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

Yujie Pan[1,3], Kefeng Zhu[1,4], Ming Xue[1,2,4*], Xuguang Wang[1,2], Ming Hu[5,6], Stanley G. Benjamin[5], 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]Nanjing University, Nanjing, China

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

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

Submitted to Monthly Weather Review
July 25, 2013
Final Revision: May 2014

*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 EnSRF-En3DVar hybrid data assimilation (DA) 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 an ensemble square root filter (EnSRF) system for RAP, which provides ensemble perturbations. Recursive filters (RF) are used to localize ensemble covariance in both horizontal and vertical within the En3DVar.

The coupled En3DVar hybrid system is evaluated with 3-hour 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 or about 40-km grid spacing, and its performance is compared to parallel GSI 3DVar and EnSRF 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 3DVar and EnSRF. When the recursive filter coefficients are tuned to achieve a similar height-dependent localization as in the EnSRF, the En3DVar results using pure ensemble covariance are close to EnSRF. Two-way coupling between EnSRF 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 3DVar 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., Purser et al. 2003a; Wu et al. 2002), 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 weight through a least square approach. 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; Hamill et al. 2001; Anderson 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 (Bishop et al. 2001; Anderson 2001; Hunt et al. 2007; Whitaker and Hamill 2002; Evensen 1994) 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., Raynaud et al. 2011; Bonavita et al. 2012; Hamill et al. 2011b; 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. 2014). 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 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 determining optimal localization that alleviates the rank deficiency problem in concert with appropriate covariance inflation makes tuning expensive (Anderson 2007; Anderson 2012). Another approach that can help alleviate this problem is to combine the full-rank static BEC with the rank-deficient ensemble BEC, creating a so-called hybrid[1] algorithm.

[1] In this study, we use the word 'hybrid' to refer to a combination of the static and ensemble-derived flow-dependent covariances, i.e., the hybrid covariance.

Hamill and Snyder (2000) were the first to propose a 3DVar-based hybrid scheme in which 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 additional equation constraints in the cost function (e.g., Ge et al. 2012; Kleist et al. 2009b). 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. (2010a, b), 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 than those initialized using 3DVar. Buehner et al. (2010a, b) 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. (2008a, b) 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 (LETKF), and the regular EnKF for ensemble perturbation updating, respectively, and WRF-ARW as the prediction model, for a hurricane period. In general, the introduction of flow-dependent ensemble covariance

[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. (2008) 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 (2013) their algorithm is renamed 4DEnVar, to better differentiate the algorithm from traditional 4DVar. Our current usage is also consistent with the conventions used by papers at the recent WMO Data Assimilation Symposium of Maryland, United States in October 2013.

91 into 3DVar or 4DVar improves the forecast results. In fact, for the NCEP operational Global
92 Forecasting System (GFS), an EnKF-En3DVar hybrid DA system (Whitaker et al. 2011; Hamill et
93 al. 2011b) based on an EnKF and the operational Grid-Point Statistical Interpolation (GSI) 3DVar
94 (Kleist et al. 2009a) was developed and operationally implemented in 2012, replacing GSI 3DVar.
95 Wang et al. (2013) reported the testing results from the GSI-based En3DVar hybrid system for GFS
96 at a reduced resolution.

97 It has been a general decision at NCEP that the hybrid DA approach will be applied to its
98 regional models as well, including the North America Mesoscale (NAM) model and the recently
99 implemented (on 1 May 2012) Rapid Refresh (RAP) system, the replacement to the Rapid Update
100 Cycle (RUC, Benjamin et al. 2004). Towards this end, an EnKF system was recently established for
101 the RAP and tested at a reduced resolution by Zhu et al. (2013, hereafter Z13) using the operational
102 observation data stream of RAP. The same as for the GFS EnKF system, the ensemble square-root
103 filter (EnSRF) algorithm of Whitaker and Hamill (2002) was used in Z13. As one of the
104 deterministic ensemble filter algorithms, EnSRF avoids sampling issues associated with the use of
105 "perturbed observations" (Whitaker and Hamill 2002; Tippett et al. 2003).

106 In Z13, short-range (up to 18 hours) forecasts from 3-hourly EnSRF analyses over a 9-day
107 period were found to be consistently better than forecasts from corresponding GSI 3DVar analyses,
108 in terms of both model state forecasts and precipitation forecast skill scores. The primary goal of
109 this current work is to extend the work Z13 by establishing and testing a coupled EnSRF-En3DVar
110 hybrid DA system for RAP that can potentially be implemented operationally. As the first step, we
111 test and evaluate the hybrid DA system running at 1/3 of the native resolution of operational RAP;
112 running the EnSRF DA system at this reduced resolution is dictated by the limited operational
113 computing resources in the near future. Moreover, running the En3DVar hybrid analyses at the
114 same resolution facilitates easy and direct comparisons with the EnSRF results, and also provides us

115 with a benchmark for a future dual-resolution implementation. With a dual-resolution
116 implementation, the En3DVar analyses will be run at a higher resolution, using the reduced-
117 resolution ensemble perturbations (as is with the operational GFS hybrid DA system). In this paper,
118 we focus on documenting and comparing the results obtained from all three systems, i.e., the GSI
119 3DVar, EnSRF and En3DVar hybrid, at the reduced, 40-km grid spacing.

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

125 **2. GSI-based EnSRF-En3DVar hybrid system for Rapid Refresh**

126 *a) The Rapid Refresh System*

127 The operational hourly-updated RUC system was designed to improve short-range weather
128 forecasting through frequent updating of initial conditions with the latest observations (Benjamin et
129 al. 2004). The RAP is a replacement of the RUC system and is based on the non-hydrostatic WRF-
130 ARW dynamic core (Skamarock et al. 2005). RAP became operational at NCEP on May 1, 2012
131 using the GSI 3DVar as the data assimilation system. In February, 2014, the system was upgraded
132 to incorporate BEC derived from ensemble perturbations from the 80-member GFS EnSRF system
133 that feeds the operational hybrid En3DVar DA system of the GFS model. The GSI is an unified DA
134 framework for both global and regional models (Kleist et al. 2009a). The horizontal grid spacing of
135 RAP is ~13 km and has 50 vertical levels extending up to 10 hPa at the model top. Compared to the
136 RUC, the RAP system is capable of assimilating more observations, including satellite radiance
137 data, and has a larger domain which covers the entire North America. The physics options used by

138 the operational RAP include the Grell-G3 cumulus parameterization, Thompson microphysics,
139 RRTM longwave radiation, Goddard shortwave radiation, MYJ turbulent mixing, RUC-Smirnova
140 land-surface model. Details on these schemes can be found in Benjamin et al. (2009).

141 As with the RUC, the RAP employs a digital filter initialization (DFI) to reduce high-
142 frequency noise during the initial period of model integration. In the operational RAP system, twice
143 DFI (TDFI) (Lynch and Huang 1992), which applies the DFI twice, once on the adiabatic
144 backward time integration and once on the full-physics forward time integration, is used.
145 Considering that for high-resolution applications where diabatic processes are more important,
146 adiabatic integration can introduce significant errors, Z13 chose to employ the digital filter
147 launching (DFL) procedure (Lynch and Huang 1994) instead in their EnSRF system for RAP. DFL
148 applies the DFI only once, on the forward integration time series. In this study, the same procedure
149 is followed by the EnSRF and En3DVar hybrid experiments. In our tests with 3-hourly cycles
150 reported in this paper, the DFL employs a 40-minute filter window centered at 20 minutes of
151 forecast time, and used Dolph filter (Lynch 1997) with a cutoff half width of 20 minutes.

152 *b) The coupled EnSRF-En3DVar hybrid system for RAP*

153 As mentioned earlier, our En3DVar hybrid system is based on the operational GSI 3DVar
154 system for RAP and it uses the operational data stream of RAP. To facilitate direct comparisons
155 with the RAP EnSRF and GSI 3DVar systems as reported in Z13, we run our hybrid tests also at the
156 reduced resolution of ~40 km grid spacing with 3-hourly assimilation cycles instead of the ~13 km
157 grid spacing and hourly cycles of the operational RAP. The use of the reduced-resolution EnSRF
158 system is due to the expected constraint in available operational computational resources. The
159 choice of 3-hourly cycles is to enable us to run a larger number of experiments and for more rapid
160 prototyping of the system. The running of the continuously cycled experiments over a 9-day period

161 is computationally expensive in terms of both CPU and storage requirements. Extensive
162 experimentation and tuning were required to arrive at quasi-optimal configurations of the RAP
163 EnSRF system, including configurations of covariance inflation and localization. For future
164 operational implementation, it is desirable to run the En3DVar at the native RAP resolution, while
165 using lower-resolution EnSRF perturbations in a dual-resolution model to save computational cost.
166 The implementation and testing of the dual-resolution coupled hybrid system for RAP will be done
167 in the future.

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

175 Fig. 1 shows a flowchart for both one-way and two-way coupled EnSRF-En3DVar
176 analysis-forecast cycle as employed in this paper. For 1-way coupled En3DVar, the EnSRF system
177 provides the background ensemble forecast perturbations to the ECV-based En3DVar hybrid
178 variational analysis, but does not re-center the EnSRF analyses on the En3DVar analysis. Two-way
179 coupling includes an additional step that re-centers the EnSRF analysis ensemble on the En3DVar
180 control analysis (the thick black arrows and bold black box in Fig. 1). The two-way coupling
181 implicitly assumes that the En3DVar control analysis is better than the EnSRF ensemble mean
182 analysis, and the re-centering should help prevent the divergence between the EnSRF and En3DVar
183 analyses so that the ensemble perturbations can sample the control forecast uncertainty well.

184 Divergence between the two systems tends to become more serious when continuous cycles are run
185 for a long period of time.

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

189 $$\delta\mathbf{x} = \delta\mathbf{x}_1 + \sum_{k=1}^{K} (\mathbf{a}_k \circ \mathbf{x}_k^{'}) \quad , \tag{1}$$

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

195 $$\begin{aligned} J(\delta\mathbf{x}_1, \mathbf{a}) &= \beta_1 J_b + \beta_2 J_e + J_o \\ &= \frac{1}{2}\beta_1 \delta\mathbf{x}_1^T \mathbf{B}^{-1} \delta\mathbf{x}_1 + \frac{1}{2}\beta_2 \mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \frac{1}{2}[\mathbf{y}_o - H(\mathbf{x}_b + \delta\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}_o - H(\mathbf{x}_b + \delta\mathbf{x})], \end{aligned} \tag{2}$$

196 which gives the solutions of partial increment $\delta\mathbf{x}_1$ and ECV **a** . Vector **a** is formed by
197 concatenating *K* vectors $\mathbf{a}_k$. Compared to a traditional 3DVar cost function, a weighted sum of J_b
198 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
199 background term associated with static covariance **B**, J_o is the observation term as in traditional
200 3DVar. J_e is the additional term associated with flow-dependent covariance for the ECV.
201 Weighting factors β_1 and β_2 are placed in front of J_b and J_e terms, respectively, and they are
202 constrained by

203 $$\frac{1}{\beta_1} + \frac{1}{\beta_2} = 1, \tag{3}$$

204 to conserve the total variances in current GSI hybrid implementation.

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

215 Apart from the variational minimization of the En3DVar hybrid cost function given by Eq.
216 (2), a major component of the overall coupled EnSRF-En3DVar hybrid DA system is an ensemble
217 DA system that provides the perturbations. In our case, the EnSRF system used is the one described
218 in Z13. The configuration settings follow the control experiment of Z13. To facilitate fair
219 comparisons between the En3DVar hybrid and EnSRF experiments, the CLSs in the En3DVar
220 hybrid system are specified to match the CLSs used by the EnSRF as closely as possible in the
221 control experiments, and the vertical and horizontal scales are measured in the natural log of
222 pressure and kilometers, respectively. The *e-folding* distance from the Gaspari and Cohn (1999)
223 localization function is $\sqrt{2}\sqrt{0.3}S_{GC}$ (where S_{GC} is cut-off radii in the EnSRF), while *an e-folding*
224 distance from the recursive filter is $2\sqrt{2}S_{RF}$ (Barker et al. 2004; Wang et al. 2008b) (where S_{RF} is
225 recursive filter covariance localization length scale). Thus, to keep the same *e-folding* distance for

226 both EnSRF and En3DVar, the cut-off radii in the EnSRF S_{GC} can be converted to the recursive
227 filter localization length scale S_{RF} in hybrid according to

228 $$S_{RF} = \sqrt{0.15} S_{GC} / \sqrt{2} . \qquad (4)$$

229 **3. Experiment designs**

230 *a. Model, observations, ensemble configuration and verification techniques*

231 The test period, model domains and boundary conditions used in this study are the same as
232 in Z13. DA experiments at ~40 km grid spacing are run in continuous 3-hourly cycles throughout
233 the 9-day retrospective testing period from May 8 to 16, 2010; the cycles start at 0000 UTC 8 May
234 2010 and end at 2100 UTC 16 May 2010. The 40 km model domain (as shown in Fig. 2) covers
235 North America with 207x207 grid points. A slightly smaller domain at ~13 km grid spacing, as
236 indicated by the bold rectangle in Fig. 2a, is used for forecasts at the native RAP resolution and for
237 precipitation verification. The domains have 50 vertical levels. Eighteen-hour deterministic
238 forecasts (after applying DFL) are launched every three hours from the En3DVar hybrid control
239 analyses as well as EnSRF ensemble mean analyses on the 40 km domain. Three-hourly ensemble
240 forecasts are produced within the assimilation cycles of EnSRF, which are fed into the En3DVar
241 control analysis (Fig. 1). Two outer loops and 50 iterations, the same as in operational RAP GSI
242 3DVar, were utilized for all the En3DVar and GSI 3DVar experiments. The 13 km deterministic
243 forecasts start from interpolated 40 km analyses at 0000 and 1200 UTC for precipitation forecast
244 evaluation. The lateral boundary conditions for both grids come from operational GFS forecasts.
245 Perturbations created using the random-CV method in the WRF 3DVar (Barker et al. 2004) are
246 added to GFS forecast boundary conditions for the ensemble forecasts and to the GFS analysis
247 initial condition at 0000 UTC May 8, 2010 to start the initial ensemble of EnSRF.

