An OSSE-based Evaluation of Hybrid Variational-Ensemble Data Assimilation for the NCEP GFS, Part II: 4DEnVar and Hybrid Variants

Daryl T. Kleist[1]
Department of Atmospheric and Oceanic Science
University of Maryland, College Park, Maryland

Kayo Ide
Department of Atmospheric and Oceanic Science, and Earth System Science Interdisciplinary Center, and Institute for Physical Science and Technology, and Center for Scientific Computation and Mathematical Modeling
University of Maryland, College Park, Maryland

Revised 12 October 2014

Monthly Weather Review

[1] *Corresponding author address*:
Dr. Daryl T. Kleist
Computer and Space Sciences Building
University of Maryland
College Park, MD 20742
E-mail: daryl.kleist@noaa.gov

Abstract

This work describes the formulation of a hybrid 4DEnVar algorithm and initialization options utilized within the National Centers for Environmental Prediction global data assimilation system. Initialization schemes that are proposed for use are the tangent linear normal mode constraint, weak constraint digital filter, and a combination thereof.

An observing system simulation experiment is carried out to evaluate the impact of utilizing hybrid 4DEnVar with various initialization techniques. The experiments utilize a dual resolution configuration, where the ensemble is run at roughly half the resolution of the deterministic component. It is found that by going from 3D to 4D, analysis error is reduced for most variables and levels. The inclusion of a time-invariant static covariance when used without a normal mode based strong constraint is found to have a small, positive impact on the analysis. The experiments show that the weak constraint digital filter degrades the quality of analysis, due to the use of hourly states to prescribe high frequency noise. It is found that going from 3D to 4D ensemble covariances has relatively larger impact in the extratropics, whereas the original inclusion of ensemble-based covariances was found to have the largest impact in the tropics. The improvements found in going from 3D to 4D covariances in the hybrid EnVar formulation are not as large as was found in Part I from the original introduction of the hybrid algorithm. The analyses generated by the 4D hybrid scheme are found to yield slightly improved extratropical height and wind forecasts, with smaller impacts on other variables and in general in the tropics.

1. Introduction

Four-dimensional variational data assimilation (4DVar) techniques that use tangent linear (Lewis and Derber 1985; Courtier et al. 1994) or linear perturbation models (Rawlins et al. 2007) and their corresponding adjoints have been shown to be powerful natural extensions to the 3DVar technique. In fact, 4DVar is the method of choice for initialization of single deterministic numerical weather prediction (NWP) applications at many operational centers (Rabier et al. 2000; Rosmond and Xu 2006; Gauthier et al. 2007; Rawlins et al. 2007). One attractive feature of 4DVar is that a dynamic model is used to help impose temporal smoothness and physical constraints on the analysis. Additionally, 4DVar allows for the simultaneous assimilation of asynchronous observations throughout a window at their appropriate times by producing a 4D analysis trajectory (Lorenc and Rawlins 2005). This is in contrast with the 3DVar-FGAT method (Rabier et al. 1998; Lawless 2010), which employs 4D model states at the appropriate time to compute innovations but only solves for a solution at a single time, typically at the center of a window. The major drawbacks to the 4DVar technique are the computational cost, complications related to developing and maintaining linearized forecast models and their corresponding adjoints, and basic assumption of linearity for the incremental formulation which may be particularly problematic for high resolution.

Much like 3DVar, 4DVar typically assumes a static error covariance but valid at the beginning of the assimilation window. The tangent linear (TL) and adjoint (AD) models then implicitly evolve this background error covariance as part of the variational solver. This procedure does allow for some flow-dependence, though the quality of the initial static background error covariance can still play a crucial role for short assimilation windows. Further, the use of a dynamic model to constrain the solution does help improve the multivariate aspects.

Much like for the 3D case, the development and application of a hybrid En4DVar technique, which utilizes ensemble-based covariances for helping to prescribe the background error covariance at the beginning of the assimilation window, has been shown to be beneficial (Buehner et al. 2010a,b; Zhang and Zhang 2012; Clayton et al. 2013; Kuhl et al. 2013). The drawbacks of such a method are the same as those for 4DVar, namely computational cost and the need for TL and AD models. As in Lorenc (2013), hybrid is used to describe a blended covariance and 4DVar implies the use of TL and AD models.

Along the lines of the 4D ensemble-based techniques (Hunt et al. 2004) such as the 4D-Local Ensemble Transform Kalman Filter (LETKF; Hunt et al. 2007), several methods expanding on the idea introduced by Lorenc (2003) have recently been proposed to utilize 4D ensemble perturbations within a variational framework (EnVar, Lorenc 2013) to help solve for a 4D-analysis increment without the need for TL and AD models (Liu at al. 2008; Tian et al. 2008; Liu et al. 2009; Tian et al. 2011; Buehner et al. 2010a,b; Liu and Xiao 2013). While most of the methods in these previous studies rely exclusively on an ensemble-based error covariance, several recent studies did combine the ensemble covariances with time-invariant static covariances in their 4DEnVar, i.e. hybrid 4DEnVar (Buehner et al. 2013; Desroziers et al. 2014; Lorenc et al. 2014; Wang and Lei 2014). Formulating the problem in the variational framework allows one to take full advantage of the many developments that have taken place over the years, such as dynamic constraints (Gauthier and Thépaut 2001; Kleist et al. 2009) and variational bias correction (Derber and Wu 1998; Dee 2005; Zhu et al. 2014).

In Part I (Kleist and Ide 2014), it was demonstrated that including ensemble covariances in a variational-based hybrid algorithm yielded improvements in the quality of analyses and subsequent forecasts for the National Centers for Environment Prediction (NCEP) global

67 forecast system (GFS) model in the context of an observing system simulation experiment
68 (OSSE). The experiments were performed using 3DVar and hybrid 3DEnVar, leaving
69 significant room for improvement. Without access to the TL and AD models, a natural extension
70 of the hybrid 3DEnVar to include 4D ensemble perturbations (hybrid 4DEnVar), is a logical next
71 step for improving upon the previous work.

72 Several previous studies have investigated the use of 4DEnVar for use with global
73 deterministic NWP. A 4DEnVar algorithm was implemented into a prototype observation space
74 global data assimilation system for the United States Navy in Bishop and Hodyss (2011). There,
75 they focused on an adaptive localization algorithm within the context of a single case study.
76 Buehner et al (2010b) performed an inter-comparison study for the Canadian operational global
77 NWP model, and found that the 4DEnVar improved upon their operational, non-hybrid 4DVar in
78 the tropics and southern hemisphere, but not in the northern hemisphere. It was also found that
79 4DEnVar performed slightly worse than a hybrid 4DVar. Buehner et al. (2013), again using the
80 operational Canadian system, found that while the use of 4D instead of 3D ensemble covariances
81 did result in small, consistent improvements in their EnVar for deterministic NWP, the gains
82 were not as large as found when going from 3DVar to 4DVar. They also found that the
83 performance of their 4DEnVar was comparable to or better than their 4DVar, except for the
84 extratropical summer regions. Using the United Kingdom Met Office global system, Lorenc et
85 al. (2014) found that both hybrid 4DVar and 4DEnVar beat their 3D counterparts in NWP trial
86 runs. However, it was also found that their hybrid 4DEnVar did not perform as well as their
87 hybrid 4DVar, due to the heavy reliance on the climatological covariance that the 4DVar
88 algorithm can propagate through the assimilation window whereas the EnVar algorithm cannot.
89 Lastly, Wang and Lei (2014) performed a comparison study of hybrid 4DEnVar with 3DEnVar

using the NCEP GFS (model) at low resolution. They found that 4DEnVar was better than 3DEnVar, with a larger impact in the extratropical troposphere than in the tropics. They also found that analysis increments from the 4DEnVar algorithm were more balanced than 3DEnVar, and as in a companion 3DEnVar study (Wang et al. 2013), the use of a dynamic constraint was valuable for improving the forecast skill of the 4DEnVar initialized forecasts in the extratropics.

