GSI-based four dimensional ensemble-variational (4DEnsVar) data assimilation: formulation and single resolution experiments with real data for NCEP Global Forecast System

Xuguang Wang and Ting Lei

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

September, 2013

Submitted to *Monthly Weather Review*

Revised April 2014

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

A four-dimensional (4D) ensemble-variational data assimilation (DA) system (4DEnsVar) was developed, building upon the infrastructure of the gridpoint statistical interpolation (GSI) based hybrid DA system. 4DEnsVar used ensemble perturbations valid at multiple time levels throughout the DA window to estimate 4D error covariances during the variational minimization avoiding the tangent linear and adjoint of the forecast model. The formulation of its implementation in GSI was described. The performance of the system was investigated by evaluating the global forecasts and hurricane track forecasts produced by NCEP GFS during a 5-week summer period assimilating operational conventional and satellite data. The newly developed system was used to address a few questions regarding 4DEnsVar. 4DEnsVar in general improved upon its 3D counterpart, 3DEnsVar. At short lead times, the improvement over Northern extratropics (NH) was similar to that over Southern extratropics (SH). At longer lead times, 4DEnsVar showed more improvement in SH than in NH. 4DEnsVar showed less impact over Tropics (TR). The track forecasts of 16 tropical cyclones initialized by 4DEnsVar were more accurate than 3DEnsVar after 1-day forecast lead times. The analysis generated by 4DEnsVar was more balanced than 3DEnsVar. Case studies showed that increments from 4DEnsVar using more frequent ensemble perturbations approximated the increments from direct, nonlinear model propagation better than using less frequent ensemble perturbations. Consistently, the performance of 4DEnsVar including both the forecast accuracy and the balances of analyses was in general degraded when less frequent ensemble perturbations were used. The tangent linear normal mode constraint had positive impact for global forecast but negative impact for TC track forecasts.

1. Introduction

Data assimilation systems that bridge the gap between two traditionally parallel variational and ensemble-based methods have gained increasing interests recently in both the research and operational numerical weather prediction (NWP) communities. Instead of using a typically static covariance, the background error covariances in the variational system (Var) are estimated flow-dependently from an ensemble of background states. Such ensemble background states are typically produced by the ensemble Kalman filter (EnKF) or its simplified variants. Early studies on such coupled data assimilation include proposing, testing and demonstrating new algorithms using simple models and simulated observations (e.g., Hamill and Snyder 2000; Lorenc 2003; Etherton and Bishop 2004; Zupanski 2005; Wang et al. 2007ab, 2009; Liu et al. 2008; Wang 2010; Wang et al. 2010). More recently, this method has been implemented and successfully tested for both regional (e.g., Wang et al. 2008ab; Wang 2011; Zhang and Zhang 2012; Barker et al. 2012; Li et al. 2012) and global NWP models (e.g., Buehner 2005; Buehner et al. 2010ab, Bishop and Hodyss 2011, Clayton et al. 2013, Wang et al. 2013, Buehner et al. 2013). These studies suggest that the coupled ensemble-variational system can leverage the strengths of both a standalone EnKF and a standalone Var system, producing an analysis that can be better than either system. The potential advantages of the coupled ensemble-variational (EnsVar) data assimilation as compared to a standalone Var and EnKF are discussed in Wang (2010). Briefly, compared to a stand-alone Var, EnsVar can benefit from flow-dependent ensemble covariance as EnKF. Compared to a stand-alone EnKF, EnsVar can be more robust for small ensemble size or large model errors (e.g., Hamill and Snyder 2000, Etherton and Bishop 2004, Wang et al. 2007b, 2009, Buehner et al. 2010b), benefit from dynamic constraints during the variational minimization (e.g., Wang et al. 2013), and take advantage of the established capabilities such as the variational data quality control and outer loops to treat nonlinearity in Var. Although in theory EnKF can adopt model space localization, to save computational costs, EnKF often adopts serial observation or batch observation processing algorithms (Houtekamer and Mitchell 2001) and the covariance localization is often conducted in observation space. In EnsVar, ensemble covariance localization is often conducted in model space rather than observation space which may be more appropriate for observations without explicit position (e.g., Campbell et al. 2010). Motivated by these earlier studies, several operational NWP centers in the world have implemented or are implementing the ensemble-variational data assimilation system operationally (e.g., Buehner et al. 2010ab, Clayton et al. 2013, Wang et al. 2013, Kuhl et al. 2013).

Although the ensemble-variational data assimilation system documented in these studies share the same spirit to incorporate flow-dependent ensemble covariance into variational systems, the specific implementation can be different in several aspects. Such differences include if the ensemble covariance is incorporated into a three-dimensional variational system (3DVar) or a four-dimensional variational system (4DVar); if the background covariance is fully or partially replaced by the ensemble covariance and if the tangent linear adjoint of the forecast model was used in the four-dimensional variational minimization. Appendix A summarizes various flavors. In this study the abbreviation of the names of various ensemble-variational data assimilation experiments follow those defined in Appendix A.

A 3DVar-based ensemble-variational (3DEnsVar) hybrid data assimilation system was recently developed based on the Gridpoint Statistical Interpolation (GSI) data assimilation system operational at the National Centers for Environmental Prediction (NCEP), and was first tested for the Global Forecast System (GFS). It was found that the new 3DEnsVar hybrid system produced more accurate forecasts than the operational GSI 3DVar system for both the general global forecasts (Wang et al. 2013) and the hurricane forecasts (Hamill et al. 2011). Wang et al. (2013) also found that GSI-based 3DEnsVar without inclusion of the static covariance outperformed GSI based EnKF due to the use of tangent linear normal mode constraint in the variational system. The 3DEnsVar hybrid system was implemented operationally for global numerical prediction at NCEP beginning May 2012.

The current GSI based 3DEnsVar and 3DEnsVar hybrid did not account for the temporal evolution of the error covariance within the assimilation window. A GSI based four–dimensional variational (4DVar) data assimilation (DA) system where the innovation is propagated in time using a tangent linear and adjoint (TLA) of the forecast model is being developed. However, efforts are needed to improve the computational efficiency of the TLA model before systematic tests can be conducted (Rancic et al. 2012). In this study, an alternative method to account for the temporal evolution of the error covariance within the GSI system was implemented. In this method, the ensemble perturbations valid at multiple time levels within the DA window are used during the variational minimization. Effectively, the four-dimensional (4D) background error covariance was estimated by the ensembles, avoiding the need of the TLA of the forecast model. Hereafter, the method is referred as "4D-Ensemble-Var (4DEnsVar)".

Incorporating the ensemble perturbations spanning the DA window in the variational framework to avoid the TLA model have been proposed and implemented in different ways in early studies. Qiu et al. (2007), Tian et al. (2008) and Wang et al. (2010) proposed methods to reduce the dimension of the problem by reducing the ensemble perturbations produced by the Monte Carlo methods or historical samples to a set of base vectors during the variational minimization. Liu et al. (2008, 2009) implemented the method in a one-dimensional shallow water model by directly ingesting the ensemble perturbations and then increased the size of the ensemble perturbation by applying the covariance localization matrix outside the variational minimization to alleviate the sampling error issue associated with a limited ensemble size. Buehner et al. (2010ab) implemented the method to the Meteorological Service of Canada's operational data assimilation system where covariance localization was adopted within the variational minimization following Buehner (2005), and systematically compared it with their EnKF and 4DVar. Bishop and Hodyss (2011) implemented 4DEnsVar to the Naval Research Laboratory (NRL) 4DVar system called Atmospheric Variational Data Assimilation System–Accelerated Representer (NAVDAS-AR; Xu et al. 2005) and proposed and tested an adaptive covariance localization method in the context of 4DEnsVar using a single case study. It is noted that the model-space 4DEnsVar algorithm is a natural extension of earlier proposed 3DEnsVar (Lorenc 2003; Buehner 2005; Wang et al. 2007a; Wang et al. 2008a, Wang 2010). One critical component in these algorithms is to incorporate ensemble covariances in the variational minimization through augmenting the control variables. In 3DEnsVar the ensemble perturbation at a single time level, e.g., the center of the assimilation window is used whereas in 4DEnsVar, ensemble perturbations at multiple time levels spanning the assimilation window are used.

4DEnsVar implemented within Meteorological Service of Canada's 4DVar system (Buehner et al. 2010a) takes the model-space based minimization formula with the variational minimization pre-conditioned upon the square root of the background error covariance. 4DEnsVar implemented within NAVDAS-AR is based on the observation-space minimization formula (Bishop and Hodyss 2011). Different from these systems, operational GSI minimization is preconditioned upon the full background error covariance matrix (Derber and Rosati 1989). Therefore the formulations of implementation of 4DEnsVar where the minimization is preconditioned upon the full background error covariance is described in this paper. The performance of the newly developed GSI-based 4DEnsVar system is evaluated by comparing with GSI-based 3DVar and 3DEnsVar . In addition to examining the performance of the system for general global forecasts, the performance of the 4DEnsVar system is studied for hurricane track forecasts for the first time. Using the newly developed GSI-based 4DEnsVar system, a few other questions were investigated. As far as the authors are aware, these questions have not been documented in previously published studies on 4DEnsVar in real data context. In 4DEnsVar, temporal evolution of the error covariance is approximated by the covariances of ensemble perturbations at discrete times. How is the performance of 4DEnsVar dependent on the temporal resolution of or the number of time levels of ensemble perturbations? In 4DEnsVar, the temporal propagation through covariance of ensemble perturbations contains linear assumption. How is the linear approximation compared to the full nonlinear model propagation? Will using 4D ensemble covariances to fit the model trajectory to observations distributed within a finite assimilation window improve the balance of the analysis and how is the balance of the analysis dependent on the temporal resolution of the ensemble perturbations? 4DEnsVar is implemented such that the tangent linear normal mode constraint (TLNMC; Kleist et al. 2009) within the GSI is allowed. What is the impact of such balance constraint on the 4DEnsVar analysis and forecast and how is that dependent on different types of forecasts such as the general global forecast or hurricane track forecasts? How does including multiple time levels of perturbations impact the convergence rates of the minimization? These questions will be addressed in a real data context where operational observations from NCEP are assimilated.

The resolution of the operational implementation of the 3DEnsVar hybrid at NCEP is T254 (triangular truncation at total wavenumber 254) for the ensemble and T574 for the variational analysis. Lei and Wang (2014) found that with this dual-resolution configuration, including the static covariance (i.e., 3DEnsVar hybrid defined in Table A1 of this paper) significantly improved the performance compared to without including the static covariance. Therefore in the operational implementation, 3DEnsVar hybrid was adopted. Here we present the evaluation results and address the aforementioned questions using experiments conducted at a reduced spectral resolution of T190 for both the ensemble and the variational analyses (hereafter single resolution experiments). Wang et al. (2013) compared the GSI-based 3DEnsVar and 3DEnsVar hybrid using the same single resolution configuration and an 80-member ensemble. It was found that the inclusion of the static covariance component in the background-error covariance did not improve the forecast skills beyond using the full ensemble covariance as the background-error covariance. Given this result and that the current study represents a first step of testing the newly extended system using real data, this study focuses on the impact of 4D extension of the ensemble covariance in a single resolution configuration and without involving the static covariance. This single-resolution configuration is different from Buehner et al. (2010b) where the ensemble was run at a reduced resolution as compared to the variational analysis (termed as dual resolution experiments). One method to include the static covariance in 4DEnsVar without involving the TLA is proposed in Buehner et al. (2013). In this method, the same static background error covariance is used at all time levels (e.g., Buehner et al. 2013). Investigation of the impact of including the static covariance using such method in GSI-based 4DEnsVar (i.e., 4DEnsVar hybrid) is ongoing and will be documented in future papers (D. Kleist, personal communication, 2013).

The rest of the paper is organized as followed. Section 2 and Appendix B describe the formulations and implementation of 4DEnsVar within the GSI. Section 3 describes the design of experiments. Section 4 discusses the experiment results and section 5 concludes the paper.

2. GSI-based 4DEnsVar formulation and implementation

Specific formulation of implementing 4DEnsVar within GSI is given in this section and Appendix B. Different from other variational systems, the minimization in the operational GSI is preconditioned upon the full background error covariance matrix. Wang (2010) describes the mathematical details on how the ensemble covariance is incorporated in the GSI 3DVar through the use of the augmented control vectors (ACV) using such preconditioning method. As shown in Wang et al. 2007a, effectively, the static covariance in GSI 3DVar was replaced by and linearly combined with the ensemble covariance in 3DEnsVar and 3DEnsVar hybrid respectively. As discussed in the introduction, the current study focuses on the impact of 4D extension of the ensemble covariance without involving the static covariance. Therefore the formula shown in this section and in Appendix B excludes the static covariance. Formulations including the static covariance will follow similar lines. Below describes the formulas of 4DEnsVar following the notation of Wang (2010). Further mathematical details of implementing 4DEnsVar in the GSI variational minimization framework are provided in Appendix B.

In 4DEnsVar, the analysis increment $\mathbf{x}'_t$ at time level *t* is defined as

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

$(\mathbf{x}_k^e)_t$ is the *k*th ensemble perturbation at time *t* 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 ° denotes the Schur product. The four-dimensional analysis increment $\mathbf{x}'_t$ is obtained by minimizing the following cost function

$$J(\mathbf{a}) = \frac{1}{2}(\mathbf{a})^T\mathbf{A}^{-1}(\mathbf{a}) + \frac{1}{2}\sum_{t=1}^{L}(\mathbf{y}_t^{o\prime} - \mathbf{H}_t\mathbf{x}'_t)^T\mathbf{R}_t^{-1}(\mathbf{y}_t^{o\prime} - \mathbf{H}_t\mathbf{x}'_t) \ (2).$$

Comparing equations (1) and (2) with the 3DEnsVar formula in Wang 2010, 4DEnsVar is a natural, temporal extension of its 3D counterpart. In 3DEnsVar, ensemble perturbations at a single time, the center of the assimilation window was incorporated. In comparison, ensemble perturbations at multiple time levels $t = 1 \dots L$ within the assimilation window were incorporated in 4DEnsVar.