248 The observations used in this study are the same as those used in the operational RAP except
249 for the exclusion of satellite radiance data. The realtime RAP system collects data from 1.5 h before
250 and 0.5 h after the time of analysis. However, for 0000 and 1200 UTC it waits half an hour longer
251 for more data (such as sounding data) to arrive. In our tests, the data sets assimilated at 3 hourly
252 intervals are the data sets collected and used by the operational hourly RAP system; as a result,
253 observations that arrived in realtime outside the 2 hours (2.5 hours for 0000 and 1200 UTC)
254 windows are not used. They include surface observations (land reporting stations, mesonets, ships,
255 and buoys, etc.), upper air observations (radiosondes, aircrafts, wind profilers, VAD data and
256 satellite retrieval winds) and GPS precipitable water (PW), the same as in Z13 except for the
257 exclusion of PW data there. The exclusion of the PW data in Z13 was due to an initial problem with
258 the EnSRF code, which has since been fixed. The distributions of most major observation types are
259 shown in Fig. 2. The satellite radiance data are not included in the experiments reported here
260 because our preliminary tests suggested that the bias correction remains an important issue within
261 the system that would require careful treatment for positive impacts. Our most recent tests with the
262 radiance data using the EnSRF show small positive impacts, and the results will be reported
263 separately in the future. Initial studies of EnKF for NCEP GFS global model also excluded satellite
264 radiance data (Whitaker et al. 2008).

265 The short-range deterministic forecasts from the En3DVar, EnSRF ensemble mean and GSI
266 3DVar analyses are verified against surface and sounding observations. The Model Evaluation
267 Tools (MET) developed by the Developmental Testbed Center (DTC) (Brown et al. 2009) are
268 employed here. MET contains comprehensive verification metrics for both deterministic and
269 probabilistic forecasts. Root-mean square error (RMSE) is used as the primary verification metric
270 for the 40 km deterministic forecasts here. The RMSEs for temperature (T), relative humidity (RH),
271 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 2014). 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 are statistically significant, the mean RMSE differences between En3DVar and GSI 3DVar together with a 90% confidence interval are 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 EnSRF_Ctl to determine whether the improvement of En3DVar over EnSRF 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 2014; 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 Z13 although they only presented the GSSs.

b. Assimilation experiments

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

295 1-way (Hybrid1W_Ctl) and 2-way coupled (Hybrid2W_Ctl), EnSRF (EnSRF_Ctl) control and GSI
296 3DVar (GSI3DVar) experiments are compared. The EnSRF control experiment, EnSRF_Ctl, uses
297 40 ensemble members and corresponds to experiment EnKF_CtrHDL from Z13 except for
298 additional GPS PW data in this study, and uses a single suite of physics parameterizations in the
299 ensemble to keep the setup simple (so that the EnSRF, GSI 3DVar and the En3DVar experiments
300 all use the same set of physics in the forecast model). The En3DVar hybrid control experiment
301 assigns equal weights ($1/\beta_2 = 0.5$) to the static and ensemble BECs. The EnSRF codes and
302 configurations are the same as the EnSRF control experiment in Z13, except for the exclusion of
303 GPS PW data there. A combination of static and adaptive covariance inflation is applied in EnSRF
304 as in Z13.

305 There are mainly two sets of tunable parameters in the En3DVar hybrid scheme. One set is
306 the covariance weighting factors, which define the weights placed on the BECs. Four sensitivity
307 experiments test the relative weights given to the static and ensemble BECs, with $1/\beta_2$=0.1, 0.5,
308 0.9, 1.0 (Hybrid01, Hybrid05, Hybrid09, Hybrid10) corresponding to 10, 50, 90, 100% weight
309 given to the ensemble BEC, respectively.

310 The other set of tunable parameters includes the horizontal and vertical CLSs applied to the
311 covariances. For weighting factor $1/\beta_2$=0.5 with 1-way coupling, we test three horizontal CLSs S_h
312 =192, 300 and 356 km in Hybrid_HS, Hybrid1W_Ctl, and Hybrid_HL, respectively (corresponding
313 to cut-off radii of 700, 1095, 1300 km according to Eq. (4)); three vertical CLSs S_v= -0.1, -0.3 and -
314 0.5 (corresponding to cut-off radii of 0.36, 1.1 and 1.8 according to Eq. (4)) are tested in
315 Hybrid_VS, Hybrid1W_Ctl and Hybrid_VL, respectively. The minus sign is due to the use of ln(p)
316 as the length measure. To facilitate the comparison with control experiment Hybrid1W_Ctl, the
317 mean domain-average RMSE difference, defined as

318 $$D = \frac{1}{N}\sum_{k=1}^{N}(RMSE^{k}{}_{Hybrid*} - RMSE^{k}{}_{Benchmark}) \ , \qquad (5)$$

319 where N is the total number of cycles and k refers to the k^{th} cycle, is calculated between experiment
320 $Hybrid*$; the benchmark experiment is Hybrid1W_Ctl here and $Hybrid*$ refers to one of
321 En3DVar sensitivity experiments.

322 All CLSs used in the En3DVar hybrid experiments described above are constant with
323 height. However, the cut-off radii used in the well-tuned EnSRF control experiment of Z13
324 (EnKF_CtrHDL in their paper) are height- and observation-type dependent based on the vertical
325 position of the observations. These localization settings are shown in Fig. 3. The horizontal cut-off
326 radius r_{cut} at the model top is 1.5 times the value at the surface for all state variables; as shown in
327 Fig. 3a, r_{cut} increases from 700 km at the surface to 1050 km at the model top. The vertical cut-off
328 radius $\ln(p_{cut})$ is not only height dependent, but also observation-type dependent. For *RH* and *T*
329 observations (solid line in Fig. 3b), the vertical cut-off radii at the model top and surface are set to a
330 quarter of 1.1 and half of 1.1, respectively. For wind observations (dash line in Fig. 3b), $\ln(p_{cut})$ is
331 twice as large as that for *RH* and *T* observations. For surface pressure observations and GPS PW
332 data (which are most strongly linked to low-level moisture), their vertical localization radii are set
333 to a constant value of 1.6. These settings were used in the control experiment of Z13, and their
334 choices were guided by the correlation scales found in the NMC-method-derived error statistics
335 used by GSI 3DVar and were further tuned based on sensitivity experiments.

336 In the En3DVar system, height-dependent localization is straightforward to implement, but
337 not observation-type-dependent localization, because unlike the serial EnSRF scheme, En3DVar
338 analyzes all observations simultaneously and the localization is performed in the state instead of the
339 observation space (Campbell et al. 2010). Theoretically, if the localization treatment was the same
340 for the EnSRF 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 EnSRF 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 potentially improve the performance of En3DVar as in EnSRF. These are examined in the next four experiments (EnSRF_Con, Hybrid_Con, Hybrid_HD and Hybrid3G), all performed with 100% ensemble covariance and all used one-way coupling.

EnSRF_Con and Hybrid_Con use constant horizontal and vertical localization radii which are close to the corresponding cut-off radii used by EnSRF_Ctl at the model top. Unlike the other En3DVar experiments, the ensemble perturbations for Hybrid_Con were provided by EnSRF_Con instead of EnSRF_Ctl for consistency.

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

The only way to apply different localization to different observations in En3DVar is to break the analysis into multiple steps, with each step analyzing a sub-set or a sub-group of observations. To do this, the corresponding EnSRF analysis that provides the ensemble perturbations also needs to be broken up into multiple steps and the EnSRF and En3DVar need to be run in alternating order. Doing so significantly increases the overall computational costs for operational implementation since the disk I/O associated with the reading and writing of ensembles and with the cost function minimizations are done multiple times per analysis cycle but is doable in a research mode. Towards this end, experiments EnSRF3G 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 EnSRF ensemble analysis is

365 followed by an En3DVar hybrid analysis step using the latest EnSRF-updated ensemble
366 perturbations.

367 Because the EnSRF includes both static and adaptive covariance inflation (Z13), it is
368 difficult to maintain the same amount and effects of inflation when each EnSRF analysis is broken
369 into three steps. Applying the static inflation every EnSRF sub-step can over-inflate the covariance,
370 while applying it only in the last step would change the overall behavior of the filter. Because our
371 primary goal here is to determine if the difference between the EnSRF and En3DVar analyses (with
372 100% ensemble covariance) is primarily caused by the observation-based localization, to avoid the
373 above issue, we run EnSRF3G without any covariance inflation and examine the RMSE differences
374 between the EnSRF and En3DVar analyses. We just need to find out if the En3DVar hybrid
375 analyses are closer to the EnSRF analyses when observation-type dependent localization is similarly
376 used in the En3DVar through the split-step procedure.

377 The mean domain average absolute RMSE difference, defined as

378 $$DB = \frac{1}{N}\sum_{k=1}^{N}\left|RMSE^{k}{}_{Hybrid*} - RMSE^{k}{}_{Benchmark}\right| , \qquad (6)$$

379 is used to measure how close the En3DVar and EnSRF analyses are. The differences between
380 Hybrid_Con and EnSRF_Con, Hybrid_Con and EnSRF_Ctl, Hybrid_HD and EnSRF_Ctl,
381 Hybrid3G and EnSRF3G (Table 2) will be calculated to examine the impacts of constant
382 localization, height-dependent localization, and observation-dependent localization, respectively.
383 The statistical significance of DB is also determined by using bootstrap resampling. The DBs at
384 cycles are randomly selected 3000 times. For this sample, a mean is calculated, along with a two-
385 tailed 90% confidence interval from 5% to 95%. If the error bars from the experiments pair do not
386 overlap, the differences between En3DVar and EnSRF are significantly reduced at the 90%
387 confidence level.

4. Results of experiments

a. GSI 3DVar, EnSRF 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), EnSRF (EnSRF_Ctl) control experiments, and those of the GSI 3DVar (GSI3DVar) experiment. Similar to Z13, single observation tests were first performed to examine the general behaviors of the En3DVar system and compare to the EnSRF results. The tests used the analysis of EnSRF_Ctl after 5 days of 3-hourly cycles as background and the analysis increments appear reasonable. To save space, the results are not shown here.

The RMSEs of the 3-hour forecasts at different height levels verified against sounding data are shown in Fig. 4. These forecasts were launched from the GSI 3DVar, EnSRF 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 EnSRF_Ctl are overall lower than those of GSI 3DVar 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 3DVar, and are also generally better than EnSRF_Ctl except for *RH* above 500 hPa, *V* at 100 hPa, and *T* below 900 hPa.

410 The average RMSEs for all levels over the entire domain are shown in Fig. 5 for forecast
411 hours 3 through 18. Generally, both EnSRF and En3DVar hybrid significantly outperform GSI
412 3DVar for all the variables throughout the forecast period at the 90% confidence level (the intervals
413 of error differences do not include zero). For *RH*, the average RMSEs of En3DVar hybrid are
414 slightly higher than those of EnSRF_Ctl by 9 hours, which appears to be related to the larger errors
415 at 3 hours at the upper levels (Fig. 4a). The RMSEs become slightly smaller after 9 hours.
416 However, the improvement of En3DVar hybrid over EnSRF for *RH* is not statistically significant.
417 For *T* and *U*, the domain-averaged RMSEs of En3DVar hybrid are significantly and consistently
418 smaller than those of GSI 3DVar and EnSRF throughout the forecast period (Fig. 5). For *V*, the
419 errors of the En3DVar and EnSRF are very similar and are all clearly lower than those of GSI
420 3DVar. The reason that En3DVar performs better than EnSRF for *U* may relate to the dominance of
421 the east-west flows that may increase the validity of the static covariance. Overall, the En3DVar
422 hybrid out-performs GSI 3DVar and EnSRF for *T* and *V* for the 18 hours of the forecast.
423 Fig. 6 shows the average RMSEs for 3-18 hour forecasts against surface observations. For 2
424 m *T* and 10 m *U*, the EnSRF and En3DVar outperform the GSI 3DVar at all forecast hours
425 significantly, with the EnSRF significantly outperforming the En3DVar hybrid at most forecast
426 hours. For 2 m *RH* and 10 m *V*, EnSRF occasionally underperforms GSI 3DVar slightly, but at most
427 forecast hours it is better. The En3DVar hybrid schemes improve over EnSRF further, enough to
428 ensure better or equal performance than GSI 3DVar for all hours, and more clearly so for *RH*. For
429 surface pressure, EnSRF underperforms GSI 3DVar initially, but becomes better after 9 hours;
430 throughout the forecast period, the En3DVar hybrid outperforms both GSI 3DVar and EnSRF
431 significantly. In general, there is little difference between the 1-way and 2-way En3DVar hybrid
432 schemes. If the cycles were run for a much longer time period, a larger divergence between the
433 EnSRF and En3DVar hybrid may develop in a 1-way coupling mode. In that case, 2-way coupling

434 would then show a bigger advantage. When the En3DVar hybrid runs at a higher resolution than the
435 EnSRF in a dual-resolution mode, there may also be more benefit from 2-way coupling.

436 Overall, the En3DVar hybrid schemes significantly outperform GSI 3DVar for all the
437 variables at all forecast hours for sounding and surface observations. Compared to EnSRF, their
438 performances are comparable, or even better for some variables. The results indicate the benefit of
439 combining the static and flow-dependent covariances. In the next section, the sensitivity to the
440 covariance weighting factors is examined.

441 Finally, one may have concern that the 9-day cycled assimilation period is not long enough
442 for the ensemble DA system to spin up (over the course of evaluating and testing our EnSRF and
443 En3DVar hybrid systems, we had run over 100 cycled experiments so extending the experiment
444 period would be expensive). To answer this question, we examine how the short-range forecast
445 errors evolve through the 9-day period. Fig. 7 shows the domain-averaged 3-hour forecast RMSEs
446 verified against sounding data at 0000 and 1200 UTC through the test period. We can see that the
447 relative performances of GSI3DVar, EnSRF and En3DVar hybrid do not change much throughout
448 the 9-day period, even in the earlier days. These results indicate that the ensemble system had spun-
449 up rather quickly.

