEC experiment in some layers (e.g., 3-h forecasts below 950 hPa and above 400 hPa, 6-h forecasts below 850 hPa and above 500 hPa) show significant degradation compared to the other two experiments (Figs. 8e,f). The 6-h relative humidity forecasts below 850 hPa and between 650 and 200 hPa is also worse for the 100% ensemble BEC experiment (Fig. 8d). The only improvement from the 100% ensemble BEC experiment is in the shorter-range 3-h wind forecast (Fig. 8a). Based on these results, a weight of 75% for the ensemble BEC was chosen for the GSI-hybrid assimilation for RAP version 3 implemented at NCEP in August 2016. The static BEC still has some positive effect in the GSI-hybrid analysis application for RAP (Wu et al. 2017; Kleist and Ide 2015a).

b. Different horizontal and vertical localization scales

The RAP analysis uses many high-resolution observations and focuses on short-range mesoscale weather forecasts. Consistent with that, the horizontal and vertical localization scales used in the RAP GSI-hybrid are considerably (5–10 times) smaller than those used in other NCEP forecast systems using GSI-hybrid. To investigate sensitivity to these localization scales, a series of retrospective experiments were conducted with different horizontal and vertical localization scales (Table 1).

The effect of the horizontal localization value (110 km) used in the operational RAP version 2 was compared with the effect of using alternative larger values of 160, 220, and 330 km (Fig. 9). The upper-air RMSE profiles for 6- and 12-h forecasts of wind, relative humidity, and temperature were examined. In the 6-h forecast, the 110-km localization scale experiment has slightly smaller RMSE for winds between 850 and 450 hPa (Fig. 9a) and for moisture between 650 and 400 hPa (Fig. 9c). Other levels of the wind and moisture RMSE show little difference when different localization scales are used. The 6-h temperature forecasts for all horizontal localization scales generally produce the same temperature RMSE (Fig. 9e). For the 12-h forecasts, the error differences between the four horizontal localization scale experiments are even more subtle
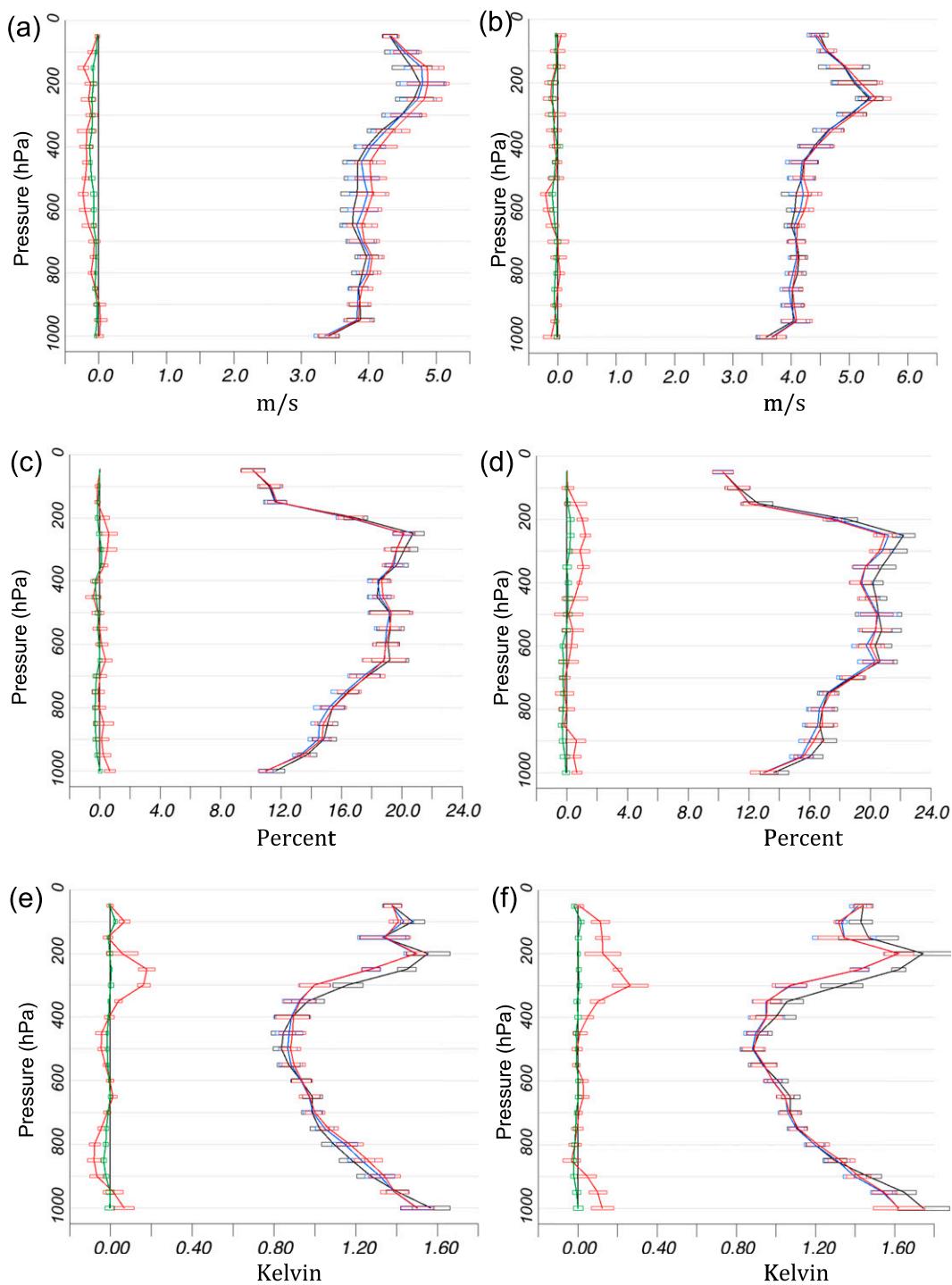
FIG. 8. As in Fig. 6, but for three RAP retrospective experiments with different BEC configurations. Exp 1 (red) 50% ensemble BEC and 50% static BEC (as in RAPv2). Exp 2 (blue) 75% ensemble BEC with 25% static BEC. Exp 3 (black) 100% ensemble BEC with no static BEC. Difference of Exp 2 − Exp 1 is shown in green. Difference of Exp 3 − Exp 1 is shown in red in the vicinity of the zero vertical line. Boxes show 95% confidence.
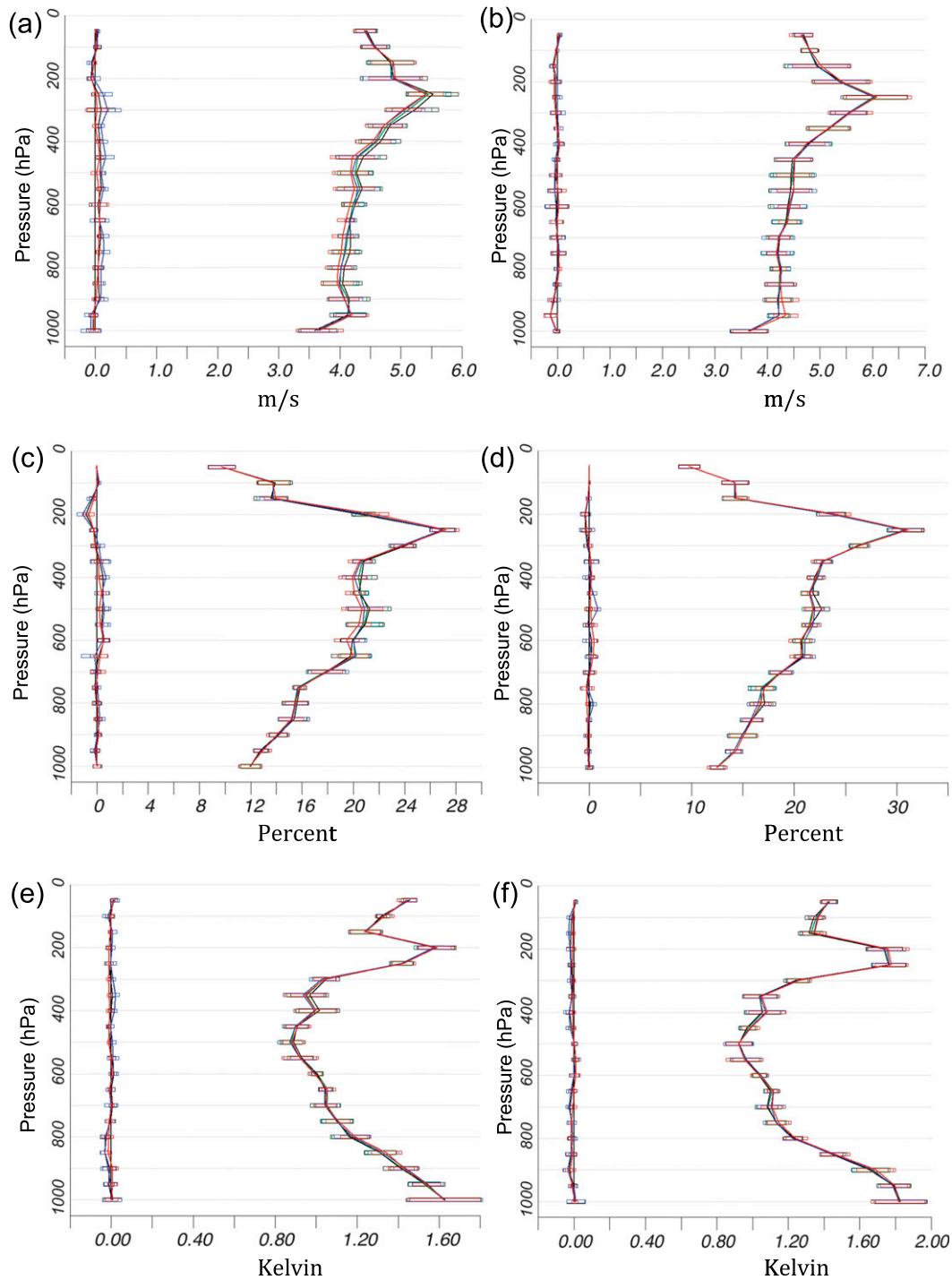
FIG. 9. As in Fig. 6, but for (left) 6- and (right) 12-h forecasts and for four RAP retrospective experiments with different horizontal localization scales: Exp 1 (red) 110 km, Exp 2 (blue) 160 km, Exp 3 (green) 220 km, and Exp 4 (black) 330 km. Difference of Exp 2 − Exp 1 is shown in red. Difference of Exp 3 − Exp 1 is shown in black. Difference of Exp 4 − Exp 1 is shown in blue in the vicinity of the zero vertical line. Boxes show 95% confidence.