The first term in Eq. (2) is associated with the augmented control vector, $\mathbf{a}$, which is formed by concatenating K vectors $\mathbf{a}_k$, $k = 1, \cdots K$. These augmented control vectors are constrained by a block-diagonal matrix $\mathbf{A}$, which defines the horizontal and vertical covariance localization on the ensemble covariance. In the current implementation, each $\mathbf{a}_k$, $k = 1, \cdots K$, is a three-dimensional field co-located at the model grid points. Each $\mathbf{a}_k$ varies in both the horizontal and vertical direction so that spatial localization is applied both horizontally and vertically. The same three-dimensional fields of $\mathbf{a}_k$ are applied for all variables and all time levels. Therefore in the current implementation of 4DEnsVar, the number of augmented control variables used is the same as that used in 3DEnsVar. For the implementation of 4DEnsVar with GFS, $(\mathbf{x}_k^e)_t$ of eq. (1) on which $\mathbf{a}_k$ are applied include ensemble perturbations of surface pressure, wind, virtual temperature, relative humidity, cloud water mixing ratio and ozone mixing ratio at different time t. The covariance localization in 4DEnsVar follows the same method adopted in GSI-based 3DEnsVar described in Wang et al. (2013). The vertical covariance localization part ($\mathbf{A}_v$) of the localization matrix $\mathbf{A}$ is realized through a recursive filter transform (Hayden and Purser 1995) with the distance measured either in scaled heights (i.e., natural log of the pressure) or in the number of model levels. For GFS, the horizontal localization is realized through a spectral filter transform. Specifically, the horizontal localization part ($\mathbf{A}_h$) of matrix $\mathbf{A}$ 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. E-folding distances equivalent to 1600 km and 1.1 scaled 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. These localization radii follow those adopted in the 3DEnsVar and 3DEnsVar hybrid experiments in Wang et al. (2013) where the same experiment configurations were used.

The last term of eq. (2) is the observational term as in the traditional 4DVar except that $\mathbf{x'}_t$ is defined by (1). $\mathbf{y}_t^{o\prime}$ and $\mathbf{H}_t$ are the innovation and linearized observation operator at time level t.

3. Experiment design

The data assimilation cycling experiments were conducted during a 5-week period, 0000 UTC 15 August 2010 ~ 1800 UTC 20 September 2010. The operational data stream including conventional and satellite data were assimilated every 6 hours. A list of types of operational conventional and satellite data are found on the NCEP website[1]. 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 following Wang et al. (2013). The same observation forward operators and satellite bias correction algorithms as in the operational Global Data Assimilation System (GDAS) 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. The digital filter (Lynch and Huang 1992) was applied during the GFS model integration following the operational configuration. Verification was conducted using data collected during the last four weeks of the experiment period. Verification of general global forecasts against European Centre for Medium-Range Weather Forecasts (ECMWF) analyses and in-situ observations were conducted. Statistical significance test using the paired t-test (Wilks, 1995, Page 121) was conducted for these verifications. A significance level of 95% was used to define if the differences seen in the comparison are statistically significant or not. Hurricane track forecasts for cases during the verification period were verified against the NHC (National Hurricane Center) best track data.

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

Following Figure 1a of Wang et al. (2013), a one-way coupled 4DEnsVar was adopted. The ensemble supplied to 4DEnsVar was initialized by an Ensemble Kalman filter (EnKF). The ensemble square root filter algorithm (EnSRF, Whitaker and Hamill 2002; Whitaker et al. 2008) was adopted. A recent implementation of EnSRF for GFS was described more fully in Hamill et al. (2011) and Wang et al. (2013). This EnKF code has been directly interfaced with GSI by using 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 and mis-representation of model errors, covariance localizations, multiplicative and additive inflation were applied. The detailed treatments and parameters used follow those in Wang et al. (2013).

A few experiments were designed to address the questions proposed in the introduction. Table 1 summarizes all experiments and their acronyms. To investigate the sensitivity of the performance of 4DEnsVar to the temporal resolution of ensemble perturbations spanning the assimilation window (i.e., t, in eq. 1 and 2), two experiments, one with hourly ensemble perturbations (4DEnsVar) and the other with two-hourly ensemble perturbations (4DEnsVar-2hr) were conducted. Specifically, in the "4DEnsVar" experiment, $L = 6$. Denote the time valid at the center of the data assimilation window as t=0. Forecast ensembles valid at the t=-3h, -2h, -1h, 0, 1h, 2h lead times were used. In the 4DEnsVar-2hr experiment, $L = 3$. Forecast ensembles valid at the t=-2h, 0 and 2h lead times were used. To study the impact of tangent linear normal mode balance constraint (TLNMC) on the 4DEnsVar analysis and forecast and how the impacts depend on different types of forecasts such as general global forecasts and hurricane track forecasts, experiments withholding the TLNMC (4DEnsVar-nbc) were conducted. In addition, studies with single observation and a case study assimilating full observations at a particular time were conducted to explore the downstream and upstream impacts of the 4D ensemble covariances and how well the linear propagation through ensemble covariances approximates the full nonlinear propagation.

4. Results

a. Single observation experiments

1) DOWNSTREAM AND UPSTREAM IMPACTS OF 4D ENSEMBLE COVARIANCES

A single observation experiment was conducted to illustrate the impact of the temporal evolution of the background error covariance in the newly developed 4DEnsVar. The observations were at the same location but valid at three different times: the beginning (t=-3h), middle (t=0) and end (t=+3h) of the 6-h assimilation window. Their increments valid at the analysis time which was the middle of the assimilation window (t=0) were compared. The observed variable was temperature at 700 hPa. The value of the temperature observation was set to be 1 degree warmer than the corresponding background value and the observation error standard deviation was set to be 1 degree. In the first experiment, a single temperature observation at t=-3h was assimilated. Figure 1a and 1d show the resulting analysis increment of temperature and geopotential height at 700 hPa valid at t=0. Relative to the observation location, the center of the maximum increment was displaced downstream toward the east and northeast. This result was consistent with that the analysis time was 3 hours later than the observation time and the prevailing background wind was blowing eastward. The second experiment was identical to the first except that the observation time was at t=0. The analysis increments valid at t=0 was plotted in Fig. 1b and 1e. Different from Fig. 1a and 1d, the center of the maximum increment was now more closely co-located with the location of the observation. For the third experiment, the observation was at t=3h. Relative to the observation location, the center of the maximum increment was displaced upstream westward of the observation location as shown in Fig.1c and 1f. As a comparison, the analysis increment from assimilating the single temperature observations valid at three different times was also computed using 3DEnsVar. Because of the absence of the temporal evolution of the background error covariance, the analysis increments produced by 3DEnsVar were independent of observation time. The analysis increments were exactly equal to that produced by 4DEnsVar when the observation was at t=0 (Fig. 1b and 1e).

2) COMPARISON WITH FULL NON-LINEAR MODEL PROPAGATION

In 4DEnsVar, the temporal propagation of observation information within the data assimilation window is effectively achieved through covariance of ensemble perturbations at discrete times. Although the ensemble forecasts were generated by full nonlinear model integrations, the temporal propagation through covariance of ensemble perturbations contains a linear assumption. Another single observation experiment was conducted to illustrate how well the linear propagation compared with the full nonlinear model propagation and how such comparison depended on the number of time levels of ensemble perturbations used in 4DEnsVar. Figure 2 illustrates such experiment where a tropical cyclone (Hurricane Daniel, 2010) was contained in the background forecast. A single meridional wind observation at 850 hPa was assimilated. The value of the wind observation was set to be $5ms^{-1}$ stronger than the background value with an observation error standard deviation of $1ms^{-1}$. The observation was valid at the beginning of the assimilation window (t=-3h). The resulting analysis increments of geopotential height at the middle of the assimilation window (t=0) at 850 hPa were shown in Fig. 2. Fig. 2a shows the increment by using the full nonlinear model propagation. First, the single wind observation at t=-3h was assimilated to update the state valid at t=-3h. Two 3-hour forecasts were then launched. These two forecasts were initialized by the states at t=-3h with and without assimilating the single observation respectively. The difference between the two forecasts was shown in Fig. 2a. Such difference reflected the actual increments valid at t=0 by propagating the increment at t=-3h through nonlinear model integration (Huang et al. 2009) and therefore can be served as the verification of increments generated by 4DEnsVar and 3DEnsVar. The spatial pattern of the increment through nonlinear model propagation consisted of a dipole structure with a negative increment and a positive increment located to the southwest and northeast side of the hurricane eye respectively. Such increment suggested the assimilation of the single wind observation at t=-3h corrected the position of the tropical cyclone in the background forecast valid at t=0 by moving the vortex south-westward. The increments from 4DEnsVar using hourly ensemble perturbations, 4DEnsVar with 2-hourly ensemble perturbations and 3DEnsVar are shown in Fig. 2b, 2c and 2d respectively. The increments from both 4DEnsVar experiments better approximated the increment from nonlinear model propagation than 3DEnsVar. While the dipole pattern of the increments by 4DEnsVar suggested that the position of the tropical cyclone in the background was shifted to the west or southwest, the increment from 3DEnsVar was dominated by a negative increment that was nearly centered at the eye with a slight positive increment on the west side of the eye. 4DEnsVar using hourly ensemble perturbations (4DEnsVar) approximates the increments from nonlinear propagation more closely than using 2-hourly ensemble perturbations. For example, the negative increment on the west side of the eye was too strong in 4DEnsVar-2hr than in 4DEnsVar. In addition, while 4DEnsVar corrected the vortex location by moving it to the southwest similar to the nonlinear propagation, 4DEnsVar-2hr moved vortex more to the west. Quantitative and systematic comparisons of 4DEnsVar using hourly and 2-hourly ensemble perturbations are included in section 4g.

b. Verification of forecasts against the ECMWF analyses

To evaluate the performance of the 4DEnsVar system, its forecast quality is measured by computing verification scores against the ECMWF analyses. Figure 3 shows the anomaly correlation (AC) of the geopotential height, temperature and wind forecasts over the globe verified against the ECMWF analysis. The ECMWF analysis data were obtained from the historical THORPEX Interactive Grand Global Ensemble (TIGGE) data archive hosted by the National Center for Atmospheric Research [2]. The anomaly correlation was calculated following this formula:

$$\mathrm{AC} = \frac{\overline{[(x^F - x^c) - \overline{(x^F - x^c)}][(x^{ec} - x^c) - \overline{(x^{ec} - x^c)}]}}{\sqrt{\overline{[(x^F - x^c) - \overline{(x^F - x^c)}]^2}}\sqrt{\overline{[(x^{ec} - x^c) - \overline{(x^{ec} - x^c)}]^2}}}. \quad (3)$$

In eq. (3), x^F, x^{ec} and x^c denote the forecast variable, the ECMWF analysis variable, and the corresponding variable from the climatological average. The over-bar denotes the areal mean (i.e., average over the domain considered). The climatological average was obtained by averaging the NCEP-NCAR reanalysis data over 1981-2010 (Kalnay et al. 1996). All data were first bi-linearly interpolated to a common grid with a 2.5-degree resolution before calculating the AC. Equation (3) was applied to each model level and each forecast during the verification period. The arithmetic average for all levels and forecasts is shown in Fig. 3. The verification started at the 2-day lead time to reflect that it was more appropriate to use the analyses to verify longer forecasts (Kuhl et al. 2013). Forecasts from 3DEnsVar are more skillful than GSI3DVar,

[2] http://tigge.ucar.edu/home/home.htm

consistent with Wang et al. (2013). 4DEnsVar further improves the skill of the forecasts compared to 3DEnsVar. The improvement of 4DEnsVar relative to 3DEnsVar is smaller than the improvement of 3DEnsVar relative to GSI3DVar. The AC was also calculated at Northern Hemisphere (NH) and Southern Hemisphere (SH) extratropics and tropics. The absolute improvement of 4DEnsVar relative to 3DEnsVar tends to be slightly larger in the SH extratropics than in the NH extratropics especially at longer lead times (e.g. Figure 4). The statistical significance of the differences of the anomaly correlations among different experiments shown in Fig. 3 and 4 are calculated using the paired t-test for each forecast lead time. The samples were accumulated by pairs of ACs from forecasts initialized at different times and located at different model levels. The differences between 4DEnsVar and GSI3DVar and between 4DEnsVar and 3DEnsVar for the lead times considered in Fig. 3 and Fig. 4 are all statistically significant (i.e., greater than 95% confidence level).

c. Verification of forecasts against in-situ observation

The 4DEnsVar system is also evaluated by comparing with the radiosonde observations. Figure 5 shows the root mean square fit (RMSF) of 6-hour forecasts to in-situ observations from marine and land surface stations, rawinsondes and aircrafts. Statistical significance of the difference between 4DEnsVar and 3DEnsVar is calculated using the paired t-test for each level. The samples were accumulated by pairs of RMSFs from forecasts initialized at different times. A blue cross is marked at the level where the difference is significant at and above the 95% level. A black cross is marked when 4DEnsVar was statistically significantly more accurate than 3DEnsVar averaged for all levels. Wind and temperature forecasts from 3DEnsVar and 4DEnsVar experiments are more accurate than GSI3DVar at most levels over NH, SH and TR. More appreciable improvement is seen in the wind forecasts than in the temperature forecasts.