450 *b. Sensitivity to covariance weighting factors*

451 We perform a set of four 1-way-coupled En3DVar hybrid experiments with $1/\beta_2$ =0.1, 0.5,
452 0.9 and 1.0, which are the weights given to the ensemble covariance. The one with $1/\beta_2$ =0.5,
453 called Hybrid05 here, is the same as experiment Hybrid1W_Ctl discussed earlier (Table 1). The 3
454 hour forecast RMSEs at different height levels verified against sounding data are shown in Fig. 8
455 for these four experiments, GSI 3DVar, and EnSRF_Ctl. It can be seen that the En3DVar hybrid
456 and EnSRF schemes generally outperform GSI 3DVar, except for *RH* from 700 to 400 levels for

457 Hybrid09 and Hybrid10, i.e., the hybrid scheme with 90 or 100% ensemble covariances.
458 Introducing 10% static covariance into the En3DVar hybrid framework reduces the error slightly
459 (comparing Hybrid09 to Hybrid10 for *RH*), and further increasing it to 50% brings the *RH* errors
460 below those of GSI 3DVar at all levels (Fig. 8a).

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

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

473 It can also be noticed from Fig. 8 that EnSRF_Ctl outperforms Hybrid10, except for wind
474 between 500 and 200 hPa levels. As pointed out earlier, if covariance localization treatments were
475 the same in EnSRF and Hybrid10, their results should be very close given that the ensemble
476 covariance is used at 100% in both cases. The use of height- and observation-type-dependent
477 covariance localization in the EnSRF but not in the En3DVar hybrid is suspected to be the main
478 cause of the differences. It had been found to help improve the EnSRF results in Z13 but is not used
479 within the En3DVar hybrid. In the following section, we examine whether doing something similar
480 within the En3DVar hybrid framework can improve the En3DVar hybrid results too.

c. Sensitivity to ensemble covariance localization

In Z13, several tests with the horizontal and vertical covariance localizations were performed. In this paper, the EnSRF experiment uses the same configuration as experiment EnKF_CtrlHDL of Z13, 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 impossible to use observation-type-dependent localization unless different observations are analyzed separately.

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-averaged forecast RMSE differences between 3 hour forecasts and those of Hybrid1W_Ctl are shown in Fig. 10. 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. 10). However, even though reduced CLSs are not preferred according to Fig. 10, 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 EnSRF (Z13). Doing so may also help further improve the En3DVar hybrid performance.

For a fair comparison with EnSRF, EnSRF_Con and Hybrid_Con, which use the same constant horizontal and vertical cut-off radii/length scales, are compared when ensemble covariance is used at 100%. As shown in Fig. 11, the significantly greater RMSEs of EnSRF_Con than EnSRF_Ctl suggest the height- and observation-type-dependent localization is a key reason for improving the performance of EnSRF. To see the effect of height-dependent localization

505 (observation-type-dependent localization is not possible in a single step analysis) on En3DVar, we
506 introduce it into Hybrid_HD which uses 100% ensemble covariance. As shown in Fig. 11,
507 Hybrid_HD outperforms Hybrid_Con and is much closer to EnSRF_Ctl for *RH*, *U* and V. For wind,
508 Hybrid_HD is even slightly better than EnSRF_Ctl at the middle levels (Fig. 11) while
509 Hybrid_Con is poorer than EnSRF_Ctl at all levels. For *RH*, EnSRF_Ctl still has smaller RMSEs
510 than hybrid_HD above 700 hPa. The greater RMSEs from EnSRF_Con suggest that the smaller
511 localization radii used in EnSRF_Ctl at the higher levels are beneficial.

512 The cutoff radii used in the EnSRF_Ctl are also observation-type dependent. As discussed
513 earlier, since the En3DVar algorithm analyzes all observations simultaneously by variational
514 minimization in the state space, making it impossible to apply observation-type-dependent
515 localization within a single analysis step. Experiments EnSRF3G and Hybrid3G break each analysis
516 cycle into three sub-steps of coupled EnSRF-En3DVar analyses, with each step analyzing a sub-
517 group of observations that share the same height-dependent localization scales. Here, we use the
518 absolute RMSE differences between pairs of En3DVar and EnSRF experiments (Table 2) together
519 with the 90% confidence interval as determined by the bootstrap resampling procedure to determine
520 the statistical significance of the differences. When the error bars from different experiment pairs do
521 not overlap, the RMSE differences between En3DVar and EnSRF are considered statistically
522 significant. As shown in Fig. 12, the RMSE differences between Hybrid_HD and EnSRF_Ctl
523 (labeled 1GHD) are reduced, by about 1/4 to 1/3 for *RH*, *U* and V, compared to the differences
524 between Hybrid_Con and EnSRF_Ctl (labeled 1GC1). The reduction in the difference is smaller but
525 still statistically significant for *T*. When constant localization is used in both EnSRF_Con and
526 Hybrid_Con, the differences (labeled 1GC2) are also reduced compared to 1GC1 case, although the
527 differences are slightly larger than the 1GHD case.

When observation-type-dependent localization is used in Hybrid3G and EnSRF3G, the differences between them (3DHD) are even smaller for *RH*, *U* and *V*. For *RH* (which has the largest RMSE difference between En3DVar and EnSRF according to Fig. 11a), the RMSE difference of *RH* is about 0.5% versus the 1.25% for the constant localization case. The reductions for *T*, *U* and *V* are smaller but still evident. Clearly, the differences between En3DVar and EnSRF are much smaller when height- and observation-type-dependent localization is used in both algorithms, especially for humidity.

Fig. 13 shows the RMSE differences together with the 90% confidence interval at different height levels. In reference to Fig. 11, those levels where domain average absolute RMSE differences between Hybrid_HD and EnSRF_Ctl (1GHD) are greater than those between Hybrid_Con and EnSRF_Ctl (1GC1) correspond to the levels where Hybrid_HV outperforms Hybrid_Con, given that EnSRF_Ctl is generally the best among the three experiments. The RMSE differences of 1GC2 are also smaller than 1GC1, but the constant localization degrades the performance of EnSRF and En3DVar for almost all the variables and at all levels in reference to Fig. 11. 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 1GC1. These results show that when similar height- and observation-type-dependent covariance localization is used in the En3DVar framework using 100% ensemble covariance, differences between EnSRF and En3DVar are significantly reduced, and such localization treatment generally brings the En3DVar results closer to the better EnSRF results. The reduction in the RMSE differences for *RH* is greater than those for *T*, *U* and *V*. Because the humidity field tends to contain smaller scale structures than other fields, it appears to benefit from tighter localization more when using height- and observation-type-dependent localization. However, because there are still differences between the EnSRF and

552 En3DVar algorithms, some differences still exist between their results, as indicated by the green
553 bars in Fig. 12. When the ensemble covariance is used at 50%, height-dependent localization did
554 not improve the En3DVar hybrid results as much as in the 100% case (not shown).

555 In summary, the use of height-dependent localization in the En3DVar hybrid framework
556 when using full ensemble covariance improves the resulting model forecasts at almost all levels and
557 forecast hours. Height- and observation-type-dependent localizations used in EnSRF are responsible
558 for about half of the differences between the EnSRF and the En3DVar with full ensemble
559 covariance. Unfortunately, observation-type-dependent localization is difficult or expensive to
560 implement with the En3DVar algorithm.

561 **5. Precipitation forecast skills on 13-km grid**

562 In this section, precipitation forecasts on the 13 km grid initialized from the 40 km GSI
563 3DVar, EnSRF_Ctl ensemble mean, Hybrid1WCtl and Hybrid2WCtl analyses (Table 3) are
564 compared. Considering extensive CPU and storage requirements, we launched the forecasts only
565 twice a day at 00 and 12 UTC. The precipitation forecasts are verified against the NCEP Stage IV
566 precipitation data. GSSs calculated for the 0.1, 1.25 and 2.5 mm h^{-1} thresholds are calculated as in
567 Z13.

568 The GSSs and BIASs for the forecasts are shown in Fig. 14. Both EnSRF and En3DVar
569 hybrid outperform GSI 3DVar on average for all forecast hours and thresholds shown. EnSRF13
570 has higher GSSs than Hybrid for 0.1 mm h^{-1} after 3 hours. For greater thresholds of 1.25 and 2.5
571 mm h^{-1}, forecasts of Hybrid1W13 are comparable to EnSRF13 by 7 hours, and are better than
572 Hybrid2W13 during the first four hours, which is consistent with the domain-averaged RMSEs of
573 *RH* shown in Fig. 5a. Fig. 14 show that EnSRF generally has the highest positive BIASs. The
574 hybrid schemes have the lowest BIASs in the first 5 hours, and values between those of GSI 3DVar

575 and EnSRF after 5 hours. The BIAS differences are relatively small and BIAS is always positive for
576 both thresholds examined.

577 From sensitivity experiments, we learned that the constant localization used in En3DVar
578 hybrid is one of the reasons for the deterioration of humidity forecasts compared to EnSRF. To help
579 further understand the impact of localization on precipitation forecast skill, the forecasts on the 13
580 km grid initialized from EnSRF_Ctl and EnSRF_Con ensemble mean, and Hybrid_Con analyses are
581 also compared. Without height- and observation-type-dependent localization in EnSRF13Con, its
582 GSSs are lower than those of EnSRF13 in the first 9 hours and are close to those of
583 Hybrid1W13Con from 4 to 7 hours, for the 0.1 mm h^{-1} threshold (Fig. 15), indicating the role of
584 height- and observation-type-dependent localization. The differences among the experiments are
585 smaller for larger thresholds (not shown).

586 **6. Summary and discussions**

587 In this paper, a coupled EnSRF-En3DVar hybrid data assimilation system based on the
588 NCEP operational GSI variational framework is established and tested for the Rapid Refresh (RAP)
589 forecasting system. It uses a recently developed, well-tuned, 40-member EnSRF system, as
590 documented in Z13, to update and provide the ensemble perturbations. A 9-day spring period
591 starting from May 8, 2010 that contains active convection is used to examine the performance of the
592 system through comparisons with parallel experiments using EnSRF and GSI 3DVar. The En3DVar
593 hybrid, EnSRF and GSI 3DVar experiments use the same observational data sets as the operational
594 RAP system except for the exclusion of satellite radiance data. The experiments are performed at a
595 reduced resolution of ~40 km grid spacing with 3-hourly assimilation cycles rather than at the
596 native 13 km grid spacing with hourly cycles of the operational RAP. The systems are evaluated
597 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 and covariance localization configurations are examined through sensitivity experiments.

With equal weighting for the ensemble and static covariances, the En3DVar hybrid scheme outperforms GSI 3DVar for all standard variables at all levels with statistical significance, and is slightly better than EnSRF, especially for later forecast hours, except for moisture during the earlier forecast hours. Apparently, 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 covariance localization scales are worse than those from pure EnSRF 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 EnSRF, especially at the high and low levels. Using similar height-dependent localization, En3DVar with 100% ensemble covariance became much closer to pure EnSRF. When using similar observation-type-dependent covariance localization in En3DVar, by running the coupled EnSRF-En3DVar analyses in three steps with each analyzing a subset of observation variables (in a similar way as in EnSRF), the results of En3DVar, with 100% ensemble covariance, become even closer to those of EnSRF. The benefit of height- and observation-type-dependent localization is negligible when the ensemble covariance is used at 50%. The multi-step EnSRF-En3DVar analysis procedure is, unfortunately, not very practical due to much increased computational costs. It is straightforward for pure EnSRF because the algorithm is serial, where observations are assimilated sequentially.

622 Previous studies (e.g., Hamill and Snyder 2000; Wang et al. 2008b) had found that the
623 En3DVar hybrid system is more robust than EnSRF when the ensemble size is small or model error
624 is large. In our study, the EnSRF, En3DVar hybrid 1-way and 2-way using 20 instead of 40
625 ensemble members for control experiments were also run and compared. In such a case, EnSRF and
626 En3DVar are both degraded, but the En3DVar hybrid is now consistently better than EnSRF for all
627 variables and all forecast hours (not shown). That indicates a larger benefit of the static covariance
628 when the ensemble is small, agreeing with earlier findings. While further specific tuning of the 20-
629 member EnSRF may improve the results somewhat, we do not expect the general conclusion to
630 change.

631 Deterministic forecasts were launched on a 13 km grid from interpolated 40-km En3DVar
632 hybrid control, EnSRF ensemble mean and GSI 3DVar analyses at 0000 and 1200 UTC of each
633 day. Hourly accumulated precipitation is better predicted in the En3DVar hybrid and EnSRF
634 experiments than GSI 3DVar, but for light precipitation, En3DVar hybrid does not perform as well
635 as EnSRF, which is consistent with the slightly worse humidity forecasts of En3DVar. When
636 constant covariance localization is used in EnSRF, its precipitation forecast skills become closer to
637 those of En3DVar using 100% ensemble covariance, indicating the covariance localization
638 difference between the EnSRF and En3DVar as a key cause for the performance differences.

639 Despite the encouraging results, the En3DVar hybrid system we have established for RAP
640 still has much room for further improvement. Adding satellite and radar data and examining their
641 impacts are among the desired tasks, as is a dual-resolution implementation where the En3DVar is
642 run at the native RAP resolution. These aspects are being pursued and the results will be reported in
643 future papers. A further refined and tested version, initially in a dual-resolution mode coupling a 13
644 km En3DVAR with a 40 km EnSRF, will likely become operational in the future, replacing the
645 GFS EnSRF system used in the recently implemented operational RAP hybrid DA system.

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.

652 **References**

653 Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*,
654 **129,** 2884-2903.

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

657 Anderson, J. L., 2012: Localization and sampling error correction in ensemble Kalman filter data
658 assimilation. *Mon. Wea. Rev.*, **140,** 2359-2371.

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

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

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

669 Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble
670 transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129,** 420-436.

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

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

676 Brown, B. G., J. H. Gotway, R. Bullock, E. Gilleland, and D. Ahijevych, 2009: The Model
677 Evaluation Tools (MET): Community tools for forecast evaluation. *25th Conf. Int.*
678 *Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and*
679 *Hydrology*, Paper 9A.6.
680 Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error
681 covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*,
682 **131,** 1013-1043.
683 Buehner, M. and A. Mahidjiba, 2010: Sensitivity of Global Ensemble Forecasts to the Initial
684 Ensemble Mean and Perturbations: Comparison of EnKF, Singular Vector, and 4D-Var
685 Approaches. *Mon. Wea. Rev.*, **138,** 3886-3904.
686 Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of
687 variational data assimilation and the ensemble Kalman filter for global deterministic NWP.
688 Part I: Description and single-observation experiments. *Mon. Wea. Rev.*, **138,** 1550-1566.
689 Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010b: Intercomparison of
690 variational data assimilation and the ensemble Kalman filter for global deterministic NWP.
691 Part II: One-month experiments with real observations. *Mon. Wea. Rev.*, **138,** 1567-1586.
692 Burgers, G., P. J. v. Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman
693 filter. *Mon. Wea. Rev.*, **126,** 1719-1724.
694 Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite
695 radiances in ensemble Kalman filters. *Mon. Wea. Rev.*, **138,** 282-290.
696 Candille, G., C. Côté, P. L. Houtekamer, and G. Pellerin, 2007: Verification of an Ensemble
697 Prediction System against Observations. *Mon. Wea. Rev.*, **135,** 2688-2699.