(Figs. 9b,d,f). With the benefits of using a 110-km horizontal localization scale for 6-h forecasts and virtually no change in skill for 12-h forecasts, that value was retained in the RAP version 3 (B16; Table 1).

There are two options to specify the vertical localization scale in GSI, using the number of levels or using fixed vertical pressure depth, which represent Gaussian half-width distances. The operational RAP version 2 used three levels as the vertical localization scale. Here, RAP retrospective experiments using nine levels and 100 hPa (both broader) as the vertical localization, respectively, were conducted to compare with the RAP version 2.

Figure 10 shows the comparison of upper-air RMSE profiles of 0- and 6-h forecast fit to wind, relative humidity, and temperature sounding observations between experiments with vertical localization scale of three levels (red lines) and nine levels (blue lines). Increasing the number of levels for vertical localization, by definition, spreads the vertical impact of observations and decreases the fit of the analysis (0-h forecast) to the sounding observations (Figs. 10a,c,e). For 6-h forecasts, skill using nine levels as the localization scale is worse for low- and midlevel winds (Fig. 10b) and is slightly worse for temperature and moisture (Figs. 10d and 10f).

Another experiment using 100 hPa as the vertical localization scale has roughly similar results as the three-level localization scale experiment. Using the vertical localization method as defined by the number of vertical levels is potentially more desirable because the depth of model levels varies, with more levels in the atmospheric layers with finer sigma-coordinate stratification (e.g., boundary layer; B16 see their Table 7). It can be argued that observations in the regions of the atmosphere with potential large vertical gradients (i.e., shallow boundary layer) should have a narrower vertical impact than observations in other layers. Results from the vertical localization scales of three levels and 100 hPa are shown in Fig. 11 (similar to Fig. 10). The RMSE profiles for the two experiments are similar in 0-h analyses and 6-h forecasts except for low-level wind. The three-level vertical localization, again not surprisingly, results in a closer analysis fit to sounding observations for low-level wind, moisture, and temperature. The difference is due to the model levels being denser at low levels. For 6-h forecasts, there is very little difference in skill between these two experiments (Figs. 11b,d,f). The main goal of the RAP data assimilation is to provide the initial field that can make the best short-range forecast. At the same time, it is also desirable to have the analysis fit to the observations closer to reflect the current weather status for situational awareness. Therefore, the vertical localization scale of three levels has been retained in RAP

version 3. Overall, these horizontal and vertical localization experiments indicate that the RAP localization values used since the initial implementation of the hybrid analysis are appropriate for that NWP system.

