GSI 3DVar-based Ensemble-Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single Resolution Experiments

Xuguang Wang
School of Meteorology, University of Oklahoma and Center for Analysis and Prediction of Storms, Norman, OK

David Parrish and Daryl Kleist
National Center for Environmental Prediction, Environmental Modeling Center, Camp Springs, MD

Jeffrey Whitaker
Earth System Research Laboratory, Physical Sciences Division, Boulder, CO

Submitted on May 4 2012

Revised on April 18 2013

Monthly Weather Review

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

Abstract

An ensemble Kalman filter-variational hybrid data assimilation system based on the grid point statistical interpolation (GSI) three dimensional variational (3DVar) system was developed. The performance of the system was investigated using the National Centers for Environmental Prediction (NCEP) Global Forecast System model. Experiments covered a 6-week Northern Hemisphere winter period. Both the control and ensemble forecasts were run at the same, reduced resolution. Operational conventional and satellite observations along with an 80 member ensemble were used. Various configurations of the system including one-way or two-way couplings, with zero or non-zero weights on the static covariance were inter-compared and compared with the GSI 3DVar system. It was found that the hybrid system produced more skillful forecasts than the GSI 3DVar system. The inclusion of a static component in the background-error covariance and re-centering the analysis ensemble around the variational analysis did not improve the forecast skill beyond the one-way coupled system with zero weights on the static covariance. The one-way coupled system with zero static covariances produced more skillful wind forecasts averaged over the globe than the EnKF at the 1-day to 5-day lead times and more skillful temperature forecasts than the EnKF at the 5-day lead time. Sensitivity tests indicated that the difference may be due to the use of the tangent linear normal mode constraint in the variational system. For the first outer loop, the hybrid system showed a slightly slower (faster) convergence rate at early (later) iterations than the GSI 3DVar system. For the second outer loop, the hybrid system showed a faster convergence.

1. Introduction

Variational data assimilation (Var) systems have been in use operationally at the National Centers for Environmental Prediction (NCEP) and most other numerical weather prediction (NWP) centers for at least a decade. Three-dimensional variational (3DVar) systems, such as the operational global statistical interpolation system (GSI; Wu et al., 2002, Kleist et al. 2009b) adopted by NCEP, use a background error covariance matrix that is either completely static or only weakly coupled to the dynamics of the forecast model. Four-dimensional variational data assimilation (4DVar) systems that use a tangent-linear version of an often simplified forecast model implicitly evolve the background error covariance over the assimilation window, starting from a typically static estimate of the covariance at the beginning of the window (e.g. Courtier et al. 1994). In comparison, ensemble Kalman filter (EnKF; e.g., Houtekamer et al. 2005; Whitaker et al. 2008, Szunyogh et al. 2005) data assimilation systems can utilize fully flow-dependent background error covariances, estimated from an ensemble of short range forecasts with the full nonlinear forecast model.

A hybrid analysis method has been proposed (e.g., Hamill and Snyder 2000; Lorenc 2003; Etherton and Bishop 2004; Zupanski 2005; Wang et al. 2007a; Wang 2010) and implemented for regional (e.g., Wang et al. 2008ab; Wang 2011; Zhang and Zhang 2012; Barker et al. 2012; Li et al. 2012) and global (e.g., Buehner 2005; Buehner et al. 2010ab, Bishop and Hodyss 2011, Clayton et al. 2012) NWP. In a hybrid method, the variational framework is typically used to calculate the analysis increment using an ensemble-based, flow-dependent estimate of the background-error covariance. The ensemble can be generated from an EnKF. Recent studies have suggested that hybrid systems may be optimal by combining the best aspects of the Var and EnKF systems (e.g., Wang et al. 2007b, 2009; Buehner et al. 2010b; Zhang and

Zhang 2012). The potential advantages of a hybrid system as compared to standalone Var and EnKF systems were summarized in Wang (2010).

A hybrid EnKF-Var data assimilation system was recently developed based on the operational GSI 3DVar system at NCEP, and was first tested for the Global Forecast System (GFS). The resolution of the operational implementation was T254 (triangular truncation at total wavenumber 254) for the ensemble and T574 for the variational analysis. These results will be documented in a forthcoming paper. Here we present the results of experiments conducted with this system at a reduced spectral resolution of T190 for both the ensemble and the variational analyses (hereafter single resolution experiments). The performances of the GSI 3DVar, the hybrid and the EnKF systems were investigated. The impacts from three aspects of the ensemble-variational coupled system were investigated. These aspects included the weights of the flow-dependent and static components in the background-error covariance, re-centering the analysis ensemble around the variational analysis, and the tangent linear normal mode constraint in the minimization. This paper will focus on describing the results of the hybrid system developed based on the GSI 3DVar system. Formulation, implementation, and results of the four-dimensional extension of the system called, "Four-dimensional Ensemble-Variational (4DEnsVar) system", will be reported in forthcoming papers. Section 2 describes the GSI 3DVar-based ensemble-variational hybrid data assimilation system (hereafter, GHDA). Section 3 describes the design of the experiments. Section 4 discusses the experiment results and section 5 concludes the paper.

2. GSI 3DVar-based EnKF-variational hybrid data assimilation system

For the one-way coupled GHDA as shown in Fig. 1a, each cycle consists of the following three steps:

1) Update the background forecast, using ensemble perturbations to estimate the background error covariance. This is achieved by using the augmented control vector (ACV) method as described below. Hereafter, GSI with the ACV is denoted as "GSI-ACV".

2) Update the forecast ensemble to generate the analysis ensemble using an EnKF.

3) Make ensemble and control forecasts to advance the state to the next analysis time.

For a two-way coupled GHDA as shown in Fig. 1b, step 2 is modified by re-centering the analysis ensemble generated by the EnKF around the control analysis to produce the final analysis ensemble. One motivation for such a modification is to allow the EnKF perturbations to evolve with the trajectory of the control forecast so that the ensemble covariance may potentially better represent the error statistics of the control forecast.

One component in the GHDA is the "GSI-ACV" (Fig. 1). Wang (2010) described the mathematical details on how the ensemble covariance was implemented in the GSI variational minimization through the ACV, where the minimization was preconditioned upon the full background error covariance. Below we briefly describe the formulas following the notation of Wang (2010). Similar notation was used in Lorenc (2003) and Buehner (2005). In the GHDA, the analysis increment $\mathbf{x}'$ is defined as

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

The first term $\mathbf{x}'_1$ is the increment associated with the static covariance. The second term is the flow-dependent increment associated with the ensemble covariance. $\mathbf{x}_k^e$ is the kth ensemble perturbation normalized by $\sqrt{K-1}$, where K is the ensemble size. The vectors $\mathbf{a}_k$, $k = 1, \cdots K$, denote the augmented control vectors for each ensemble member. The symbol $\circ$ denotes the

Schur product. The analysis increment $\mathbf{x}'$ is obtained by minimizing the following hybrid cost function

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

The first term on the right hand side is the traditional 3DVar background term with the static covariance $\mathbf{B}_1$. The last term is the observational term, which is the same as in a traditional 3DVar system except that $\mathbf{x}'$ is defined by (1).

In the second term, $\mathbf{a}$ is a vector formed by concatenating K unitless vectors, $\mathbf{a}_k$, $k = 1, \cdots K$. These augmented control vectors are constrained by a block-diagonal matrix, $\mathbf{A}$, which defines the localization applied to the ensemble covariance. In the current implementation, each $\mathbf{a}_k$, $k = 1, \cdots K$, is a three-dimensional field located at the model grid points. Each $\mathbf{a}_k$ varies in both the horizontal and vertical directions so that the spatial localization is applied both horizontally and vertically. The same three-dimensional fields, $\mathbf{a}_k$, are applied for all variables. In other words, the cross-variable covariance calculated from the ensemble is not modified by the localization. Specifically, in eq. (1) the variables on which $\mathbf{a}_k$ are applied include surface pressure, wind, virtual temperature, relative humidity, cloud water mixing ratio and ozone mixing ratio. In GHDA, the vertical covariance localization part ($\mathbf{A}_v$) of matrix $\mathbf{A}$ is realized through a recursive filter transformation (Hayden and Purser 1995) with the distance measured either in scale heights (i.e., natural log of the pressure) or number of model levels (see $\mathbf{A}$ in eq. (16) of Wang 2010 on where the localization is implemented during the minimization). For the GFS, the horizontal localization is realized through a spectral filter transform. Specifically, the horizontal localization part ($\mathbf{A}_h$) of matrix $\mathbf{A}$ in eq. (16) of Wang (2010) is converted into the spectral space by $\mathbf{A}_h = \mathbf{S}^{-1}\mathbf{A}_s\mathbf{S}$, where $\mathbf{S}$ represents the transformation from horizontal grid space to spectral space and $\mathbf{S}^{-1}$ is the inverse spectral transformation. $\mathbf{A}_s$ is a diagonal matrix containing the spectral coefficients corresponding to the horizontal localization function predefined in model grid-space. No further spectral truncations such as those in Buehner (2005) are implemented here and the minimization is still conducted with respect to the augmented control vectors in the model grid space with the full horizontal resolution. E-folding distances equivalent to 1600 km and 1.1 scale height (natural log of pressure is equal to 1.1) cut-off distances in the Gaspari-Cohn (1999) localization function were adopted for the horizontal and vertical localizations respectively in the current study.

There are two factors β_1 and β_2 whose inverses define the weights placed on the static covariance and the ensemble covariance respectively. In the current implementation, these two weighting factors satisfy $\frac{1}{\beta_1}+\frac{1}{\beta_2}=1$. Wang et al. (2007a) proved that the method of using augmented control vectors to incorporate ensembles in the variational framework as in eq. (1) and (2) and the method of directly combining the ensemble covariance with the static covariance are theoretically equivalent. Wang et al. (2007a) also described the relationship between the weighting factors applied on the covariances such as implemented in the current study and in Wang et al. (2008a) to those applied on the increments such as implemented in Lorenc (2003) and Buehner (2005). Note that eq. (2) provides a generic form of the hybrid cost function. After plugging in $\mathbf{z}$ in eq. (11) of Wang (2010) and $\mathbf{x}$ immediately defined after eq. (11) of Wang (2010) to either equation (6) in Wang (2010) or eq. (2) in this paper, the inverse of $\mathbf{B}_1$ and $\mathbf{A}$, and the weighting factors β_1 and β_2 , do not explicitly appear in the cost function. For further details please refer to Wang (2010).

Another component of the GHDA is the ensemble update, which is achieved by using an EnKF. An ensemble smoother version (i.e., a version taking into account the four-dimensional ensemble covariance within the assimilation window) of the square root filter algorithm

(Whitaker and Hamill 2002) was adopted. A recent implementation of an EnKF for the GFS was described more fully in Hamill et al. (2011). This EnKF code has been efficiently parallelized following Anderson and Collins (2007) and directly interfaced with the GSI by using the GSI's observation operators, pre-processing and quality control for operationally assimilated data. In the EnKF, to account for sampling errors due to the limited ensemble members, cut-off distances of 1600 km in the horizontal direction and 1.1 scale heights in the vertical direction were used for the localization for all observations except the surface pressure and satellite radiance observations, where vertical localization was prescribed to be 2.2 and 3.3 scale heights respectively to account for the non-local nature of these observations. Temporal localization using a 16-hour cut-off distance was also implemented[1]. To account for the deficiency in the spread of the first guess ensemble from the EnKF, both multiplicative and additive inflation were applied in the EnKF. For the multiplicative inflation, an adaptive algorithm proposed by Whitaker and Hamill (2012) was adopted which inflated the posterior ensemble in proportion to the amount of the reduction of the ensemble variance due to the assimilation of observations. This algorithm resulted in a larger inflation in regions of dense observations. In this study, the inflation was performed by relaxing the posterior ensemble variance to 90% of the prior ensemble variance. For the additive inflation, the additive noise was drawn from a full year's inventory of differences between 48-hour and 24-hour forecasts valid at the same time. A factor of 40% was applied to the differences before being added to the posterior ensemble. These parameters were tuned so that the average background ensemble spread matched the average background errors. The additive perturbations were applied to the analysis rather than

[1] The data assimilation window is defined as extending from 3 hours before to 3 hours after the center of the assimilation window. The bell-shape Gaspari-Cohn localization function tapers from the center of the assimilation window and reaches zero16 hours away from the center of the assimilation window.

background ensemble so that the flow dependent structure could be established for the additive perturbations during the 6-hour model integration.