This work is similar to the aforementioned studies that investigated the use of 4DEnVar for initializing global determinist NWP. Here, the NCEP GFS is utilized as in Wang and Lei (2014), however, an OSSE is used instead of real observations as in Part I (Kleist and Ide 2014). OSSEs have the distinction of allowing for the calculation of actual analysis error since the "truth" is known. Furthermore, the experiments utilize a dual-resolution paradigm as in Part I. Given the importance of initialization within the context of 4DEnVar for the NCEP GFS system (Wang and Lei 2014) and the Met Office system (Lorenc et al. 2014), we also aim to corroborate previous findings as well as explore alternate, computationally efficient initialization options. The remainder of the manuscript is organized as follows. Section 2 describes the implementation of the 4D extension to the hybrid including a time-invariant static error covariance supplement. Section 3 then follows with a description of various OSSE-based experiments that demonstrate the impact of utilizing 4DEnVar and hybrid variants relative to the 3D hybrid experiments that were carried out in Part I. Several experiments are carried out to demonstrate the impact of including a static error covariance in the dual-resolution 4DEnVar paradigm, as well as to show the impact of various dynamic constraints. A summary and motivation for future work then follows.

2. 4D Extensions of 3D Hybrid

A. GSI-BASED HYBRID

Traditionally, incremental 4DVar involves solving for the optimal solution ($\mathbf{x}_0'$) at the beginning of a time window, obtained by minimizing a cost function

$$J(\mathbf{x}_0') = \frac{1}{2}(\mathbf{x}_0')^{\mathrm{T}} \mathbf{B}_{\mathrm{f}}^{-1}(\mathbf{x}_0') + \frac{1}{2}\sum_{k=1}^{K}(\mathbf{H}_k\mathbf{x}_k' - \mathbf{y}_k')^{\mathrm{T}} \mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{x}_k' - \mathbf{y}_k') \qquad (1)$$

$$\mathbf{x}_k' = \mathbf{M}_k\mathbf{x}_0' \qquad (2)$$

where $\mathbf{B}_{\mathrm{f}}$ is the static background error covariance, $\mathbf{R}$ the observation error covariance, $\mathbf{H}$ the linearized observation operators, and $\mathbf{y}'$ the observation innovations. Note that preconditioning is being ignored for simplicity. The solution is determined by using the asynchronous observations up to K-time levels and the TL model ($\mathbf{M}_k$) for each time-level index (k). The AD model (transpose of $\mathbf{M}_k$) is necessary in order to obtain the gradient for the minimization. This cost function can then be extended to a hybrid 4DVar cost function by including an ensemble control variable ($\boldsymbol{\alpha}^n$) for the ensemble contribution to the analysis increment at the beginning of the window:

$$\begin{aligned} J(\mathbf{x}_{\mathrm{f}}', \boldsymbol{\alpha}) = \beta_{\mathrm{f}} \frac{1}{2}(\mathbf{x}_{\mathrm{f}}')^{\mathrm{T}} \mathbf{B}_{\mathrm{f}}^{-1}(\mathbf{x}_{\mathrm{f}}') + \beta_{\mathrm{e}} \frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{\alpha}^n)^{\mathrm{T}} \mathbf{L}^{-1}(\boldsymbol{\alpha}^n) + \\ \frac{1}{2}\sum_{k=1}^{K}(\mathbf{H}_k\mathbf{x}_k' - \mathbf{y}_k')^{\mathrm{T}} \mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{x}_k' - \mathbf{y}_k') \end{aligned} \qquad (3)$$

where $\mathbf{L}$ denotes the error covariance for the ensemble control variable, specified to be unit amplitude and Gaussian in structure, thereby acting to enforce localization of the ensemble based error covariance. Here, the localization is assumed to be static and does not evolve with the flow, which Fairbairn et al. (2014) demonstrates to be potentially detrimental to 4DEnVar

relative to 4DVar using a toy model. Just as in the 3D hybrid case, $\mathbf{x}_0'$ becomes a linear combination of that which is derived from a static error covariance ($\mathbf{x}_\mathrm{f}'$) and that which is derived from the ensemble perturbations ($\mathbf{x}_\mathrm{e}^n$) as prescribed by the control variable ($\boldsymbol{\alpha}^n$) for an ensemble of size N:

$$\mathbf{x}_0' = \mathbf{x}_\mathrm{f}' + \mathbf{T}\sum_{n=1}^{N}\left(\boldsymbol{\alpha}^n \circ \left(\mathbf{x}_\mathrm{e}\right)_0^n\right) \quad (4)$$

Here, both $\mathbf{x}_\mathrm{f}'$ and $\left(\mathbf{x}_\mathrm{e}\right)_0^n$ are valid at t=0 and the hybrid analysis increment is propagated to each time level (k) by the TL as in Eq. (2). For a dual-resolution configuration, $\mathbf{T}$ represents the interpolation (or other manipulation) from ensemble to deterministic resolution. This hybrid configuration will hereafter be referred to as H-4DVar, recalling that it requires the AD (and forward TL) model within the minimization.

As pointed out in previous studies (Buehner 2013; Desroziers et al. 2014; Lorenc et al. 2014; Wang and Lei 2014), this can be further manipulated to solve a similar 4D cost function as in (4) without the need for the TL and the AD by utilizing 4D ensemble perturbations (hybrid 4DEnVar), where the increment valid at each of the observation time levels can be prescribed as

$$\mathbf{x}_k' = \mathbf{M}_k\mathbf{x}_\mathrm{f}' + \mathbf{T}\sum_{n=1}^{N}\left(\boldsymbol{\alpha}^n \circ \left(\mathbf{x}_\mathrm{e}\right)_k^n\right). \quad (5)$$

For hybrid 4DEnVar, $\mathbf{M}_k$ is simply chosen to be the identity model, where a single static contribution is valid through the whole window as in hybrid 3DEnVar. The ensemble contribution to the increment now utilizes 4D, nonlinear perturbations associated with each of the observation bins. In this particular formulation, $\boldsymbol{\alpha}^n$ is assumed to be the same throughout the assimilation window, analogous to the weights in a 4D-LETKF without temporal localization. More details regarding the GSI 4DEnVar formulation can be found in Wang and Lei (2014).

Alternate formulations and practical implementations of 4DEnVar are provided in Desorziers et al. (2014).

To allow for a single time level, the time invariant contribution, $\mathbf{x}'_{\mathrm{f}}$ is simply chosen to be prescribed in the center of the window and $\mathbf{M}_k$ is set to be an identity model. One could formulate an alternate hybrid and utilize the TL and AD models to evolve the static contribution *only*, prescribe $\mathbf{x}'_{\mathrm{f}}$ to be at the beginning of the window, and linearly combine with the 4D ensemble contribution, as in Eq. (5). For the purposes of this work, only the former possibility where the identity model is used and the static contribution is defined to be valid at the center of the window is considered due to the computational advantage. The special case where the static contribution is set to zero ($\beta_{\mathrm{f}}^{-1}=0$) will be referred to as 4DEnVar, whereas any configuration that utilizes a time-invariant static contribution will be referred to as hybrid 4DEnVar (or H-4DEnVar). Given a configuration that utilizes a dual-resolution option, the static contribution is aimed at accomplishing two things: 1) filling the null space at higher spatial frequencies that the low resolution ensemble cannot resolve, assuming there are observations at such frequencies, and 2) providing more search directions for the minimization, complementing the relative few such search directions available via localized linear combinations of the ensemble perturbations.

The necessary components to solve the GSI using the formulation that is described by Eqs. (3) and (5), 4DEnVar, are already in place. The ensemble control variable for the hybrid is developed and implemented as part of the 3D-hybrid work described in Part I. The 4D ensemble perturbations are derived from a serial square root filter form of an ensemble Kalman filter (EnKF; Whitaker and Hamill 2002; Whitaker et al. 2008). A 4DVar capability within the GSI was previously developed through collaboration with colleagues at NASA GMAO (Todling and Trémolet, personal communication). A key component of the extension of 3DVar to 4DVar

within the GSI was the addition of new observation handling features and the inclusion of time binning. All configurations (3D and 4D) utilize similar default observation selection. With these pieces in place, the 3D hybrid capability is extended to allow for 4D ensemble perturbations and a 4DEnVar option.