5. Evaluation of temporal matching for flow-dependent contribution to RAP forecast skill

Applying the GSI-hybrid assimilation with GDAS ensemble to the RAP substantially improves the forecasts of upper-air wind, moisture, and temperature, despite the GDAS ensemble data only being available four times per day (section 3). In the RAP configuration for hourly analyses, GSI uses the latest GDAS ensemble forecast with valid time closest to the analysis time. The impact of using the off valid time GDAS ensemble forecast in GSI-hybrid analysis was already investigated and discussed in section 3 through comparing the EnVar and EnVAR-hourly GFS experiments. The results indicate there is no degradation to RAP forecast skill from using 6-hourly available GDAS ensemble forecasts instead of the hourly ensembles in GSI-hybrid analysis for the RAP system. This result is understandable from the fact that the GDAS ensemble is a coarser-resolution global forecast to represent large-scale weather patterns that do not change substantially within a 6-h period.

Still, a question remains on the sensitivity of RAP GSI-hybrid analysis to time-matching the global ensemble for the ensemble-based BEC component with the RAP analysis time. In this section, we describe experiments that exaggerate that time difference. A new set of four retrospective experiments from 15 to 21 June 2014 was conducted based on the updated RAP version 3 configuration. The first retrospective run used the GSI-3DVAR, and the second used the GSI-hybrid analysis with the GDAS ensemble that has a valid time close to the analysis time (time matching difference always less than 6 h). A third experiment used the GSI-hybrid analysis with a fixed ensemble forecast (i.e., the same set of GDAS ensemble forecasts valid near the time of the first RAP cycle on 15 June 2014 was used for all following hourly RAP cycles for the next 6 days through 21 June 2014). With the fixed ensemble forecast, the flow-dependent component should fade away after couple of days of cycling, and large-scale waves could have larger phase errors compared to the real-time ensembles. The fourth experiment, most extreme, also used the same hybrid assimilation but using global ensemble data from 6 months later (December 2014).