3. Experiment design

The data assimilation cycling experiments were conducted during a 6-week period, 0000 UTC 15 December 2009 ~ 1800 UTC 31 January 2010. The operationally available observations including conventional and satellite data were assimilated every 6 hours. A list of types of the operational conventional and satellite data are found on the NCEP website[2]. The operational NCEP Global Data Assimilation System (GDAS) consisted of an "early" and a "final" cycle. During the "early" cycle, observations assimilated had a short cutoff window. The analyses were then repeated later including the data that had missed the previous "early" cutoff to provide the "final" analyses for the 6-h forecast which was used as the first guess of the next "early" cycle. As a first test of the newly developed hybrid system, only observations from the "early" cycle were assimilated. The same observation forward operators and satellite bias correction algorithms as in the operational GSI 3DVar system were used. The quality control decisions from the operational GDAS were adopted for all experiments. The GFS model was configured the same way as the operational GFS except that the horizontal resolution was reduced to T190 to accommodate the sensitivity experiments using limited computing resources. The model contained 64 vertical levels with the model top layer at 0.25 hPa. An 80-member ensemble was run following the operational configuration. A digital filter (DFI; Lynch and Huang 1992) was applied during the GFS model integration for all experiments following the operational configuration. For all of the experiments presented in this work, the same model

[2] http://www.emc.ncep.noaa.gov/mmb/data_processing/prepbufr.doc/table_2.htm and table18.htm.

configuration was adopted, and the same observations were ingested, except that the EnKF excluded satellite derived precipitation rates. This exclusion was because the proper observation space vertical covariance localization adopted by the EnKF for observation types such as the precipitation rates was still under research. Earlier work by Treadon et al. (2002) also reported little impact of satellite derived rain rates assimilated by the GSI 3DVar system on the global forecasts. Verification was conducted using data collected during the last 4 weeks of the experiment period.

Since the operational static covariance was derived from GFS forecasts at higher resolution, both the correlation length scales and the magnitude of the error variances of the control variables were tuned for the lower resolution experiments. The tuning was achieved by incrementally changing the correlation length scale and the error variance by 10% and running the standalone GSI 3DVar system over the 6-week period until the performance of the GSI 3DVar system converged. The final, tuned static covariance, whose error variance and horizontal length scales were 20% larger than the operational covariance, was used in the following experiments.

A few sensitivity tests were conducted for the hybrid system. Both one-way and two-way coupling experiments were conducted. Additionally, two different sets of background covariance weighting factors ($\frac{1}{\beta_1} = 0\ and\ 0.5$) were adopted. The former used 0% static background error covariance and 100% ensemble covariance, and the latter used a blend of 50% static and 50% ensemble background error covariances. The impact of applying the tangent linear normal mode balance constraint (TLNMC) during the variational minimization where the background ensemble was purely from the ensemble covariance was also investigated. The one-way coupled system with and without the use of the TLNMC was compared with the EnKF. In addition, the impact of the inclusion of an ensemble covariance on the minimization convergence rates was investigated. For all of the GSI 3DVar and the GHDA experiments, two outer loops were used following the operational configuration. Table 1 lists the experiments conducted along with naming conventions.

4. Results

a. Comparison of various configurations of the hybrid system and the GSI 3DVar system

1). FITS OF ANALYSES TO OBSERVATIONS

A series of experiments assimilating a single observation were conducted to verify that the GSI-ACV ingested the flow-dependent ensemble covariance properly. In contrast to the GSI 3DVar whose increment was quasi-isotropic, flow-dependent increments similar to Fig. 4 of Wang et al. (2008a) were found for the GSI-ACV (not shown). In this subsection, the ensemble-variational coupled experiments with various configurations (3DEnsVar1way, 3DEnsVar2way and Hybrid1way0.5 in Table 1) and the GSI 3DVar experiment are compared.

Figure 2 shows the root-mean-square fit of the analysis to rawinsonde observations averaged over the experiment period. The analyses from 3DEnsVar1way and 3DEnsVar2way fit the observations similarly. The analyses from 3DEnsVar1way and 3DEnsVar2way fit the temperature observations more (less) closely than GSI3DVar above (below) 550 hPa[3]. The analyses from 3DEnsVar1way and 3DEnsVar2way fit the wind observations more (less) closely than GSI3DVar above (below) 850 hPa. The analyses from Hybrid1way0.5 fit the observations more closely than GSI3DVar throughout all vertical levels. Compared to 3DEnsVar1way and

[3] Note that the fit of the analyses to observations assimilated is not a measure of the accuracy of the analyses.

3DEnsVar2way, the analyses using 50% static and 50% ensemble covariances (Hybrid1way0.5) fit the observations less (more) closely above (below) 250 hPa. Wang et al. (2008b) found that analyses from 3DVar for the Weather Research and Forecast (WRF) model fit the observations more closely than the WRF Ensemble Transform Kalman Filter (ETKF)-3DVar hybrid. The relative difference of the fits of the analysis to observations between the hybrid and 3DVar algorithms may therefore be dependent on the specific configuration of the data assimilation and forecast system. In general, the fit of the analyses to observations is determined by the combined effects of the relative magnitude of the background and observation error variance, the degrees of freedom and the accuracy of the background error covariance, and the accuracy of the background forecast. To confirm the impact of the magnitude of the background error variance and the degrees of freedom of the background error covariance, the fits of the analyses to observations from differently configured GSI3DVar experiments were compared. In these experiments, the background error variance and the correlation length scale were varied. It was found that for smaller background error variances or larger correlation scales, the analyses tended to fit the observations less (not shown).

2). VERIFICATION OF FORECASTS

The root mean square errors (RMSEs) of wind and temperature forecasts verified against the rawinsonde data at different forecast lead times over the globe were calculated. As shown in Figure 3, the forecasts produced by the various configurations of the ensemble-variational coupling experiments (3DEnsVar1way, 3DEnsVar2way and Hybrid1way0.5) are more skillful than that of the GSI3DVar experiment (similar results were found at 6-hour lead time). Relative to the variation of the errors in the vertical, which determines the range on the x-axis[4], the improvement of temperature forecasts relative to GSI3DVar increases whereas the improvement of wind forecasts decreases from the 24-hour to 120-hour lead time. Figure 4 shows the RMSEs of the wind and temperature forecasts verified against the rawinsonde data at the 72-hour lead times over the Northern Hemisphere (NH) extratropics, Tropics and Southern Hemisphere (SH) extratropics. Relative to the variation of errors in the vertical, the improvement relative to GSI3DVar is larger over the extratropics than the Tropics.

The variously configured ensemble-variational coupling experiments were also inter-compared amongst each other. Figures 3 and 4 show that in general the performance of the two-way coupled system (3DEnsVar2way) is not better than the one-way coupled system (3DEnsVar1way)[5]. The inclusion of the static covariance with a 50% weight (Hybrid1way0.5) does not improve the performance beyond the use of the full ensemble covariance (3DEnsVar1way). Reducing the weight on the static covariance from 50% to 25% does not improve the performance beyond 3DEnsVar1way (not shown). Earlier studies (e.g., Wang et al. 2007b) suggested that the optimal weight placed on the static covariance depended on the relative quality of the static and ensemble covariance estimates. For example, Wang et al. (2007b) showed that when the size of the ensemble was decreased, the optimal weight applied on the static covariance was increased. It is expected that for the GHDA with a smaller ensemble size or with the ensemble run at a lower resolution than the control forecast (hereafter dual-resolution experiment), the inclusion of the static covariance would have a positive impact.

---

[4] At different lead times, the magnitude and range of the errors in general increase. Such measure provides an assessment of the improvement relative to the range of the errors at the corresponding lead times.

[5] The difference between one-way and two-way 3DEnsVar in the mid-troposphere in Fig. 3e was not significant as when the number of samples was reduced the difference became smaller.

Research on comparing the hybrid under the single and dual-resolution configurations and the impact of the static covariance in these configurations is being conducted. Our initial results showed that for a dual resolution configuration using an 80-member ensemble where the EnKF was run at a half of the resolution of the deterministic 3DVar, the combination of the static and ensemble covariances significantly improved the performance relative to using the ensemble covariance alone, and the hybrid improved upon the 3DVar with the dual resolution configuration (not shown). It is also expected that in the dual-resolution configuration, re-centering the coarser resolution analysis ensemble around the higher resolution control analysis (i.e., two-way coupling) would improve the forecast than without re-centering (i.e., one-way coupling) since the higher resolution control analysis is supposed to provide more accurate analyses.

Analyses of wind, temperature and specific humidity from ECMWF were used as independent verifications (available from http://tigge.ecmwf.int). Forecast lead times at and beyond 72-hour were chosen to reflect that it would be more appropriate to use the analyses to verify longer forecasts than short forecasts. Consistent with Fig. 3, the forecasts from various ensemble-variational coupling configurations generally fit the ECMWF analyses more closely than those from GSI3DVar. Relative to the variation of the errors in the vertical, the improvement of temperature forecasts increases or remains similar whereas the improvement of wind and specific humidity forecasts decreases from the 72-hour to 120-hour lead time (not shown). Further verification with respect to different parts of the globe (Figure 5) shows that relative to the variation of errors in the vertical, the improvement relative to GSI3DVar is larger over the extratropics than the Tropics for wind and temperature forecasts, consistent with the verification against the rawinsonde observations. For specific humidity forecasts, the improvement relative to GSI3DVar in the Tropics is comparable to or larger than the extratropics. Also consistent with the verification against rawinsonde observations, the inclusion of the static covariance with a 50% weight (Hybrid1way0.5) and the use of two-way coupled hybrid (3DEnsVar2way) generally do not further improve the performance beyond the one-way coupled system with a full ensemble covariance (3DEnsVar1way).

b. Verification of background ensemble spread

As mentioned in section 2, both multiplicative and additive inflation were implemented in the EnKF to alleviate the deficiency of the ensemble in accounting for system errors. In this section, the relationship of the 6-hour background ensemble spread to the 6-hour background forecast error is evaluated. Figure 6 shows the square root of the ensemble variance plus the observation-error variance, and the root-mean-square fit of the first guess to the rawinsonde observation. For the theory behind the use of the above metrics to verify the ensemble spread, please refer to Gelb (1974, Eqs. 9.1-15, Page 318), Houtekamer et al. (2005), Wang et al. (2008b) and Whitaker et al. (2008). For both temperature and wind forecasts, the ensemble is under-dispersive in the lower and upper troposphere and is over-dispersive in the middle of the troposphere. The same pattern is found for other configurations of the hybrid system (not shown). A similar pattern was found in Whitaker et al. (2008) where the EnKF was tested in GDAS at T62 resolution assimilating only conventional observation and in Wang et al. (2008b) where the ensemble Transform Kalman filter (ETKF; Wang and Bishop 2003, Wang et al. 2004; 2007b) was used to produce the ensemble for the WRFVAR based hybrid system. The fact that the vertical structures of the spread and skill do not match suggests that the multiplicative inflation and additive noise methods that aim to parameterize system errors are deficient and therefore do not correctly represent the vertical structure of the actual system errors. In both system error parameterizations, there is only one tunable parameter. It is possible that the spread-skill consistency may be improved if more level dependent tunable parameters are introduced on the additive noise methods. The ensemble spread is also decaying during the first 6 hours of model integration, which suggests that other methods to account for the system errors should be explored. For example, one can explore the use of multiple parameterizations, stochastic physics (Buizza et al. 1999) and Stochastic Kinetic energy backscatter schemes (Shutts 2005) to account for model errors. It is expected that the performance of the GHDA will be further improved when the deficiency of the ensemble spread is further alleviated.

c. Impact of TLNMC balance constraint

Imbalance between variables introduced during data assimilation can degrade the subsequent forecasts. The TLNMC was implemented in the GSI minimization to improve the balance of the initial conditions. Kleist et al. (2009a) showed that the impact of the TLNMC resulted in substantial improvement in the forecasts initialized by the GSI 3DVar system. In the GHDA, the static background error covariance as shown by Wang et al. (2007a; 2008a) was effectively replaced by or was weighted with the flow-dependent ensemble covariance. The mass-wind relationship in the increment associated with the ensemble was defined by the multivariate covariance inherent in the ensemble perturbations. The background ensemble covariance could also become more balanced due to the 6-hour spin up during the forecast steps of the data assimilation cycling. On the other hand, the covariance localization applied on the ensemble covariance could degrade balance (e.g., Lorenc 2003; Kepert 2009; Holland and Wang 2013). The impact of the TLNMC on the ensemble increment was therefore investigated.