B. SINGLE OBSERVATION EXAMPLE

To demonstrate the impact of the 4DEnVar algorithms, a set of experiments that assimilate only a single simulated observation is performed. A summary and description of the experiments is available in Table 1. The two hybrid cases that include a static and ensemble contribution to the increment are done so with $\beta_f^{-1}=0.25$ and $\beta_e^{-1}=0.75$. For the two 4DVar variants, the TL and AD that are utilized in the inner loop are adaptations of a simple model that has been under development at NCEP for the sole purpose of running 4DVar. Although not yet ready for large problems or cycling due to its current computational inefficiencies, it is still quite useful for evaluating low-resolution single observation experiments that require only a few iterations to minimize. For computational reasons, a single resolution of T126 with a 40 member ensemble is utilized for the single observation example for demonstration. All of the solutions that have an ensemble contribution to the increment use Gaussian localization with a cutoff of 1200 km in the horizontal and 1.2 scale heights in the vertical. The single observation chosen for the experiments is a mid-tropospheric temperature observation taken at the beginning of a 6-hr assimilation window, three hours prior to the "update" analysis time, and is assigned a negative two degree departure from the background at that time (a 3-hr forecast) and a one degree observation error standard deviation. The observation is chosen to reside upstream from the base

of a shortwave trough, similar to the single observation experiments shown in Buehner et al. (2010a) for comparison.

The resultant analysis increment at the middle of the assimilation window, three hours after the observation is taken, for the various 4D configurations is shown in Fig. 1. All four experiments show the maximum increment downstream from where the observation was taken consistent with the northwesterly background flow. The 4DVar experiment results in a spatially broad, quasi-Gaussian temperature increment (Fig. 1, upper left). This is not terribly surprising given that only three hours elapse between the time that the observation was taken at the beginning of the window and the analysis time at the center of the window. The three experiments that utilized ensemble covariances exhibit a temperature increment that is stretched along the height gradient as would be expected (Fig. 1). All of the experiments show a cyclonic wind response to the cold temperature observation and increment, with the ensemble- and hybrid-based experiments showing a stronger wind response than the 4DVar case. It is clear that the 4DEnVar case suffers from sampling (spurious correlations) more so than the two hybrid variants. This is not surprising either, in this case, given that a 40-member ensemble is utilized in combination with quite broad localization. However, adding a time-invariant, static contribution to the increment helps to reduce the impact of these apparent problems substantially, without hurting the 4D nature of the increment. In fact, the H-4DEnVar increment is qualitatively and quantitatively very similar to that from the H-4DVar experiment, despite the fact that one uses a fairly simply dynamic model while the other utilizes 4D, nonlinearly evolved ensemble perturbations. This is in contrast to the single observation experiments carried out in Lorenc et al. (2014), where the more heavily reliance on a static, time-invariant contribution to the increment was shown to be problematic in the H-4DEnVar paradigm.

The same single observation test can be utilized to further investigate how the algorithms handle the propagation of information through the assimilation window, by visualizing the analysis increment at various k-time levels. Recall that for the 4DVar cases, this involves an explicit propagation of the increment through the use of a linear, dynamic model, i.e. Eq. (2). For the 4DEnVar variants, the propagation of information is achieved implicitly through correlations contained within the 4D ensemble perturbations, i.e. Eq. (5) with $\mathbf{M}_k=\mathbf{I}$. The time evolution of the temperature and wind increments at 500 hPa in 3-hr intervals for the H-4DVar and H-4DEnVar cases is shown in Fig. 2. First, note that the solutions of the two experiments are identical at the beginning of the window, the time at which the observation was taken. This is expected given the same ensemble perturbations, relative weights, and static error covariance, since the TL and/or nonlinear models have not evolved the increments in the two experiments. This is not the case for observations taken at other times in the assimilation window (not shown), where for the H-4DVar case, the effective "static" error covariance is evolved forward in time consistent with the linear dynamic model. For other times in the six hour assimilation window, middle and end respectively, the temperature and wind increments for the two hybrid cases are strikingly similar both qualitatively and quantitatively. By the end of the window, the H-4DEnVar case has more small amplitude features away from the location of the original observation and has a slightly tighter structure aligned with the height contours. The H-4DVar case, on the other hand, has somewhat broader features at the end of the window. It is quite encouraging that the computationally cheaper ensemble method can mimic quite nicely the evolution of the increment within the assimilation window. Although not shown, other single observation experiments for other variables and levels reveal similar results. A similar comparison for different single observation tests can be found in Lorenc et al. (2014).

3. Constraints on high frequency noise

The 4DEnVar (and hybrid) option is implemented in such a way as to allow for the application of many of the standard features included in the 3D GSI such as variational quality control, variational satellite bias correction, the tangent linear normal mode constraint (TLNMC; Kleist et al. 2009), and various weak constraints such as digital filter (Gustafsson 1993; Polavarapu et al. 2000; Gauthier and Thépaut 2001). However, the 4DEnVar solution to the analysis problem requires special attention when it comes to the application of such constraints given that the 4D aspect of the problem is obtained through a dot product of the weights and the ensemble perturbations in a discrete manner without the explicit use of a dynamic model. The TLNMC and the weak constraint digital filter, as well as the combination of the two, are explored within the context of 4DEnVar. In this work, the constraints are imposed on the high-resolution deterministic analysis. While the EnKF analysis is re-centered about the filtered, deterministic anlaysis, the individual members are only explicitly filtered using the full-field digital filter of the GFS model.

A. TANGENT LINEAR NORMAL MODE CONSTRAINT

For a dynamic constraint such as the TLNMC, the minimization procedure of the cost function within the 4D context can be prohibitively expensive if one were to use hourly time levels and a six-hour assimilation window, as an example. Furthermore, the original implementation of the normal mode constraint for application within hybrid 3DEnVar requires the filter to be applied to the total analysis increment as a sum of static and ensemble contributions without much flexibility, as in Eq. (2) from Part I (Kleist and Ide 2014). For these

reasons, several options related to the application of various constraints are explored and implemented into the analysis code.

Building on previous successes in implementing the TLNMC in the 3DVar (Kleist et al. 2009) and hybrid 3DEnVar (Part I) context, various possibilities for application of such a constraint in the 4DEnVar and hybrid context are developed and tested. The first possibility is to apply the TLNMC to only the static contribution to the increment:

$$\mathbf{x}'_k = \mathbf{C}\,\mathbf{x}'_{\mathrm{f}} + \mathbf{T}\sum_{n=1}^{N}\left(\boldsymbol{\alpha}^n \circ \left(\mathbf{x}_{\mathrm{e}}\right)_k^n\right). \quad (6)$$

This requires a single constraint operation per iteration, making the cost equivalent to 3D applications. Note that this capability can also be applied within the 3D hybrid context, though previous results demonstrate that the inclusion of the constraint to the total increment aids in reducing potential imbalances introduced by the ensemble based increment (Wang et al. 2013; Wang and Lei 2014). This formulation, given by Eq. (6), is motivated by the desire to allow for the use of the unfiltered ensemble covariance while still filtering the static contribution to the increment.

In order to help mitigate imbalances that can be introduced through sampling and localization, two additional possibilities to apply the TLNMC to the total increment are explored in the H-4DEnVar (and 4DEnVar) context. One is to simply apply the constraint over all k-time levels:

$$\mathbf{x}'_k = \mathbf{C}_k\left[\mathbf{x}'_{\mathrm{f}} + \mathbf{T}\sum_{n=1}^{N}\left(\boldsymbol{\alpha}^n \circ \left(\mathbf{x}_{\mathrm{e}}\right)_k^n\right)\right]. \quad (7)$$

Here, the balance operator $\mathbf{C}_k$ is also denoted with a time level index, since the possibility exists for linearization about the background for each time level, as the constraint is in fact tangent-linear. However, for practical reasons such as computational cost and memory, the background state used for linearization in $\mathbf{C}$ is taken from the center of the observation window and applied

291 to all k time levels. Wang and Lei (2014) utilized the constraint in this manner for their
292 4DEnVar experiments and found improvements in forecast skill in the extratropical troposphere.
293 Even still, the application of the constraint in this manner can be prohibitively expensive. For a
294 six-hour window with hourly time levels and hourly ensemble perturbations, the computational
295 cost of the analysis goes up substantially. For this reason, another possibility to apply the full
296 constraint only to the solution in the middle of the window is considered. The advantage of such
297 a method is that the increment that is applied to the background and used to restart the model can
298 be filtered explicitly reducing spin-up and spin-down, without the considerable cost of having to
299 filter all time levels. This of course introduces inconsistencies between the incremental solution
300 in the center of the window and the other time-levels, which could have undesirable
301 consequences. All applications of the TLNMC referred to hereafter utilize the default 8 vertical
302 modes and a single iteration as in Kleist et al. (2009) and Part I (Kleist and Ide 2014).
303