The results from first three experiments described above for flow-dependence study are shown in Figs. 12 and 13. Figure 12 shows upper-air RMSE profiles and Fig. 13 shows RMSE time series of 3- and 6-h forecast
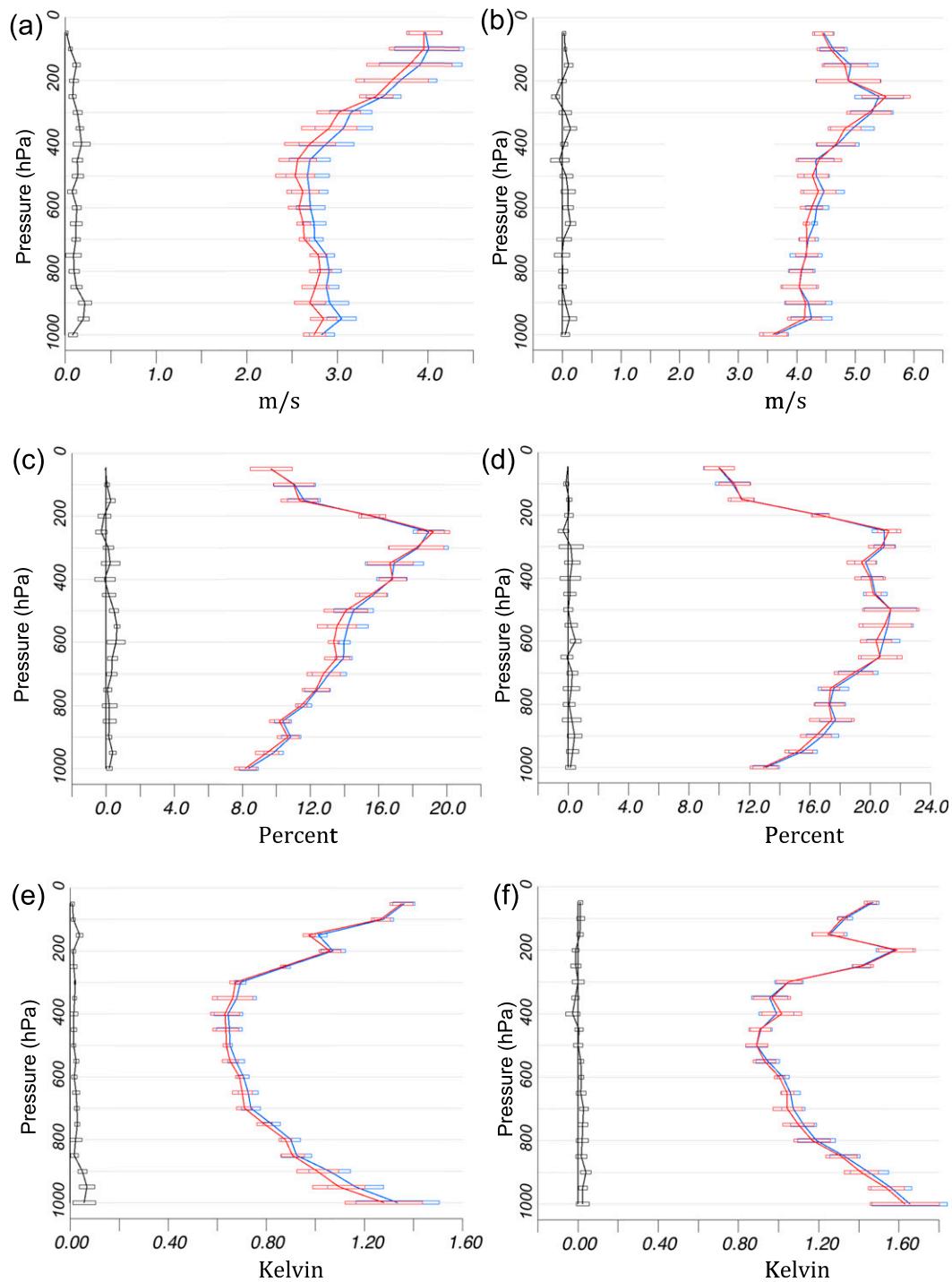
FIG. 10. Upper-air RMSE profiles of (left) 0- and (right) 6-h forecasts for (a),(b) wind (m s^{-1}); (c),(d) relative humidity (%); and (e),(f) temperature (K) against sounding observations in 1000–100 hPa. Two RAP retrospective experiments test the impact of vertical localization scales: three levels (red line) and nine levels (blue line). Difference (nine levels − three levels) is plotted in black. Boxes show 95% confidence.
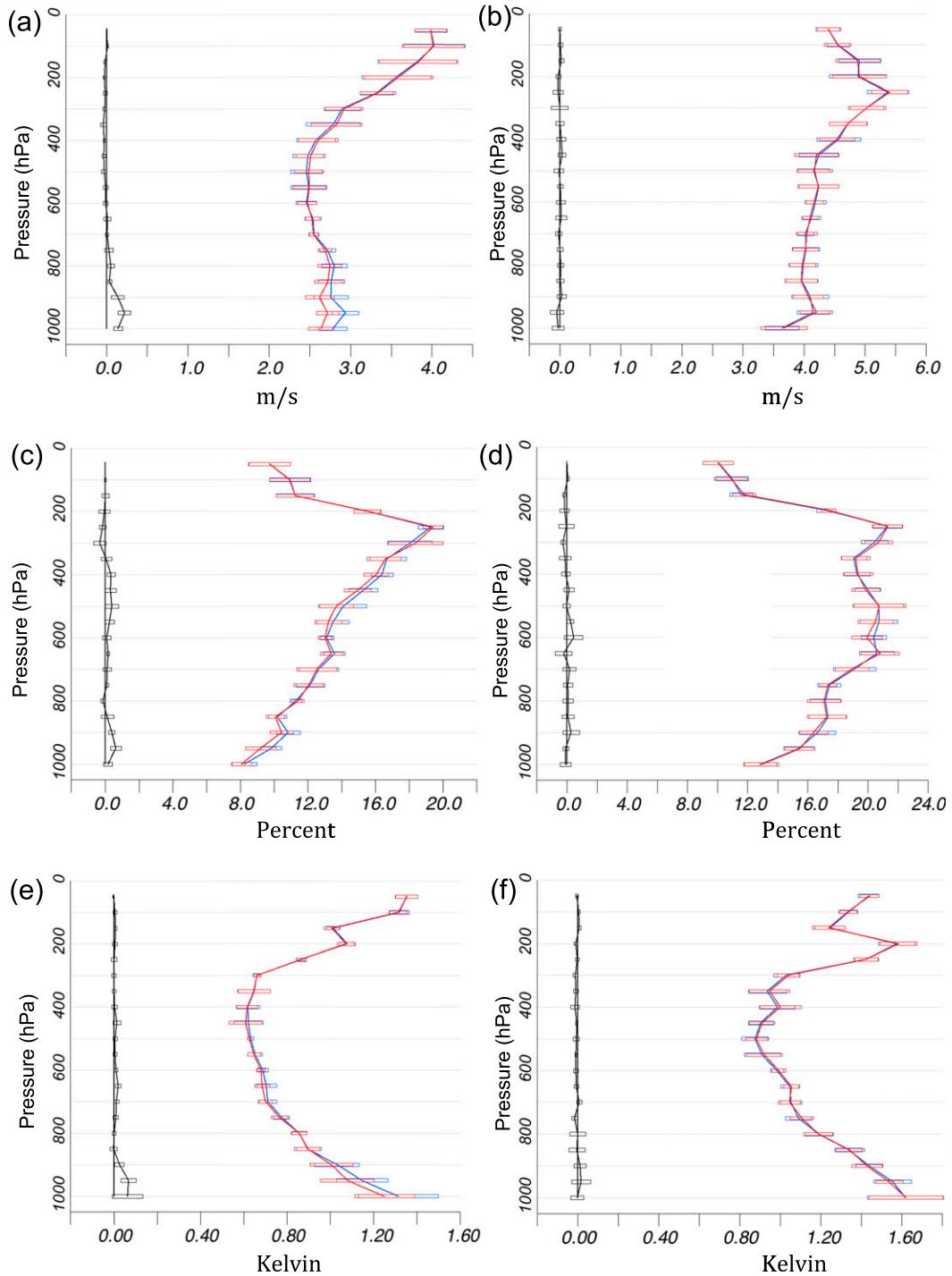
FIG. 11. Upper-air RMSE profiles of (left) 0- and (right) 6-h forecasts for (a),(b) wind (m s^{-1}); (c),(d) relative humidity (%); and (e),(f) temperature (K) against sounding observations in 1000–100 hPa. Two RAP retrospective experiments test the impact of vertical localization scales: three levels (red line) and fixed length option −0.15, which is about 100 hPa (blue line). Difference (fixed length option −0.15 − 3 levels) is plotted in black. Boxes show 95% confidence.