Over NH and SH, 4DEnsVar shows consistent improvement relative to 3DEnsVar for wind forecasts and neutral or slightly positive impact for temperature forecast. Over TR, 4DEnsVar shows mostly neutral impact compared to 3DEnsVar for both wind and temperature forecasts.

Forecasts at longer lead times were also verified against in-situ observations (Fig. 6). Same statistical significance tests as Fig. 5 were conducted. Temperature forecasts from 4DEnsVar show overall positive impact relative to 3DEnsVar for both NH and SH at the 4-day lead time. 4DEnsVar shows neutral impact on wind forecasts over NH and positive impact over SH at the 4-day lead time. Over TR, 4DEnsVar shows positive impact relative to 3DEnsVar only for wind forecasts at low levels. These results are in general consistent with those found in Buehner et al. (2010b) except that Buehner et al. (2010b) found that the positive impact of 4DEnsVar relative to 3DEnsVar in NH was similar to that in SH at longer lead time. Such differences could be because our experiment was conducted during NH summer whereas Buehner et al. (2010b) conducted the experiments during NH winter or because of the differences in numerical models. It could also be because other differences between the two data assimilation systems such as the methods employed by each system in treating wind-mass imbalance during the variational minimization.

d. Verification of hurricane track forecasts

Early studies have shown that ensemble-based data assimilation may be particularly helpful with hurricane initialization due to the use of flow dependent estimates of the background error covariances (Torn and Hakim 2009, Zhang et al. 2009, Wang 2011). Several studies in particular explored the use of 3DEnsVar hybrid DA in hurricane forecasts (Wang 2011, Hamill et al. 2011, Li et al. 2012). They have found that deterministic forecasts from the 3DEnsVar hybrid were superior to those initialized from 3DVar. To date, however, no experiments have been performed with a 4DEnsVar applied for hurricane predictions. As shown in Fig. 2, application of 4DEnsVar for hurricane initialization and predictions can be particularly interesting because of the temporal variation of the error covariance associated with TC structure and location changes within the DA window. In this section, the performance of 4DEnsVar for hurricane forecasts was evaluated. Given that the experiments were conducted at the reduced resolution, only the hurricane track forecasts were verified.

1) REVIEW OF HURRICANE CASES DURING THE EXPERIMENT PERIOD

A total of 16 named storms (eight storms from the Atlantic basin and eight storms from the Pacific basin) during the 2010 hurricane season occurred in the verification period. During the experiment verification period, for the Atlantic basin as shown in Fig.7a, Hurricane Danielle, Earl and Julia and Igor reached category 4. Igor was the strongest tropical cyclone of the Atlantic Basin during the 2010 season. In addition to the above 4 hurricanes in the Atlantic Basin, 3 storms reached the tropical storm category. In the East Pacific, Frank, a category 1 hurricane, was close to the southwest coast of Mexico. In the West Pacific, during the experiment verification time, as shown in Fig.7b, Typhoon Kompasu made landfall at South Korea. Typhoon Fanapi caused heavy rainfall in Taiwan and Southern China. According to the hurricane forecast verification reports by the National Hurricane Center (NHC)[3] and Joint Typhoon Warning Center (JTWC)[4], the official hurricane track forecasts were more accurate during the 2010 season than the average of previous years.

2) COMPARISON OF TRACK FORECASTS

[3] www.nhc.noaa.gov/verification/pdfs/Verification_2010.pdf
[4] http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/atcr/2010atcr.pdf

The cyclones in the forecasts were tracked using the NCEP tropical cyclone tracker (Marchok 2002). To ensure a head to head comparison among forecasts initialized by different data assimilation methods, the following criteria were followed to include a particular forecast in the verification sample pool: (i) Forecasts must have been available for all systems involved in comparison; (ii) the cyclone must have been reported in TCVITALS[5] at the initial time of the forecast; (iii) the observed TC must have been a tropical cyclone or a subtropical cyclone at the lead time being evaluated following the NHC practice[6].

Figure 8a shows the root mean square error of the track forecasts from 4DEnsVar, 3DEnsVar and GSI3DVar. 3DEnsVar outperforms GSI3DVar, consistent with the results in Hamill et al. (2011). Track forecasts by 4DEnsVar are more accurate than 3DEnsVar after the 2-day forecast lead time. The statistical significance of the differences of the track forecast errors among different experiments shown in Fig. 8 are calculated using the paired t-test for each forecast lead time. The samples were accumulated by pairs of track errors from forecasts initialized at different times. The differences among 3DEnsVar and GSI3DVar are statistically significant for all lead times considered. The differences among 4DEnsVar and 3DEnsVar are statistically significant after 1-day lead time. In addition to examining the averaged track forecast errors, a separate measure of the performance of the track forecast following Zapotocny et al. (2008) was adopted to further examine the robustness of the difference seen in Fig. 8a. In this measure, the percentage of forecasts from one DA method that was better than forecasts from GSI3DVar was computed. Figure 8b shows the percentage of forecasts from 3DEnsVar and 4DEnsVar that were better than GSI3DVar. 60~68% of the forecasts from 3DEnsVar are better than GSI3DVar for the forecast lead times considered. For 4DEnsVar, 68-80% of the

[5] http://www.emc.ncep.noaa.gov/mmb/data_processing/tcvitals_description.htm.
[6] http://www.nhc.noaa.gov/verification/verify2.shtml

forecasts are better than GSI3DVar. Comparing 3DEnsVar and 4DEnsVar shows that the percentage of better forecasts by 4DEnsVar is larger than that of 3DEnsVar especially after the 1-day lead time. This result is consistent with that in Fig. 8a.

e. Impact of 4D ensemble covariances on convergence rate during the variational minimization and discussion on the second outer loop

With a similar experiment configuration, Wang et al. (2013) found that 3DEnsVar showed a slightly slower (faster) convergence rate at early (later) iterations than GSI 3DVar for the first outer loop, and a faster convergence rate for the second outer loop. Compared to 3DEnsVar, ensemble perturbations at multiple time levels were used during the variational minimization in 4DEnsVar. To investigate the impact of including multiple time levels of perturbations on the convergence of the minimization, the convergence rates of 3DEnsVar, and 4DEnsVar were compared. Figure 9 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. Following the configuration of the operational GSI, two outer loops were used during the variational minimization. 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 9 also shows that the iterations were terminated at the similar level of the ratio of gradient norm for the 3DEnsVar and 4DEnsVar experiments. For the first outer loop, 4DEnsVar shows slightly a slower convergence rate than 3DEnsVar. For the second outer loop, 4DEnsVar shows faster convergence than 3DEnsVar. For the experiments conducted in this study, the cost of 4DEnsVar variational minimization is approximately 1.5

times of that of 3DEnsVar. Tests comparing the computational costs have shown that 4DEnsVar is about one-order of magnitude less expensive than the TLA 4DVar being developed (Rancic et al. 2012).

As shown in Fig. 9, in the operational implementation of GSI, two outer loops were adopted to treat the nonlinearity during the assimilation. In GSI 3DVar, the implementation of the outer loops follows the same method in the incremental 4DVar in Courtier and Hollingsworth (1994) and Lawless et al. (2006). The only difference is that in GSI 3DVar, the mapping from the control variable to the observations does not involve the component of the tangent linear model. Compared to the first outer loop, in the second outer loop of GSI 3DVar, the innovation was updated by using the analysis resultant from the first outer loop as the background and the reference state for the linearization of the observation operator was changed from the first guess to the analysis resultant from the first outer loop. The equivalence between such outer loop implementation and the Gauss-Newton method for solving the nonlinear assimilation problem (Bjorck A 1996) was shown in Lawless et al (2006). In incremental 4DVar, the background error covariance at the beginning of the DA window is the same in the first and second outer loops. However, the reference state upon which the error covariance is propagating across the DA window is updated by using the analysis resultant from the first outer loop (Courtier and Hollingsworth 1994, Jazwinski (1998; pgs. 279-281)). Following incremental 4DVar, in the current implementation of GSI-based 4DEnsVar, ensemble forecasts should be re-run before the second outer loop. Specifically, ensemble perturbations used in the first outer loop valid at t=-3h should be maintained. These perturbations will then be added to the analysis from the first outer loop valid at t=-3h to form the new ensemble analyses at t=-3h. New ensemble forecasts within the DA window will then be initialized by this set of ensemble analyses. This procedure propagates the ensemble covariance following the trajectory defined by the analysis resultant from the first outer loop. However, due to the computational cost of re-running the ensemble, this step was omitted in this study and the ensemble perturbations throughout the DA window used for the second outer loop were the same as the first outer loop. An attempt was made to illustrate the impact of using the updated trajectory to evolve the ensemble covariance through a single observation experiment using the same hurricane case as in Fig. 2. The result (not shown) suggested a slightly improved increment using re-evolved ensemble perturbations when using the increment from nonlinear propagation as verification. Future work is needed to systematically explore the impact of the second outer loop in 4DEnsVar and the impact of using re-evolved ensemble perturbations in the second outer loop.

f. Impact of 4D ensemble covariances on balance

Imbalance between variables introduced during data assimilation can degrade the subsequent forecasts. The mass-wind relationship in the increment associated with the ensemble-based method was defined by the multivariate covariance inherent in the ensemble perturbations. Such inherent relationship can be altered by the commonly applied covariance localization (e.g., Lorenc 2003; Kepert 2009; Holland and Wang 2013). Compared to 3D analysis methods, one attractive aspect of the analysis produced by a 4D method is the temporal smoothness. In 4DVar, this is achieved through the explicit use of a dynamic model. In 4DEnsVar, instead, the 4D increments were obtained through the Schur product of extended control variables and ensemble perturbations valid at discrete times. The balance of the analysis produced by 4DEnsVar is investigated in this section. The mean absolute tendency of surface pressure (Lynch and Huang 1992) is a useful diagnostic metric to show the amount of imbalance for an analysis generated by a data assimilation system. The hourly surface pressure tendency averaged over the experiment period was calculated and summarized in Table 2. For all hemispheres, the forecasts initialized by 4DEnsVar are slightly more balanced than the 3DEnsVar. Note that for all the experiments, following the operational configuration of GFS, the digital filter was applied during the model integration. In this study, the digital filter was configured with a 4-hour filtering window where the forecast state at the center of the window was replaced by the weighted average of forecast states spanning the 4-hour window. The impact of the digital filter on the forecasts started from the second hour of the model integration. The results in Table 2 suggest that the forecasts initialized by 4DEnsVar were still more balanced than 3DEnsVar even when DFI was applied.

g. Quantitative evaluation of the sensitivity to the number of time levels of the ensemble perturbations

In a typical 4DVar, the analyses are obtained via fitting the model trajectory to observations distributed within a finite assimilation window through the use of the tangent linear and adjoint of the forecast model. In 4DEnsVar, the 4D analyses are obtained through variational cost function minimization within the temporally evolved ensemble forecast perturbation space spanning the assimilation window. Effectively, the four-dimensional (4D) background error covariance of a nonlinear system was approximated by the covariances of ensemble perturbations at discrete times. Using a single observation experiment, section 4a illustrates how the 4DEnsVar increments approximate the increments made by the nonlinear model propagation and how such approximation depends on the number of time levels of ensemble perturbations used in 4DEnsVar. This section provides further quantitative investigation on how the performance of 4DEnsVar depends on the number of time levels at which the ensemble perturbations are sampled. To evaluate the linear approximation quantitatively, the correlation of the increments from nonlinear model propagation and 4DEnsVar was calculated. Figure 10 shows an example for a case where the center of the assimilation window was at 6 UTC on August 25 in 2010. All operational observations within the first hour (between t=-3h and t=-2h) of the 6-hour DA window were assimilated. The increments valid at t=3h, the end of the assimilation window, were evaluated. Following the same method in Fig. 2a of section 4a, the true increment at t=3h was calculated through nonlinear model propagation. First, all observations within the first hour were assimilated to update the state valid at t=-3h. Two 6-hour forecasts were then launched. These two forecasts were initialized by the states at t=-3h with and without assimilating those observation respectively. Such difference reflected the actual increments valid at t=3h by propagating the increment at t=-3h through nonlinear model integration. Increments by 4DEnsVar and 4DEnsVar-2h were evaluated by computing the spatial correlation of these increments with the true increments. Fig. 10 shows such correlation for different state variables at various model levels. It is found that increments by 4DEnsVar using hourly ensemble perturbations correlate with the true increments more than using the 2-hourly ensemble perturbations for most of the model levels and variables considered.

Another way to systematically evaluate the performance of 4DEnsVar as a function of the number of time levels of ensemble perturbations is to compare the performance of forecasts initialized by 4DEnsVar using hourly ensemble perturbations versus 4DEnsVar using 2-hourly ensemble perturbations. Therefore a separate data assimilation cycling and forecast experiment where the ensemble perturbations were sampled at 3 time levels instead of 6-time levels were conducted during the whole experiment period. In other words, in this experiment the ensemble perturbations were sampled every two hours. This experiment was named as 4DEnsVar-2hr. Figure 11 shows that the performance of forecasts initialized by 4DEnsVar was degraded when less frequent ensemble perturbations were used especially at longer forecast lead times. Similar statistical significance tests as Fig. 3 were conducted for the results in Fig. 11. It was found that such degradation is statistically significant at the 72-h and 96-h lead times for geopotential height and meridional wind forecasts, and at 96-h lead time for the temperature and zonal wind forecasts. The AC calculated for NH, SH and TR showed similar results (not shown).

The balance of the 4DEnsVar analyses to the temporal resolution of the ensemble perturbations was also examined. Table 2 shows that using less frequent ensemble perturbations, the 4DEnsVar analyses became less balanced.