698 Courtier, P., E. Andersson, E. Keckley, J. Pailleux, D. Vasiljevic, M. Hamurd, A. Hollingsworth, F.
699 Rabier, and M. Fisher, 1998: The ECMWF implementation of three-dimensional variational
700 assimilation (3D-Var). I: formulation. *Quart. J. Roy. Met. Soc.*, **124,** 1783-1808.
701 Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using
702 Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99,** 10143-10162.
703 Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical
704 implementation. *Ocean Dynamics*, **53,** 343-367.
705 Fujita, T., D. J. Stensrud, and D. C. Dowell, 2007: Surface data assimilation using an ensemble
706 Kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea.*
707 *Rev.*, **135,** 1846-1868.
708 Gandin, L. S. and A. H. Murphy, 1992: Equitable skill scores for categorical forecasts. *Mon. Wea.*
709 *Rev.*, **120,** 361-370.
710 Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational
711 data analysis method with recursive filter for Doppler radars. *J. Atmos. Ocean. Technol.*, **21,**
712 457-469.
713 Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three
714 dimensions. *Quart. J. Roy. Meteor. Soc.*, **125,** 723-757.
715 Ge, G., J. Gao, M. Xue, and K. K. Droegemeier, 2012: Diagnostic pressure equation as a weak
716 constraint in a storm-scale three dimensional variational radar data assimilation system. *J.*
717 *Atmos. Ocean. Tech.*, **29,** 1075-1092.
718 Hamill, T. M. and C. Snyder, 2000: A hybrid ensemble Kalman filter - 3D variational analysis
719 scheme. *Mon. Wea. Rev.*, **128,** 2905-2919.
720 Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background
721 error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129,** 2776-2790.

722 Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011a: Global ensemble predictions
723 of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.*, **139,**
724 668-688.
725 Hamill, T. M., J. S. Whitaker, D. T. Kleist, M. Fiorino, and S. G. Benjamin, 2011b: Predictions of
726 2010's tropical cyclones using the GFS and ensemble-based data assimilation methods.
727 *Mon. Wea. Rev.*, **139,** 3243-3247.
728 Hayden, C. M. and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields:
729 Applications to NESDIS operational processing. *J. Appl. Meteor.*, **34,** 3-15.
730 Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter
731 technique. *Mon. Wea. Rev.*, **126,** 796-811.
732 Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen,
733 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real
734 observations. *Mon. Wea. Rev.*, **133,** 604-620.
735 Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal
736 chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, **230,**
737 112-126.
738 Jung, Y., M. Xue, and M. Tong, 2012: Ensemble Kalman filter analyses of the 29-30 May 2004
739 Oklahoma tornadic thunderstorm using one- and two-moment bulk microphysics schemes,
740 with verification against polarimetric data. *Mon. Wea. Rev.*, **140,** 1457-1475.
741 Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009a: Introduction of
742 the GSI into the NCEP global data assimilation system. *Wea. Forecasting*, **24,** 1691-1705.
743 Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, R. M. Errico, and R. Yang, 2009b: Improving
744 incremental balance in the GSI 3DVAR analysis system. *Mon. Wea. Rev.*, **137,** 1046-1060.

745 Le Dimet, F. X. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of
746 meteorological observations: Theoretical aspects. *Tellus*, **38A,** 97-110.
747 Li, Y., X. Wang, and M. Xue, 2012: Assimilation of radar radial velocity data with the WRF
748 ensemble-3DVAR hybrid system for the prediction of hurricane Ike (2008). *Mon. Wea. Rev.*,
749 **140,** 3507-3524.
750 Lin, Y. and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses:
751 Development and applications. *19th Conf. Hydrology*, San Diego, CA, Amer. Meteor.Soc.,
752 Paper 1.2.
753 Liu, C. and Q. Xiao, 2013: An Ensemble-Based Four-Dimensional Variational Data Assimilation
754 Scheme. Part III: Antarctic Applications with Advanced Research WRF Using Real Data.
755 *Mon. Wea. Rev.*, **141,** 2721-2739.
756 Liu, C., Q. Xiao, and B. Wang, 2008: An ensemble-based four-dimensional variational data
757 assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.*,
758 **136,** 3363-3373.
759 Lorenc, A., 2003: The potential of the ensemble Kalman filter for NWP - a comparison with 4D-
760 Var. *Quart. J. Roy. Meteor. Soc.*, **129,** 3183-3204.
761 Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor.*
762 *Soc.*, **112,** 1177-1194.
763 Lynch, P., 1997: The Dolph-Chebyshev window: A simple optimal filter. *Mon. Wea. Rev.*, **125,**
764 655-660.
765 Lynch, P. and X.-Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon.*
766 *Wea. Rev.*, **120,** 1019-1034.
767 Lynch, P. and X.-Y. Huang, 1994: Diabatic initialization using recursive filters. *Tellus*, **46A,** 583-
768 597.

769 Meng, Z. and F. Zhang, 2007: Tests of an ensemble Kalman filter for mesoscale and regional-scale
770 data assimilation. Part II: Imperfect model experiments *Mon. Wea. Rev.*, **135,** 1403–1423
771 Mizzi, A. P., 2012: Comparison of a GSI/ETKF Regional Hybrid Using WRF/ARW with
772 GSI/LETKF and GSI/EnKF Regional Hybrids. *16th Conference on Integrated Observing*
773 *and Assimilation Systems for Atmosphere, Oceans, and Land Surface*, New Orleans, LA,
774 Amer. Met. Soc.
775 Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-
776 interpolation analysis system. *Mon. Wea. Rev.*, **120,** 1747-1763.
777 Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003a: Numerical aspects of the
778 application of recursive filters to variational statistical analysis. Part II: Spatially
779 inhomogeneous and anisotropic general covariances. *Mon. Wea. Rev.*, **131,** 1536-1548.
780 Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003b: Numerical aspects of the
781 application of recursive filters to variational statistical analysis. Part I: Spatially
782 homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131,** 1524-1535.
783 Putnam, B. J., M. Xue, Y. Jung, N. A. Snook, and G. Zhang, 2014: The analysis and prediction of
784 microphysical states and polarimetric variables in a mesoscale convective system using
785 double-moment microphysics, multi-network radar data, and the ensemble Kalman filter.
786 *Mon. Wea. Rev.*, **142,** 141-162.
787 Rabier, F., H. Jarvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF
788 operational implementation of four-dimensional variational assimilation. I: Experimental
789 results with simplified physics. *Quart. J. Roy. Met. Soc.*, **126,** 1143-1170.
790 Raynaud, L., L. Berre, and G. Desroziers, 2011: An extended specification of flow-dependent
791 background error variances in the Meteo-France global 4D-Var system. *Quart. J. Roy.*
792 *Meteor. Soc.*, **137,** 607-619.

793 Schwartz, C. S. and Z. Liu, 2014: Convection-permitting forecasts initialized with continuously
794 cycling limited-area 3DVAR, ensemble Kalman filter, and “hybrid” variational–ensemble
795 aata assimilation systems. *Mon. Wea. Rev.*, **142,** 716–738.
796 Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. D. Powers,
797 2005: A Description of the Advanced Research WRF Version 2, 88 pp.
798 Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with
799 the adjoint vorticity equation. Part I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113,** 1311-1328.
800 Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble
801 square root filters. *Mon. Wea. Rev.*, **131,** 1485-1490.
802 Tong, M. and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a
803 compressible nonhydrostatic model: OSS Experiments. *Mon. Wea. Rev.*, **133,** 1789-1807.
804 Wang, X., 2010: Incorporating ensemble covariance in the Gridpoint Statistical Interpolation
805 variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138,** 2990-2995.
806 Wang, X., 2011: Application of the WRF hybrid ETKF–3DVAR data assimilation system for
807 hurricane track forecasts. *Wea. Forecasting*, **26,** 868-884.
808 Wang, X., C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently
809 proposed ensemble/VAR hybrid analysis schemes. *Mon. Wea. Rev.*, **135,** 222–227.
810 Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A hybrid ETKF-3DVAR data
811 assimilation scheme for the WRF model. Part I: Observing system simulation experiment.
812 *Mon. Wea. Rev.*, **136,** 5116-5131.
813 Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008b: A hybrid ETKF-3DVAR data
814 assimilation scheme for the WRF model. Part II: Real observation experiment. *Mon. Wea.*
815 *Rev.*, **136,** 5132-5147.

816 Wang, X., T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2009: A Comparison of the Hybrid and
817 EnSRF Analysis Schemes in the Presence of Model Errors due to Unresolved Scales. *Mon.*
818 *Wea. Rev.*, **137,** 3219-3232.
819 Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-Based Ensemble–Variational
820 Hybrid Data Assimilation for NCEP Global Forecast System: Single-Resolution
821 Experiments. *Mon. Wea. Rev.*, **141,** 4098-4117.
822 Whitaker, J., D. T. Kleist, X. Wang, and T. Hamill, 2011: Tests of a hybrid variational-ensemble
823 global data assimilation system for hurricane prediction. *24th Conf. Wea. Forecasting/20th*
824 *Conf. Num. Wea. Pred.*, Amer. Meteor. Soc., Paper J16.2.
825 Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations.
826 *Mon. Wea. Rev.*, **130,** 1913-1924.
827 Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with
828 the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136,** 463-482.
829 Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with
830 spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130,** 2905-2916.
831 Xue, M., Y. Jung, and G. Zhang, 2010: State estimation of convective storms with a two-moment
832 microphysics scheme and an ensemble Kalman filter: Experiments with simulated radar data
833 *Q. J. Roy. Meteor. Soc*, **136,** 685-700.
834 Xue, M., J. Schleif, F. Kong, K. K. Thomas, Y. Wang, and K. Zhu, 2013: Track and intensity
835 forecasting of Hurricanes: Impact of cloud-resolving resolution and ensemble Kalman filter
836 data assimilation on 2010 Atlantic season forecasts. *Wea. Forecasting*, **28,** 1366-1384.
837 Zhang, F., M. Zhang, and J. Poterjoy, 2013: E3DVar: Coupling an ensemble Kalman filter with
838 three-dimensional variational data assimilation in a limited-area weather prediction model
839 and comparison to E4DVar. *Mon. Wea. Rev.*, **141,** 900-917.

840 Zhang, M. and F. Zhang, 2011: E4DVar: Coupling an ensemble Kalman filter with four-
841 dimensional variational data assimilation in a limited-area weather prediction model. *Mon.*
842 *Wea. Rev.*, **140,** 587-600.
843 Zhu, K., Y. Pan, M. Xue, X. Wang, J. S. Whitaker, S. G. Benjamin, S. S. Weygandt, and M. Hu,
844 2013: A Regional GSI-Based Ensemble Kalman Filter Data Assimilation System for the
845 Rapid Refresh Configuration: Testing at Reduced Resolution. *Mon. Wea. Rev.*, **141,** 4118-
846 4139.
847

List of figures

Fig. 1. Flowchart of a full GSI-based EnSRF-En3DVar hybrid data assimilation cycle, with one-way or two-way coupling between the EnSRF (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 EnSRF in the two-way coupling procedure, when the En3DVar hybrid control analysis is used to replace the ensemble mean of the EnSRF 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 Z13.

Fig. 3. Profiles of (a) horizontal and (b) vertical cut-off radii for the EnSRF 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. Mean 3-hour forecast RMSEs at different height levels 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. 5. The 9-day and domain-averaged forecast RMSEs verified against sounding data (upper panel in each frame) and the 90% confidence interval of the RMSE differences between En3DVar hybrid experiments *and* GSI 3DVar/EnSRF_Ctl (lower panel) for (a) *RH,* (b) *T*, (c) *U, and* (d) V, for different forecast hours.