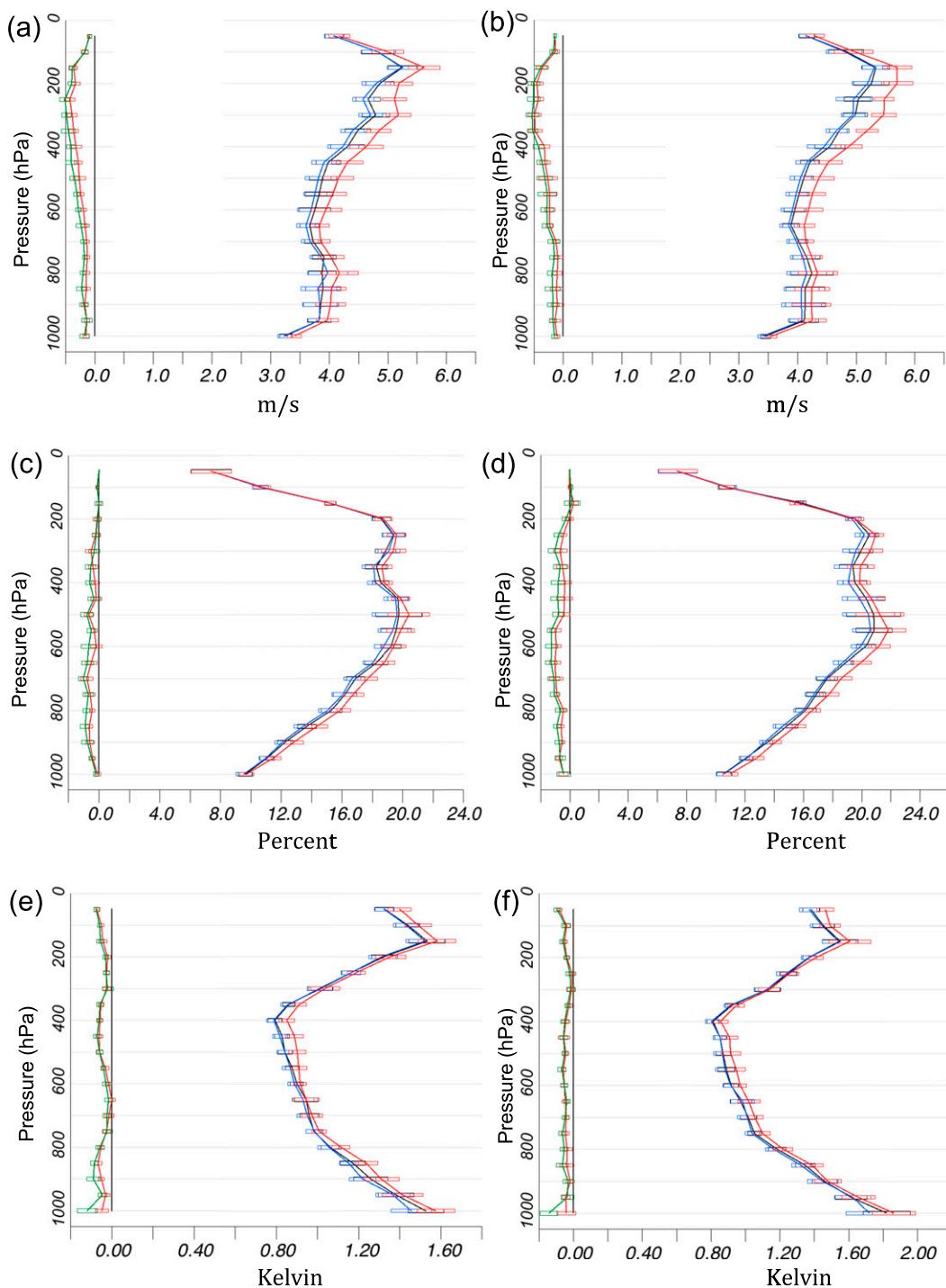
FIG. 12. Upper-air RMSE profiles of (left) 3- and (right) 6-h forecast errors for (a),(b) wind (m s^{-1}); (c),(d) relative humidity (%); and (e),(f) temperature (K) against sounding observations in 1000–100 hPa. Three RAP retrospective experiments running from 15 to 21 Jun 2014 were conducted to investigate the contribution from the flow dependence of the ensemble BEC: RAP using GSI 3DVAR (red line), RAP GSI-hybrid analysis using correct GFS ensemble forecast (blue line), and RAP GSI-hybrid analysis using fixed GFS ensemble forecast valid at the beginning of the test period (black line). Difference of hybrid analysis using corrected GDAS ensemble minus 3DVAR is plotted in green close to 0 and difference of hybrid analysis using fixed GDAS ensemble minus 3DVAR is plotted in red close to 0. Boxes show 95% confidence.
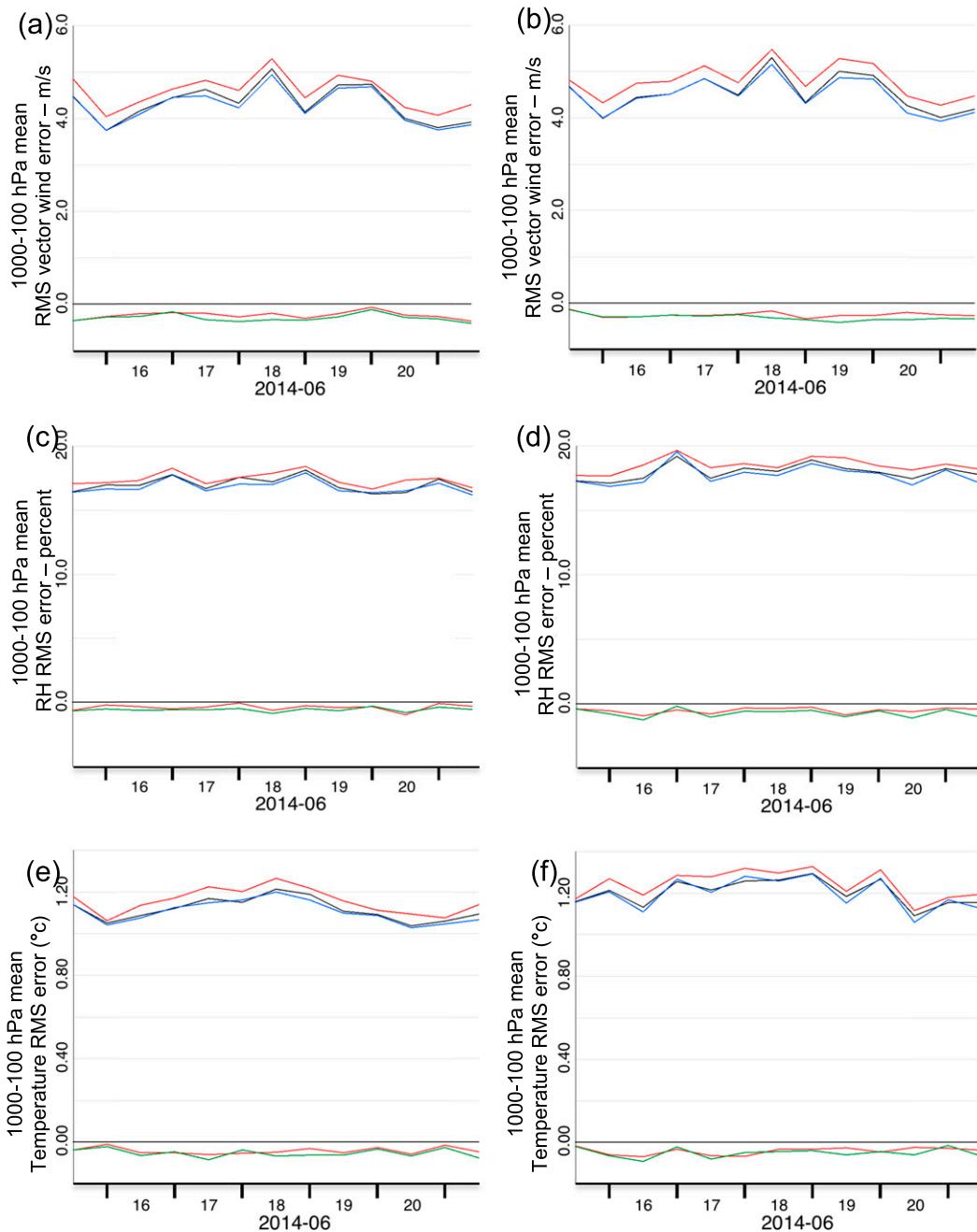
FIG. 13. The same experiments as in Fig. 12, but for upper-air RMSE time series of (left) 3- and (right) 6-h forecasts for (a),(b) wind; (c),(d) relative humidity; and (e),(f) temperature against sounding observations in 1000–100 hPa.