The hurricane track forecast was also degraded after the one-day lead time when less frequent ensemble perturbations were used (Figure 12). Similar statistical significance tests as Fig. 8 was conducted for the results in Fig. 12. The degradation was statistically significant after 1-day lead time. This result is further confirmed by calculating the percentage of forecasts with hourly ensemble perturbations that were better than the forecasts with 2-hourly ensemble perturbations (Fig. 12b). These results are consistent with the expectation that the temporal evolution of the error covariance with the assimilation window is better approximated with more frequent ensemble perturbations.

h. Impact of TLNMC

The tangent linear normal mode constraint (TLNMC) was implemented in the GSI minimization to improve the balance of the initial conditions. The TLNMC operator was applied to the analysis increment during the variational minimization. The operator contained 3 steps including calculating the tangent linear tendency model, projecting the tendency onto the gravity modes and reducing the gravity mode tendencies. For simplicity, the tendency model was obtained from a tangent linear version of a general, hydrostatic, adiabatic primitive equation model. The tendency model used for the TLNMC purpose also did not include parameterized physics. More details on TLNMC implemented in GSI 3DVar were provided in Kleist et al. (2009). Kleist et al. (2009) showed that the impact of TLNMC resulted in substantial improvement in the global forecasts initialized by GSI 3DVar. The TLNMC was applied on the analysis increments associated with the ensemble covariances when 3DEnsVar and 4DEnsVar were implemented within GSI. Wang et al. (2013) found that TLNMC improved the global forecasts initialized by 3DEnsVar and also concluded that the better performance of 3DEnsVar relative to EnKF was due to the ability of 3DEnsVar in using such constraint during the variational minimization.

In 4DEnsVar, the TLNMC operator was applied to the analysis increment at different time levels, $\mathbf{x}'_t$. In the current implementation, the reference state for the tangent linear tendency model was assumed to be time-invariant throughout the assimilation window. In addition, as discussed in section 4f, one attractiveness of 4DEnsVar analyses compared to its 3D counterpart is the temporal smoothness, which itself can lead to more balanced analyses. Given the simplification and assumptions made in TLNMC and the inherently more balanced analyses in 4DEnsVar, the impact of further applying TLNMC within 4DEnsVar was examined in this study. Experiments configured to be the same as 4DEnsVar-2hr, but without the use of the TLNMC were conducted (hereafter, the experiment is named as 4DEnsVar-nbc). The impact of TLNMC on the balance of the analyses is first measured following the same method in section 4f. It was found that TLNMC resulted in substantial decrease of surface pressure tendency and therefore more balanced 4DEnsVar analyses (Table 2). The accuracy of the forecasts initialized by 4DEnsVar withholding the TLNMC was also evaluated. Figure 11 shows that TLNMC yields significant positive impacts measured by the global AC for the forecast lead times considered. Similar statistical significance test as in Fig. 3 was conducted. Such test revealed that the positive impact of TLNMC was statistically significant for the lead times and variables considered. Further calculating the AC in NH, SH, and TR showed that most of the positive impact of TLNMC was from SH. The TLNMC showed neutral impact over the TR. For NH, slight positive impacts were found at early lead times and slight negative impacts were found after the 3-day lead time. The slight negative impact over NH at longer lead times could be due to the time-invariant reference state used when the tendency model was calculated. This hypothesis was consistent with a neutral or slight positive impact of TLNMC on the 3DEnsVar over NH (not shown).

The TLNMC implemented in GSI does not include the diabatic processes in the tendency model. Therefore it may not be appropriate for the forecasts associated with strong moist processes such as the tropical cyclone forecasts. Therefore the impacts of TLNMC on the TC forecasts were examined also. Figure 12 shows that the TLNMC showed negative impact on TC track forecasts. Similar statistical significance test as in Fig. 8 was conducted for results in Fig. 12. It was found that the negative impact of TLNMC on TC track forecasts was statistically significant for all the lead times in Fig. 12a. This result is further confirmed by calculating the percentage of forecasts without TLNMC that were better than the forecasts with TLNMC (Fig. 12b). The TLNMC showed negative impact on TC track forecasts initialized by 3DEnsVar also (not show). Withholding TLNMC, 4DEnsVar showed improved TC track forecast than 3DEnsVar even with reduced number of levels of perturbations (not shown). These results suggest that further development of the balanced constraints by considering the moisture processes are needed for forecasts with strong moisture processes.

5. Conclusion and discussion

A GSI based Four-dimensional ensemble-variational data assimilation system (4DEnsVar) was developed. Different from its 3D counterpart (3DEnsVar), the ensemble perturbations valid at multiple time levels throughout the DA window were effectively used to estimate the 4D background error covariances during the variational minimization. The TLA of the forecast model was conveniently avoided. Different from 4DEnsVar implemented in other systems, 4DEnsVar implemented within GSI minimization was preconditioned upon the full background error covariance matrix. The specific formulations and implementations of 4DEnsVar within GSI were introduced first. Using the newly developed GSI-based 4DEnsVar system, a few questions were investigated. What is the value of using the 4D ensemble covariance in 4DEnsVar for general global forecasts and for hurricane track forecasts? In 4DEnsVar, temporal evolution of the error covariance is approximated by the covariances of ensemble perturbations at discrete times. How is the performance of 4DEnsVar dependent on the temporal resolution of or the number of time levels of ensemble perturbations? How is this linear approximation compared to the full nonlinear model propagation? Will using 4D ensemble covariances to fit the model trajectory to observations distributed within a finite assimilation window improve the balance of the analysis and how is the balance of the analysis dependent on the temporal resolution of the ensemble perturbations? What is the impact of further applying the tangent linear normal mode balance constraint on the 4DEnsVar analysis and forecast and how is that dependent on different types of forecasts such as the general global forecast or hurricane track forecasts? How does including multiple time levels of perturbations impact the convergence rates of the minimization?

The performance of the system and aforementioned questions were investigated using NCEP GFS at a reduced resolution. The ensemble was supplied by an EnKF. The experiments were conducted over a summer month period assimilating NCEP operational conventional and satellite data. The findings from these experiments in addressing those aforementioned questions are summarized below. A series of single observation experiments revealed that the newly developed 4DEnsVar was able to reflect the temporal evolution of the background error covariance in the DA window. The global forecasts were verified against both the in-situ observations and the ECMWF analyses. 4DEnsVar in general improved upon 3DEnsVar. At short lead times, the improvement of 4DEnsVar relative to 3DEnsVar over NH was similar to that over SH. At longer forecast lead times, 4DEnsVar showed more improvement in SH than NH. The improvement of 4DEnsVar over TR was neutral or slightly positive when forecasts were verified against the in-situ observations. Track forecasts of 16 named tropical cyclones during the verification period were verified against the NHC best track data. The track forecasts initialized by 4DEnsVar were more accurate than 3DEnsVar after the 1-day forecast lead time. A single observation case study where Hurricane Daniel 2010 was the background and a case study assimilating all operational observations at the beginning of the assimilation window were conducted to reveal how well covariance of ensemble perturbations approximated the propagation using the full nonlinear model both qualitatively and quantitatively. It was found that increments from 4DEnsVar using more frequent ensemble perturbations approximated the increments from direct, nonlinear model propagation better than using less frequent ensemble perturbations. Consistently, using experiments over the full experiment period, it was found that when less frequent ensemble perturbations were used in the assimilation window, the performance of the forecasts initialized by 4DEnsVar was degraded especially after 2-3 day lead times for global forecast and after 1-day lead time for hurricane track forecasts. Analyses generated by 4DEnsVar were more balanced than those by 3DEnsVar. 4DEnsVar using more frequent ensemble perturbations produced analyses that were more balanced than using less frequent ensemble perturbations. TLNMC showed positive impact on 4DEnsVar for global forecasts verified using the anomaly correlation metric and negative impact for hurricane track forecasts. For the first outer loop, 4DEnsVar showed slightly a slower convergence rate than 3DEnsVar. For the second outer loop, 4DEnsVar showed slightly a faster convergence than 3DEnsVar.

As discussed in the Introduction section, in this study, as a first step of testing the newly developed 4DEnsVar for GSI, experiments were conducted where the single control forecast and the ensemble were run at the same, reduced resolution. Wang et al. (2013) found little impact of including the static covariance in the background error covariance when comparing 3DEnsVar and 3DEnsVar hybrid with similar experiment settings as this study. Therefore in the current study no static covariance was included. Recent experiments (Lei and Wang 2014) found that with the dual resolution configuration where the ensemble was run at a reduced resolution compared to the control forecast and analysis, the static covariance showed a significant positive impact. Buehner et al. (2013) found that when the static covariance was included in the Canadian 4DEnsVar system, the 4D ensemble covariance only resulted in small improvement in forecast quality. Further work is needed to compare GSI-based 4DEnsVar and 3DEnsVar at dual resolution mode and to explore the impact of the 4D extension of the ensemble covariance relative to its 3D counterpart when a static covariance is included.

In the current study, no temporal covariance localization was applied on the ensemble covariance in 4DEnsVar. A temporal localization is being developed within GSI-based 4DEnsVar. Preliminary tests showed positive impact of the temporal localization on the performance of 4DEnsVar. Future work is needed to further explore such impact. As discussed in section 4e, in 4DEnsVar currently implemented in GSI, to save computational cost, during the second outer loop, the same evolved ensemble perturbations as in the first outer loop were used although the trajectory was updated during the first outer loop. Future work is needed to explore the impact of the second outer loop and the impact of re-centering the ensemble perturbations on the new trajectory during the second outer loop in 4DEnsVar.

Acknowledgements. The study was supported by NOAA THOPREX grant, NA08OAR4320904, NASA NIP grant NNX10AQ78G, and NOAA HFIP grant NA12NWS4680012. Jeff Whitaker and Daryl Kleist are acknowledged for helpful discussion.

APPENDIX A

Acronyms for coupled ensemble-variational data assimilation

Table A1 here

APPENDIX B

Mathematical framework for implementing 4DEnsVar in GSI variational minimization

Different from other variational data assimilation systems, operational GSI minimization is preconditioned upon the full background error covariance matrix. Wang (2010) and Wang et al. (2013) introduced and described the formulas of 3DEnsVar hybrid in GSI. In other words, those formulas describe how the ensemble covariance is implemented in the GSI 3DVar variational minimization through the augmented control vectors (ACV) under such preconditioning. In this section, we further extend this framework and derive formulas to show how the 4DEnsVar is implemented within GSI. The key of this derivation is that the minimization of the new cost function (1) and (2) can be preconditioned in the same way as shown in Wang (2010). In other words, the same conjugate gradient minimization procedure from the original GSI-based 3DEnsVar will be followed for GSI-based 4DEnsVar.

Denote the new control variable as

$$\mathbf{x} = \mathbf{a}. \tag{3}$$

The increment in 4DEnsVar can be expressed as

$$\mathbf{x'}_t = \sum_{k=1}^{K} \mathbf{a}_k \circ (\mathbf{x}_k^e)_t = [diag[(\mathbf{x}_1^e)_t] \quad \cdots \quad diag[(\mathbf{x}_K^e)_t]]\mathbf{a}, \tag{4}$$

where $diag$ is an operator that turns a vector into a diagonal matrix where the *n*th diagonal element is given by the *n*th element of the vector (Wang et al. 2007a). We further denote $\mathbf{D}_t = [diag[(\mathbf{x}_1^e)_t] \quad \cdots \quad diag[(\mathbf{x}_K^e)_t]]$. Then the increment becomes

$$\mathbf{x'}_t = \mathbf{D}_t\mathbf{a} = \mathbf{D}_t\mathbf{x}. \tag{5}$$

Denote the new background error covariance as

$$\mathbf{B} = \mathbf{A}. \tag{6}$$

As in the original GSI 3DVar and 3DEnsVar, 4DEnsVar is also preconditioned by defining a new variable

$$\mathbf{z} = \mathbf{B}^{-1}\mathbf{x} = (\mathbf{A}^{-1}\mathbf{a}). \tag{7}$$

In the rest of the derivation, we will show that $\nabla_z J = \mathbf{B}\nabla_x J$, and therefore the minimization for the 4DEnsVar cost function can follow the same conjugate gradient method used by the original GSI, which then concludes the derivation. The rest of the terms in eq. (3)-(7) are defined the same as in eq. (1)-(2).

First, we derive the gradient of the hybrid cost function with respect to $\mathbf{x} = \mathbf{a}$. The gradients of the new cost function with respect to the extended control variables $\nabla_{\mathbf{a}} J$ are given as

$$\nabla_{\mathbf{x}} J = \nabla_{\mathbf{a}} J = \mathbf{A}^{-1}\mathbf{a} + \sum_{t=1}^{L} \mathbf{D}_t^T \mathbf{H}_t^T \mathbf{R}_t^{-1}(\mathbf{H}_t \mathbf{x}'_t - \mathbf{y}_t^{o\prime}) = \mathbf{z} + \sum_{t=1}^{L} \mathbf{D}_t^T \mathbf{H}_t^T \mathbf{R}_t^{-1}(\mathbf{H}_t \mathbf{x}'_t - \mathbf{y}_t^{o\prime}) \tag{8}$$

Next we derive the gradient of the hybrid cost function with respect to $\mathbf{z}$. The gradients of the new cost function with respect to $\boldsymbol{z} = \mathbf{A}^{-1}\mathbf{a}$ are given by

$$\nabla_{\mathbf{z}} J = \nabla_{\mathbf{A}^{-1}\mathbf{a}} J = \mathbf{a} + \mathbf{A}\sum_{t=1}^{L} \mathbf{D}_t^T \mathbf{H}_t^T \mathbf{R}_t^{-1}(\mathbf{H}_t \mathbf{x}'_t - \mathbf{y}_t^{o\prime}) = \mathbf{x} + \mathbf{B}\sum_{t=1}^{L} \mathbf{D}_t^T \mathbf{H}_t^T \mathbf{R}_t^{-1}(\mathbf{H}_t \mathbf{x}'_t - \mathbf{y}_t^{o\prime}). \tag{9}$$

Comparing $\nabla_{\mathbf{x}} J$ in eq. (8) and $\nabla_{\mathbf{z}} J$ in eq. (9), we thus obtain

$$\nabla_{\mathbf{z}} J = \mathbf{B}\nabla_{\mathbf{x}} J. \tag{10}$$

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*, in press.