Experiments configured to be the same as GSI3DVar and 3DEnsVar1way, but without the use of the TLNMC were conducted. Figure 7 shows that the TLNMC yields significantly positive impact for forecasts from both GSI3DVar and 3DEnsVar1way over the globe, especially after 1-day forecast lead time. Relative to the vertical variation of the errors, the impact of the TLNMC on GSI3DVar and 3DEnsVar1way is comparable. Figure 8 shows the impact of the TLNMC decomposed into the extratropics and tropics at the 120-hour lead time. The TLNMC shows positive impact in both NH and SH extratropics, and mostly neutral impact in the tropics except the positive impact for GSI3DVar at the middle to lower levels. At the 120-hour lead time, the positive impact of the TLNMC is comparable between the NH and SH extratropics. At shorter lead times (e.g., 72-hour, not shown), the positive impact of the TLNMC is larger in the SH than the NH extratropics.

d. Measure of imbalance

The mean absolute tendency of surface pressure (Lynch and Huang 1992) is a useful diagnostic for showing the amount of imbalance for an analysis generated by a data assimilation system. Figure 9a shows the mean absolute surface pressure tendency calculated using the GFS output at every model integration time step (two minutes) for 3DEnsVar1way with and without the use of the TLNMC, and GSI3DVar with and without the use of the TLNMC up to the 9-hour lead time. A representative case during the experiment period was selected. For both GSI3DVar and 3DEnsVar1way, applying the TLNMC results in more balanced analyses and forecasts throughout the 9-hour period. The analyses generated by GSI3DVar are more balanced than 3DEnsVar1way especially when the TLNMC is not applied.

Note that for all of the experiments, following the operational configuration of the GFS, a digital filter was applied during the model integration. In this study, the digital filter was configured so that its impact on the forecasts started from the second hour of the model integration. Figure 9b shows the mean absolute surface pressure tendency for the same case as in Figure 9a except that the DFI is turned on at the second hour of the model integration. For all experiments, the use of the DFI improves the balance of the forecasts starting from the second hour. Since the hourly GFS output where DFI was applied at the second hour was readily available for the whole experiment period, the hourly surface pressure tendency averaged over the experiment period was calculated and summarized before and after the second hour (Table 2). For both GSI3DVar and 3DEnsVar1way, applying the TLNMC results in more balanced forecasts even after the DFI is applied. However, the difference is smaller compared to when the DFI is not used. The analyses generated by GSI3DVar are still more balanced than 3DEnsVar1way after the DFI is applied, although the difference is smaller compared to when the DFI is not used.

Note that although the imbalance decreases quickly after the DFI is applied, errors due to the imbalance can grow with time and lead to a difference in the forecast accuracy at longer lead time as seen in Figure 8. As described in section 2, the covariance localization transform was performed on the augmented control variables and these control variables were used to modulate the ensemble perturbations in the space of surface pressure, wind, virtual temperature, relative humidity, cloud water and ozone mixing ratios. As discussed in Kepert (2009) and Clayton et al. (2012), covariance localization conducted in a space such as stream function and velocity potential can potentially better preserve balance. Further investigation of applying the localization on different variable spaces and their interaction with the TLNMC is left for future study.

e. Impact on convergence during the variational minimization

In addition to the description in section 2, the detailed formula to implement the ensemble covariance, the covariance localization and the weighting factors in the GSI minimization are found in Wang (2010). Different from Lorenc (2003) and Buehner (2005), the weighting factors in the GHDA were applied on the penalty terms associated with the static and ensemble covariances rather than the increments. Different from Lorenc (2003), Buehner (2005), and Wang et al. (2008a), the covariance localization in the GHDA was implemented to be in compliance with the full background covariance preconditioning in the GSI. Please refer to Wang (2010) for details. To investigate the impact of the inclusion of the ensemble covariance in the GSI minimization, the convergence rates of 3DEnsVar1way, and Hybrid1way0.5 were compared with that of GSI3DVar. Figure 10 shows the level of convergence measured by the ratio of the gradient norm relative to the initial gradient norm during the variational minimization averaged over the experiment period. For the first outer loop, 3DEnsVar1way and Hybrid1way0.5 shows a slightly slower convergence rate at early iterations and a slightly faster convergence rate at later iterations than GSI3DVar. For the second outer loop, 3DEnsVar1way and Hybrid1way0.5 shows faster convergence than GSI3DVar. In the current experiments, the maximum iteration steps were 100 and 150 for the first and second outer loops for all experiments. The same numbers were used in the operational system. The minimization was terminated at the maximum iteration step in most cases. Figure 10 also shows that the iterations are terminated at the similar level of the ratio of gradient norms for the GSI3DVar, 3DEnsVar1way and Hybrid1way0.5 experiments. The convergence rate is not sensitive to whether a 100% or a 50% weight are applied on the ensemble covariance. For the experiments conducted in this study, the cost of the hybrid and EnKF analysis updates were comparable and were about twice that of the GSI 3DVar update.

f. Comparison of 3DEnsVar with EnKF

Figure 11 shows the root mean square error of the wind and temperature forecasts verified against rawinsonde data at different forecast lead times over the globe for EnKF and 3DEnsVar1way. 3DEnsVar1way was selected given its generally better performance than the other configurations of the ensemble-variational coupling system. Here, the EnKF forecasts were single forecasts from the EnKF mean analyses rather than the mean of the ensemble forecasts. Figure 11 shows that wind forecasts from 3DEnsVar1way fit the observations better than EnKF. For temperature forecasts, 3DEnsVar1way fit the observations averaged over the globe similarly to EnKF at shorter lead times and fit the observations more closely than EnKF at longer lead times (e.g., 120-hour). Further decomposition of the RMSEs into NH and SH extratropics and Tropics shows that such differences are mostly from the NH extratropics (Figure 12). In the SH extratropics, 3DEnsVar1way shows consistent improvement over EnKF only for the wind forecasts. No consistent, appreciable difference between EnKF and 3DEnsVar1way is found in the tropics. The relative performance between EnKF and 3DEnsVar1way verified against the ECMWF analyses shows similar results to those results verified against the observations (not shown). As in Whitaker et al. (2008), EnKF performs generally better than GSI3DVar. Since EnKF supplies the ensemble covariance to the hybrid system, the better performance of EnKF relative to GSI3DVar also explains why the hybrid system is better than the GSI 3DVar system.

There were several methodological and implementation differences between EnKF and 3DEnsVar1way: (1) 3DEnsVar1way adopted "model space" covariance localization where the localization was applied on the covariance of the model space vector. In comparison, EnKF adopted "observation space" localization where the localization was applied on the covariance between observation space vector and the model state vector. Campbell et al. (2010) suggested such a difference could lead to performance differences when observations representing integrated measures were assimilated. To alleviate the potential problems associated with the observation space localization when integrated measures were assimilated, EnKF adopted larger vertical localization scales for satellite radiance and surface pressure observations (section 2). (2) EnKF assimilated observations sequentially whereas 3DEnsVar1way assimilated all observations simultaneously. A recent study by Holland and Wang (2013) suggested that the simultaneous/sequential assimilation in combination with different covariance localization methods could lead to performance differences in the ensemble based data assimilation. (3) The ensemble smoother version of EnKF was adopted where effectively the four-dimensional ensemble covariance was utilized during the 6-hour assimilation window. The current 3DVar-based hybrid experiments used the three-dimensional ensemble covariance centered at the middle of the assimilation window and therefore did not account for the temporal dimension of the error covariance. (4) The hybrid adopted two outer loops to treat nonlinearity during the variation minimization whereas the EnKF did not apply an equivalent procedure. (5) The TLNMC was applied during the minimization of the hybrid whereas EnKF did not apply an equivalent procedure.

An in-depth investigation and understanding of the contribution of the aforementioned factors to the performance differences between EnKF and 3DEnsVar1way are needed in future work. A preliminary investigation by comparing experiments of the hybrid with one outer loop and two outer loops [6]showed no appreciable degradation of 3DEnsVar1way with only one outer loop (not shown). An extension of the current hybrid system where the four-dimensional ensemble covariance was utilized during the 6-hour assimilation window (i.e. like the four-dimensional ensemble-variatioanl (4DEnsVar) system in Buehner et al. 2010a) showed appreciable improvement relative to the current three-dimensional hybrid system (to be shown in forthcoming papers). Therefore, the aforementioned factor 3 did not explain the difference between the 3DEnsVar1way and EnKF experiments seen in Figures 11 and 12. Further comparisons were conducted between EnKF and 3DEnsVar1way with and without the use of the TLNMC. Figures 7 and 8 show that the performance of 3DEnsVar1way is degraded when the TLNMC is withheld. Comparing the experiments of 3DEnsVar1way withholding the TLNMC (3DEnsVar1way_nbc) with EnKF shows that after withholding the TLNMC, the EnKF and the 3DEnsVar1way_nbc performed similarly (Fig. 13). This result suggests that the TLNMC implemented in the variational minimization of 3DEnsVar1way (although the DFI is already applied for both 3DEnsVar1way and the EnKF experiments) could be one cause as to the better forecast performance of 3DEnsVar1way than EnKF as seen in Fig. 11. Consistently, Table 2 shows that during the model integration after the DFI is applied the EnKF forecast is less balanced than the 3DEnsVar1way forecast where the TLNMC is implemented.

5. Conclusion and discussion

A GSI 3DVar-based ensemble-variational hybrid data assimilation system was developed. In the hybrid system, flow-dependent ensemble covariances were estimated from an

[6] Note that since this is only 3DVar (not 4DVar), only the non-linear observation operators are re-linearized as part of the outer loop, not the forecast model.