errors for wind, relative humidity, and temperature against sounding observations. In Fig. 12, the RMSEs from two experiments using the hybrid analysis (blue and black lines) are close to each other but both are significantly smaller than those from experiment using 3DVAR (red lines). In other words, the hybrid EnVar analysis in RAPv3 continued to produce improved short-range forecasts over the 3DVAR with fixed BEC,

the same results as in experiments with RAPv2 (Fig. 3). As in the experiments in section 3, the most significant improvements from a hybrid analysis are in the wind forecasts, followed by moisture and then temperature. Profiles from the two hybrid experiments indicate that the hybrid analysis with the correct ensemble (blue lines) valid close to the analysis time does give slightly smaller RMSE. The extra improvement from using the

correct (or real-time flow dependent) ensemble is most clear in wind and moisture forecasts from 900 to 200 hPa in both 3- and 6-h forecast (Figs. 12a–d) and very small for temperature forecasts (Figs. 12e,f) at most levels. Also, the hybrid experiment with the correct ensemble gives smaller RMSE for temperature forecasts near the surface. This may come from the fact that the near-surface moisture and temperature could change a lot with different weather patterns (e.g., with/without cloud/precipitation).

The time evolution of the RMSE (Fig. 13) for these experiments with different degrees of ensemble-analysis time-matching supports the profile results shown in Fig. 12. Consistent in time, the RMSE from the two experiments with hybrid analysis (blue and black lines) are close to each other and both are smaller than ones from 3DVAR (red lines). The experiment with the correct ensemble (blue line) has slightly smaller RMSE at most times compared to the experiment with the fixed-in-time ensemble (black line). The benefit of the flow-dependent information can be seen in wind forecast skill as the time offset between the fixed ensemble forecast valid time (15 June) and the analysis time becomes longer (Figs. 13a,b). The 6-h wind forecast RMSEs from the two hybrid experiments on 15 and 16 June are almost identical and then grow further apart after 17 June. The moisture and temperature forecasts from using the fixed-in-time ensemble (at 15 June) are slightly degraded compared to results from use of the real-time ensemble forecast for assimilation (Figs. 13c–f).

It is somewhat surprising that the RAP GSI-hybrid analysis is not highly sensitive to the valid time of the global ensemble, and the experiment with an ensemble valid close to the analysis time gives only a slightly better forecast compared to the experiment with fixed-in-time ensemble. Both hybrid experiments are able to produce more accurate short-range forecasts than the 3DVAR experiment. Quantitatively, the difference between the RMSE profiles in Fig. 12 and the time series in Fig. 13 show that using the fixed ensemble can provide about 80% of the improvement that the experiment with the correct ensemble can give over 3DVAR, at least over this 6-day test period. These results suggest that the GFS ensemble BEC provides a time-invariant benefit to help improve the forecast for the large-scale weather system. To further improve the forecast to local- and storm-scale system, high-resolution ensemble forecast from RAP ensemble or high-resolution global ensemble are needed to provide enough details in ensemble spread space on those local and storm systems.

To further investigate this topic, a fourth RAP retrospective experiment was conducted using GSI-hybrid analysis with a GDAS ensemble from a completely different season than the test period. We assume that this off-season (6 months later) GFS ensemble represents a totally different weather pattern from the retrospective period, and includes little or no time- and flow-dependent information. The results from the new experiment (green lines) were plotted in Fig. 14 with the results from the other three experiments. RMSE profiles from the new off-season-ensemble experiment (green lines) mostly overlay with the profile from the fixed ensemble experiment (black lines). This confirms that the slight improvement from the correct ensemble experiment is from the real-time flow-dependent information in the real-time ensemble that is valid close to the analysis time.

The RAP static BEC was calculated using the NMC method based on perturbations from a year-long period and it reflects long-term average features of the model error covariance. When an ensemble is used in GSI-hybrid analysis, the distribution of analysis increment, ratio of analysis fitting to the observation, and the balance among different analysis fields are mainly calculated based on the ensemble perturbations, which reflect the status of the background error covariance more accurately than static BEC in term of the weather pattern and dynamical and thermodynamic relations among difference variables. The clear improvements from hybrid assimilation with both fixed and off-season global ensemble perturbations indicate that balances including cross-variable covariances from those ensembles are still much better than those from the static BEC. Further, the portion of the weather pattern with strong flow-dependent features, like fronts and jets, only exist in limited areas of the large RAP domain at any given time. Thus, most of the domain has atmospheric conditions in which balances among different analysis variables can be represented well by a GDAS-based ensemble from any time.

These GSI-hybrid analysis experiments provide insights on the setup for the RAP system: we can dismiss the requirement that the ensemble forecast must be valid within 6 h of the analysis time. Still the RAP should try to use the ensemble BEC valid as close to the analysis time as possible.

6. Summary and discussion

Since the Rapid Refresh became operational at NCEP in 2012, it has played an important role in decision-making related to severe weather, aviation, and renewable energy (Wilczak et al. 2015). The data assimilation component of the RAP, GSI, has been developed, refined, and tested to better benefit these important forecast applications.