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

Bjorck Å (1996) Numerical methods for least squares problems. SIAM, Philadelphia

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.

−, P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010b: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic

NWP. Part II: One-Month Experiments with Real Observations. *Mon. Wea. Rev.*, **138**, 1567-1586.

Buehner, M., J. Morneau and C. Charette, 2013: Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlinear Processes in Geophysics*, in press.

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.

Courtier, P., Thépaut, J.N. and Hollingsworth, A., 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Q.J.R.M.S.*, **120**, 1367-1387.

Derber, John, Anthony Rosati, 1989: A Global Oceanic Data Assimilation System. *J. Phys. Oceanogr.*, **19**, 1333–1347.

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

Clayton, A. M., Lorenc, A. C. and Barker, D. M., 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. Q.J.R. Meteorol. Soc.., **139**, 1445-1461.

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

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.

Houtekamer, P. L., Herschel L. Mitchell, 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Jazwinski, Andrew H. (1970). *Stochastic Processes and Filtering Theory*. New York: Academic Press.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.

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

Kuhl, D. D., T. E. Rosmond, C. H. Bishop, J. McLay, N. L. Baker, 2013: Comparison of Hybrid Ensemble/4DVar and 4DVar within the NAVDAS-AR Data Assimilation Framework. *Mon. Wea. Rev.*, **141**, 2740–2758.

Lawless, A., Gratton, S. and Nichols, N. (2006) *An investigation of the convergence of incremental 4D-Var.* In: The Fourth WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography, 2006.

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.* 3507-3524.

Lei, T. and X. Wang, 2014: A comparison of GSI-based hybrid data assimilation for NCEP GFS at single and dual resolution. *Mon. Wea. Rev*., in preparation.

Liu, C., Q. Xiao and B. Wang, 2008: An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test. *Mon. Wea. Rev.*, **136**, 3363–3373.

Liu,C., Q. Xiao and B. Wang, 2009: An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part II: Observing System Simulation Experiments with Advanced Research WRF (ARW). *Monthly Weather Review* **137**:5, 1687-1704.

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

Marchok, T., 2002: How the NCEP tropical cyclone tracker works. Preprints, 25th Conf. on Hurricanes and Tropical Meteorology, San Diego, CA, Amer. Meteor. Soc., P1.13. [Available online at http://ams.confex.com/ams/pdfpapers/37628.pdf.]

Qiu,C, L. Zhang and A. Shao, 2007: An explicit four-dimensional variational data assimilation method. Science in China Series D: Earth Sciences, 50,1232-1240.

Rancic, M., D. Kleist, and R. Todling, 2012: Development of a 4DVar version of GSI at NCEP. American Meteorological Society 92nd annual meeting, New Orleans, LA, https://ams.confex.com/ams/92Annual/webprogram/Paper199871.html

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.

Tian, X.*,* Z. Xie*, and* A. Dai *(*2008*),* An ensemble-based explicit four-dimensional variational assimilation method*,* J. Geophys. Res.*,* 113*, D21124, doi:*10.1029/2008JD010358.

Torn, Ryan D., Gregory J. Hakim, 2009: Ensemble Data Assimilation Applied to RAINEX Observations of Hurricane Katrina (2005). *Mon. Wea. Rev.*, **137**, 2817–2829.

Wang, B., J. Liu, S. Wang, W. Cheng, J. Liu, C. Liu, Q. Xiao and Y. Kuo, 2010: An economical approach to four dimensional variational data assimilation. *Adv. Atmos. Sci*., 27, 715-727.

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

Wang, X., 2011: Application of the WRF Hybrid ETKF–3DVAR Data Assimilation System for Hurricane Track Forecasts. *Weather and 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.

−, D. Parrish, D. Kleist and J. S. Whitaker, 2013: GSI 3DVar-based Ensemble-Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single Resolution Experiments. *Mon. Wea. Rev.*, **141**, 4098–4117.

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

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

Xu, L., T. Rosmond, and R. Daley, 2005: Development of NAVDAS-AR: Formulation and initial tests of the linear problem. *Tellus*, **57A**, 546–559.

Zapotocny, T. H., J. A. Jung, J. F. Le Marshall, and R. E. Treadon, 2008: A two-season impact study of four satellite data types and rawinsonde data in the NCEP Global Data Assimilation System. Wea. Forecasting, **23**, 80–100.

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.

Zhang, F., Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-Resolving Hurricane Initialization and Prediction through Assimilation of Doppler Radar Observations with an Ensemble Kalman Filter. *Mon. Wea. Rev.*, **137**, 2105–2125.

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 for Northern Extratropics (NH), Tropics (TR) and Southern Extratropics (SH) for the 3DEnsVar, 4DEnsVar, 4DEnsVar-2hr and 4DEnsVar-nbc experiments respectively.

Table A1. Characteristics of different flavors of coupled ensemble-variational data assimilation and their acronyms.

Figure Captions

Figure 1. Temperature increments (red contours in a-c; units: K) and geopotential height increments (red contours in d-f; units: m) at 700 hPa valid at the middle of a 6-hour assimilation window after assimilating a single temperature observation valid at three different times. Black contours are the background temperature (a-c) and height fields (d-f) valid at the analysis time. The observation was located at the same place denoted by the "+" sign, but at three different times: the beginning (a, d), middle (b,e) and end of the 6-hour assimilation window (c,f).

Figure 2. Geopotential height increments (color shades, units: m) at 850 hPa valid at the middle of a 6-hour assimilation window after assimilating a single meridional wind observation. Black contours are background height fields valid at the middle of the assimilation window. The observation was located at the "+" sign and valid at the beginning of the assimilation window. (a) increment by model integration; (b) increment by 4DEnsVar, (c) increment by 4DEnsVar-2h and (d) increment by 3DEnsVar.

Figure 3. The globally averaged anomaly correlation of geopotential height (a), temperature (b), zonal (c) and meridional wind (d) forecasts verified against the ECMWF analyses. Solid, dotted and dashed lines are for GSI3DVar, 3DEnsVar and 4DEnsVar experiments respectively. The "+" right below the upper x-axis denotes the lead time when the difference between 3DEnsVar and GSI3DVar is statistically significant. The "+" right above the lower x-axis denotes the lead time when the difference between 4DEnsVar and 3DEnsVar is statistically significant.

Figure 4. Averaged anomaly correlation of geopotential height at Northern Extratropics (NH,a), Southern Extratropics (SH, b) and Tropics (TR, c). Lines and symbol "+" are defined the same as Fig. 3.

Figure 5. The root-mean-square fit of the 6-hour forecasts to the in-situ observations for temperature (left column) and wind (right column) as a function of pressure for the Northern Hemisphere extra-tropics (a,b), Southern Hemisphere extra-tropics (c,d) and tropics (e, f) for the GSI3DVar, 3DEnsVar, 4DEnsVar experiments. Line definition is the same as Figure 3. Blue "+" indicates the levels where 4DEnsVar is statistically significantly better than 3DEnsVar. Black "+" and black "-" after "total" denote if 4DEnsVar is or is not statistically significantly better than 3DEnsVar respectively after averaging over all levels.

Figure 6. Same as Fig. 5 except at the 96-hour forecast lead time.

Figure 7. Tracks of tropical cyclones during the verification period in the Atlantic and East Pacific (a) and West Pacific basins (b).

Figure 8. Skill of tropical cyclone track forecasts measured by (a) the root mean square errors and (b) percentage of forecasts that were better than the reference GSI3DVar forecast for 4DEnsVar (thick solid), 3DEnsVar (dotted) and GSI3DVar (thin solid) experiments. The numbers above the x-axis of (a) denotes the statistical significant confidence level of the difference between 4DEnsVar and 3DEnsVar. The differences among 3DEnsVar and GSI3DVar

are statistically significant for all lead times considered in (a). The numbers above the x-axis of (b) denote the number of samples used in the calculation at each lead time.

Figure 9. Averaged ratios of gradient norms as a function of the number of iterations in the first and second outer loops during the variational minimization of 3DEnsVar (solid) and 4DEnsVar (dash) experiments.

Figure 10. The correlation of analysis increments by 4DEnsVar using hourly ensemble perturbations (4DEnsVar) and using 2-hourly ensemble perturbations (4DEnsVar-2hr) with the increment by nonlinear model propagation. The increments are valid at the end of the 6-hour assimilation window by assimilating all observations within the first hour of the assimilation window. See text for more details.

Figure 11. The globally averaged anomaly correlation of geopotential height (a), temperature (b), zonal (c) and meridional wind (d) forecasts verified against the ECMWF analyses. Solid, dotted and dashed lines are for 4DEnsVar-nbc, 4DEnsVar-2hr and 4DEnsVar experiments respectively. The "+" right above the lower x-axis denotes the lead time when the difference between 4DEnsVar and 4DEnsVar-2hr is statistically significant. The "+" right below the upper x-axis denotes the lead time when the difference between 4DEnsVar-2hr and 4DEnsVar-nbc is statistically significant.

Figure 12. Skill of tropical cyclone track forecasts measured by (a) the root mean square errors and (b) percentage of forecasts that were better than the reference 4DEnsVar-2hr forecast for 4DEnsVar (thick solid), 4DEnsVar-2hr (thin solid) and 4DEnsVar-nbc (dotted) experiments. The black (blue) numbers above the x-axis of (a) denote the statistical significant confidence level of the difference between 4DEnsVar and 4DEnsVar-2hr (between 4DEnsVar-2hr and 4DEnsVar-nbc). The numbers above the x-axis of (b) denote the number of samples used in the calculation at each lead time.

Experiment	Description
GSI3DVar	The GSI 3DVar experiment
4DEnsVar	4D ensemble-variational DA experiment with hourly ensemble perturbations
4DEnsVar-2hr	4D ensemble-variational DA experiment with 2-hourly ensemble perturbations
3DEnsVar	3D ensemble-variational DA experiment
4DEnsVar-nbc	Same as “4DEnsVar-2hr” except without the use of the tangent linear normal mode balance constraint (TLNMC)

Table 1. A list of experiments.

Unit: hPa hr^{-1}	NH	TR	SH
3DEnsVar	0.402	0.524	0.643
4DEnsVar	0.392	0.516	0.635
4DEnsVar-2hr	0.396	0.519	0.638
4DEnsVar-nbc	0.461	0.571	0.719

Table 2. Averaged hourly absolute surface pressure tendency during the experiment period for Northern Extratropics (NH), Tropics (TR) and Southern Extratropics (SH) for the 3DEnsVar, 4DEnsVar, 4DEnsVar-2hr and 4DEnsVar-nbc experiments respectively.

	Number of time levels of ensemble perturbations incorporated in the DA window during the variational minimization	Weights on static and ensemble covariance	Tangent linear and adjoint of the forecast model
3DEnsVar	One, usually valid at the center of the DA window	0% on static and 100% on ensemble	Not needed
3DEnsVar hybrid	One, usually valid at the center of the DA window	Nonzero on static and ensemble covariances, sum of weights is usually constrained to be equal to 1	Not needed
4DEnsVar	Multiple	0% on static and 100% on ensemble	Not needed
4DEnsVar hybrid	Multiple	Nonzero on static and ensemble covariances, sum of weights is usually constrained to be equal to 1	Not needed, same static covariance is used for multiple time levels, equivalent to assuming a numerical model of identity matrix
Ens4DVar	One, usually valid at the beginning of the DA window	0% on static and 100% on ensemble	Needed
Ens4DVar hybrid	One, usually valid at the beginning of the DA window	Nonzero on static and ensemble covariances, sum of weights is usually constrained to be equal to 1	Needed

Table A1. Characteristics of different flavors of coupled ensemble-variational data assimilation and their acronyms.