EnKF-generated ensemble and incorporated in the variational minimization by extending the control variables. The performance of the system was investigated with the NCEP GFS model where both the single control forecast and the ensemble forecasts were run at the same, reduced resolution. An 80 member ensemble was utilized. The experiments were conducted over a Northern Hemisphere winter month period assimilating the NCEP operational conventional and satellite data. Various configurations including one-way and two-way couplings, with zero and non-zero weights on the static covariance were compared with a GSI 3DVar experiment. Verification of forecasts showed that the coupled system using these various configurations produced more skillful forecasts than the GSI 3DVar system. For wind and temperature forecasts, the improvement relative to the GSI 3DVar system was larger over the extratropics than the Tropics. For specific humidity forecasts, the improvement in the Tropics was comparable to or larger than the extratropics. It was found that including a non-zero static covariance (Hybrid1way0.5) or using a two-way coupled configuration (3DEnsVar2way) did not improve beyond the one-way coupled system with the use of zero weight on the static covariance (3DEnsVar1way). 3DEnsVar1way produced more skillful wind forecasts than EnKF for the 1-day to 5-day lead times and more skillful temperature forecasts at later lead times (e.g., 120-hour) averaged over the globe. Further decomposition of the RMSEs into NH and SH extratropics and Tropics showed that such differences were mostly from the NH extratropics. In the SH extratropics, the 3DEnsVar1way experiment showed a consistent improvement over the EnKF only for the wind forecasts. No consistent, appreciable difference between EnKF and 3DEnsVar1way was found in the tropics. The spread of the first guess ensemble was evaluated and it was found that the ensemble was under-dispersive in the lower and upper troposphere and was over-dispersive in the middle of the troposphere. Further, the impacts of the tangent-linear normal-mode balance constraint (TLNMC) implemented in the variational minimization were studied. It was found that similar to the impact of TLNMC on the GSI 3DVar system, the balance constraint showed positive impacts on 3DEnsVar1way at longer forecast lead times, especially in the extratropics. The impact of the TLNMC was further diagnosed by using the mean absolute tendency of the surface pressure. For both GSI3DVar and 3DEnsVar1way, applying the TLNMC resulted in more balanced analyses. The analyses generated by GSI3DVar were more balanced than the analyses of 3DEnsVar1way. The EnKF analysis was less balanced than 3DEnsVar1way when the TLNMC was applied for the latter. Further comparisons between EnKF and 3DEnsVar1way with and without the use of the TLNMC suggested that the TLNMC could be one cause as to the better performance of 3DEnsVar1way as compared to EnKF. The convergence rates during the variational minimization were compared between the GSI3DVar and hybrid experiments. For the first outer loop, the hybrid showed a slightly slower convergence rate at early iterations and a slightly faster convergence rate at later iterations than GSI3DVar. For the second outer loop, the hybrid showed a faster convergence than GSI3DVar. The convergence rate was not sensitive whether a 100% or a 50% weight was applied on the ensemble covariance.

In this study, results for the GSI 3DVar-based ensemble-variational hybrid system were presented. An extension of the system where a four-dimensional ensemble is used in the variational minimization (e.g., Buehner et al. 2010ab), including formulations and implementation in the GSI and tests with real observation data will be reported in forthcoming articles. Research on comparing the hybrid under single and dual-resolution configurations and the impact of the static covariance in such configurations are being conducted and will be reported in future papers. Further studies on optimally determining the weights on the static and ensemble covariances are needed (e.g., Bishop and Satterfield 2013)

Acknowledgements. The study was supported by NOAA THOPREX grant NA08OAR4320904, NASA NIP grant NNX10AQ78G and NOAA HFIP grant NA12NWS4680012. Ting Lei is acknowledged for his assistances on plots. The authors thank our many collaborators at EMC, in particular John Derber, Russ Treadon, Bill Lapenta, and Steve Lord and discussion with Tom Hamill.

References

Anderson, J. L., and N. Collins, 2007: Scalable Implementations of Ensemble Filter Algorithms for Data Assimilation. *Journal of Atmospheric and Oceanic Technology*, **24**, 1452-1463.

Barker, D. and co-authors, 2012: The Weather Research and Forecasting (WRF) model's community variational/ensemble data assimilation system: WRFDA. *Bulletin of the American Meteorological Society*, **93**, 831-843.

Bishop, C. H., and D. Hodyss, 2011: Adaptive Ensemble Covariance Localization in Ensemble 4D-VAR State Estimation. *Mon. Wea. Rev*., **139**, 1241-1255.

Bishop, C. H. and E. A. Satterfield, 2013: Hidden error variance theory 1: Exposition and analytic model. *Mon. Wea. Rev*., in press.

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

−, P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments. *Mon. Wea. Rev*., **138**, 1550-1566.

−, --, --, --, and --, 2010b: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations. *Mon. Wea. Rev.*, **138**, 1567-1586.

Buizza, R. , M. Miller, and T. N. Palmer, 1999: Stochastic simulation of model uncertainties. *Quart. J. Roy. Meteor. Soc.,* **125,** 2887–2908.

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

Clayton, A. M., A. C. Lorenc and D. M. Barker, 2012: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q. J. Roy. Meteor. Soc.*, submitted.

Courtier, P., J. N. Thèpaut and A. Hollingsworth 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart.J. Roy. Meteor. Soc*., **120**, 1367- 1387.

Dee, D. P. and Uppala, S., 2009: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Q. J. R. Meteor. Soc.,* **135**, 1830–1841.

Derber, J. C. and W. S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, **126**, 2287-2299.

Etherton, B. J., and C. H. Bishop, 2004: Resilience of Hybrid Ensemble/3DVAR Analysis Schemes to Model Error and Ensemble Covariance Error. *Mon. Wea. Rev.,* **132**, 1065-1080.

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

Gelb, A., 1974: *Applied Optimal Estimation.* MIT Press, 374 pp.

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

−, J. S. Whitaker, M. Fiorino, and S. J. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev*., **139**, 668-688.

Hayden, C. M., and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields: applications to NESDIS operational processing. *J. Applied Meteorology*, **34**, pp. 3–15.

Holland, B., and X. Wang, 2013: Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter . *Q. J. R. Meteo. Soc.*, **139,** 758-770.

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

Kepert, J. D. 2009: Covariance localisation and balance in an Ensemble Kalman Filter. *Q J Roy Meteor Soc* 135: 1157-1176.

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

--, --, --, --, W. Wu and S. Lord, 2009b: Introduction of the GSI into NCEP global data assimilation system. *Wea. Forecasting*, **24**, 1691-1705.

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

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

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

Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079-3102.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott and J. A. York, 2005: Assessing a local ensemble Kalman filter: perfect model experiments with the NCEP global model. *Tellus*, **57A**, 528-545.

Treadon, R. E., H. L. Pan, W. S. Wu, Y. Lin, W. S. Olson, and R. J. Kuligowski, 2002: Global and regional moisture analyses at NCEP. *Proc. ECMWF Workshop on Humidity Analysis*, Reading, United Kindom, ECMWF, 33-47.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140-1158.

Wang, X., C. H. Bishop, and Simon J. Julier, 2004: Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, **132**, 1590-1605.

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

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

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

−, T. M. Hamill, J. S. Whitaker and C. H. Bishop, 2007b: A comparison of hybrid ensemble transform Kalman filter-OI and ensemble square-root filter analysis schemes. *Mon. Wea. Rev.*, **135**, 1055-1076.

−, D. Barker, C. Snyder, T. M. Hamill, 2008a: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116-5131.

−, −, −, −, 2008b: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: real observation experiments. *Mon. Wea. Rev*., **136**, 5132-5147.

−, T. M. Hamill, J. S. Whitaker, C. H. Bishop, 2009: A comparison of the hybrid and EnSRF analysis schemes in the presence of model error due to unresolved scales. *Mon. Wea. Rev*., **137**, 3219-3232.

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

−, −, 2012: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Mon. Wea. Rev.,* **140,** 3078–3089.

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

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

Zhang, M. and F. Zhang, 2012: E4DVar: Coupling an Ensemble Kalman Filter with Four-Dimensional Variational Data Assimilation in a Limited-Area Weather Prediction Model. *Mon. Wea. Rev*., **140**, 587–600.

Zupanski, M., 2005: Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Mon. Wea. Rev.*, 133, 1710–1726.

Table Captions

Table 1. A list of experiments.

Table 2. Averaged hourly absolute surface pressure tendency during the experiment period. The second row is the result before the second hour of the model integration when the DFI is not applied and the third row is the result after the second hour when the DFI is applied.

Figure Captions

Figure 1. Flow charts for the GSI-based EnKF-variational hybrid data assimilation system. (a) is for one-way coupled hybrid system and (b) is for two-way coupled hybrid system.

Figure 2. The root-mean-square fit of the analyses to the rawinsonde observations for temperature (a) and wind (b) as a function of pressure. Solid, dash, dotted, dash-dotted lines are for GSI3DVar, 3DEnsVar1way, 3DEnsVar2way, and Hybrid1way0.5.

Figure 3. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at 24-hour (a,b), 72-hour (c,d) and 120-hour (e, f) forecast lead times for the GSI3DVar, 3DEnsVar1way, 3DEnsVar2way, and Hybrid1way0.5 experiments. Line definition is the same as Figure 2.

Figure 4. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at the 72-hour forecast lead time for the Northern Hemisphere extra-tropics (a,b), tropics (c,d) and Southern Hemisphere extra-tropics (e,f) for the GSI3DVar, 3DEnsVar1way, 3DEnsVar2way, and Hybrid1way0.5 experiments. Line definition is the same as Figure 2.

Figure 5. The root-mean-square fit of the temperature (first column), zonal wind (second column) and specific humidity (third column) forecasts to the ECMWF analyses for Northern Hemisphere extra-tropics (first row), tropics (second row) and Southern Hemisphere extra-

tropics (third row) at the 72-hour lead time for the GSI3DVar, 3DEnsVar1way, 3DEnsVar2way, and Hybrid1way0.5 experiments. Line definition is the same as Figure 2.

Figure 6. Vertical profiles of the square root of the EnKF first guess ensemble variance plus the observation error variance (dotted) and the square root of the innovation variance (solid) for (a) temperature and (b) wind.

Figure 7. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at 24-hour (a,b), 72-hour (c,d) and 120-hour (e, f) forecast lead times. Thick solid, thick dash, thin solid and thin dash lines are for the GSI3DVar, GSI3DVar_nbc, 3DEnsVar1way, and 3DEnsVar1way_nbc experiments respectively.

Figure 8. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at the 120-hour forecast lead time for the Northern Hemisphere extra-tropics (a,b), tropics (c,d) and Southern Hemisphere extra-tropics (e,f) for the GSI3DVar, GSI3DVar_nbc, 3DEnsVar1way, and 3DEnsVar1way_nbc experiments. Line definition is the same as Figure 7.

Figure 9. The mean absolute surface pressure tendency calculated using GFS outputs every 2 minutes up to 9 hours for forecasts initialized from the GSI3DVar, GSI3DVar_nbc, 3DEnsVar1way, 3DEnsVar1way_nbc and EnKF experiments.

Figure 10. Averaged ratios of gradient norms as a function of the iterations in the first and second outer loops during the variational minimization of the GSI3DVar, 3DEnsVar1way and Hybrid1way0.5 experiments.

Figure 11. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at 24-hour (a,b), 72-hour (c,d) and 120-hour (e, f) forecast lead times. Solid and dash lines are for the 3DEnsVar1way and the EnKF experiments.

Figure 12. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at the 120-hour forecast lead time for the Northern Hemisphere extra-tropics (a,b), tropics (c,d) and Southern Hemisphere extra-tropics (e,f) for the 3DEnsVar1way and the EnKF experiments. Line definition is the same as Figure 11.

Figure 13. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (a) and wind (b) as a function of pressure at the 120-hour forecast lead time. Solid and dash lines are for 3DEnsVar1way_nbc and the EnKF experiments.

Experiment	Description
GSI3DVar	The GSI 3DVar experiment
3DEnsVar1way	The one-way coupled ensemble-variational DA experiment with 0% weight on the static covariance and 100% weight on the ensemble covariance
Hybrid1way0.5	The one-way coupled ensemble-variational DA experiment with 50% weight on the static covariance and 50% weight on the ensemble covariance
3DEnsVar2way	The two-way coupled ensemble-variational DA experiment with 0% weight on the static covariance and 100% weight on the ensemble covariance
EnKF	The EnKF experiment
GSI3DVar_nbc	Same as "GSI3DVar" except without the use of the tangent linear normal mode balance constraint (TLNMC)
3DEnsVar1way_nbc	Same as "3DEnsVar1way" except without the use of the tangent linear normal mode balance constraint (TLNMC)

Table 1. A list of experiments.

Unit: hPa hr^{-1}	**GSI3DVar**	**GSI3DVar_nbc**	**3DEnsVar**	**3DEnsVar_nbc**	**EnKF**
Before DFI	0.548	0.963	0.581	1.071	0.968
After DFI	0.510	0.536	0.539	0.573	0.546

Table 2. Averaged hourly absolute surface pressure tendency during the experiment period. The second row is the result before the second hour of the model integration when the DFI is not applied and the third row is the result after the second hour when the DFI is applied.