GSI has been developed for many years by NCEP in collaboration with other institutions, including the Global Systems Division (GSD), for operations and community
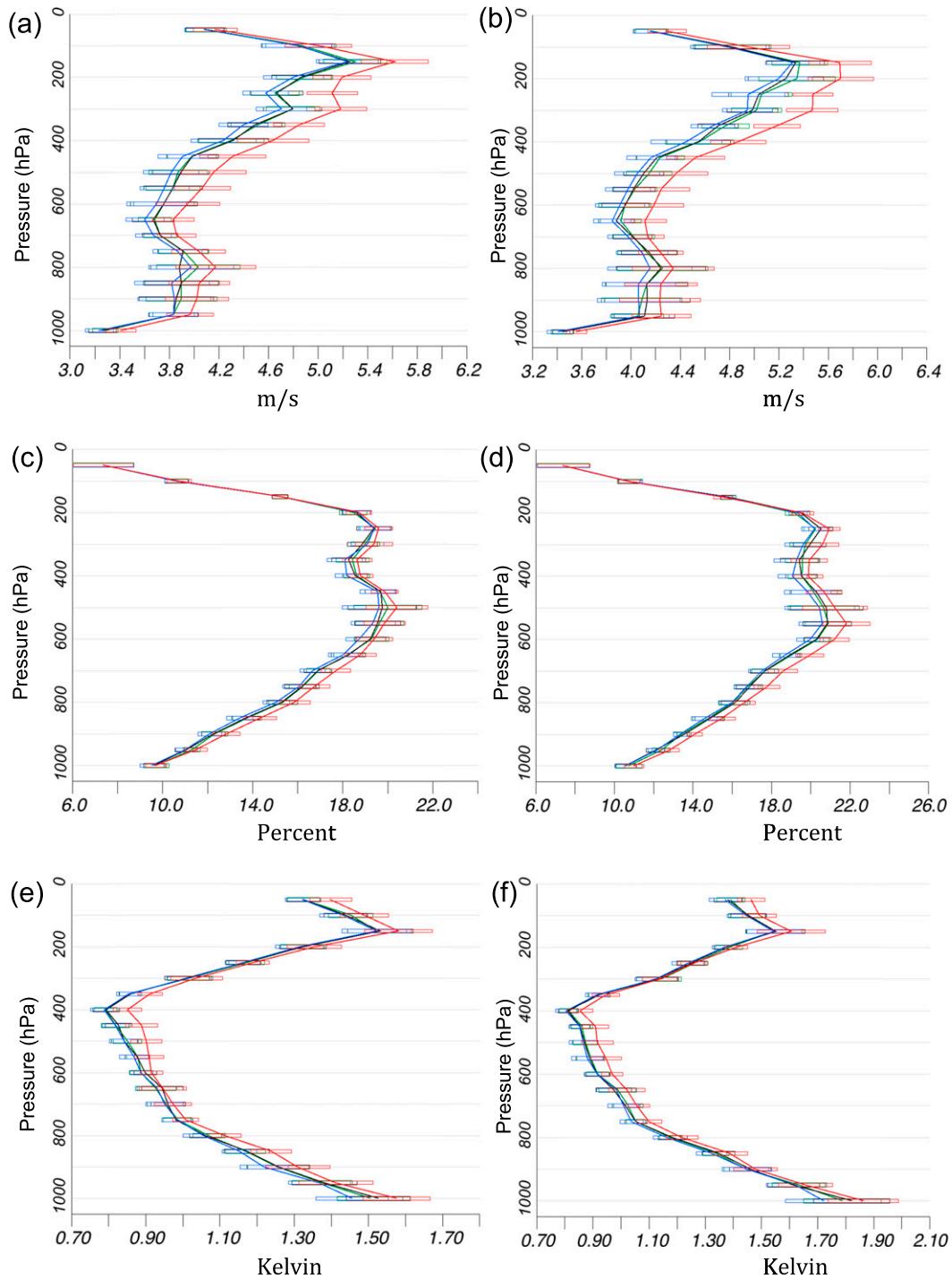
FIG. 14. As in Fig. 12 (June 2014 period), but with one additional experiment for hybrid assimilation using off-season ensemble from December 2014 (green line). Boxes show 95% confidence.

research. It now includes useful well-developed functions for data assimilation, such as forward model operators for many types of observations, detailed quality control, effective parallel computation, flexible data usage control, treatment for cloud/hydrometeors, etc. Dual-resolution

ensemble-variational hybrid analysis is one of those functions, successfully applied in the RAP system for the RAP version 2 upgrade. This article introduces the testing, evaluation, and parameter sensitivity for GSI ensemble–variational hybrid analysis for RAP operations.

Because of the latency and 6-h cycling of the real-time GDAS/EnKF ensemble, hourly cycling assimilation in the RAP can only use the global ensemble forecasts valid close to the desired analysis time instead of exactly matching the analysis time for most of its cycles. In RAP operations, the GSI analysis in six continuous cycles uses the same GDAS/EnKF ensemble for the hybrid analysis.

The tests of GSI-hybrid assimilation versus the 3DVAR shows that the use of hybrid assimilation with GDAS/EnKF ensemble data valid close to the analysis can significantly improve upper-air forecasts from the RAP system, but the impact on surface, ceiling, and precipitation forecasts appear to be neutral. The same conclusions can be drawn from two other one-week RAP retrospective experiments that compared hybrid analysis versus 3DVAR analysis in difference season (results not shown in this paper). Comparing the RAP version 2 hybrid configuration with 6-hourly GFS ensemble data versus an experiment with hourly GFS ensemble data valid exactly at analysis time indicates that application of GFS ensemble data valid at each hourly time does not improve RAP forecast skill. A similar result is found in another sensitivity test: results using ensemble grid data at 3 times coarser resolution (cheaper computationally) than data on the analysis grid show no difference in accuracy. These results indicate that while the GDAS/EnKF ensemble reflects primarily large-scale weather pattern forecast error covariance, its application in the regional RAP GSI-hybrid analysis is very effective for regional short-range forecast accuracy.