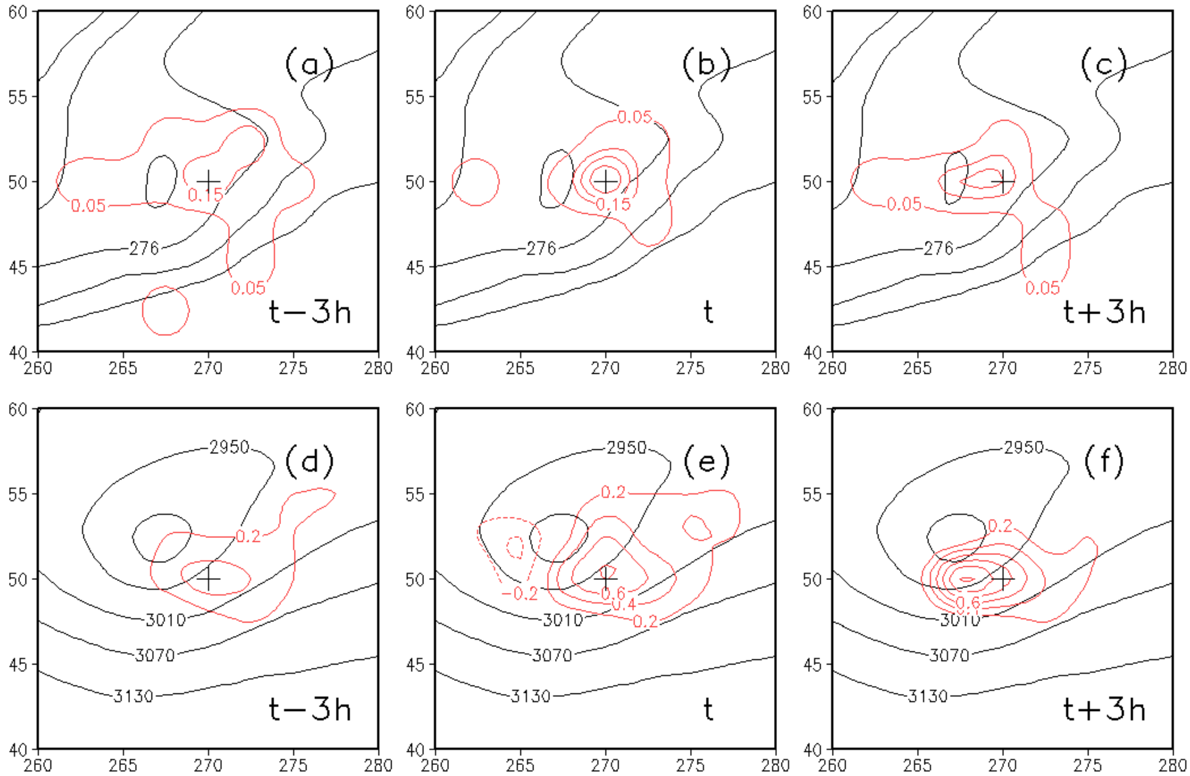

Figure 1. Temperature increments (red contours in a-c; units: K) and geopotential height increments (red contours in d-f; units: m) at 700 hPa valid at the middle of a 6-hour assimilation window after assimilating a single temperature observation valid at three different times. Black contours are the background temperature (a-c) and height fields (d-f) valid at the analysis time. The observation was located at the same place denoted by the "+" sign, but at three different times: the beginning (a, d), middle (b,e) and end of the 6-hour assimilation window (c,f).

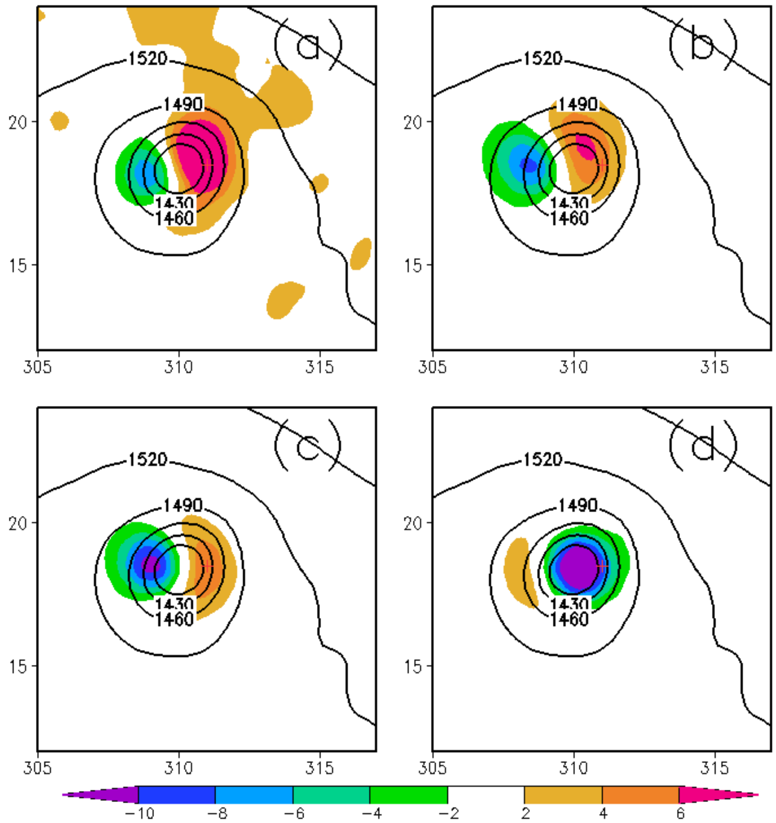

Figure 2. Geopotential height increments (color shades, units: m) at 850 hPa valid at the middle of a 6-hour assimilation window after assimilating a single meridional wind observation. Black contours are background height fields valid at the middle of the assimilation window. The observation was located at the "+" sign and valid at the beginning of the assimilation window. (a) increment by model integration; (b) increment by 4DEnsVar, (c) increment by 4DEnsVar-2h and (d) increment by 3DEnsVar.

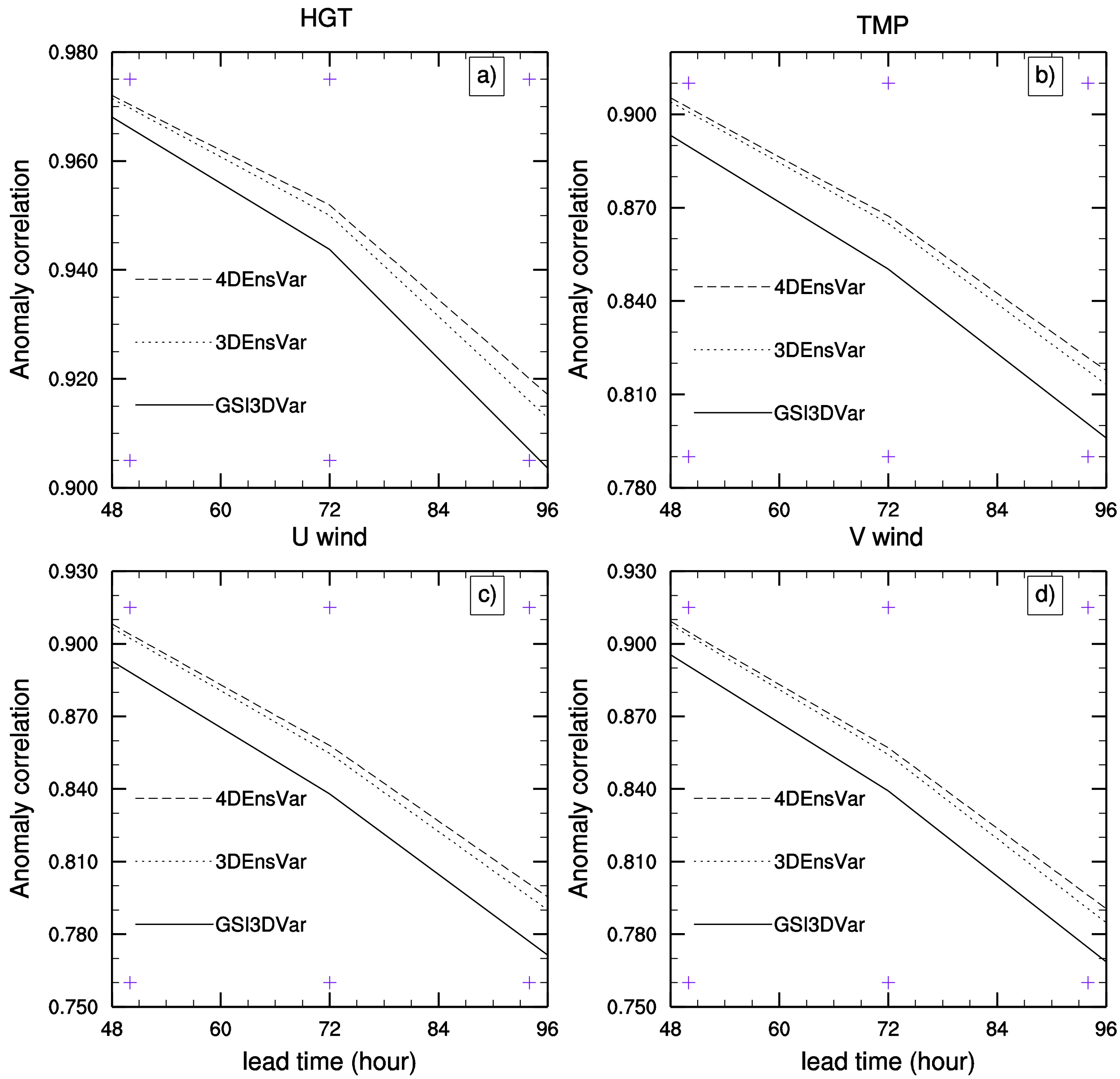

Figure 3. The globally averaged anomaly correlation of geopotential height (a), temperature (b), zonal (c) and meridional wind (d) forecasts verified against the ECMWF analyses. Solid, dotted and dashed lines are for GSI3DVar, 3DEnsVar and 4DEnsVar experiments respectively. The "+" right below the upper x-axis denotes the lead time when the difference between 3DEnsVar and GSI3DVar is statistically significant. The "+" right above the lower x-axis denotes the lead time when the difference between 4DEnsVar and 3DEnsVar is statistically significant.

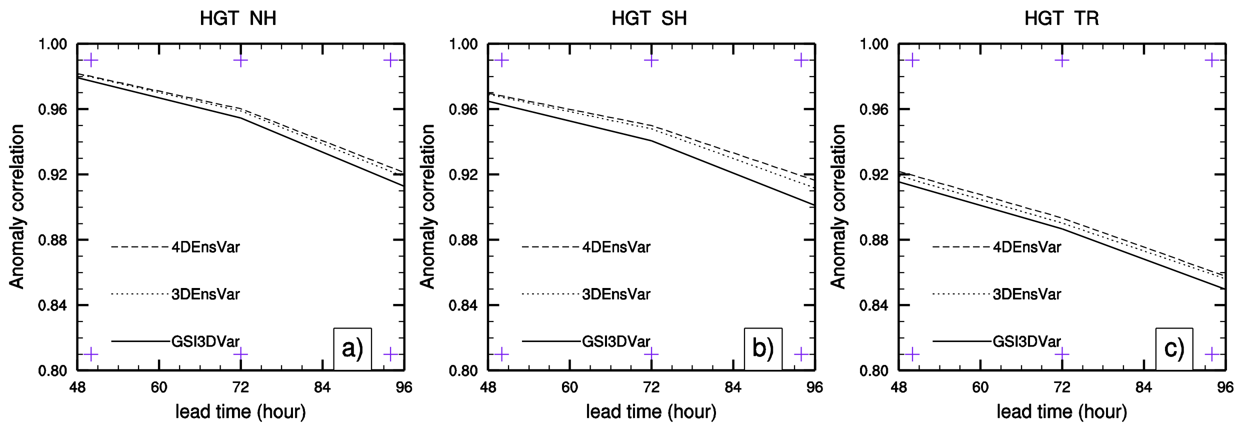

Figure 4. Averaged anomaly correlation of geopotential height at Northern Extratropics (NH,a), Southern Extratropics (SH, b) and Tropics (TR, c). Lines and symbol "+" are defined the same as Fig. 3.

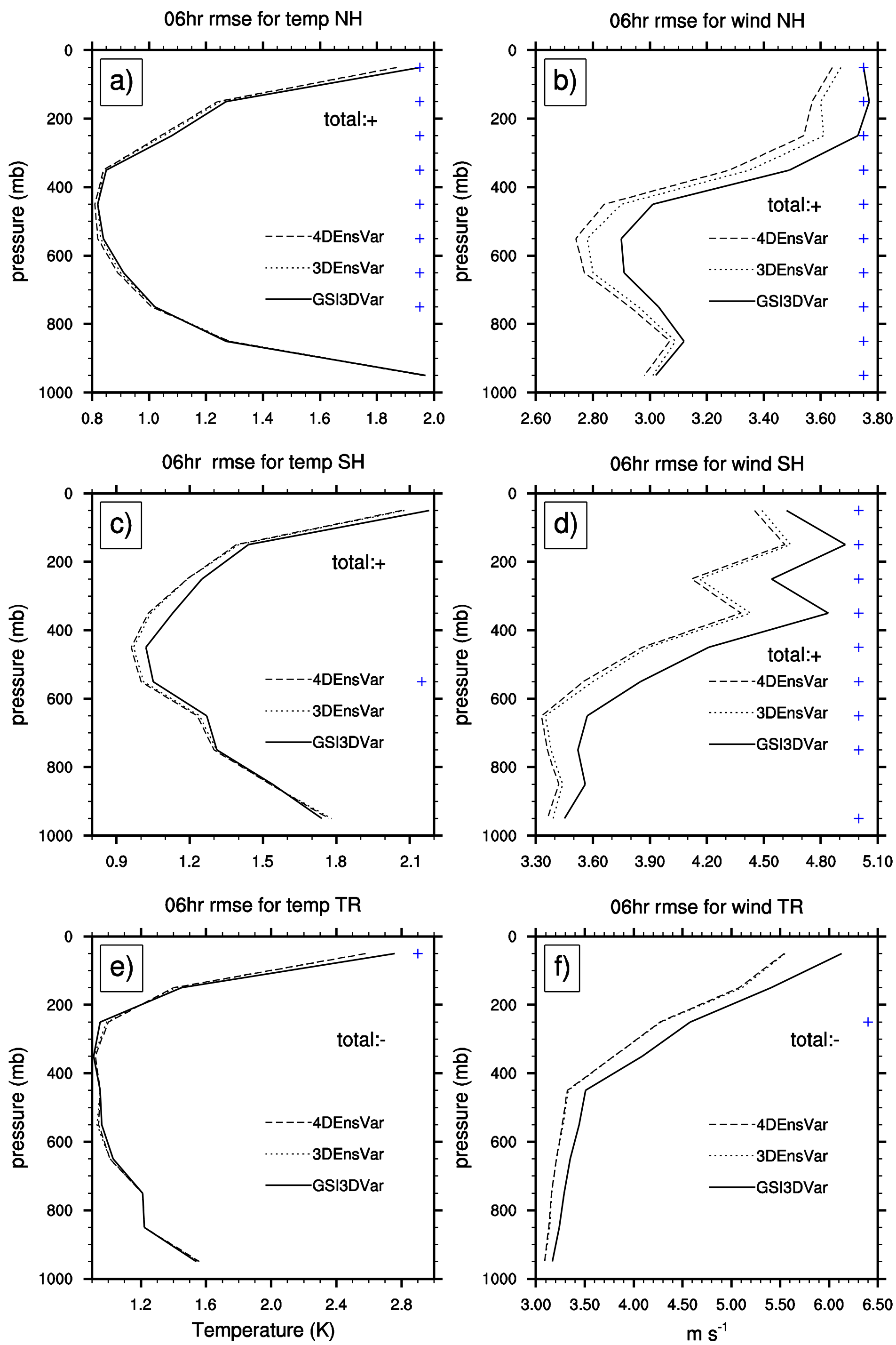

Figure 5. The root-mean-square fit of the 6-hour forecasts to the in-situ observations for temperature (left column) and wind (right column) as a function of pressure for the Northern Hemisphere extra-tropics (a,b), Southern Hemisphere extra-tropics (c,d) and tropics (e, f) for the GSI3DVar, 3DEnsVar, 4DEnsVar experiments. Line definition is the same as Figure 3. Blue "+" indicates the levels where 4DEnsVar is statistically significantly better than 3DEnsVar. Black "+" and black "-" after "total" denote if 4DEnsVar is or is not statistically significantly better than 3DEnsVar respectively after averaging over all levels.