871 Fig. 6. The 9-day and domain-averaged forecast RMSEs verified against surface observations
872 (upper panel in each frame) and the 90% confidence interval of the RMSE differences
873 between En3DVar hybrid experiments and GSI 3DVar/EnSRF_Ctl (lower panel) for (a)
874 surface pressure, (b) 2-m *RH*, (c) 2-m temperature, (d) 10-m *U*, and (e) 10-m *V* for different
875 forecast hours. The horizontal axis is forecast hour. The error bars in domain-averaged
876 forecast RMSEs represent the two-tailed 90% confidence interval.
877 Fig. 7. Domain-averaged 3-hour forecast RMSEs (upper panels in each frame) verified against
878 sounding data at 0000 and 1200 UTC through test period and the 90% confidence interval of
879 RMSE differences (lower panel of each frame) between the En3DVar hybrid and EnSRF
880 experiments and GSI 3DVar for *(*a) RH, *(*b) T, *(*c) U, and *(*d*)* V.
881 Fig. 8. Mean 3-hour forecast RMSEs at different height levels verified against sounding data for (a*)*
882 *RH*, (b) T, (*c) U*, and (d) V for experiments GSI 3DVar, Hybrid01, Hybrid05, Hybrid07,
883 Hybrid10 and EnSRF_Ctl.
884 Fig. 9. The same as Fig. 5, except for experiments GSI 3DVar, Hybrid01, Hybrid05, Hybrid07,
885 Hybrid10 and EnSRF_Ctl.
886 Fig. 10. Mean absolute forecast RMSE differences between different experiments and
887 Hybrid1W_Ctl, verified against sounding data, for 3-hour forecast averaged over the 9-day
888 forecast period over the entire model domain.
889 Fig. 11. The same as Fig. 4 but for experiments EnSRF_Ctl, EnSRF_Con, Hybrid_HD and
890 Hybrid_Con.
891 Fig. 12. Nine-day and domain-averaged absolute 3-hour forecast RMSE differences verified
892 against sounding data, where 1GC1 means difference between Hybrid_Con and EnSRF_Ctl,
893 1GC2 means difference between Hybrid_Con and EnSRF_Con, 1GHD means difference

894 between Hybrid_HD and EnSRF_Ctl, and 3GHD means difference between Hybrid3G and
895 EnSRF3G. The error bars represent the two-tailed 90% confidence interval.
896 Fig. 13. Nine-day and domain-averaged absolute RMSE differences between Hybrid_Con and
897 EnSRF_Ctl (labeled 1GC1), Hybrid_Con and EnSRF_Con (labeled 1GC2), Hybrid_HD and
898 EnSRF_Ctl (labeled 1GHD), Hybrid3G and EnSRF3G (labeled 3GHD) for (a) *RH, (*b) T, *(*c)
899 U, and (d) V at different height levels. The error bars represent the two-tailed 90% confidence
900 interval.
901 Fig. 14. Average precipitation GSSs and BIASs of 13-km forecasts as a function of forecast length
902 for thresholds (a) (b) 0.1 mm h-1, (c) (d) 1.25 mm h-1, and (e) (f) 2.5 mm h-1 for control
903 experiments.
904 Fig. 15. The same as Fig. 14, but for EnKF13, EnKF13Con and Hybrid13Con.
905

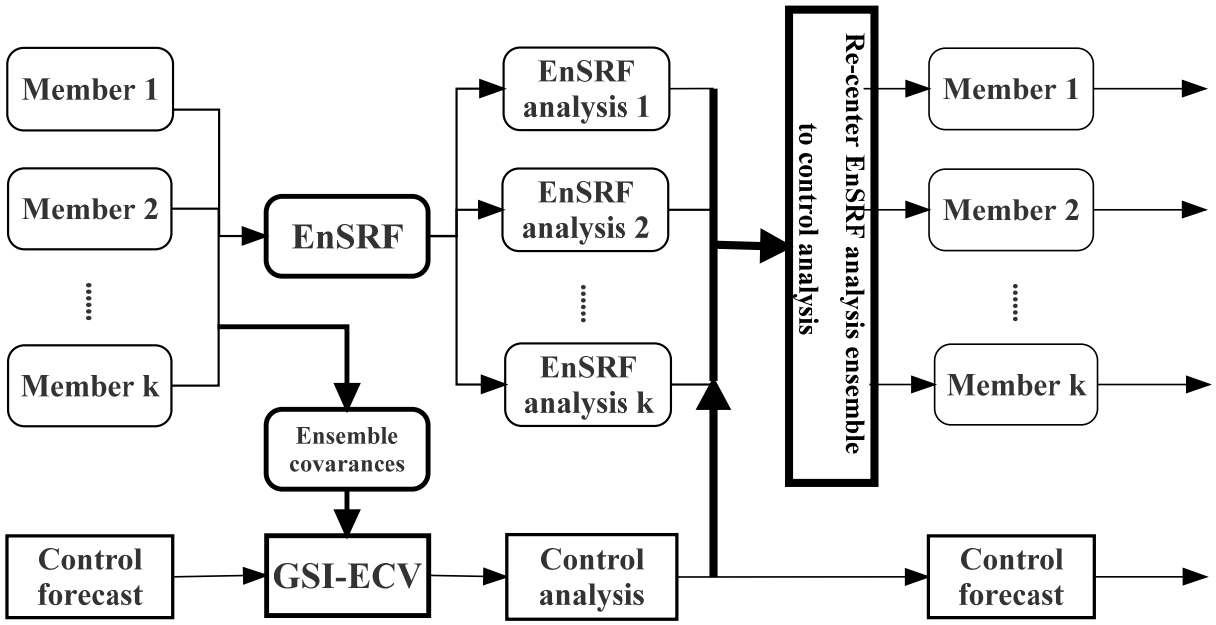
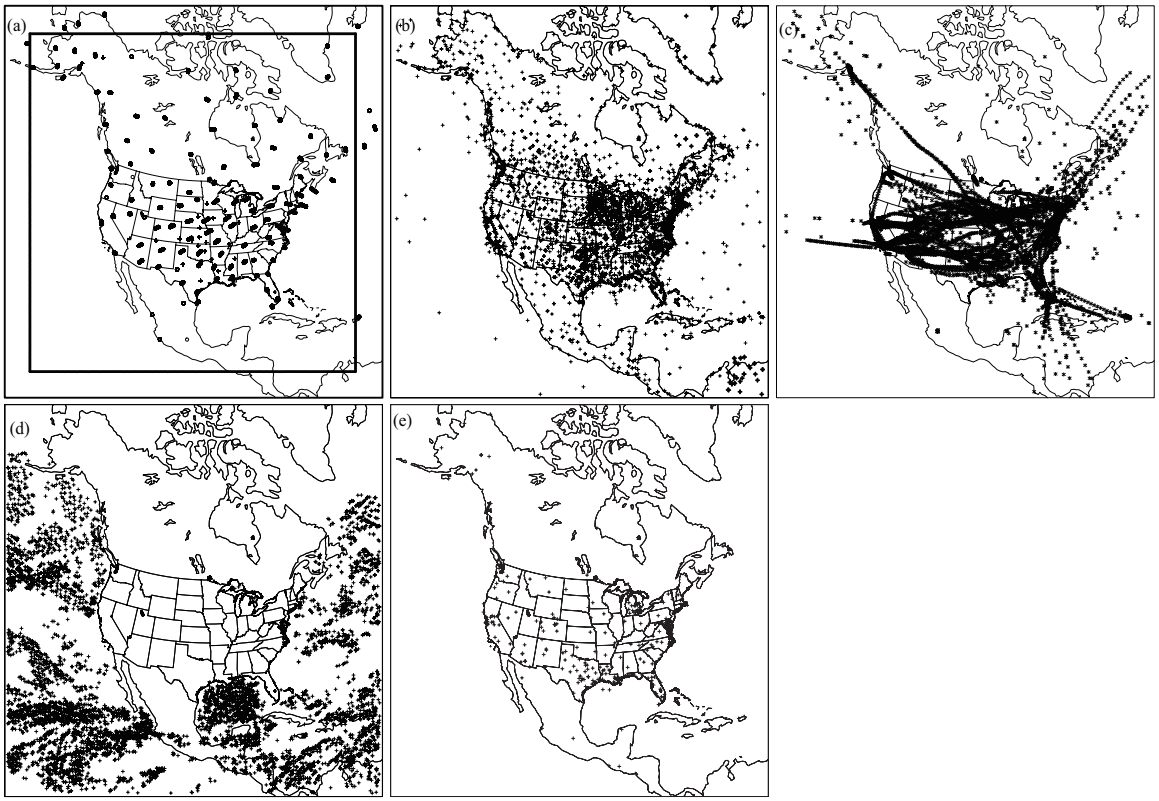


Fig. 1. Flowchart of a full GSI-based EnSRF-En3DVar hybrid data assimilation cycle, with one-way or two-way coupling between the EnSRF (upper portion) and En3DVar hybrid control analysis (lower portion denoted En3DVar). The thick upward pointing arrow indicates the feedback of the En3DVar hybrid analysis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

913



914

915

916 Fig. 2. Example of the horizontal distributions of observation at 0000 UTC May 14: (a) sounding
917 (circles) and profile (pluses), (b) surface stations over land and for ships, buoys, etc., (c) aircraft
918 observations, (d) satellite retrieval winds, and (e) GPS precipitable water (PW) data. The small box
919 in (a) is the domain used by 13 km forecasts. (a)-(d) are adopted from Z13.

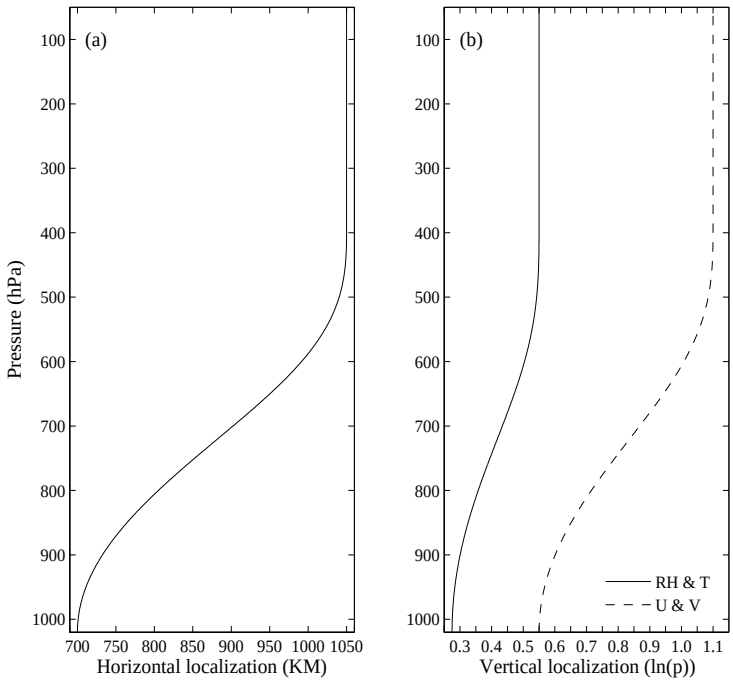
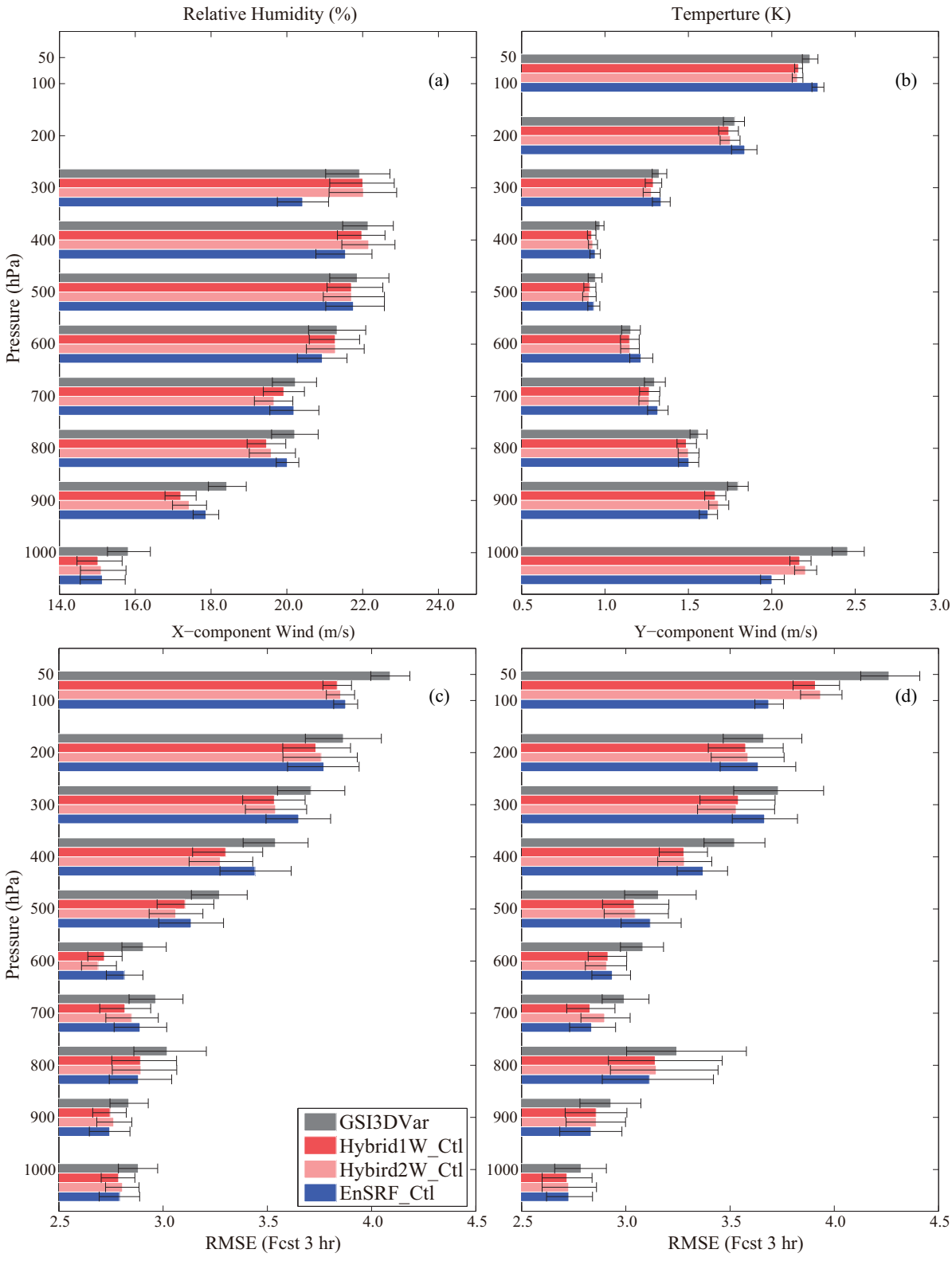


Fig. 3. Profiles of (a) horizontal and (b) vertical cut-off radii for the EnSRF control experiment. The horizontal axis is the cut-off radius of a given observation at a particular vertical position given in pressure. The vertical axis is the vertical position of observations given by the pressure.



925

926 Fig. 4. Mean 3-hour forecast RMSEs at different height levels verified against sounding data for (a)
927 *RH,* (b) T, (c) *U,* and (d) V for the labeled experiments. Error bars represent the two-tailed 90%
928 confidence interval (5% on the left and 95% on the right) using the bootstrap distribution method.