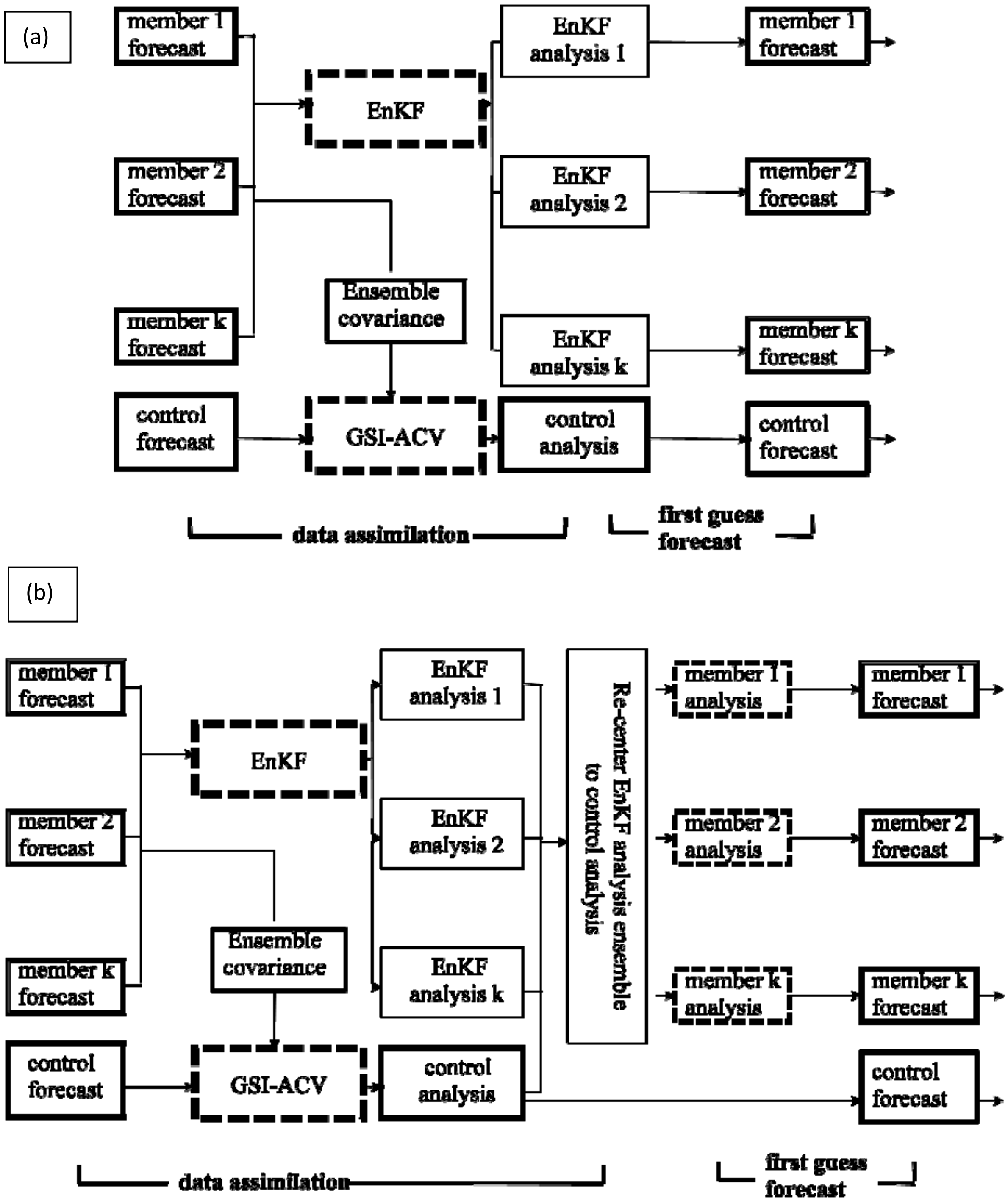

Figure 1. Flow charts for the GSI-based EnKF-variational hybrid data assimilation system. (a) is for one-way coupled hybrid system and (b) is for two-way coupled hybrid system.

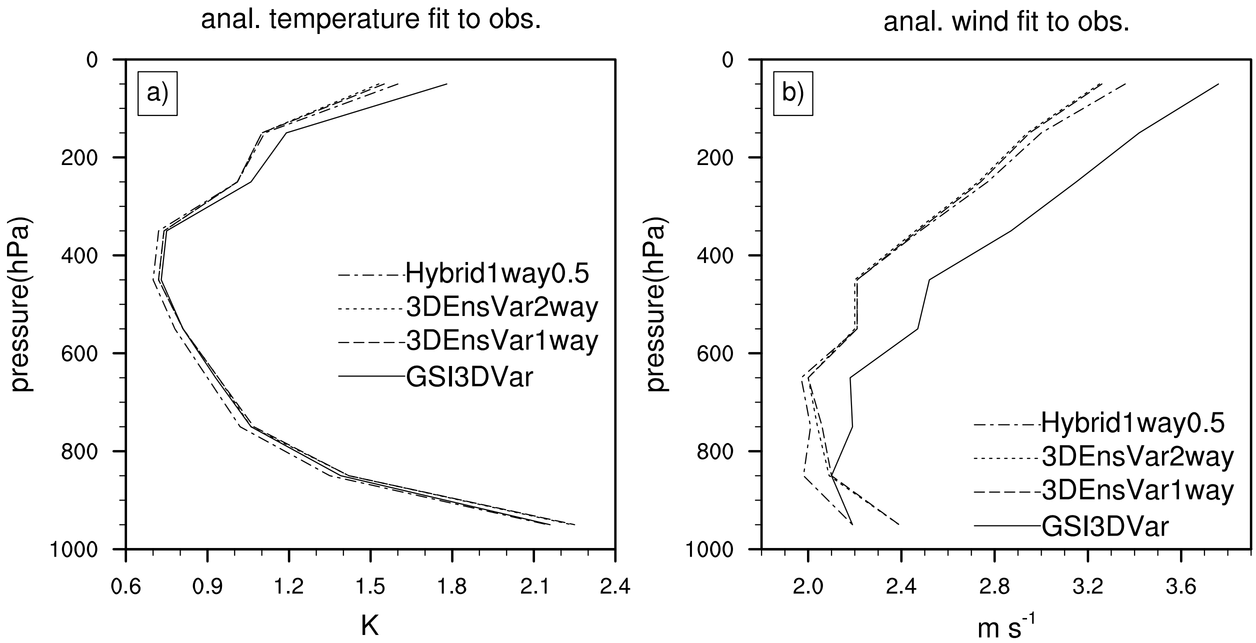

Figure 2. The root-mean-square fit of the analyses to the rawinsonde observations for temperature (a) and wind (b) as a function of pressure. Solid, dash, dotted, dash-dotted lines are for GSI3DVar, 3DEnsVar1way, 3DEnsVar2way, and Hybrid1way0.5.

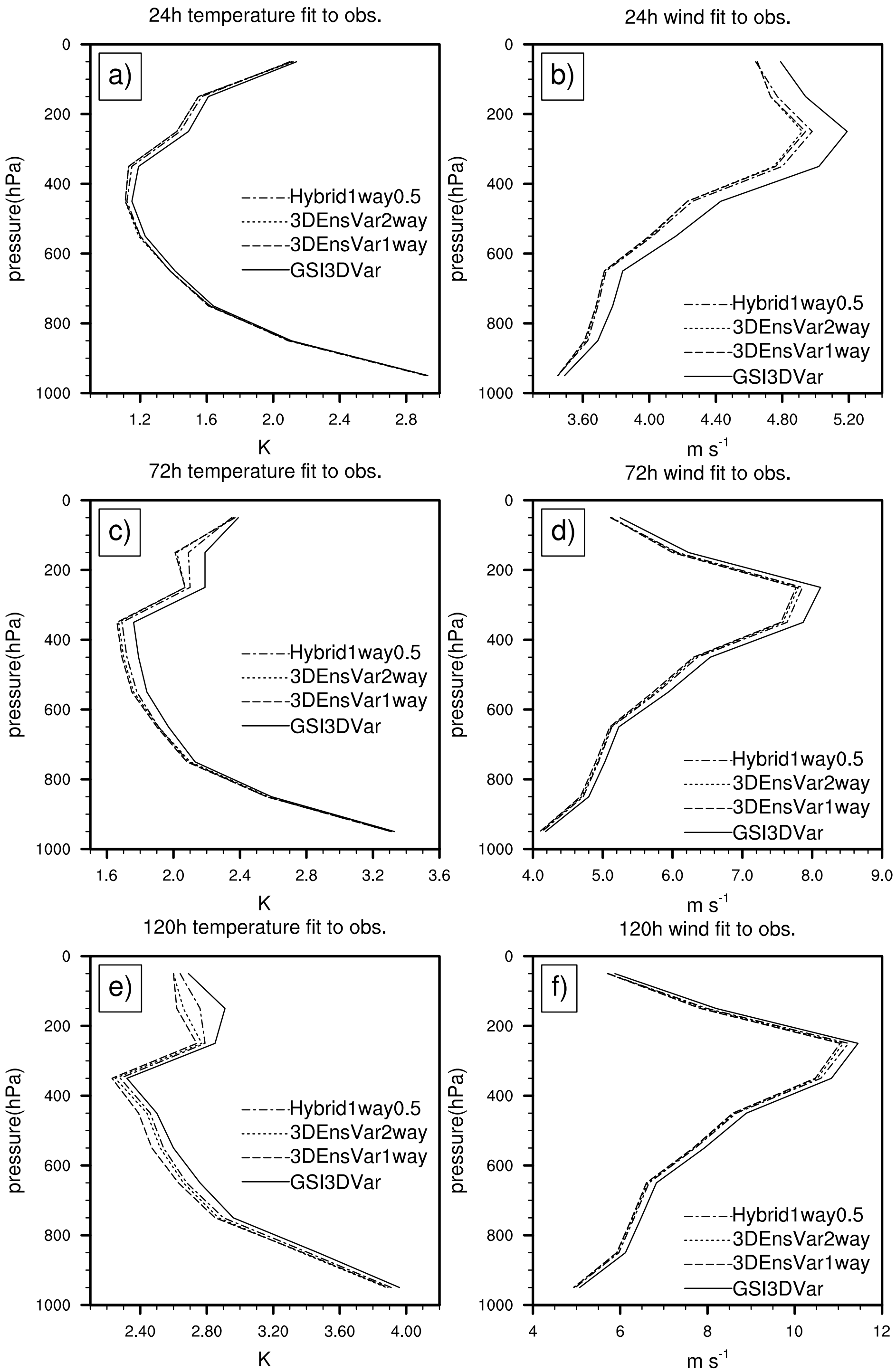

Figure 3. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at 24-hour (a,b), 72-hour (c,d) and 120-hour (e, f) forecast lead times for the GSI3DVar, 3DEnsVar1way, 3DEnsVar2way, and Hybrid1way0.5 experiments. Line definition is the same as Figure 2.