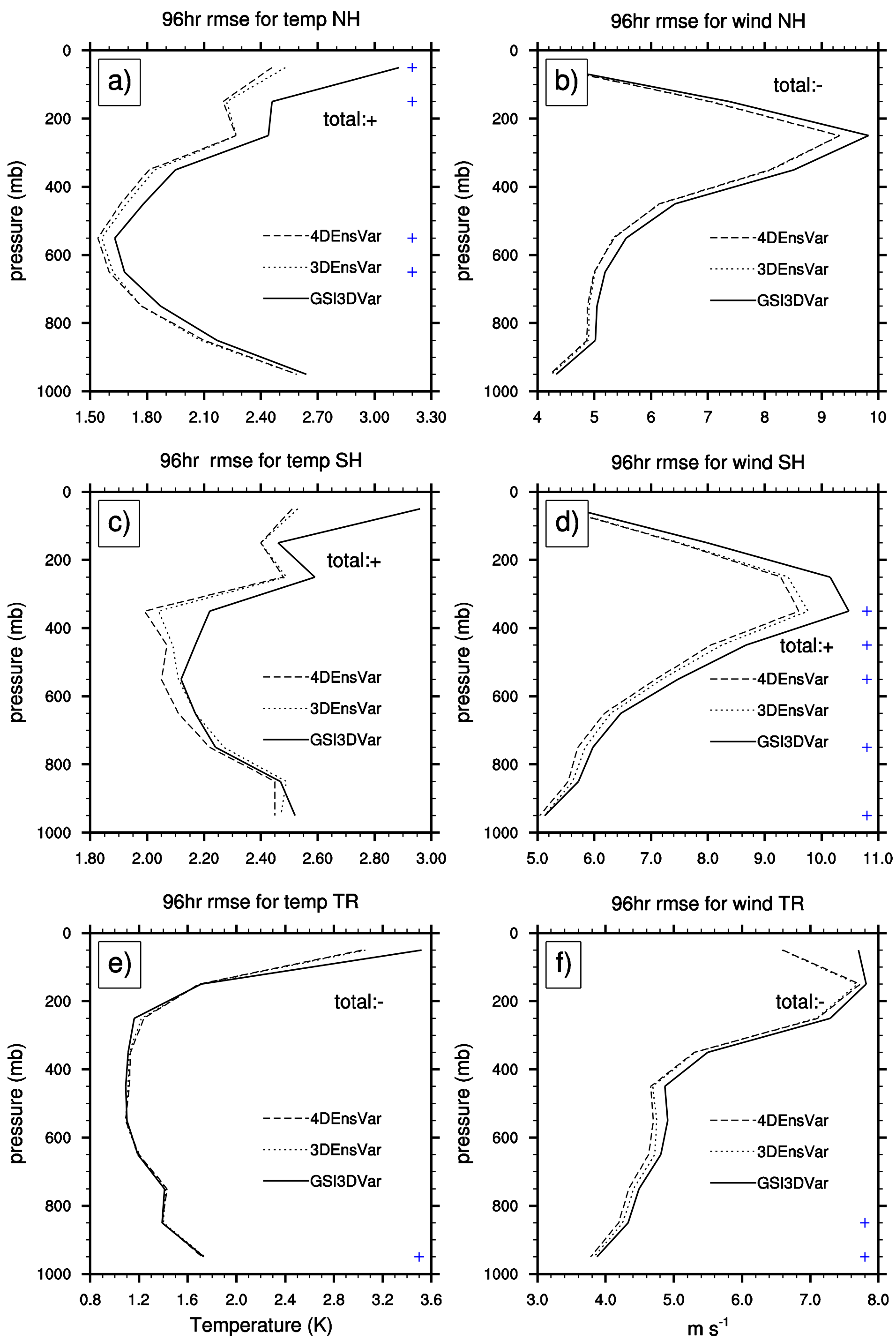

Figure 6. Same as Fig. 5 except at the 96-hour forecast lead time.

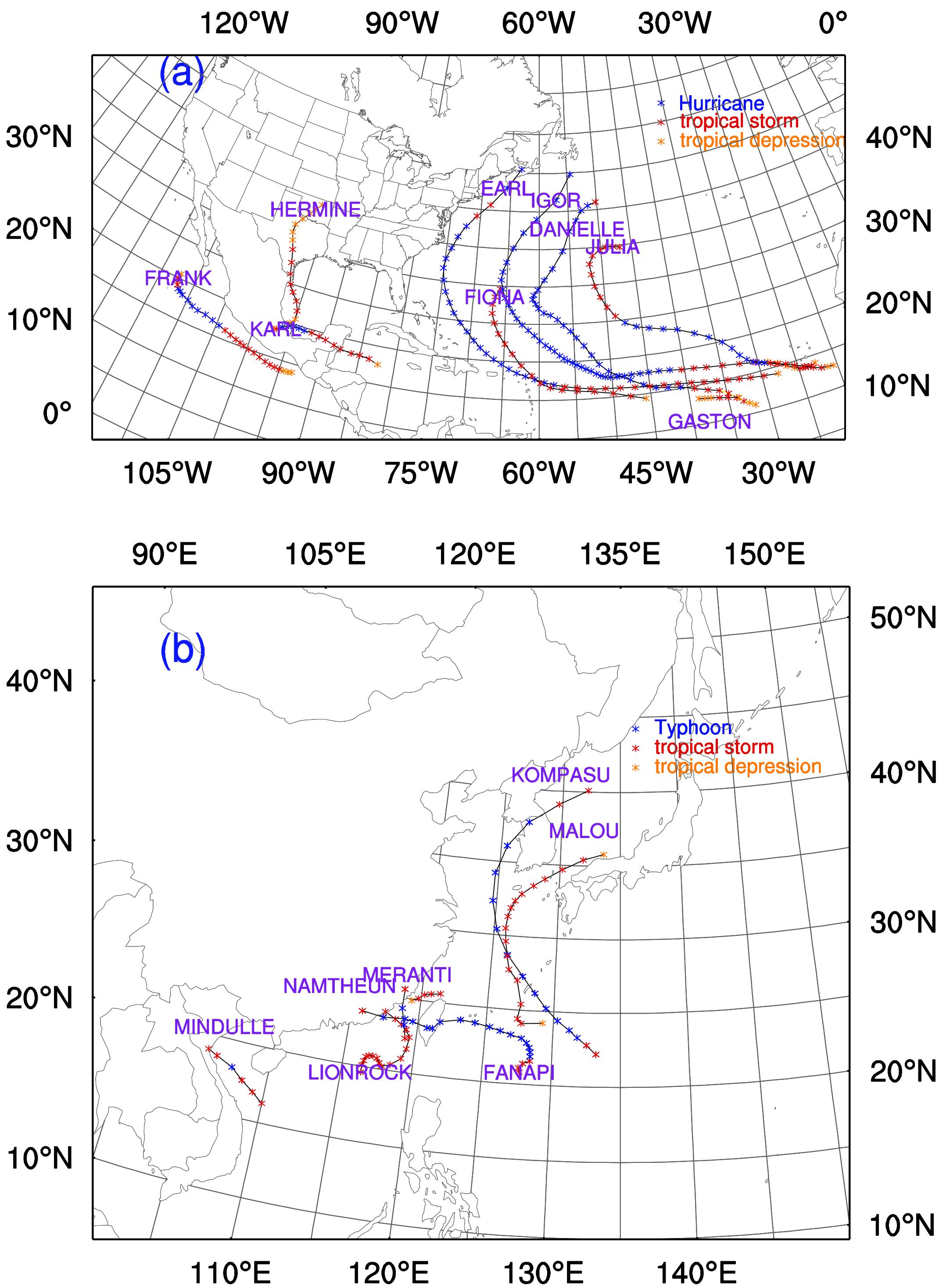

Figure 7. Tracks of tropical cyclones during the verification period in the Atlantic and East Pacific (a) and West Pacific basins (b).

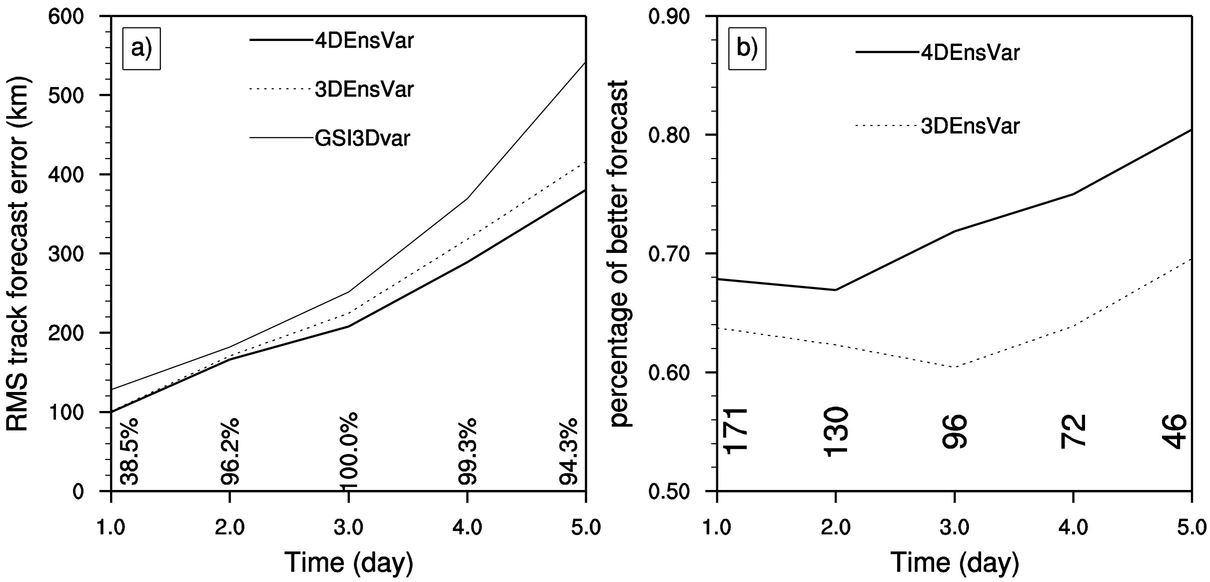

Figure 8. Skill of tropical cyclone track forecasts measured by (a) the root mean square errors and (b) percentage of forecasts that were better than the reference GSI3DVar forecast for 4DEnsVar (thick solid), 3DEnsVar (dotted) and GSI3DVar (thin solid) experiments. The numbers above the x-axis of (a) denotes the statistical significant confidence level of the difference between 4DEnsVar and 3DEnsVar. The differences among 3DEnsVar and GSI3DVar are statistically significant for all lead times considered in (a). The numbers above the x-axis of (b) denote the number of samples used in the calculation at each lead time.