304 B. WEAK CONSTRAINT DIGITAL FILTER

305 Digital filter weak constraints have been shown to be beneficial in many 4DVar
306 applications (Polavarapu et al. 2000; Gauthier and Thépaut 2001; Wee and Kuo 2004). A similar
307 constraint has previously been implemented by colleagues at NASA GMAO for the 4DVar
308 extension of the GSI.[1] We denote this digital filter initialization as the JCDFI. The formulation
309 of the JCDFI involves the addition of a new penalty term (see Gustafsson 1993 or Polavarapu et
310 al. 2000 for a more detailed derivation):

311 $$J_{\mathrm{dfi}} = \chi\left\langle \mathbf{x}_m - \mathbf{x}^{\mathrm{i}}_m, \mathbf{x}_m - \mathbf{x}^{\mathrm{i}}_m \right\rangle \quad (8)$$

[1] The JCDFI that was originally developed for 4DVar applications was not implemented for the standard double conjugate gradient minimization algorithm that is utilized for the hybrid applications. Minor modifications were necessary to adapt the code for use in this context.

where $\mathbf{x}_m^{\mathrm{i}}$ denotes a filtered or "initialized" state at time m (in the case of 4DEnVar, assumed to be the center of the assimilation window), and χ a general weighting factor. This weighting parameter is typically denoted α in the literature but avoided in this context to remove confusion with the ensemble control variable. The filtered state is constructed from the 4D increment using the same filter coefficients (h) as in the standard formulation (Lynch and Huang 1992)

$$\mathbf{x}_m^{\mathrm{i}} = \sum_{k=1}^{K} \mathrm{h}_{k-m} \mathbf{x}_k^{\mathrm{u}} \quad (9)$$

where $\mathbf{x}_k^{\mathrm{u}}$ denotes the unfiltered increment at each time level-k. The norm for the penalty function in Eq. (8) is chosen to be a dry energy norm, though the capability does exist to use a moist energy norm. Such a constraint in the 4DEnVar context, on one hand, potentially allows for noise control within the 4D increment without adding much computational cost. On the other hand, high frequency noise cannot be described by the hourly incremental states in this context. Furthermore, there are no means to correct the origin of the high frequency noise with links to the actual model equations as would be the case for a 4DVar-based formulation. Despite these limitations, it does allow for a cost effective alternative to employ some filtering of the 4D increment. Additionally, it is possible to explore the use of a combination of noise constraints, weak constraint digital filter and TLNMC, to improve the quality of analysis.

C. SINGLE ANALYSIS IMPACT

To ensure that the constraints are all properly functioning for use with the 4DEnVar paradigm within GSI, a single analysis test case that assimilates real observations including satellite radiances within a 6-hr window valid at 06 UTC 15 July 2010 is performed. The background and ensemble members are generated from an offline experiment that utilizes the dual-resolution hybrid configuration, but at higher resolution than the OSSE-based cycled

335 experiments described in Part I, with T574 for the high-resolution deterministic component and
336 T254 for the 80-member ensemble. Double inner-loops of 100 iterations each are utilized,
337 whereby the quality control, observation selection, and re-linearization are performed and the
338 nonlinear model is not rerun. The analysis is run with four separate configurations: 1) 4DEnVar,
339 2) 4DEnVar with TLNMC on the increment over all time levels, 3) 4DEnVar with JCDFI, and 4)
340 4DEnVar with the TLNMC imposed at the center of the window only combined with the JCDFI.
341 The fourth configuration that utilizes both constraints is done with computational considerations
342 in mind. In all four configurations, there is no static contribution to the solution, i.e., $\beta_f^{-1}=0$.
343 The desire is to have the TLNMC remove gravity mode tendencies in the middle of the window
344 while allowing the weak constraint digital filter to maintain the connection to the other time
345 levels through the procedure to get the filtered state, and penalize additional noise accordingly,
346 without having to use the normal mode procedure explicitly over all time levels.

347 The divergence increment for the experiments that ran with the JCDFI is damped across
348 the entire spectrum, including at large scales (Fig. 3). Some interesting behavior is observed on
349 the high spatial frequency part of the spectrum, most of which is an artifact of the dual-resolution
350 aspect of the configuration. The aliasing that results from the interpolation within the current
351 dual-resolution configuration is an area of active research and the subject of a future manuscript.
352 As expected, the more the solution is constrained, the more difficult it becomes to draw to the
353 observations (Fig. 4). At the end of the first inner loop minimization (100 iterations), the penalty
354 reduction is greatest for the 4DEnVar case with no extra constraints, followed by the
355 experiments that utilized a single constraint (TLNMC and JCDFI), followed by the configuration
356 that used both constraints. At the end of the second inner loop minimization, the order is
357 slightly changed, but the final observation penalty is within one percent for all four

configurations. The minimization appears well behaved with a steady reduction of the gradient norm in all four experiments. By design, the TLNMC has a significant impact on the incremental tendencies (Kleist et al. 2009), particularly the gravity mode tendencies (Table 2). On the other hand, while the JCDFI reduces the total tendencies by a factor of two, it has very little impact on the amount of tendencies that project onto undesirable gravity modes due to the aforementioned issues regarding the inability to filter the highest frequency noise using only hourly states. The combined (COMB) configuration that utilizes JCDFI and TLNMC in the center of the window appears to be the best compromise, by significantly reducing the total tendencies while also reducing the percentage that project onto gravity modes, all at a significantly cheaper computational cost than running the normal mode constraint on all time levels. The inclusion of a static error covariance (H-4DEnVAR) also acts to decrease the incremental tendencies slightly while improving the ratio of gravity mode to total tendencies.

4. Cycled Experiment Results

A. ANALYSIS ERROR COMPARISON

A series of experiments is designed to study the impact of the 4D-ensemble and hybrid components as an extension to the 3D experiments that were described in Part I. All of these new experiments utilize the same cold start initial condition for the control and ensemble, simulated observations from the Joint OSSE (Andersson and Masutani 2010), GFS model version and spatial resolutions (dual resolution mode, T382 deterministic with an 80 member T190 ensemble), comparable versions for the GSI and EnKF codes, identical inflation parameters (Whitaker and Hamill 2012), and cycling configuration for the data assimilation including the re-centering procedure. The experiments are designed in such a way as to infer which components of the hybrid and 4D extensions yield improvements (or degradations) to the

quality of analyses. A listing and description of the experiments can be found in Table 3. Note that the two 3D experiments were carried out for Part I, but are included here as a reference point. To reiterate, the assimilation and nature run models are not the same, and it is important to keep in mind their different handling of various components such as parameterized convection when interpreting the results.

The first experiment, H-4DENVAR_NMI, is designed to build upon the successes of the 3DHYB experiment in Part I and to test the impact of going from 3D to 4D ensemble perturbations. The assimilation window is kept fixed at six hours, though the H-4DENVAR_NMI experiment utilizes hourly time levels and therefore, hourly background and ensemble perturbations, e.g., Eqs. (3), (5), and (7). The observation selection procedure is identical in both configurations. This hybrid configuration also utilizes a 25% time-invariant static contribution to the analysis increment as well as the TLNMC over all seven time levels. The change in the analysis error for zonal wind, temperature, and specific humidity is shown in Fig. 5. Generally speaking, the analysis error is smaller in the H-4DENVAR_NMI experiment, especially for upper tropospheric extratropical winds and temperature, and lower tropospheric water vapor. It appears that by going from 3D to 4D, the temperature analysis error has actually increased over the southern polar cap in the lower troposphere. Also of note is the increase in analysis error for specific humidity in the mid-tropospheric tropics, temperature in the upper tropospheric tropics, as well temperature in the upper stratosphere. The increased analysis error is probably related to differences in topography or model physics, as well as the use of three hourly output of the nature run to generate the continuous simulated observation set (Errico et al. 2012). In particular, differences between the nature run model and assimilation model in terms of the handling of vertical mixing in very stable regimes, deep convection in the tropics, and

damping in the upper stratosphere are likely being exposed. Other 4D experiments exhibit similar behavior. By in large, however, the impact in going from 3D to 4D is generally positive, consistent with the better use of observations distributed through the assimilation windows as was demonstrated with the single observation test. However, the impact is smaller than what was found in going from 3DVar to hybrid 3DEnVar (Part I, Kleist and Ide 2014), consistent with the findings of Buehner et al. (2013) and not as impressive as the findings in Wang and Lei (2014).