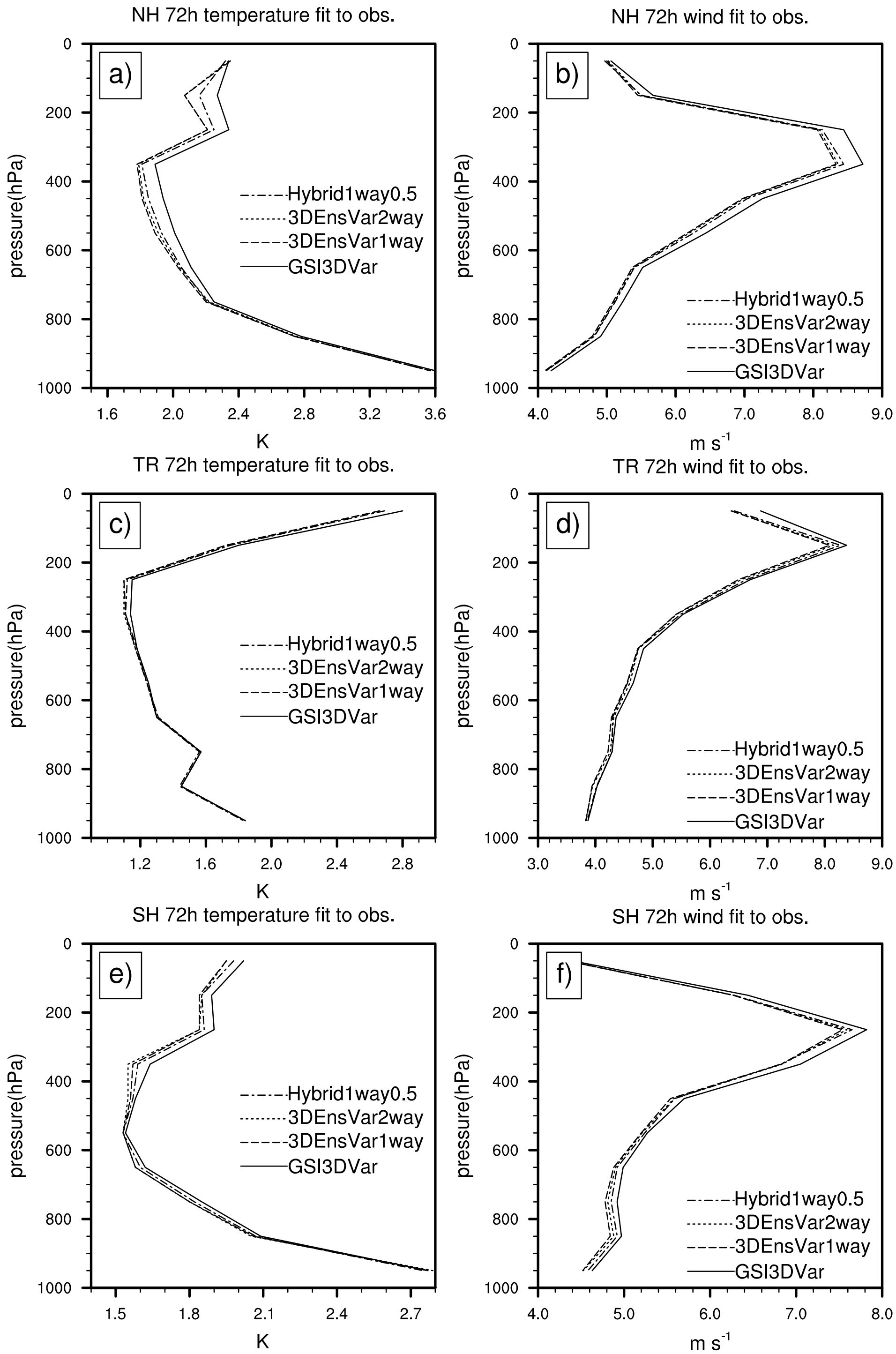

Figure 4. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at the 72-hour forecast lead time for the Northern Hemisphere extra-tropics (a,b), tropics (c,d) and Southern Hemisphere extra-tropics (e,f) for the GSI3DVar, 3DEnsVar1way, 3DEnsVar2way, and Hybrid1way0.5 experiments. Line definition is the same as Figure 2.

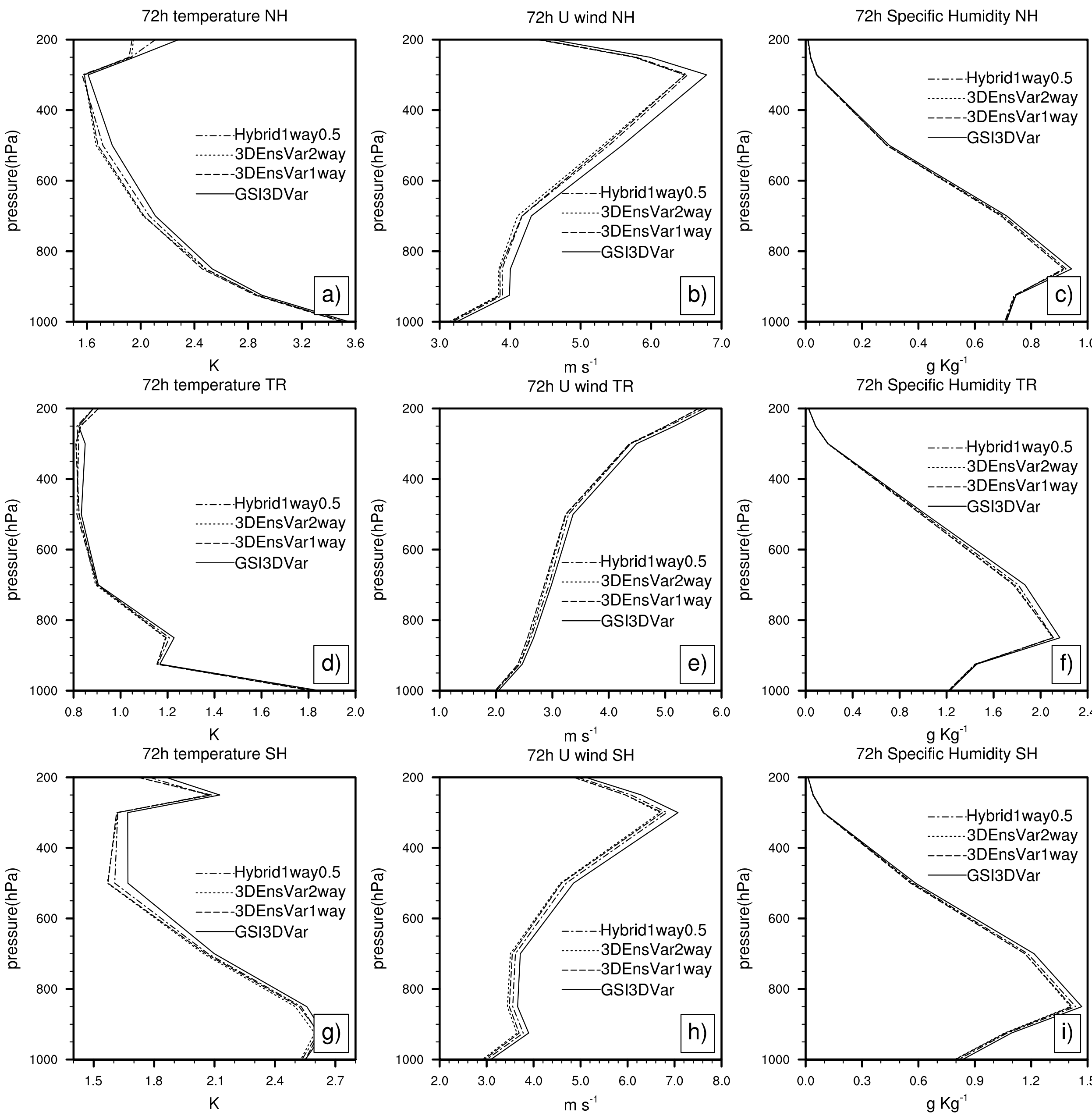

Figure 5. The root-mean-square fit of the temperature (first column), zonal wind (second column) and specific humidity (third column) forecasts to the ECMWF analyses for Northern Hemisphere extra-tropics (first row), tropics (second row) and Southern Hemisphere extra-tropics (third row) at the 72-hour lead time for the GSI3DVar, 3DEnsVar1way, 3DEnsVar2way, and Hybrid1way0.5 experiments. Line definition is the same as Figure 2.

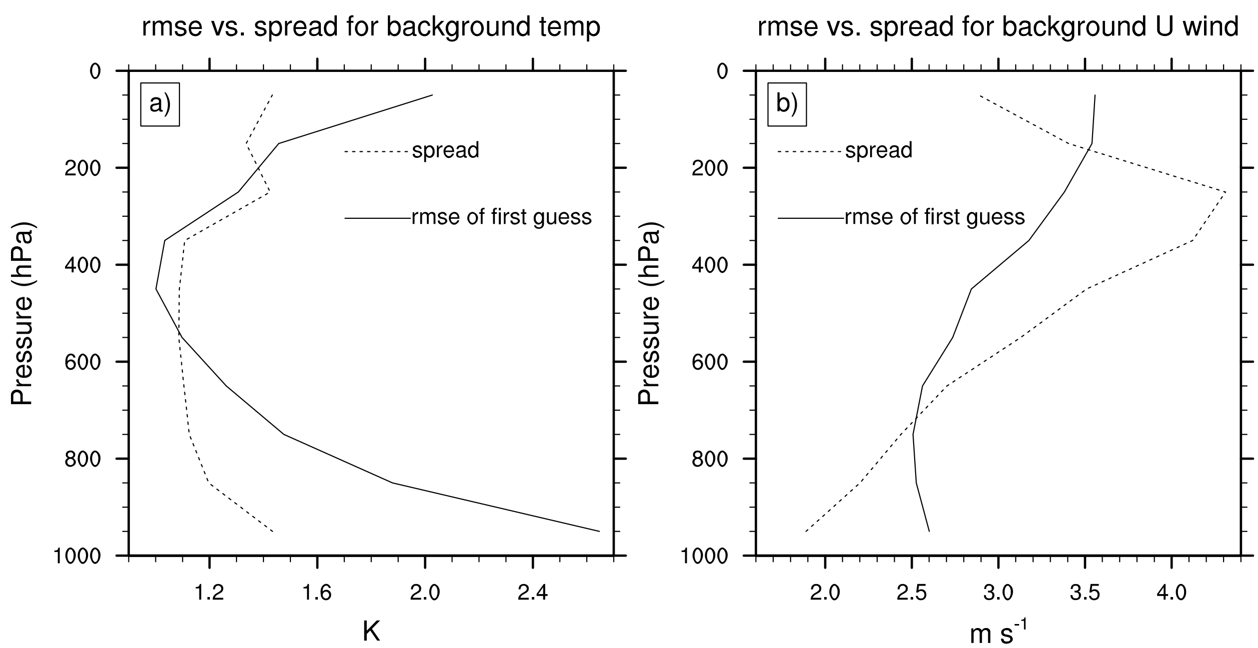

Figure 6. Vertical profiles of the square root of the EnKF first guess ensemble variance plus the observation error variance (dotted) and the square root of the innovation variance (solid) for (a) temperature and (b) wind.