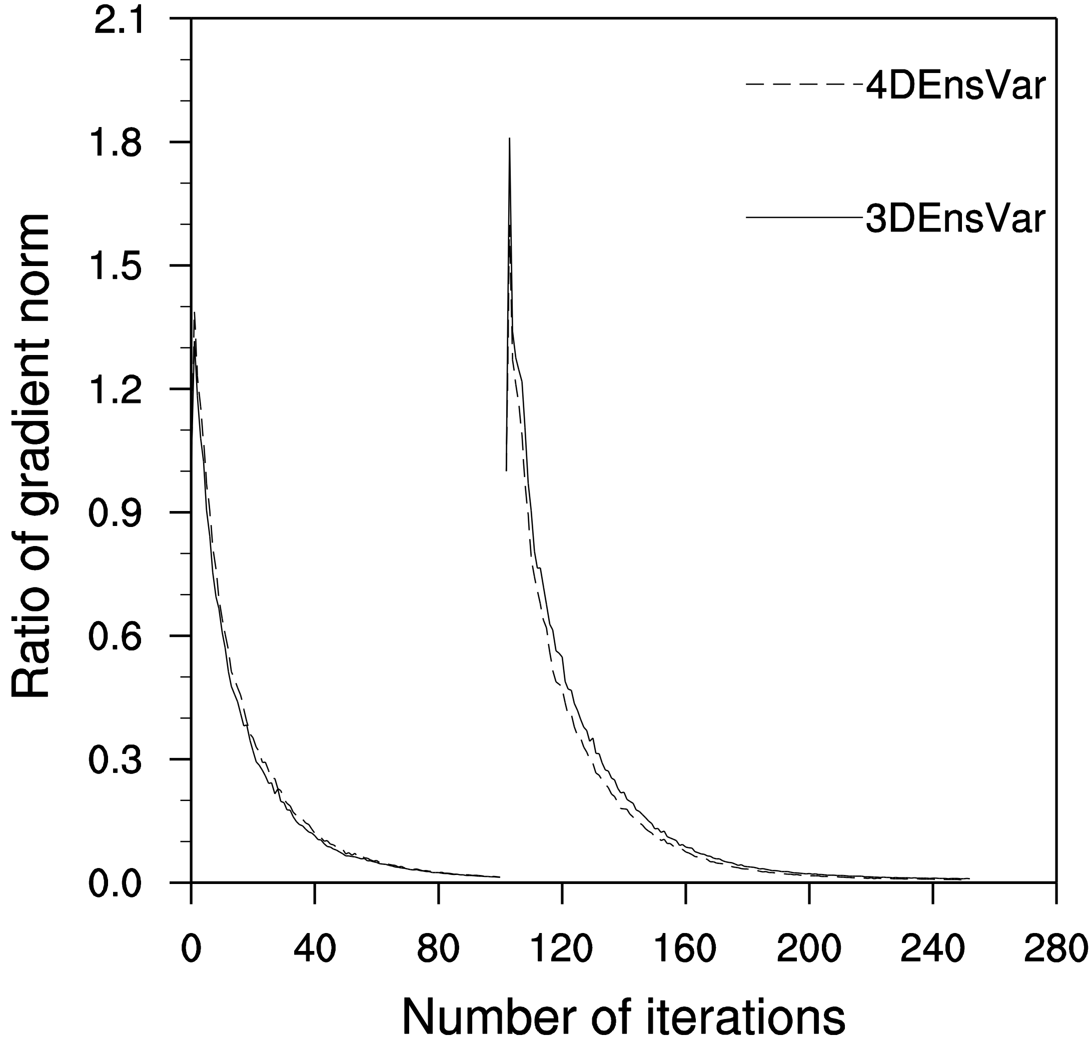

Figure 9. Averaged ratios of gradient norms as a function of the number of iterations in the first and second outer loops during the variational minimization of 3DEnsVar (solid) and 4DEnsVar (dash) experiments.

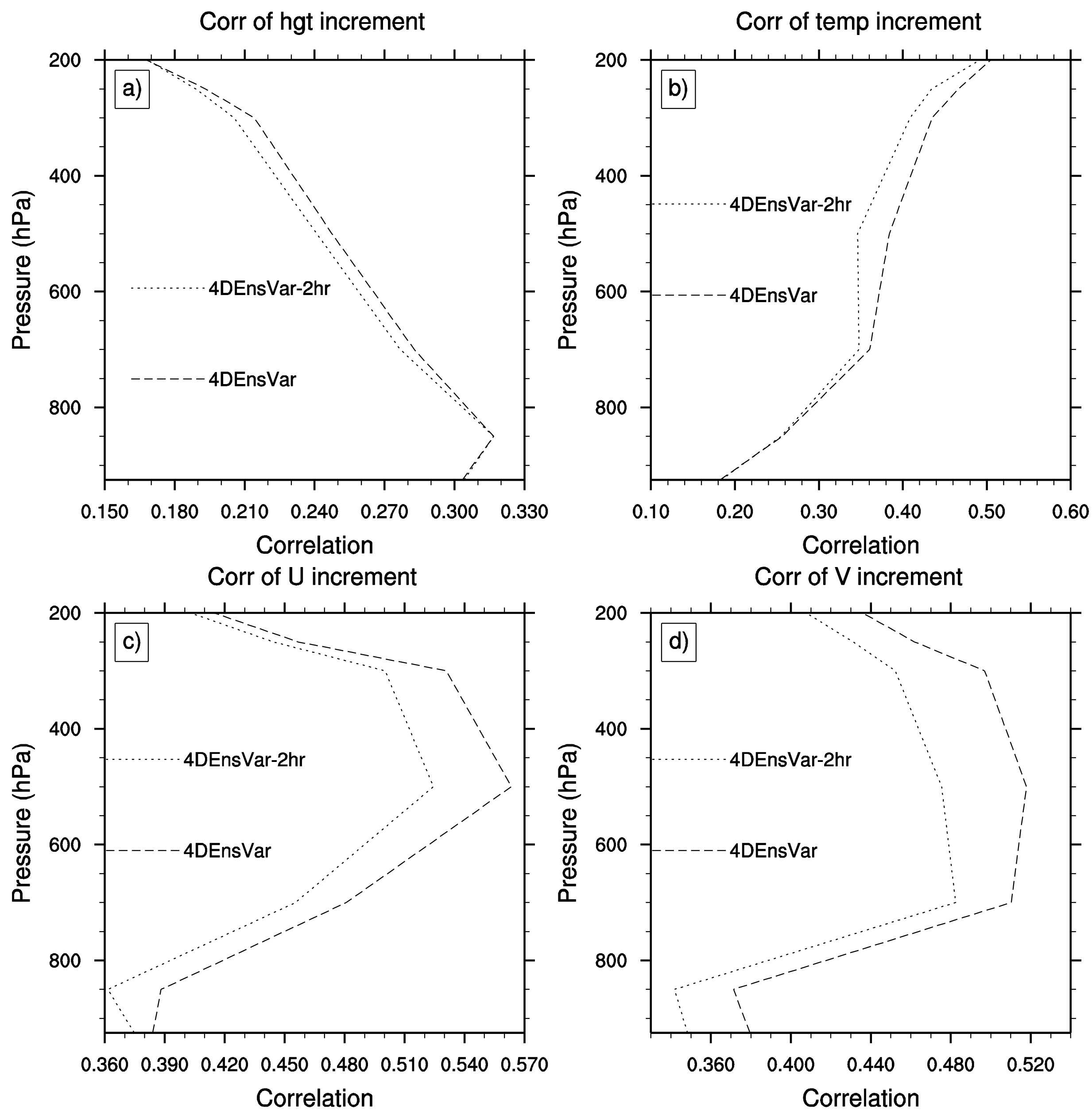

Figure 10. The correlation of analysis increments by 4DEnsVar using hourly ensemble perturbations (4DEnsVar) and using 2-hourly ensemble perturbations (4DEnsVar-2hr) with the increment by nonlinear model propagation. The increments are valid at the end of the 6-hour assimilation window by assimilating all observations within the first hour of the assimilation window. See text for more details.

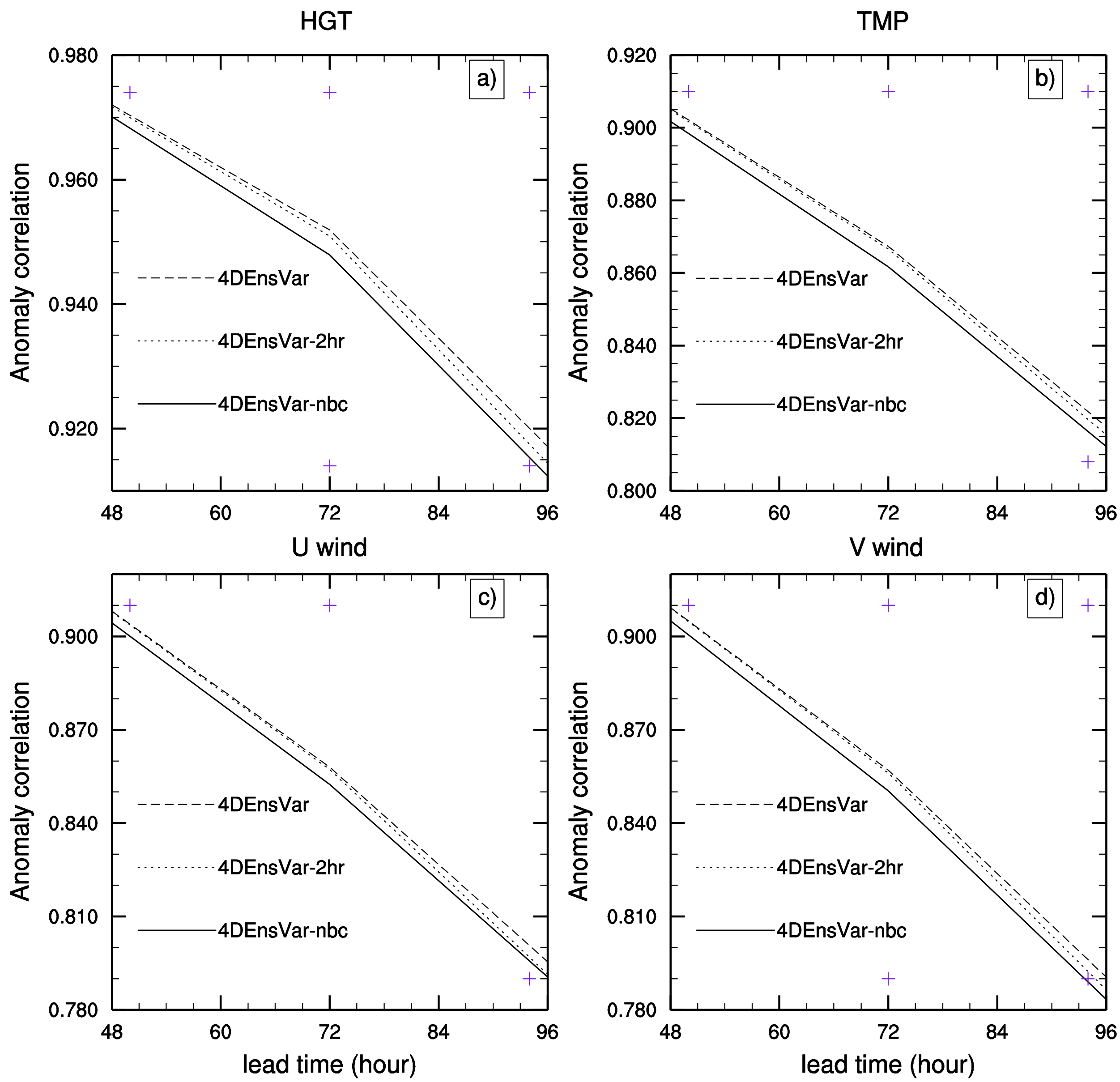

Figure 11. The globally averaged anomaly correlation of geopotential height (a), temperature (b), zonal (c) and meridional wind (d) forecasts verified against the ECMWF analyses. Solid, dotted and dashed lines are for 4DEnsVar-nbc, 4DEnsVar-2hr and 4DEnsVar experiments respectively. The “+” right above the lower x-axis denotes the lead time when the difference between 4DEnsVar and 4DEnsVar-2hr is statistically significant. The “+” right below the upper x-axis denotes the lead time when the difference between 4DEnsVar-2hr and 4DEnsVar-nbc is statistically significant.

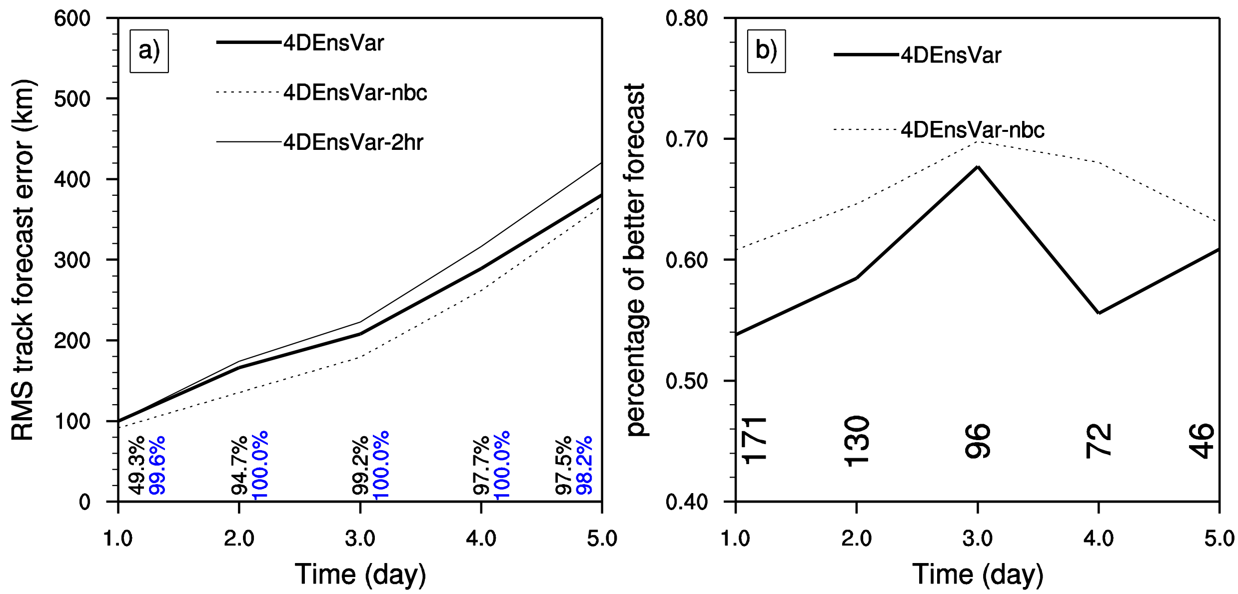

Figure 12. Skill of tropical cyclone track forecasts measured by (a) the root mean square errors and (b) percentage of forecasts that were better than the reference 4DEnsVar-2hr forecast for 4DEnsVar (thick solid), 4DEnsVar-2hr (thin solid) and 4DEnsVar-nbc (dotted) experiments. The black (blue) numbers above the x-axis of (a) denote the statistical significant confidence level of the difference between 4DEnsVar and 4DEnsVar-2hr (between 4DEnsVar-2hr and 4DEnsVar-nbc). The numbers above the x-axis of (b) denote the number of samples used in the calculation at each lead time.