Although encouraged by the fact that the H-4DEnVar_NMI yielded improved analyses relative to the original 3D hybrid experiment, the application of the TLNMC over all seven observation bins is computationally expensive, given the necessity to calculate incremental tendencies as well as grid to spectral transforms, all within the iterative scheme. An experiment is carried out to test the impact of replacing the TLNMC with a more computationally efficient JCDFI (H-4DEnVar_DFI). The digital filter term is based on a dry energy norm and utilizes a weighting parameter (χ) of 10 in Eq. (8), based on previous findings (Polavarapu et al. 2000; Gauthier and Thépaut 2001). Relative to an experiment that does not utilize any of the constraint options (H-4DEnVar), the H-4DEnVar_DFI is found to have small, consistently negative impact (Fig. 6, purple). Sensitivity experiments that varied the amplitude of the weak constraint cost function parameter did not yield significant improvements, as a large increase in the parameter amplitude results in a significantly worse fit to observations.

Similar to the findings of Kleist et al. (2009), the TLNMC does have an impact reducing the background and analysis errors for surface pressure whereas the JCDFI does not (Fig. 7). In fact, the H-4DEnVar_DFI results in increased background error for surface pressure relative to no constraint for many cycles (Fig. 7 top). The JCDFI as designed is not able to filter out higher

427 frequency noise, given that it cannot be described with a time series of hourly states (Sec. 3c).
428 Furthermore, there is no direct link to the model equations themselves as there is in the 4DVar
429 variant of the digital filter. In contrast, the H-4DEnVar_NMI results in a consistent reduction of
430 analysis error relative to the experiment with no constraint (Fig. 6, blue), consistent with the
431 findings of Wang et al. (2013), Kleist and Ide (2014), and Wang and Lei (2014).

432 Despite these shortcomings, an experiment is carried out that utilizes a combined
433 constraint similar to that which was utilized in section 3. In this new experiment (H-
434 4DEnVar_COMB), the same digital filter parameters from H-4DEnVar_DFI are utilized while
435 the TLNMC is only applied to the center of the assimilation window, or in other words, to the
436 increment that is added to the 6-hr forecast which is then passed on as the initial condition for the
437 forecast model. The JCDFI uses information from all time levels, and therefore aids to remove
438 some gravity mode tendencies for the time levels away from the center according to the filter
439 weights without having to explicitly apply the normal mode initialization over all time levels.
440 This is demonstrated by looking at the total and gravity mode tendencies from the single analysis
441 case used in section 3. The largest incremental tendencies are generated for the analysis that
442 utilized no constraint whereas the TLNMC is most efficient at reducing them (Fig. 8 top).
443 While the JCDFI does act to reduce the incremental tendencies, it also acts to increase the
444 amount of gravity mode tendencies that project onto the total tendencies (Fig. 8 bottom). What
445 is especially troubling is the fact that the JCDFI increases the ratio of gravity mode to total
446 tendencies to be larger than with no constraint at all. The combined filter is most efficient at
447 reducing the incremental tendencies at the center of the window as expected, but the ratio of
448 gravity mode to total tendencies is not as small as the TLNMC-only over all time levels. Like
449 the JCDFI itself, the combined constraint acts to degrade the quality of analyses for most levels,

450 the exception being the layer between 300 hPa and 20 hPa (Fig. 6). Additionally, the combined
451 constraint is worse than the multi time-level TLNMC for all layers except for a slight
452 improvement found between 300 hPa and 150 hPa. It is clear from this set of experiments that of
453 the constraint formulations considered, the TLNMC is the only one that consistently yields
454 improved analyses.

455 The suite of constraint experiments all included a 25% contribution from the static error
456 covariance. It was found in Part I that including a static **B** contribution had very little impact on
457 the solution. To explore this further, an additional 4DEnVar experiment is run to test the impact
458 of a static contribution in the absence of other constraints. Surprisingly, even though the static
459 contribution is time invariant, it does act to reduce analysis error in the 4D hybrid context (Fig.
460 9). The discrepancy between this result and the findings in Part I and Wang et al. (2013) and
461 Wang and Lei (2014) can be attributed to the use of the TLNMC and use of a dual-resolution
462 configuration. The TLNMC is in fact improving upon the quality of the hybrid analyses even
463 further (Fig. 6). However, since the TLNMC is actually a strong constraint on the incremental
464 solution, the inclusion of a static contribution may be unnecessary for certain configurations.
465 This encouraging result provides motivation to explore the use something like the TLNMC
466 within the EnKF analysis update.

467

468 B. FORECAST IMPACT

469 Similar to Part I, a forecast impact experiment is carried out utilizing 00 UTC analyses
470 from the H-4DEnVar_NMI configuration to initialize the GFS model for integration out to 7.5
471 days. The first two weeks of the experiment are ignored to account for spin-up. All verification
472 is done relative to the ECMWF nature run on a common grid. It was shown in Part I that

forecasts initialized from the 3DHYB analyses were generally superior to those that were initialized from 3DVar analyses. The analyses from the H-4DEnVar_NMI experiment are chosen in order to directly compare to the 3D hybrid experiments carried out in Part I, where the TLNMC was also utilized, in addition to the fact that he H-4DEnVar_NMI experiment resulted in the smallest analysis errors relative to the other 4D hybrid experiments.

A comparison of the 500 hPa geopotential height anomaly correlation die off curves reveals that the 4D hybrid experiment yields improved forecasts at this level for all lead times in both extratropical hemispheres (Fig. 10). The improvement over the 3D hybrid experiment is not as large in amplitude as was found in going from 3DVar to hybrid 3DEnVar. The 4D hybrid configuration results in improved height forecasts at other levels beyond 72-h, except for the upper stratosphere (Fig. 11). This is consistent with the increased analysis temperature errors that were found for these upper levels (Fig. 5 middle), related to the ensemble under-dispersion in terms of temperature in the upper stratosphere (noted in Part I).

The forecast impact on other variables and levels is more mixed (Fig. 12), similar to the analysis error findings (Fig. 5). In general, it can be seen that the only region that resulted in a consistent improvement in going from 3D to 4D hybrid was in the northern hemisphere troposphere (Fig. 12, left column). While there are some improvements in the wind and height forecasts in the southern hemisphere beyond 48 hours, the differences are generally not statistically significant. The impact in the tropics is generally found to be neutral to worse. This is counter to the findings in Part I in going from 3DVar to 3D hybrid (Kleist and Ide 2014), where significant improvements were found in the tropics and southern hemisphere troposphere. The use of 4D versus 3D ensembles results in only modest improvements, consistent with the findings in Buehner et al. (2013). This is somewhat counter to the findings of Wang and Lei

(2014), where large improvements were found in their 4DEnVar versus 3DEnVar in the extratropics, particularly in the Southern Hemisphere. However, they too found only very small impacts in the tropics.

5. Summary and Conclusions

An extension of the GSI-based hybrid variational ensemble algorithm to include 4D ensemble perturbations is proposed and implemented. The 4DEnVar algorithm has several advantages relative to 4D-EnKF and 4DVar algorithms that make it attractive for an operational center such as NCEP. Since 4DEnVar does not require the use of additional dynamic models (i.e., the TL and AD), it requires less in terms of computational resources than a traditional 4DVar algorithm. Since it is based on a variational algorithm, it becomes quite trivial to supplement the ensemble-based (4D) covariance with some static estimate that can help ameliorate potential sampling issues, particularly for small ensemble sizes. Furthermore, such a configuration allows for easy implementation of a dual-resolution algorithm, solving for a high-resolution increment using a low-resolution ensemble and control variable (or vice versa, analogous to the interpolation of ensemble weights algorithm, Yang et al. 2009). Lastly, variational-based methods have the advantage of physical space localization, implicit through the localization of the ensemble control variable, which can be important for observations such as satellite radiances.