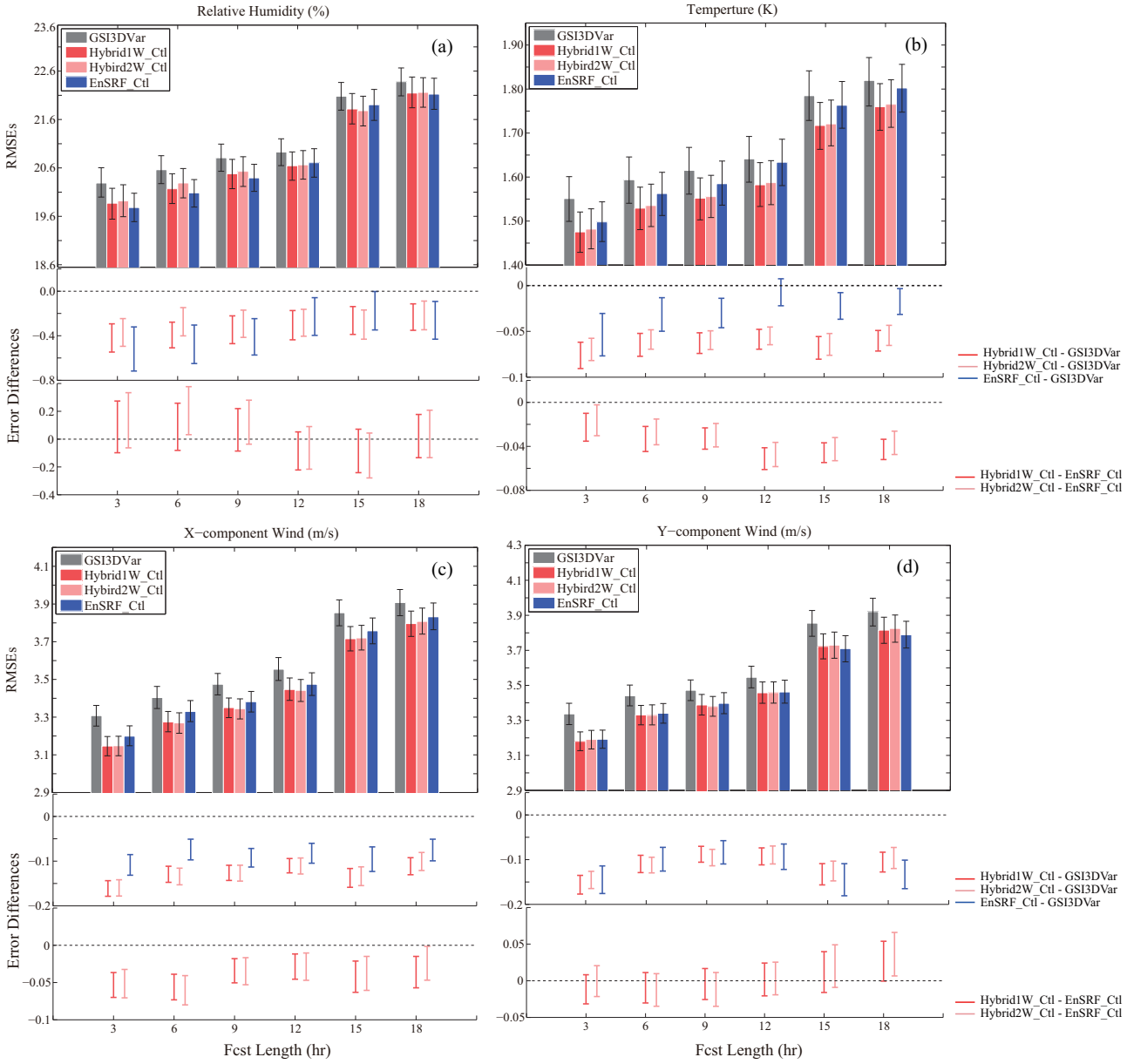


Fig. 5. The 9-day and domain-averaged forecast RMSEs verified against sounding data (upper panel in each frame) and the 90% confidence interval of the RMSE differences between En3DVar hybrid experiments and GSI 3DVar/EnSRF_Ctl (lower panel) for (a) *RH,* (b) *T*, (c) *U, a*nd (d) V, for different forecast hours. Error bars represent the two-tailed 90% confidence interval (5% at the bottom and 95% on the top) using the bootstrap distribution method.

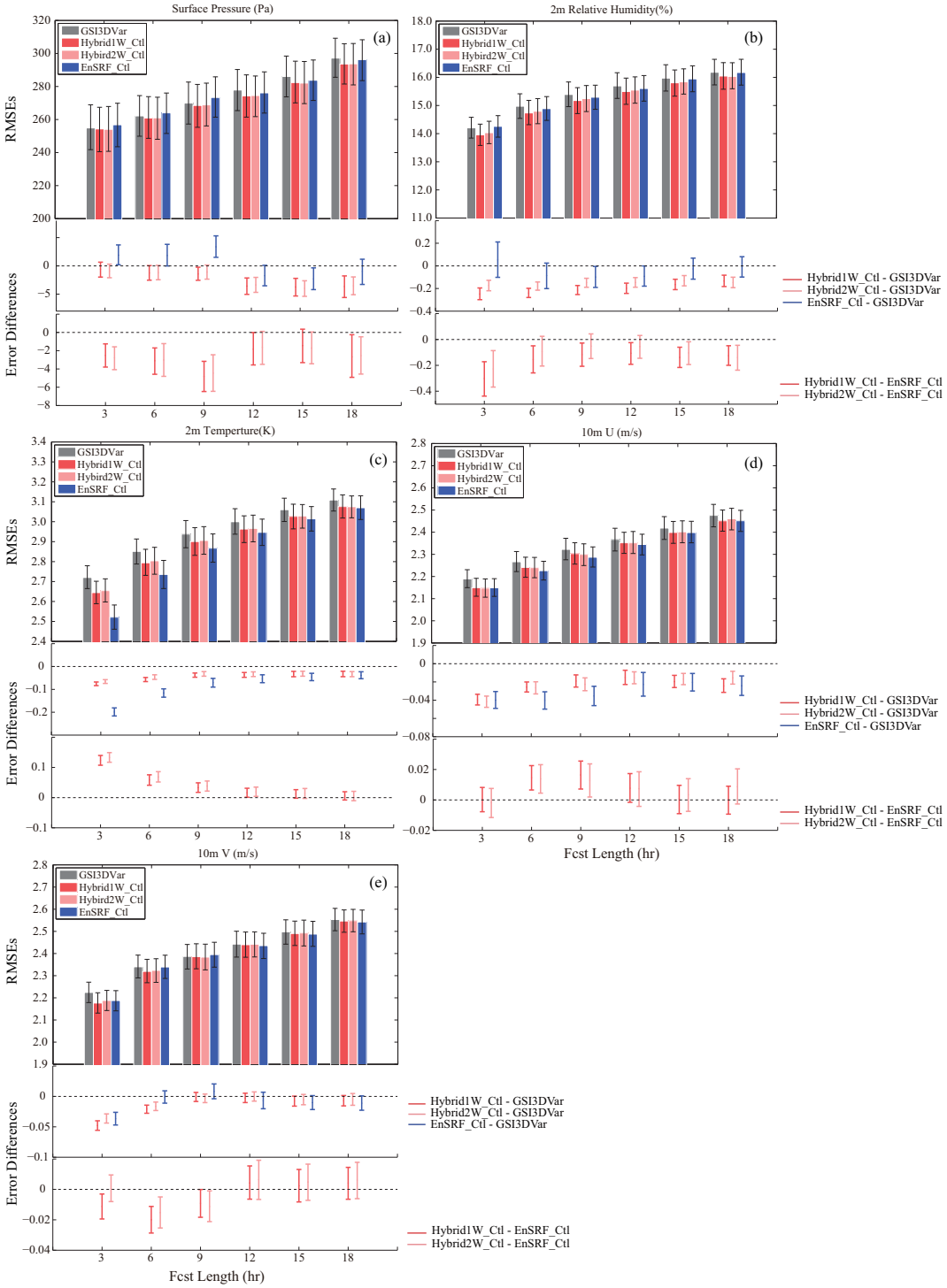


Fig. 6. The 9-day and domain-averaged forecast RMSEs verified against surface observations (upper panel in each frame) and the 90% confidence interval of the RMSE differences between En3DVar hybrid experiments and GSI 3DVar/EnSRF_Ctl (lower panel) for (a) surface pressure, (b) 2-m *RH*, (c) 2-m temperature, (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.

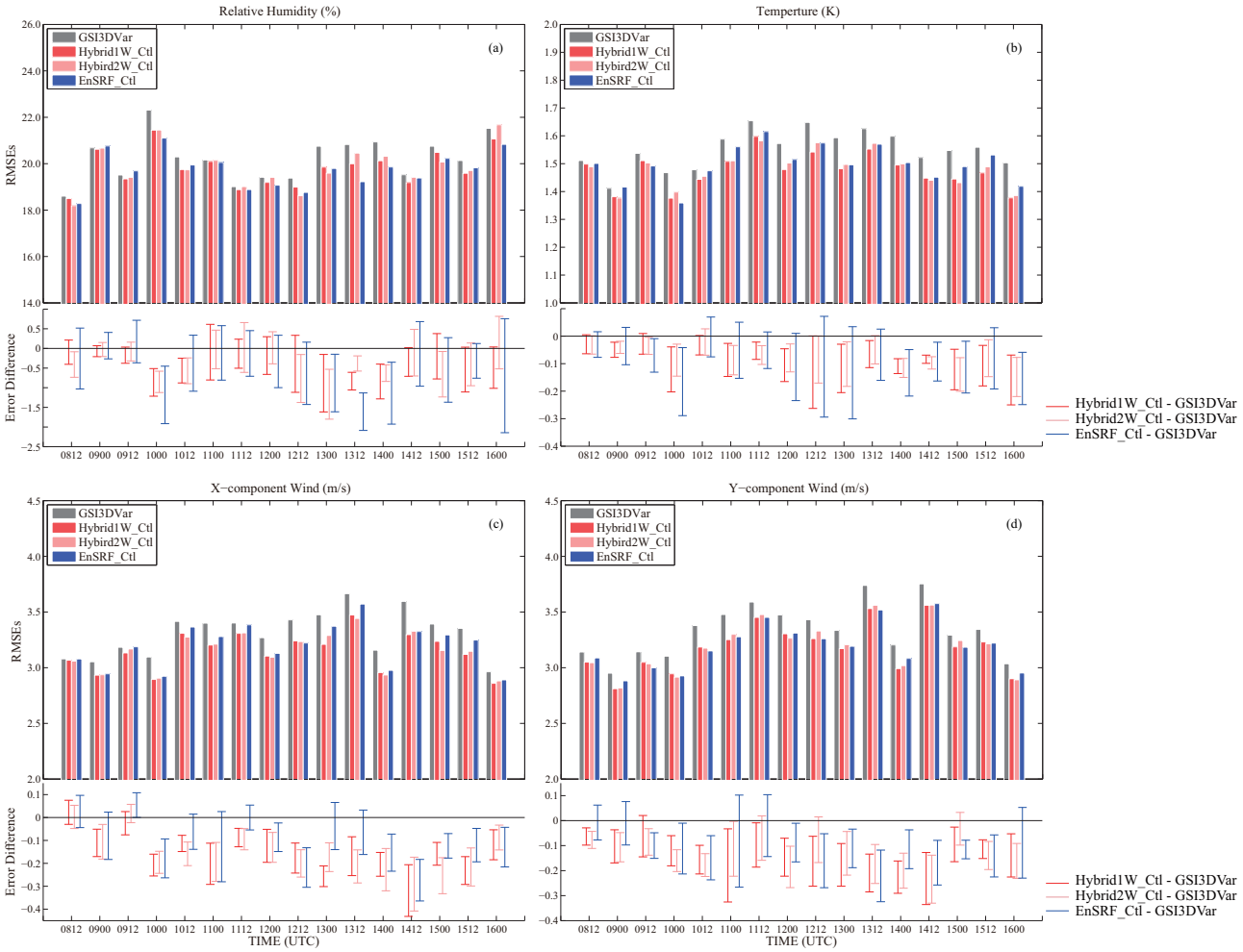


Fig. 7. Domain-averaged 3-hour forecast RMSEs (upper panels in each frame) verified against sounding data at 0000 and 1200 UTC through test period and the 90% confidence interval of RMSE differences (lower panel of each frame) between the En3DVar hybrid and EnSRF experiments and GSI 3DVar for (a) *RH*, (b) *T*, (c) *U*, and (d) *V*.

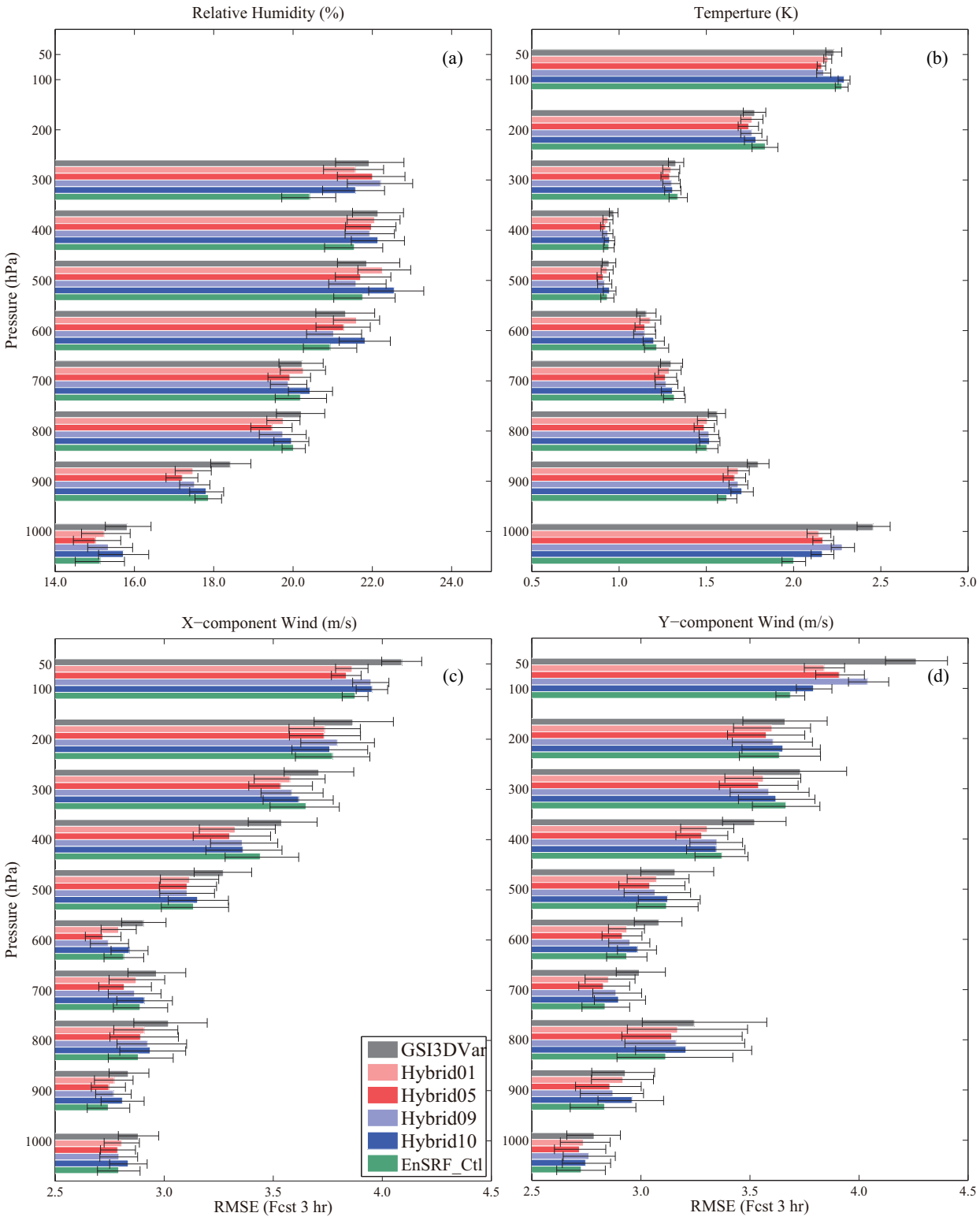


Fig. 8. Mean 3-hour forecast RMSEs at different height levels verified against sounding data for (a) *RH,* (b) *T*, (c) *U,* and (d) V for experiments GSI 3DVar, Hybrid01, Hybrid05, Hybrid07, Hybrid10 and EnSRF_Ctl. Error bars represent the two-tailed 90% confidence interval (5% on the left and 95% on the right) using the bootstrap distribution method.