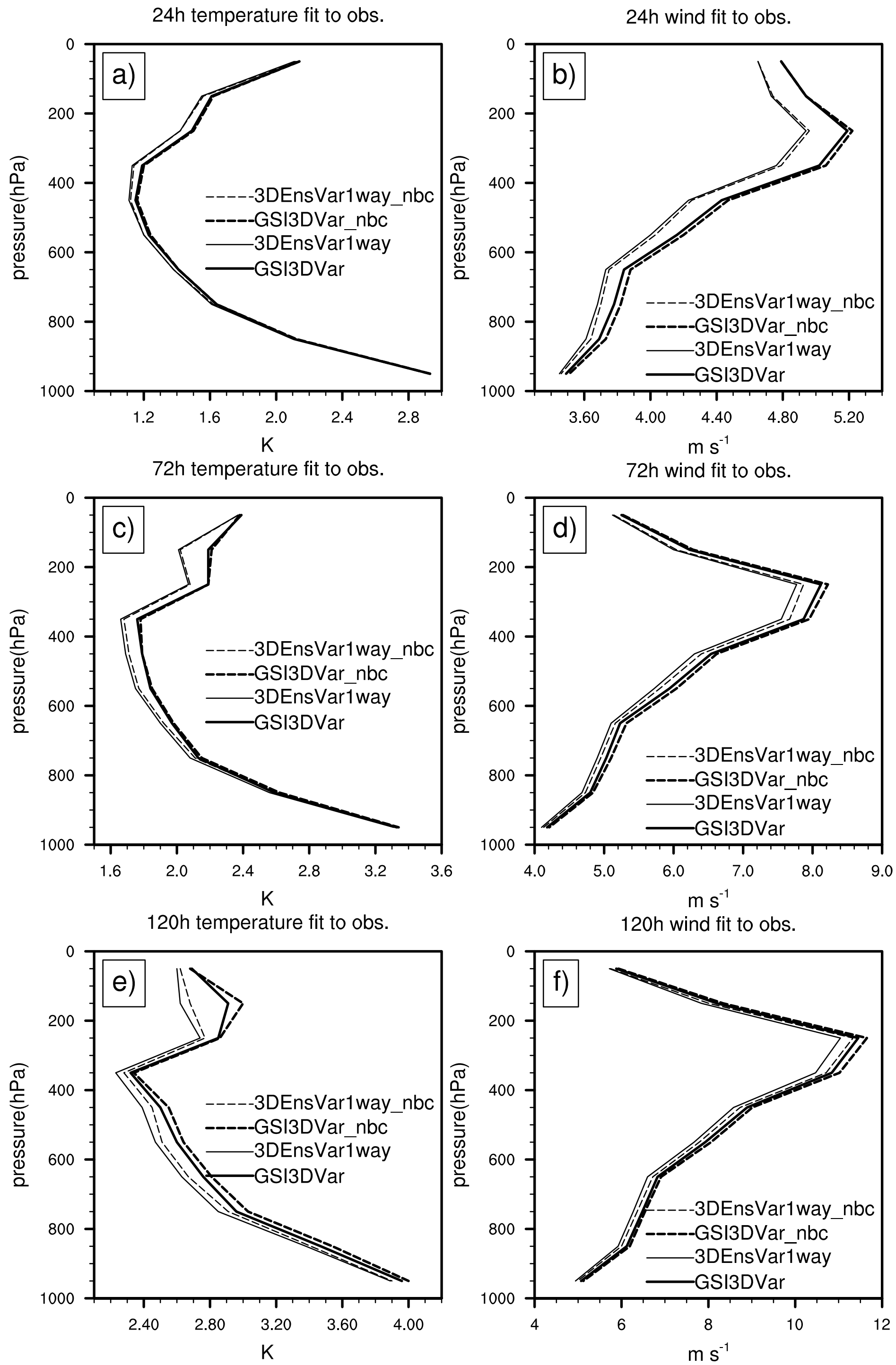

Figure 7. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at 24-hour (a,b), 72-hour (c,d) and 120-hour (e, f) forecast lead times. Thick solid, thick dash, thin solid and thin dash lines are for the GSI3DVar, GSI3DVar_nbc, 3DEnsVar1way, and 3DEnsVar1way_nbc experiments respectively.

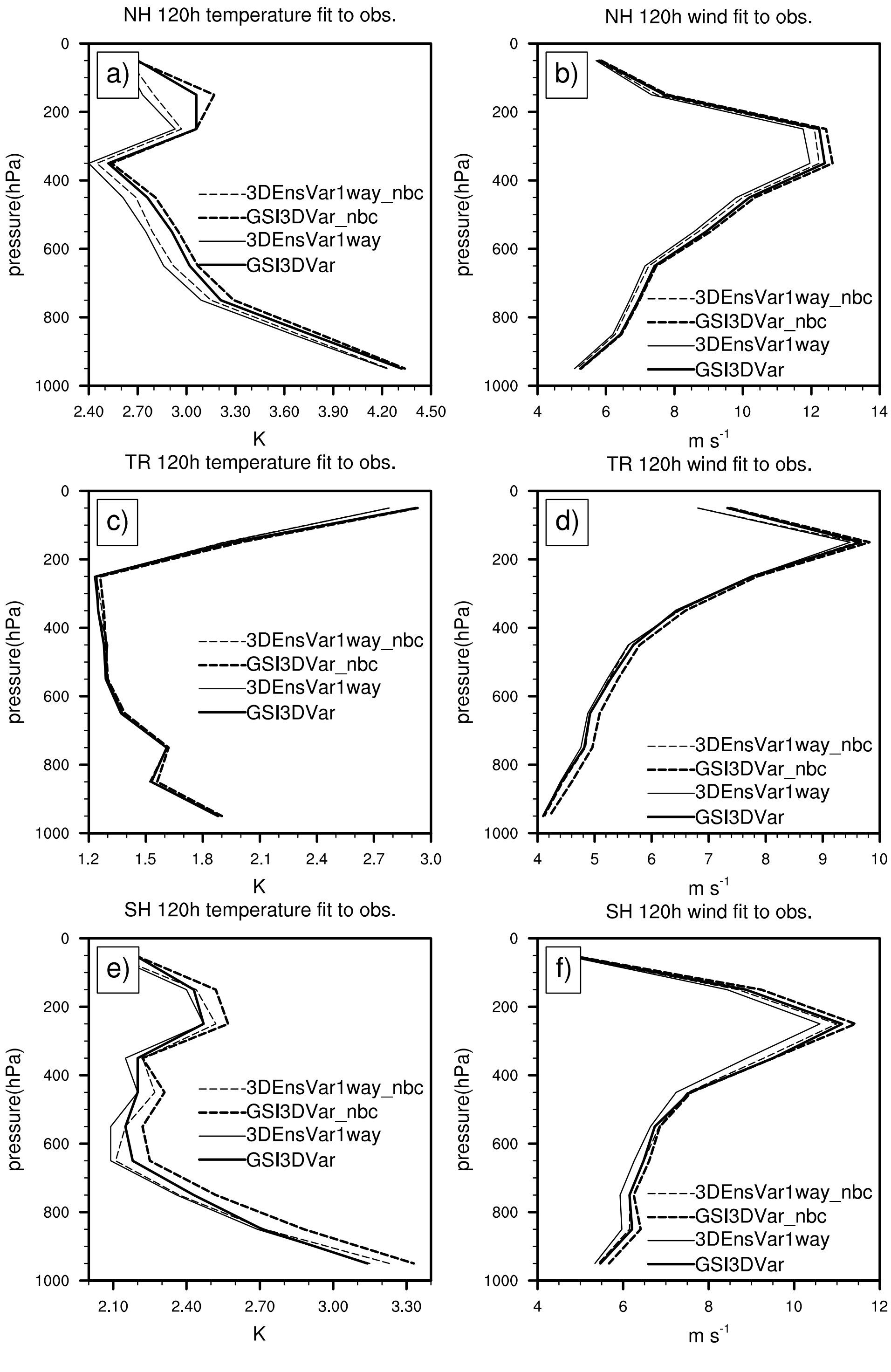

Figure 8. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at the 120-hour forecast lead time for the Northern Hemisphere extra-tropics (a,b), tropics (c,d) and Southern Hemisphere extra-tropics (e,f) for the GSI3DVar, GSI3DVar_nbc, 3DEnsVar1way, and 3DEnsVar1way_nbc experiments. Line definition is the same as Figure 7.

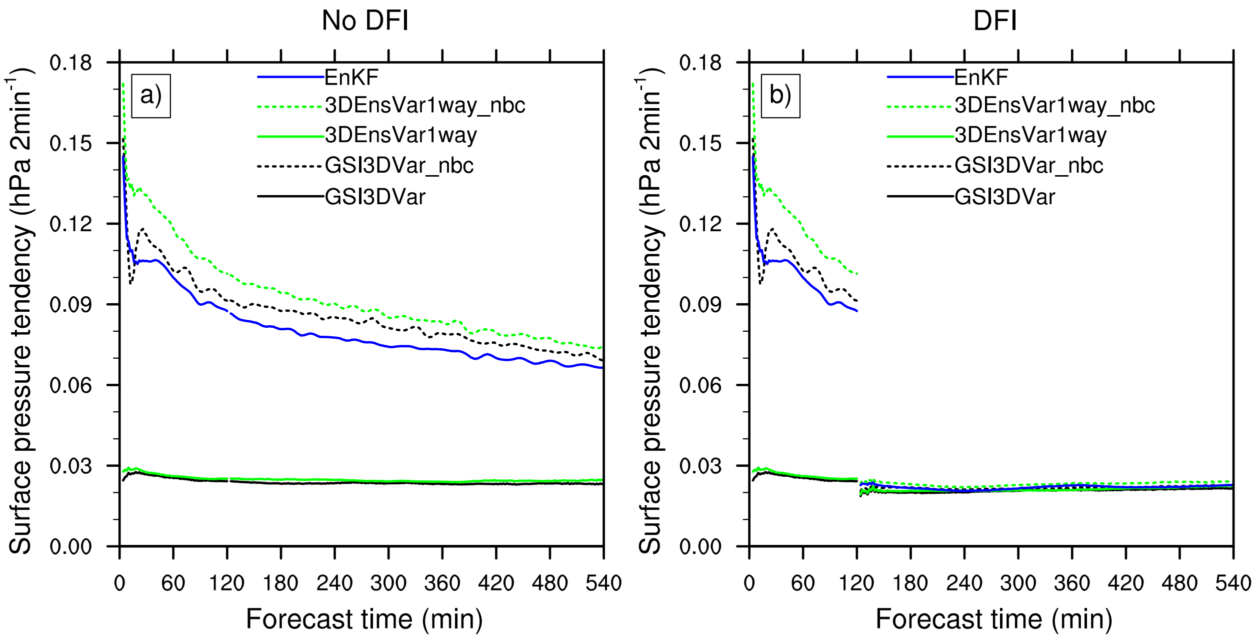

Figure 9. The mean absolute surface pressure tendency calculated using GFS outputs every 2 minutes up to 9 hours for forecasts initialized from the GSI3DVar, GSI3DVar_nbc, 3DEnsVar1way, 3DEnsVar1way_nbc and EnKF experiments.

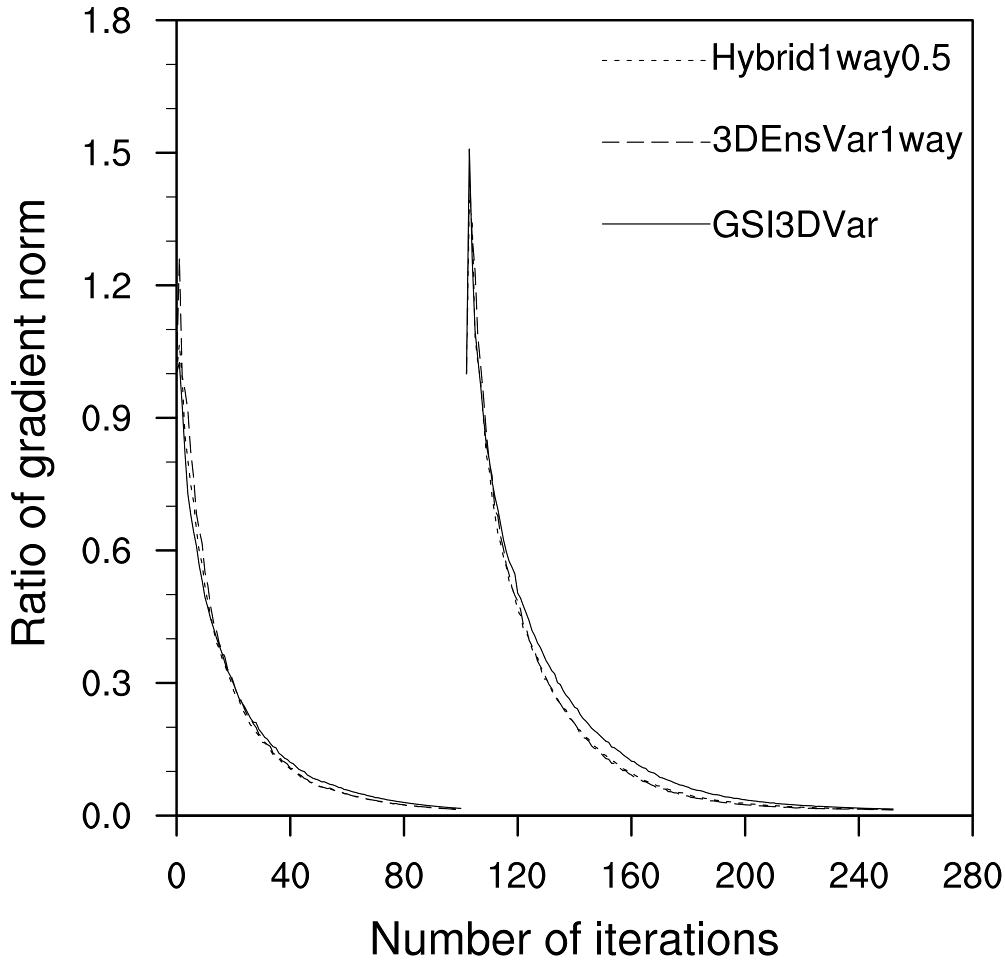

Figure 10. Averaged ratios of gradient norms as a function of the iterations in the first and second outer loops during the variational minimization of the GSI3DVar, 3DEnsVar1way and Hybrid1way0.5 experiments.