RAP retrospective experiments were conducted for additional sensitivity tests for different ratios of ensemble and static BEC, and for different values for horizontal and vertical localization scales to determine the best GSI-hybrid configuration for RAP system. The results show the original horizontal and vertical localization values used in RAP assimilation are appropriate for RAP forecast skill. In the RAP version 3 upgrade in August 2016, the fraction of ensemble BEC was increased from 50% to 75% (based on results described here), and the horizontal and vertical localization scales were kept at 110 km and three levels.

Generally, improvements from hybrid data assimilation are attributed to the real-time flow-dependent information brought in through the ensemble BEC. However, our tests with GSI-hybrid using a fixed ensemble or even from an ensemble valid 6 months later suggest that the use of any available GDAS/EnKF global ensemble BEC can help to improve RAP analysis and forecast skill over using the static BEC only. This result indicates that the impacts of regional GSI-hybrid analysis are not sensitive to the valid time of the global ensemble used, and the time-independent components of the cross-variable covariances (but not necessarily the time-dependent flow-dependent component) in the ensemble-based BEC are better than those from the NMC method. This result is helpful in understanding the practical implications of the GSI-hybrid technique. For example, the RAP should always use the hybrid ensemble–variational hybrid option within GSI, even if the GDAS/EnKF ensemble forecast may be older than 6 h from analysis time. Also, if no real-time ensemble data are available for a user (including in the research community), GSI-hybrid assimilation can still be applied with a global ensemble forecast to obtain the benefit of hybrid assimilation. Another application of this result is to generate a new static BEC based on perturbations collected from global ensemble forecast instead of the NMC method. The new static BEC should have better multiple-variate balance and further improve the GSI-hybrid analysis in RAP.

The range of the hybrid localization and hybrid weight tested in this study is limited. Most of the results in this study were from one-week long experiments, which are relatively short, but found to be consistent on a day-to-day basis. Those values should be used as initial reference for high-resolution short-range data assimilation systems and users are suggested to conduct their own experiments to find the best values for their application.

Improved short-range forecast performance from the RAP demonstrated here with the hybrid ensemble-variational analysis capability in GSI improves decision-making for many safety- and economic-related activities. In the future, GSI-hybrid analysis studies will be conducted with higher-resolution regional ensemble forecasts instead of global ensemble forecasts in an attempt to improve the surface, cloud, and precipitation forecasts. The GSI is a product of many years of development and cooperation from its combined operational and research NWP community.

regional GSI-hybrid analysis in this study was developed by David Parrish, Wan-Shu Wu, and Daryl Kleist in NOAA/NWS/NCEP/EMC.

REFERENCES

Benjamin, S. G., and Coauthors, 2004a: An hourly assimilation/ forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518, doi:10.1175/1520-0493(2004)132<0495:AHACTR>2.0.CO;2.

——, G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004b: Mesoscale weather prediction with the RUC hybrid isentropic/terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473–494, doi:10.1175/1520-0493(2004)132<0473: MWPWTR>2.0.CO;2.

——, and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, doi:10.1175/MWR-D-15-0242.1.

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, doi:10.1256/qj.04.15.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter– 3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

James, E. P., and S. G. Benjamin, 2017: Observation system experiments with the hourly updating Rapid Refresh model using GSI hybrid ensemble–variational data assimilation. *Mon. Wea. Rev.*, **145**, 2897–2918, doi:10.1175/MWR-D-16-0398.1.

Kleist, D. T., and K. Ide, 2015a: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon. Wea. Rev.*, **143**, 433–451, doi:10.1175/MWR-D-13-00351.1.

——, and ——, 2015b: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4D EnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, doi:10.1175/MWR-D-13-00350.1.

——, D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP global data assimilation system. *Wea. Forecasting*, **24**, 1691–1705, doi:10.1175/2009WAF2222201.1.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-VAR. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203, doi:10.1256/qj.02.132.

——, 2013: Recommended nomenclature for EnVar data assimilation methods. *Research Activities in Atmospheric and Oceanic Modeling,* WGNE, 2 pp., http://www.wcrp-climate.org/WGNE/ BlueBook/2013/individual-articles/01_Lorenc_Andrew_EnVar_ nomenclature.pdf.

Parrish, D., and J. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763, doi:10.1175/1520-0493(1992)120<1747: TNMCSS>2.0.CO;2.

Peckham, S. E., T. G. Smirnova, S. G. Benjamin, J. M. Brown, and J. S. Kenyon, 2016: Implementation of a digital filter initialization in the WRF Model and application in the Rapid Refresh. *Mon. Wea. Rev.*, **144**, 99–106, doi:10.1175/MWR-D-15-0219.1.

Shao, H., and Coauthors, 2016: Bridging research to operations transitions: Status and plans of community GSI. *Bull. Amer. Meteor. Soc.*, **97**, 1427–1440, doi:10.1175/BAMS-D-13-00245.1.

Wang, X., 2010: Incorporating ensemble covariance in the grid-point statistical interpolation variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995, doi:10.1175/2010MWR3245.1.

——, C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble/3D-Var hybrid analysis schemes. *Mon. Wea. Rev.*, **135**, 222–227, doi:10.1175/ MWR3282.1.

——, 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.*, **141**, 4098–4117, doi:10.1175/MWR-D-12-00141.1.

Weatherhead, E. C., and Coauthors, 1998: Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J. Geophys. Res.*, **103**, 17 149–17 161, doi:10.1029/98JD00995.

Weygandt, S. S., and S. G. Benjamin, 2007: Radar reflectivity-based initialization of precipitation systems using a diabatic digital filter within the Rapid Update Cycle. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction,* Park City, UT, Amer. Meteor. Soc., 1B.7, https://ams.confex. com/ams/22WAF18NWP/techprogram/paper_124540.htm.

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**, 463–48, doi:10.1175/ 2007MWR2018.1.

Wilczak, J., and Coauthors, 2015: The Wind Forecast Improvement Project (WFIP): A public–private partnership addressing wind energy forecast needs. *Bull. Amer. Meteor. Soc.*, **96**, 1699–1718, doi:10.1175/BAMS-D-14-00107.1.

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, doi:10.1175/ 1520-0493(2002)130<2905:TDVAWS>2.0.CO;2.

——, D. F. Parrish, E. Rogers, and Y. Lin, 2017: Regional ensemble–variational data assimilation using global ensemble forecasts. *Wea. Forecasting*, **32**, 83–96, doi:10.1175/ WAF-D-16-0045.1.