Several OSSE-based experiments are carried out to demonstrate that the 4DEnVar algorithm contributes positively to the quality of analysis relative to the 3DEnVar and 3DHYB experiments that were carried out in Part I. It is found that going from 3D to 4D ensemble perturbations reduces the analysis error for most variables and levels, though the impact is generally small as in Buehner et al. (2013), in contrast to the findings for the extratropics in

519 Wang and Lei (2014). The addition of a time-invariant static contribution to the analysis
520 increment (i.e., hybridization) does improve the quality of analyses, but only in the absence of
521 the TLNMC. Experiments that utilized the JCDFI, either alone or in combination with the
522 TLNMC, resulted in increased analysis error due to the use of hourly states in prescribing the
523 filtered state. Despite the computational cost associated with it, the use of the TLNMC over all
524 4D time levels produced the highest quality analyses. Wang and Lei (2014) also found that the
525 use of the TLNMC improved forecast quality within their 4DEnVar experiments. It is possible
526 that the JCDFI may still have utility if one were to utilize higher temporal frequency for the
527 observation bins and ensemble time levels. This will prove to be computationally expensive, but
528 may be worth pursuing as computing power continues to increase.

529 The H-4DEnVar_NMI experiment had the largest impact in the extratropical troposphere.
530 This corresponds to a more dynamically active region, where the propagation of information and
531 the usage of the appropriate time levels within the observations are more important. The
532 amplitude of the error reduction in going from the 3DHYB to H-4DEnVar_NMI is found to be
533 smaller than that when going from the 3DVar control to 3DHYB. Going from 3D to 4D in this
534 context resulted in increased analysis error for a few variables in certain regions. This could be
535 the result of the use of discrete (3 hourly) time levels from the nature run to generate the 4D
536 observations as well as differences between the assimilation and nature run models. Perhaps the
537 use of stochastic physics in place of additive inflation (and a single deterministic model) within
538 the ensemble could reduce the negative impact in going to 4D. There is work underway to
539 generate new high-resolution nature runs with better temporal output. It will be interesting if
540 some of these issues can be reduced with the new data sets. Lastly, it is found that the improved

analyses generally result in improved forecast skill, especially for extratropical tropospheric height and wind forecasts.

Although the GSI has a 4DVar capability within it, the inefficiencies of the inner loop dynamic model make it unaffordable for running fully cycled experiments. Work is ongoing to make the dynamic model more efficient. Once it is ready and can be utilized for such an experiment, it will be interesting to compare the results of the H-4DEnVar experiments with actual 4DVar, including H-4DVar.

There is also additional work that can be done on the 4DEnVar algorithm itself, such as treating the outer loop more appropriately, along the lines of the ideas proposed in Yang et al. (2012). To this point, the 4DEnVar algorithm has been executed more like a 3DVar implementation at least in terms of how the multiple inner loops operate. Perhaps some of the noted issues can be improved through a proper outer loop where the nonlinear control model is rerun but the perturbations are held fixed. An alternative would be to allow for the full reintegration of the control and the ensemble, much like running in place (Yang et al. 2012) or an iterative ensemble Kalman smoother (Bocquet and Sakov 2014). Lastly, work is underway to explore the use of a 4D incremental analysis update (IAU, Bloom et al. 1996) within the hybrid 4DEnVar context. The 4DIAU has been demonstrated to be a suitable alternative to a weak constraint digital filter for 4DEnVar (Lorenc et al. 2014). The 4DIAU is especially attractive as it provides a means for passing the 4D incremental state to the forecast model, instead of the single incremental state in the center of the assimilation window as is typically done.

In addition the aforementioned work related specifically to 4DEnVar, there is related work that could be performed to improve the system. The TL and AD models that are used as part of the TLNMC and 4DVar applications are still based on a dry, adiabatic tendency model.

564 This has been shown to have a negative impact in the tropics (Kleist et al. 2009; Wang et al.
565 2013; Wang and Lei 2014). Work is ongoing to add various moist physics to these models to
566 help improve the use of the TLNMC in the tropics. It should also be pointed out that the
567 selection of weighting parameters between the ensemble and static contribution has been
568 simplistic to this point, and could certainly be investigated further as in Bishop et al. (2013a,b).
569 One proposed modification to the application of the weighting already underway would be to
570 consider scale-dependent parameters (Chapter 4 of Kleist 2012).
571
572
573
574 *Acknowledgements*: This work was completed as part of the first author's Ph.D. thesis at the
575 University of Maryland-College Park, made possible through the support of EMC management,
576 especially Steve Lord, John Derber, and William Lapenta. The ECMWF nature run was
577 provided by Erik Andersson through arrangements made by Michiko Masutani. We are grateful
578 to the Joint Center for Satellite Data Assimilation for providing access to their supercomputers to
579 perform the experiments. The support of ONR grants (N000140910418, N0001410557) is
580 gratefully acknowledged. The authors wish to thank Nikke Privé and Ron Errico for kindly
581 providing access to the calibrated, simulated observations. The authors wish to thank Dave
582 Parrish and Jeff Whitaker for contributing to various aspects of the hybrid, EnVar, and EnKF
583 developments. Lastly, the authors acknowledge Ting Lei and Xuguang Wang for good
584 discussion and collaboration as they worked on an implementation of 4DEnVar in parallel to this
585 work. Two anonymous reviewers helped to significantly improve the manuscript.
586

References

Andersson, E., and M. Masutani, 2010: Collaboration on observing system simulation experiments (Joint OSSE), ECMWF Newsletter No. 123, Spring 2010, 14-16. [Available online at http://www.ecmwf.int/publications/newsletters/pdf/123.pdf]

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., Elizabeth A. Satterfield, 2013a: Hidden Error Variance Theory. Part I: Exposition and Analytic Model. *Mon. Wea. Rev.*, **141**, 1454–1468.

Bishop, C. H., Elizabeth A. Satterfield, Kevin T. Shanley, 2013b: Hidden Error Variance Theory. Part II: An Instrument That Reveals Hidden Error Variance Distributions from Ensemble Forecasts and Observations. *Mon. Wea. Rev.*, **141**, 1469–1483.

Bloom, S. C., L. L. Takacs, A. M. da Silva, D. Ledvina, 1996: Data Assimilation Using Incremental Analysis Updates. *Mon. Wea. Rev.*, **124**, 1256–1271.

Bocquet, M. and P. Sakov, 2014: An iterative ensemble Kalman smoother. *Quart. J. Roy. Meteor. Soc.,* **140**, 1521-1534.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.*, **138**, 1550-1566.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Wea. Rev.*, **138**, 1567-1586.

610 Buehner, M., J. Morneau, and C. Charette, 2013: Four-dimensional ensemble-variational data
611 assimilation for global deterministic weather prediction. *Nonlin. Processes Geophys.,* **20**,
612 669-682, doi:10.5194/npg-20-669-2013, 2013.
613 Clayton A. M, A. C. Lorenc and D. M. Barker, 2013: Operational implementation of a hybrid
614 ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor.*
615 *Soc.,* **139**, 1445-1461.
616 Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational
617 implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc*.,
618 **120,** 1367-1388.
619 Dee, D. P., 2005: Bias and data assimilation, *Quart. J. Roy. Meteor. Soc.* **613,** 3323-3343.
620 Derber, J. C., and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI
621 analysis system. *Mon. Wea. Rev.*, **126**, 2287–2299.
622 Desroziers, G., J.-T. Camino, and L. Berre, 2014: 4DEnVar: Link with 4D state formulation of
623 variational assimilation and different possible implementations. *Quart. J. Roy. Meteor. Soc*.,
624 in press. doi: 10.1002/qj.2325
625 Errico, R.M., R. Yang, N. Privé, K.-S. Tai, R. Todling, M.E. Sienkiewicz, and J. Guo, 2012:
626 Development and validation of observing system simulation experiments at NASA's Global
627 Modeling and Assimilation Office. *Quart. J. Roy. Meteor. Soc.,* **139**, 1162-1178.
628 Fairbairn, D., S. R. Pring, A. C. Lorenc, and I. Roulstone, 2014: A comparison of 4DVar with
629 ensemble data assimilation methods. *Quart. J. Roy. Meteor. Soc.,* **140**, 281-294.
630 Gauthier, P., and J.-N. Thépaut, 2001: Impact of the digital filter as a weak constraint in the
631 preoperational 4DVar assimilation system of Météo-France. *Mon. Wea. Rev*., 128, 2905-
632 2919.