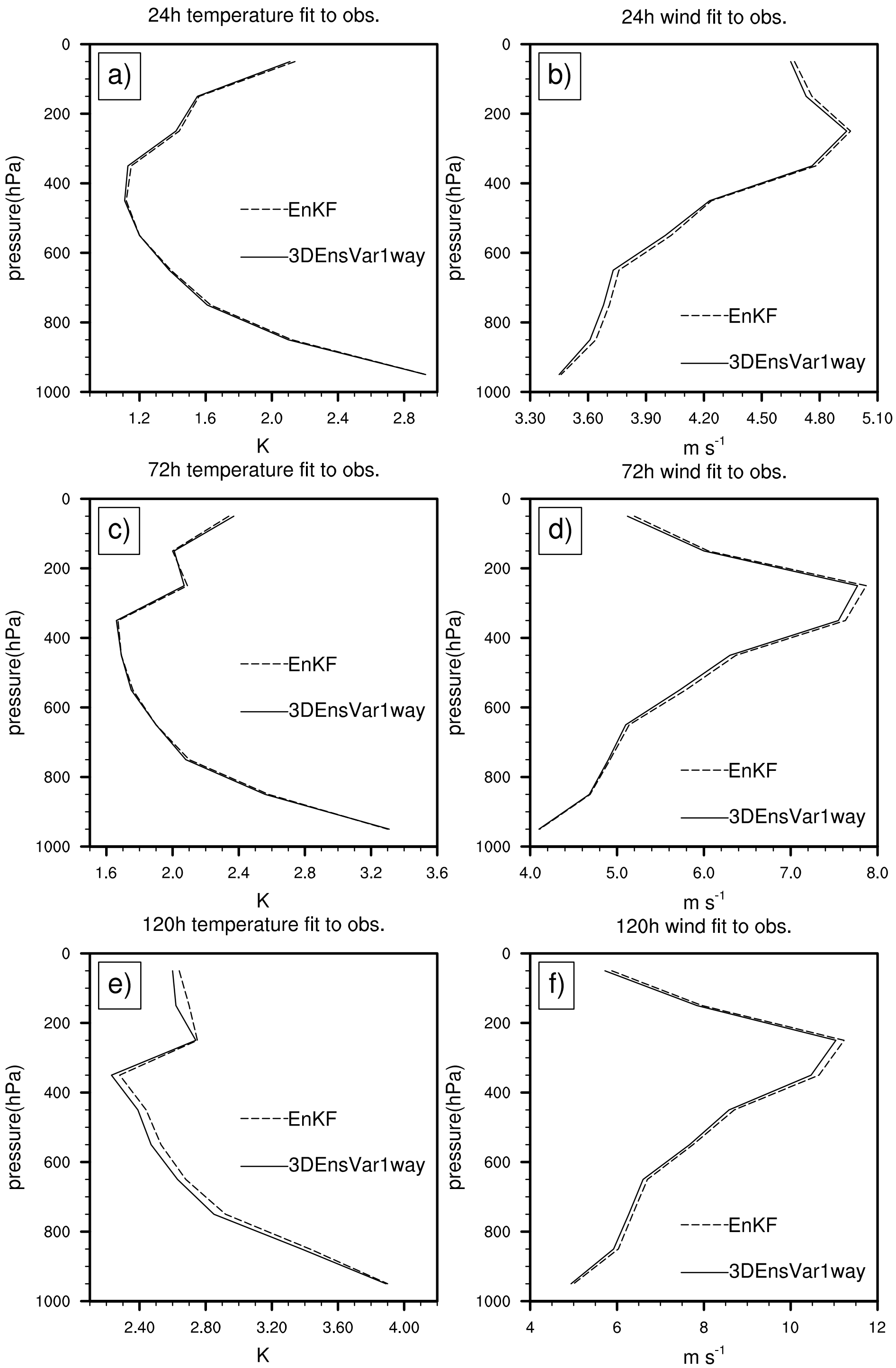

Figure 11. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at 24-hour (a,b), 72-hour (c,d) and 120-hour (e, f) forecast lead times. Solid and dash lines are for the 3DEnsVar1way and the EnKF experiments.

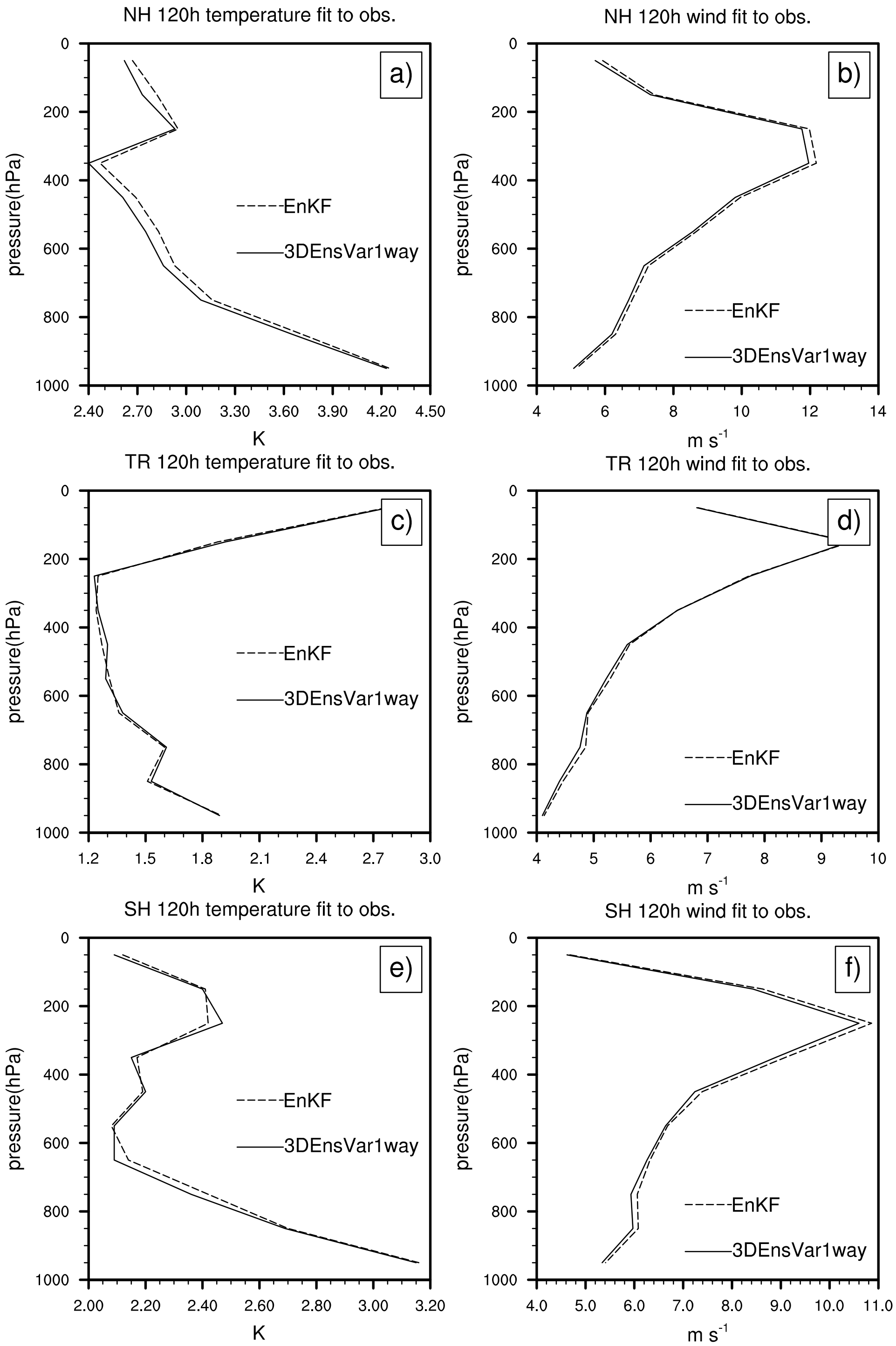

Figure 12. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (left column) and wind (right column) as a function of pressure at the 120-hour forecast lead time for the Northern Hemisphere extra-tropics (a,b), tropics (c,d) and Southern Hemisphere extra-tropics (e,f) for the 3DEnsVar1way and the EnKF experiments. Line definition is the same as Figure 11.

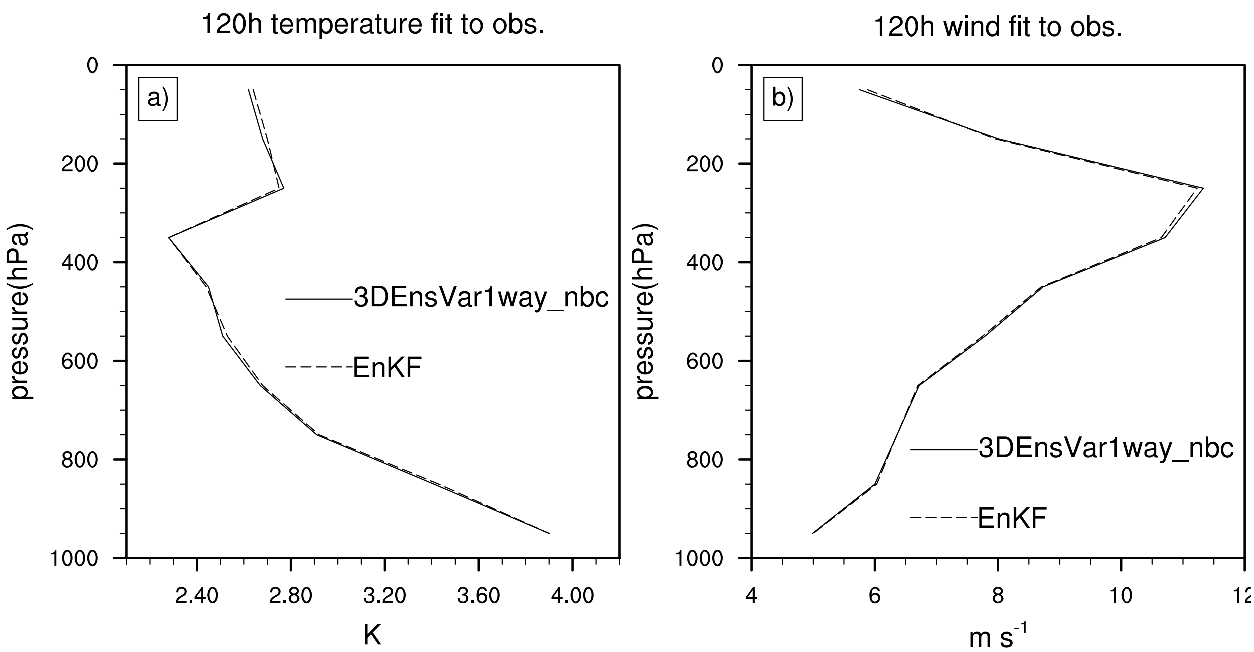

Figure 13. The root-mean-square fit of the forecasts to the rawinsonde observations for temperature (a) and wind (b) as a function of pressure at the 120-hour forecast lead time. Solid and dash lines are for 3DEnsVar1way_nbc and the EnKF experiments.