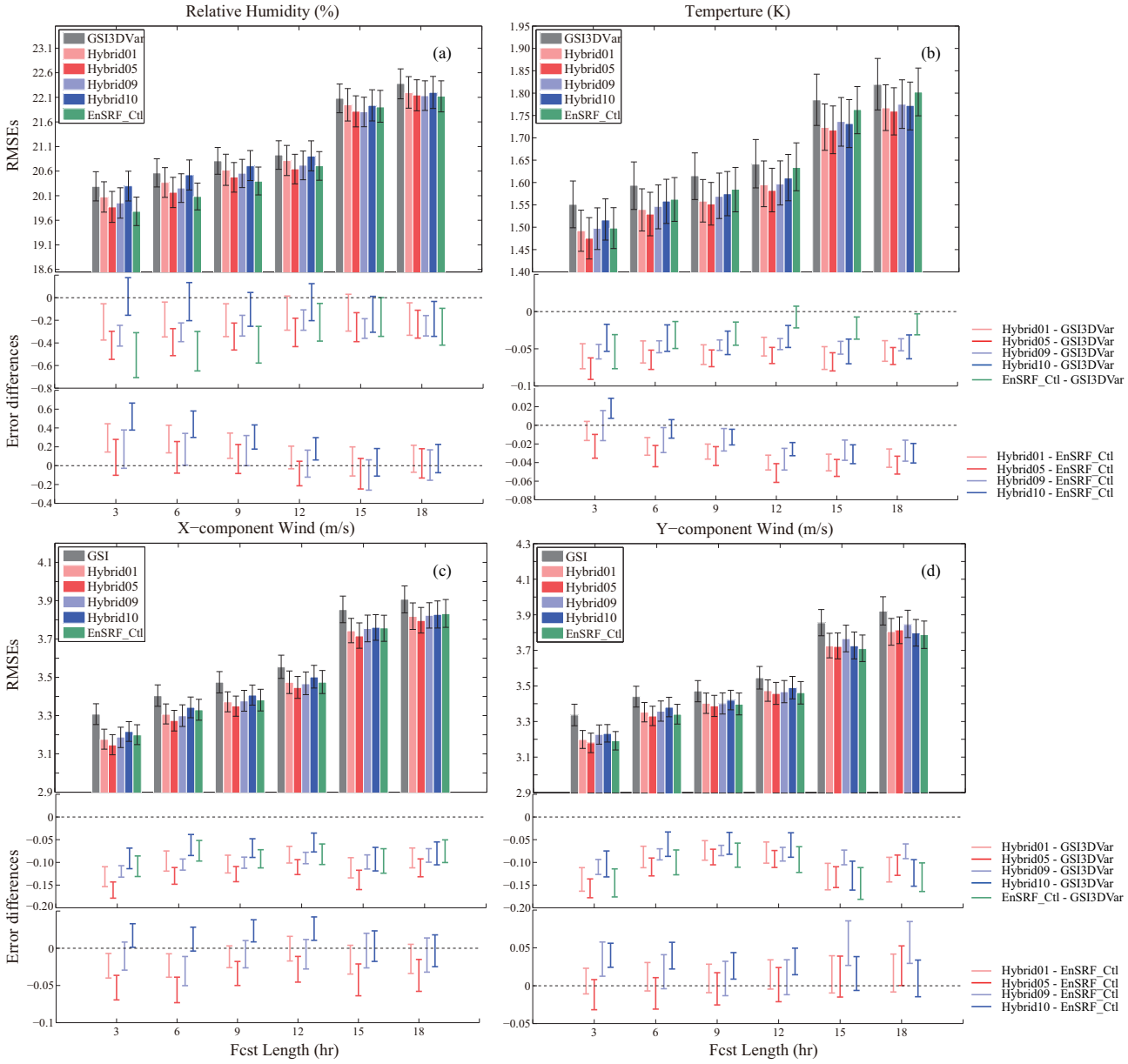
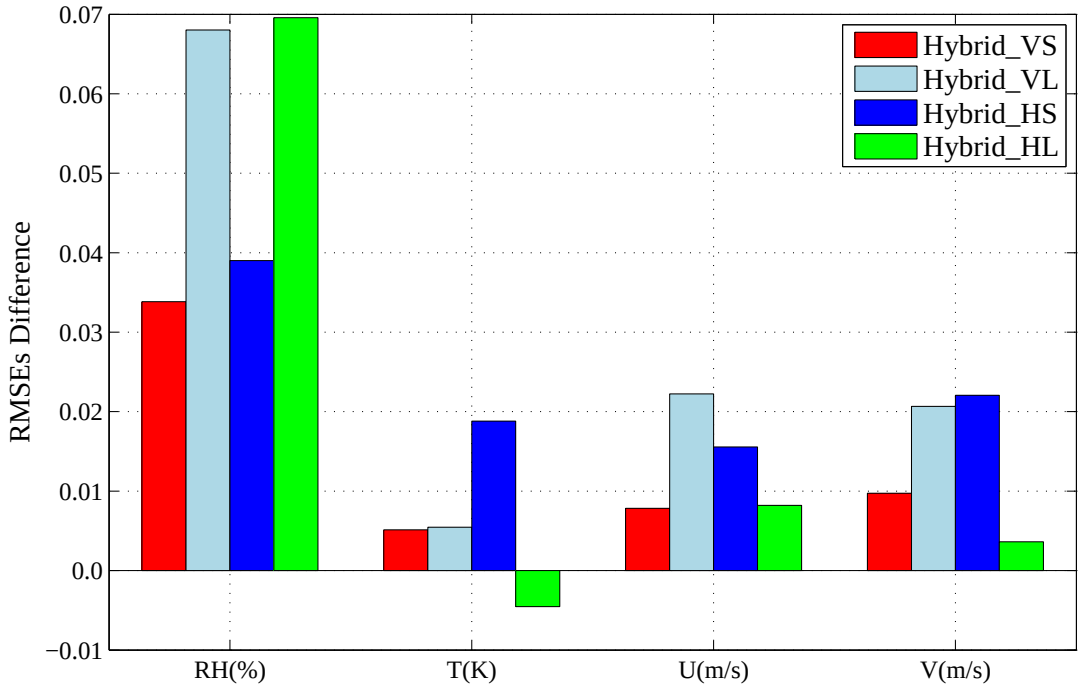
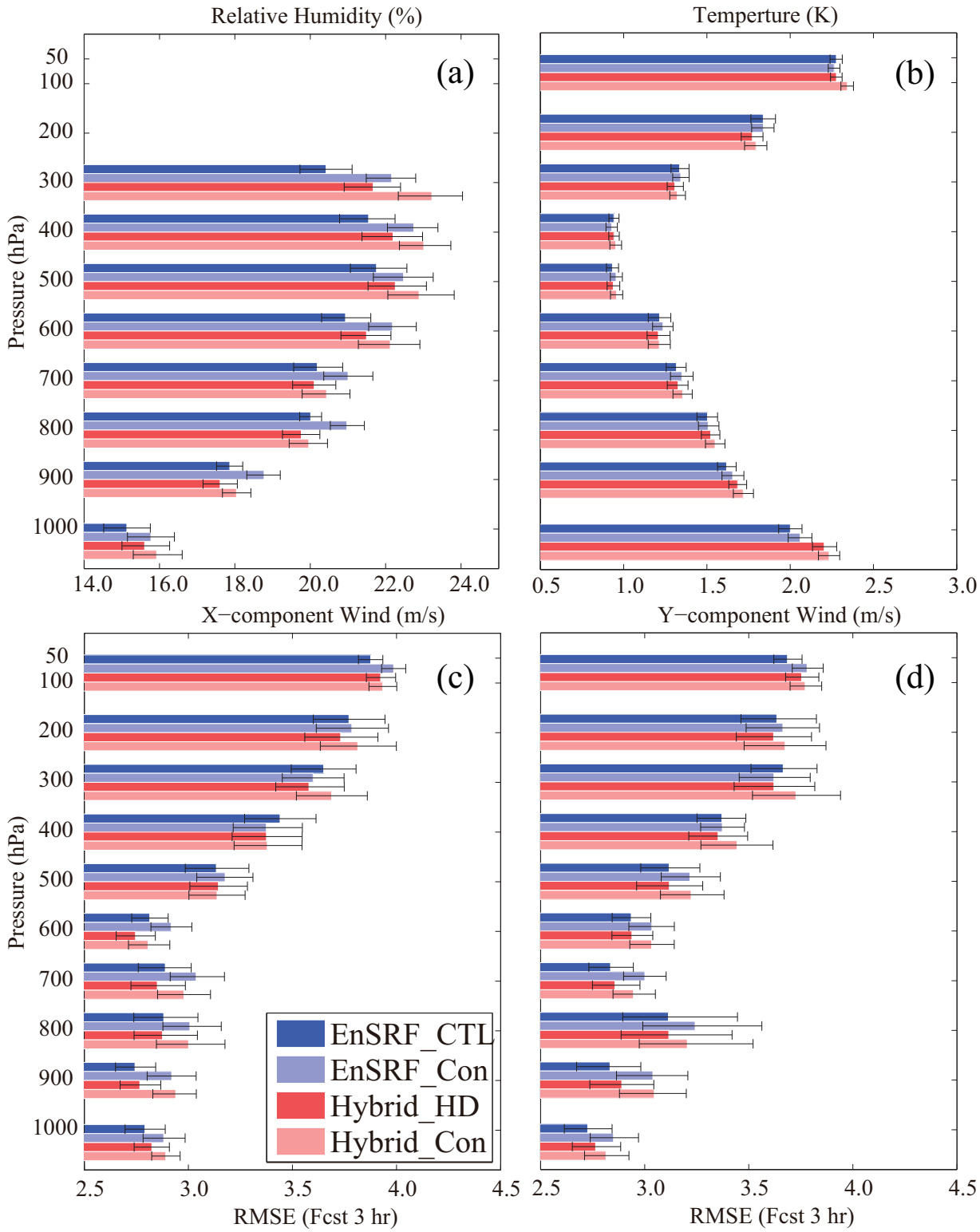


Fig. 9. The same as Fig. 5, except for experiments GSI 3DVar, Hybrid01, Hybrid05, Hybrid07, Hybrid10 and EnSRF_Ctl.



960
961 Fig. 10. Mean absolute forecast RMSE differences between different experiments and
962 Hybrid1W_Ctl, verified against sounding data, for 3-hour forecast averaged over the 9-day forecast
963 period over the entire model domain.

964



965
966
967 Fig. 11. The same as Fig. 4 but for experiments EnSRF_Ctl, EnSRF_Con, Hybrid_HD and
968 Hybrid_Con.

969

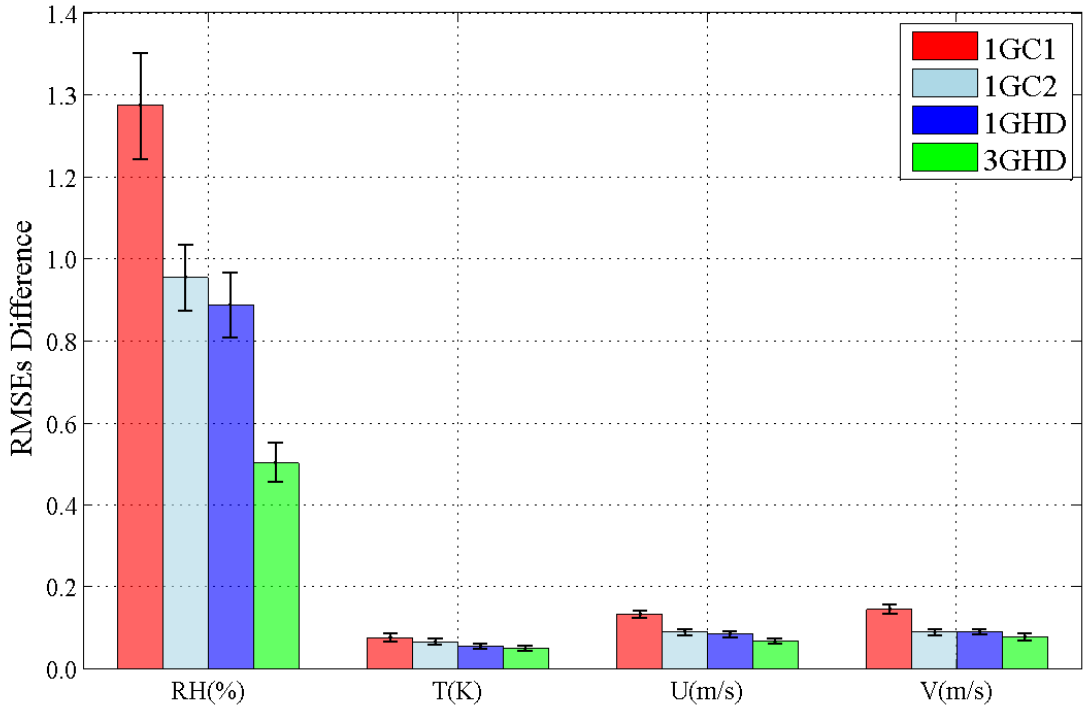


Fig. 12. Nine-day and domain-averaged absolute 3-hour forecast RMSE differences verified against sounding data, where 1GC1 means difference between Hybrid_Con and EnSRF_Ctl, 1GC2 means difference between Hybrid_Con and EnSRF_Con, 1GHD means difference between Hybrid_HD and EnSRF_Ctl, and 3GHD means difference between Hybrid3G and EnSRF3G. The error bars represent the two-tailed 90% confidence interval.