633 Gauthier, P , M. Tanguay, S. Laroche, and S. Pellerin, 2007: Extension of 3DVar to 4DVar:
634 Implementation of 4DVar at the Meteorological Service of Canada. *Mon. Wea. Rev.,* **135,**
635 2339–2364.
636 Gustafsson, N., 1993: Use of a digital filter as weak constraint in variational data assimilation.
637 *Proc. Workshop on Variational Assimilation, with Special Emphasis on Three-Dimensional*
638 *Aspects,* Reading, United Kingdom, ECMWF, 327–338.
639 Hunt, B., E. Kalnay, E. J. Kostelich, E. Ott, D. J. Patil, T. Sauer, I. Szunyogh, J. A. Yorke, and
640 A. V. Zimin, 2004: Four-dimensional ensemble Kalman filtering. *Tellus*, **56A**, 273-277.
641 Hunt, B. R., E. Kostelich, and I. Szunyogh 2007: Efficient Data Assimilation for Spatiotemporal
642 Chaos: a Local Ensemble Transform Kalman Filter, *Physica D*, **230**, 112-126.
643 Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, R. M. Errico, and R. Yang, 2009:
644 Improving incremental balance in the GSI 3DVar analysis system. *Mon. Wea. Rev*., **137,**
645 1046-1060.
646 Kleist, D. T, 2012: An evaluation of hybrid variational-ensemble data assimilation for the NCEP
647 GFS. Ph.D thesis. Dept. of Atmospheric and Oceanic Science. University of Maryland-
648 College Park. 149 pp. (Available online at http://drum.lib.umd.edu/handle/1903/13135).
649 Kleist, D.T., and K. Ide, 2014: An OSSE-based evaluation of hybrid variational-ensemble data
650 assimilation for the NCEP GFS, Part I: System description and 3D-Hybrid results. *Mon.*
651 *Wea. Rev*., under review.
652 Kuhl, D. D., T. E. Rosmond, C. H. Bishop, J. McClay, and N. L. Baker, 2013: Comparison of
653 hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework.
654 *Mon. Wea. Rev*., **141,** 2740-2758.

655 Lawless, A. S., 2010: A note on the analysis error associated with 3D-FGAT. *Quart. J. Roy.*
656 *Meteor. Soc*., **136**: 1094–1098.
657 Lewis, J. M. and J. C. Derber, 1985: The use of adjoints equations to solve a variational
658 adjustment problem with advective constraints. *Tellus.* **37A***,* 309–322.
659 Liu C., Q. Xiao, and B. Wang, 2008: An ensemble-based four dimensional variational data
660 assimilation scheme. Part I: technique formulation and preliminary test. *Mon. Wea. Rev*.,
661 **136,** 3363-3373.
662 Liu C., Q. Xiao, and B. Wang, 2009: An ensemble-based four dimensional variational data
663 assimilation scheme. Part II: Observing system simulation experiments with advanced
664 research WRF. *Mon. Wea. Rev*., **137,** 1687-1704.
665 Liu C., and Q. Xiao, 2013: An ensemble-based four dimensional variational data assimilation
666 scheme. Part III: Antarctic applications with advanced research WRF using real data. *Mon.*
667 *Wea. Rev*., **141,** 2721-2739.
668 Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP – a comparison with
669 4D-VAR. *Quart. J. Roy. Meteor. Soc*., **129**, 3183-3203.
670 Lorenc, A. C., and F. Rawlins, 2005: Why does 4D-Var beat 3D-Var?. *Quart. J. Roy. Meteor.*
671 *Soc*., **131**, 3247-3257.
672 Lorenc, A.C. 2013: Recommended nomenclature for EnVar data assimilation methods. *Research*
673 *Activities in Atmospheric and Oceanic Modeling*. WGNE, 2pp, URL: http://www.wcrp-
674 climate.org/WGNE/BlueBook/2013/individual-
675 articles/01_Lorenc_Andrew_EnVar_nomenclature.pdf)

676 Lorenc, A. C., N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, 2014: Comparison of
677 hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Wea.*
678 *Rev*., submitted.
679 Lynch, P., and X.Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter.
680 *Mon. Wea. Rev*., **120**, 1019-1034.
681 Polavarapu, S., M. Tanguay, L. Fillion, 2000: Four-dimensional variational data assimilation
682 with digital filter initialization. *Mon. Wea. Rev.*, **128**, 2491–2510.
683 Rabier, F., J.-N. Thépaut, and P. Courtier, 1998: Extended assimilation and forecast experiments
684 with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc*.,
685 **124**: 1861–1887.
686 Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF
687 operational implementation of four-dimensional variational assimiliation. Part I:
688 Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc*., **126**, 1143-1170.
689 Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and
690 T. J. Payne, 2007: The Met Office global 4-Dimensional data assimilation system. *Quart. J.*
691 *Roy. Meteor. Soc*., **133**, 347–362.
692 Rosmond, T., and L. Xu, 2006: Development of NAVDAS-AR: nonlinear formulation and outer
693 loop tests. *Tellus*, **58A**, 45-59.
694 Tian, X., Z. Xie, and A. Dai, 2008: An ensemble-based explicit four-dimensional variational
695 assimilation method. *J. Geophys. Res.*, **113**, D21124, doi: 10.1029/2008JD010358.
696 Tian, X., Z. Xie, and Q. Sun, 2011: A POD-based ensemble four-dimensional variational
697 assimilation method. *Tellus.*, **63A**, 805-816.

698 Wee, T.-K., Y.-H. Kuo, 2004: Impact of a Digital Filter as a Weak Constraint in MM5 4DVar:
699 An Observing System Simulation Experiment. *Mon. Wea. Rev.*, **132**, 543–559.
700 Wang, X., D. Parrish, D. Kleist, J. Whitaker, 2013: GSI 3DVar-Based Ensemble–Variational
701 Hybrid Data Assimilation for NCEP Global Forecast System: Single-Resolution
702 Experiments. *Mon. Wea. Rev.*, **141**, 4098–4117.
703 Wang, X. and T. Lei, 2014: GSI-based four dimensional ensemble-variational (4DEnsVar) data
704 assimilation: Formulation and single resolution experiments with real data for the NCEP
705 Global Forecast System. *Mon. Wea. Rev.*, **142,** 3303-3325.
706 Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed
707 observations. *Mon. Wea. Rev*., **130**, 1913-1924.
708 Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation
709 with the NCEP Global Forecast System. *Mon. Wea. Rev*., **136**, 463-482.
710 Whitaker, J. S. and T. M. Hamill, 2012: Evaluating methods to account for system errors in
711 ensemble data assimilation. *Mon. Wea. Rev*., **140**, 3078-3089.
712 Yang, S.-C., E. Kalnay, B. Hunt, and N.E. Bowler, 2009: Weight interpolation for efficient data
713 assimilation with the Local Ensemble Transform Kalman Filter. *Quart. J. Roy. Meteor. Soc*.,
714 **135**, 251-262.
715 Yang, S.-C., E. Kalnay, and B. Hunt, 2012: Handling nonlinearity in an Ensemble Kalman Filter:
716 experiments with the three variable Lorenz model. *Mon. Wea. Rev*., **140**, 2628-2646.
717 Zhang, M. and F. Zhang, 2012: E4DVar: Coupling an ensemble Kalman filter with four-
718 dimensional variational data assimilation in a limited-area weather prediction model. *Mon.*
719 *Wea. Rev.*, **140,** 587–600.

Zhu, Y., J. Derber, A. Collard, D. Dee, R. Treadon. G. Gayno, and J. A. Jung, 2014: Enhanced radiance bias correction in the National Centers for Environmental Prediction's Gridpoint Statistical Interpolation data assimilation system. *Quart. J. Roy. Meteor. Soc*., **140**, 1479-1492.