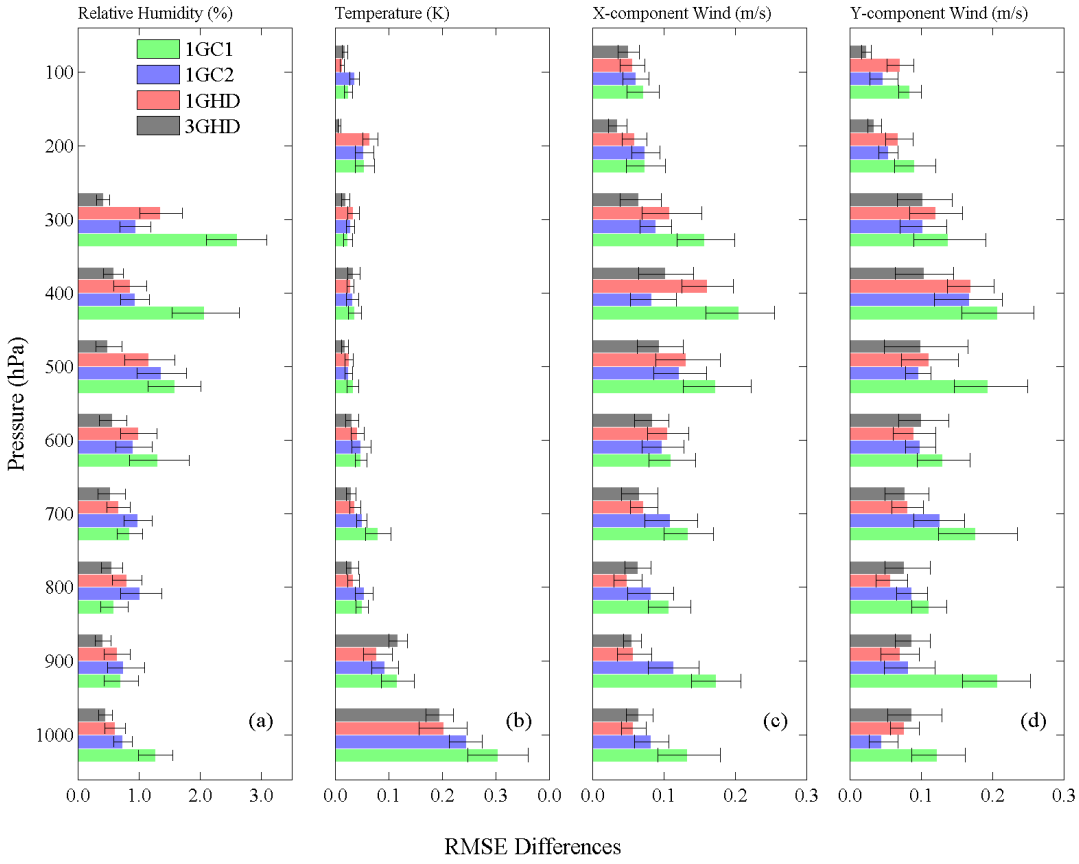
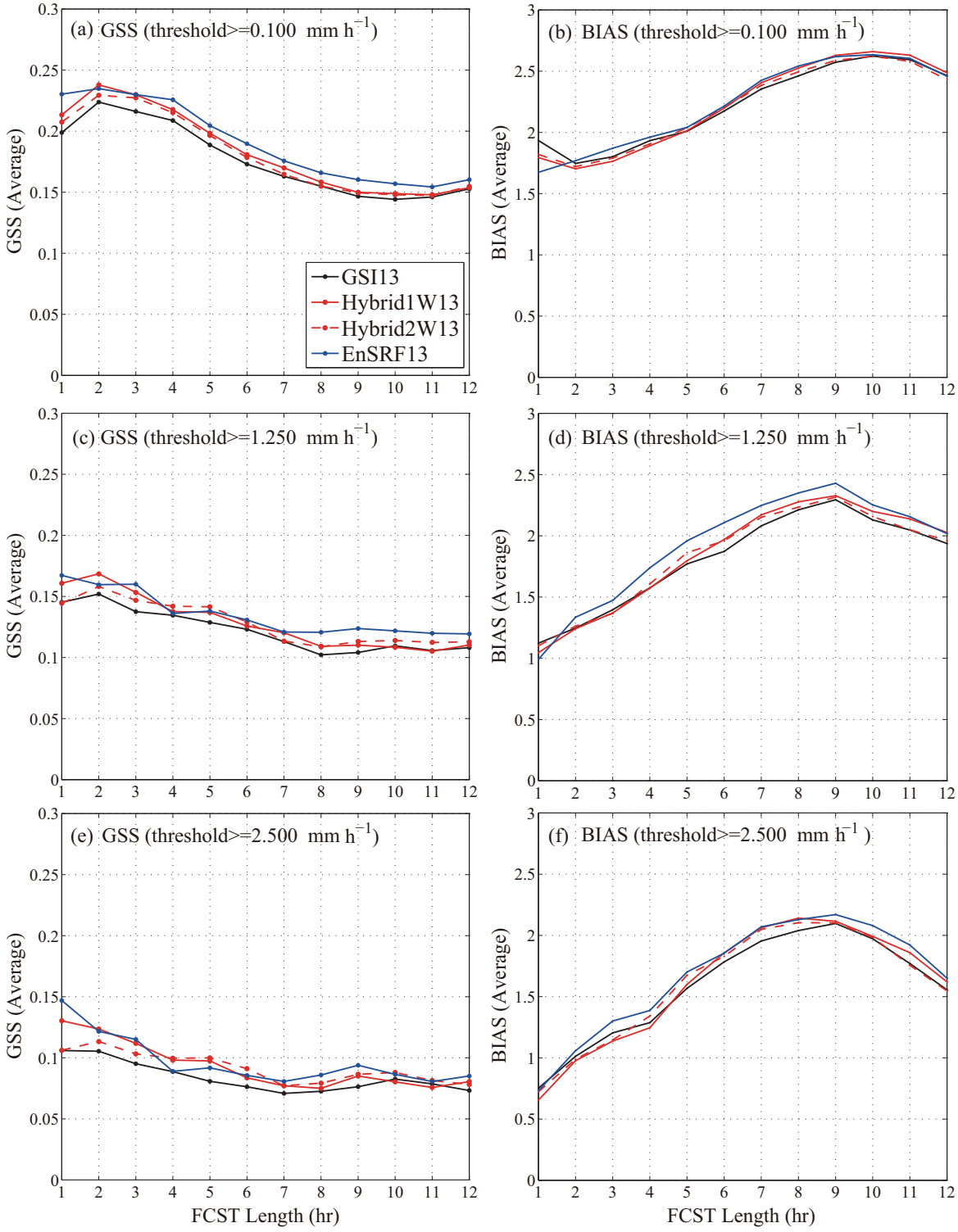


Fig. 13. Nine-day and domain-averaged absolute RMSE differences between Hybrid_Con and EnSRF_Ctl (labeled 1GC1), Hybrid_Con and EnSRF_Con (labeled 1GC2), Hybrid_HD and EnSRF_Ctl (labeled 1GHD), Hybrid3G and EnSRF3G (labeled 3GHD) for (a) *RH*, (b) *T*, (c) *U,* and (d) *V* at different height levels. The error bars represent the two-tailed 90% confidence interval.

983



984
985
986 Fig. 14. Average precipitation GSSs and BIASs of 13-km forecasts as a function of forecast length
987 for thresholds (a) (b) 0.1 mm h^{-1}, (c) (d) 1.25 mm h^{-1}, and (e) (f) 2.5 mm h^{-1} for control experiments.

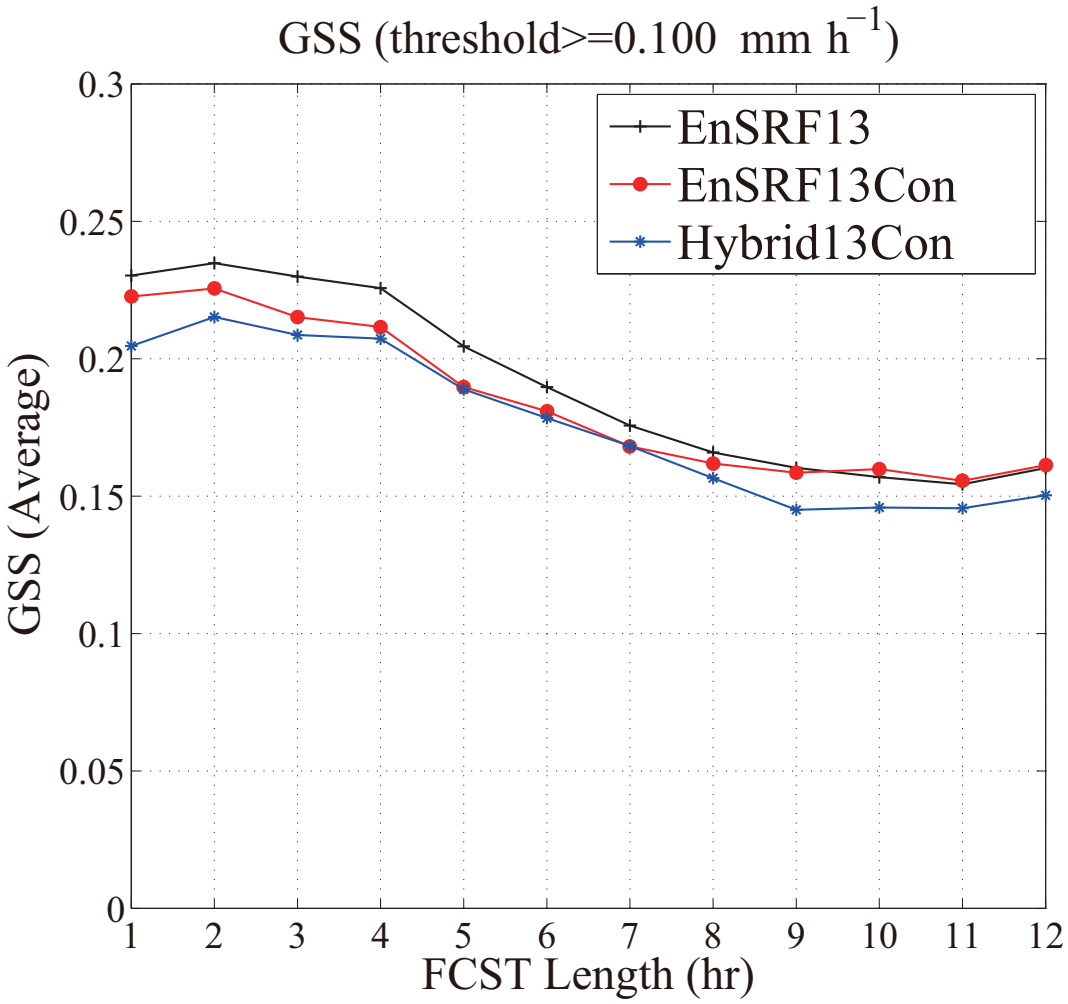


Fig. 15. The same as Fig. 14, but for EnSRF13, EnSRF13Con and Hybrid13Con.

990 Table 1. List of data assimilation experiments. In the horizontal and vertical localization columns,
991 ↗ means increasing with height.

Experiment group	Experiment (including alternative names)	Ensemble covariance weighting factor $(1/\beta_2)$	Horizontal cut-off radius for hybrid/EnSRF (km)	Vertical cut-off radius for hybrid/EnSRF in ln(p)	Ensemble size	EnSRF-En3DVar Coupling
Control experiments	GSI3DVar	N.A.				
	EnSRF_Ctl	-	700 ↗ 1050	*RH* and *T*: 1.1/4 ↗ 1.1/2 *U* and *V*: 1.1/2 ↗ 1.1 PS and PW: 1.6	40	-
	Hybrid1W_Ctl /Hybrid05	0.5	~1095	1.1	40	1-way
	Hybrid2W_Ctl	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	EnSRF_Con	-	1095	1.1	40	-
	Hybrid_Con (Perturbations from EnSRF_CON)	1.0	~1095	1.1	40	1-way
	Hybrid_HD	1.0	700 ↗ 1050	1.1/2 ↗ 1.1	40	1-way
	Hybrid3G	1.0	700 ↗ 1050	*RH* and *T*: 1.1/4 ↗ 1.1/2 *U* and *V*: 1.1/2 ↗ 1.1 PS and PW: 1.6 (observations are assimilated in 3 groups)	40	1-way
	EnSRF3G	-	700 ↗ 1050	*RH* and *T*: 1.1/4 ↗ 1.1/2 *U* and *V*: 1.1/2 ↗ 1.1 PS and PW: 1.6 (observations are assimilated in 3 groups)	40	-

992

993 Table 2. list of mean domain average absolute RMSE difference pair

Name	Hybrid*	Benchmark
1GC1	Hybrid_Con	EnSRF_Ctl
1GC2	Hybrid_Con	EnSRF_Con
1GHD	Hybrid_HD	EnSRF_Ctl
3GHD	Hybrid3G	EnSRF3G

994

995

996 Table 3. List of precipitation forecast name and corresponding data assimilation experiments on 40
997 km grid

Precipitation forecast name	Data assimilation experiment
EnSRF13	EnSRF_Ctl
Hybrid1W13	Hybrid1W_Ctl
Hybrid2W13	Hybrid2W_Ctl
EnSRF13Con	EnSRF_Con
Hybrid13Con	Hybrid_Con

998