Experiment	*Description*	*Relevant Equations*
4DVar	4DVar	Eqs. (1)-(2)
H-4DVar	Hybrid ensemble-4DVar (with adjoint) β_f^{-1}=0.25, β_e^{-1}=0.75	Eqs. (2)-(4)
4DEnVar	4DEnVar(no static **B**) β_f^{-1}=0.0, β_e^{-1}=1.0	Eqs. (3) & (5) $\mathbf{M}_k$=**I**
H-4DEnVar	Hybrid 4D-Ensemble-Var β_f^{-1}=0.25, β_e^{-1}=0.75	Eqs. (3) & (5) $\mathbf{M}_k$=**I**

743
744 **Table 1: Description of various 4D (and hybrid) single observation experiments and relevant equations for**
745 **Figs. 1 and 2.**
746

747

748

749

Experiment	*Total Tendency*	*Gravity Mode Tendency*	*Ratio*
4DEnVar	2.853×10^{-3}	2.742×10^{-3}	0.96
4DEnVar+TLNMC	3.333×10^{-4}	2.076×10^{-4}	0.62
4DEnVar+JCDFI	1.841×10^{-3}	1.797×10^{-3}	0.98
4DEnVar+COMB	1.301×10^{-4}	9.046×10^{-5}	0.70
H-4DEnVar	2.509×10^{-3}	2.329×10^{-3}	0.93

750
751 **Table 2: The root mean square sum of the incremental (spectral) tendencies (total and gravity mode) as well**
752 **as the ratio (gravity mode / total) for the eight vertical modes kept as part of the TLNMC for the single**
753 **analysis valid at 06 UTC 15 July 2010 using various 4DEnVar configurations.**
754

755

Experiment	*Control*	*Description*	*Relevant Equations*
3DVar	--	3DVar control	Part I
3DHYB	3DVar	3D-Hybrid β_f^{-1}=0.25, β_e^{-1}=0.75	Part I
4DEnVar	--	4DEnVar No Constraints β_f^{-1}=0., β_e^{-1}=1.0	Eqs. (3) & (5) $\mathbf{M}_k$=**I**
H-4DEnVar	4DEnVar	Hybrid 4DEnVar No constraints β_f^{-1}=0.25, β_e^{-1}=0.75	Eqs. (3) & (5) $\mathbf{M}_k$=**I**
H-4DEnVar_NMI	H-4DEnVar 3DHYB	H-4DEnVar + TLNMC (all observation *k*-levels)	Eqs. (3) & (5) $\mathbf{M}_k$=**I** Eq. (7)
H-4DEnVar_DFI	H-4DEnVar	H-4DEnVar + JCDFI (χ=10)	Eqs. (3) & (5) $\mathbf{M}_k$=**I** Eqs. (8)-(9)
H-4DEnVar_COMB	H-4DEnVar	H-4DEnVar + TLNMC (*k*=mid only) + JCDFI (χ=10)	Eqs. (3) & (5) $\mathbf{M}_k$=**I** Eq. (7) for *k*=*m* Eqs. (8)-(9)

Table 3: Description of various (hybrid) OSSE-based experiments.

List of Figures

Figure 1. 500 hPa Temperature (shaded, K) and vector wind (vectors, m s^{-1}) analysis increment resulting from the assimilation of a single 500 hPa temperature observation (-2K innovation, 1K error) taken -3h from analysis time (location of observation denoted with black dot) for 4DVar (upper left), H-4DVar (lower left), 4DEnVar (upper right), and H-4DEnVar (lower right). The solid contours are the background 500 hPa geopotential height (dm) valid at the center of the assimilation window (analysis time). The reference arrow is representative of 1 m s^{-1}.

Figure 2. As in Fig. 1, except for the analysis increment valid at the beginning (t-3h, top), center (t=0h, middle), and end (t+3h, bottom) of the assimilation window for the H-4DVar (left) and H-4DEnVar (right) cases. The solid contours are the background 500 hPa geopotential height (dm) valid at beginning (top), center (middle), and end (bottom) of the assimilation window.

Figure 3. Power spectra of the divergence increment for a single analysis valid at 06 UTC 15 July 2010 on model level 35 (top, approximately 200 hPa) and 14 (bottom, approximately 850 hPa) utilizing 4DEnVar (non-hybrid) with no constraints (red), tangent linear normal mode constraint (blue), weak constraint digital filter (purple), and combined normal mode and weak constraint (green).

Figure 4. Observation penalty (top) and reduction of the gradient norm (bottom) by iteration for an analysis valid at 06 UTC 15 July 2010 utilizing 4DEnVar (non-hybrid) with no constraints (red), tangent linear normal mode constraint (blue), weak constraint digital filter (purple), and combined normal mode and weak constraint (green).

784 **Figure 5.** Difference in the time-averaged zonal mean standard deviation of the analysis error
785 for the H-4DEnVar_NMI experiment relative to 3DHYB for zonal wind (top, ms^{-1}), temperature
786 (middle, K), and specific humidity (bottom, g kg^{-1}), covering the period spanning 00 UTC 1
787 August 2005 through 18 UTC 30 August 2005.
788
789 **Figure 6.** Difference in the total energy (J kg^{-1}) time-averaged global mean standard deviation
790 of the analysis error for the H-4DEnVar_NMI (blue), H-4DEnVar_DFI (purple), and H-
791 4DEnVar_COMB (green) experiments relative to H-4DEnVar, covering the period spanning 00
792 UTC 1 August 2005 through 18 UTC 30 August 2005.
793
794 **Figure 7.** Time series of the global mean surface pressure error (hPa) for the background (top)
795 and analysis (bottom) from the H-4DEnVar (red), H-4DEnVar_NMI (blue) and H-
796 4DEnVar_DFI (purple) experiments.
797
798 **Figure 8.** The root mean square sum of the incremental spectral tendencies (top) for the total
799 tendency (dashed) and gravity mode tendency for the eight vertical modes (solid) for the
800 4DEnVar (red), 4DEnVar_NMI (blue), 4DEnVar_DFI (purple) and 4DEnVar_COMB (green)
801 configurations for the single analysis valid at 06 UTC 15 July 2010 as a function of observation
802 bin (analysis relative time). Also plotted is the ratio (bottom, gravity mode/total tendency) for
803 the 4DEnVar (red), 4DEnVar_NMI (blue), 4DEnVar_DFI (purple) and 4DEnVar_COMB
804 (green) configurations.
805
806 **Figure 9.** As in Fig. 5, but for the difference H-4DEnVar-4DEnVar.

807 **Figure 10.** Time-averaged 500 hPa anomaly correlation (upper panels) for the Northern
808 Hemisphere (left) and Southern Hemisphere (right) for the 3DHYB (red), H-4DEnVar_NMI
809 (blue), and 3DVar (green) experiments for forecasts from the 00 UTC analyses as a function of
810 lead time as well as the difference (lower panels) for H-4DEnVar-3DHYB (blue) and 3DVar-
811 3DHYB (green). The 95% confidence threshold for a significance test (derived from a standard
812 t-test) is also plotted in the lower panels for the H-4DEnVar (blue bars) and 3DVar (green bars)
813 differences.
814
815 **Figure 11**. Time-averaged root mean square geopotential height errors (m) for forecasts from
816 the 00 UTC analyses in the 3DHYB experiment as a function of lead time for the Northern
817 Hemisphere (upper left) and Southern Hemisphere (lower left) verified against the ECMWF
818 nature run forecasts verifying between 27 July 2005 and 01 September 2005. The difference
819 between the two experiments (H-4DEnVar_NMI-3DHYB) for the Northern Hemisphere (upper
820 right) and Southern Hemisphere (lower right) is also plotted.
821
822 **Figure 12**. Percent change in root mean square error from H-4DENVAR_NMI minus 3DHYB
823 for the period covering 27 July 2005 and 01 September 2005 in the northern hemisphere (green),
824 tropics (red) and southern hemisphere (blue) for selected variables as a function of forecast lead
825 time. The forecast variables include 500 hPa geopotential height (upper left, upper right), 850
826 hPa vector wind (upper middle), 200 hPa vector wind (middle row), and 700 hPa specific
827 humidity (bottom row). All verification is performed against the ECMWF nature run. The error
828 bars represent the 95% confidence threshold for a significance test.

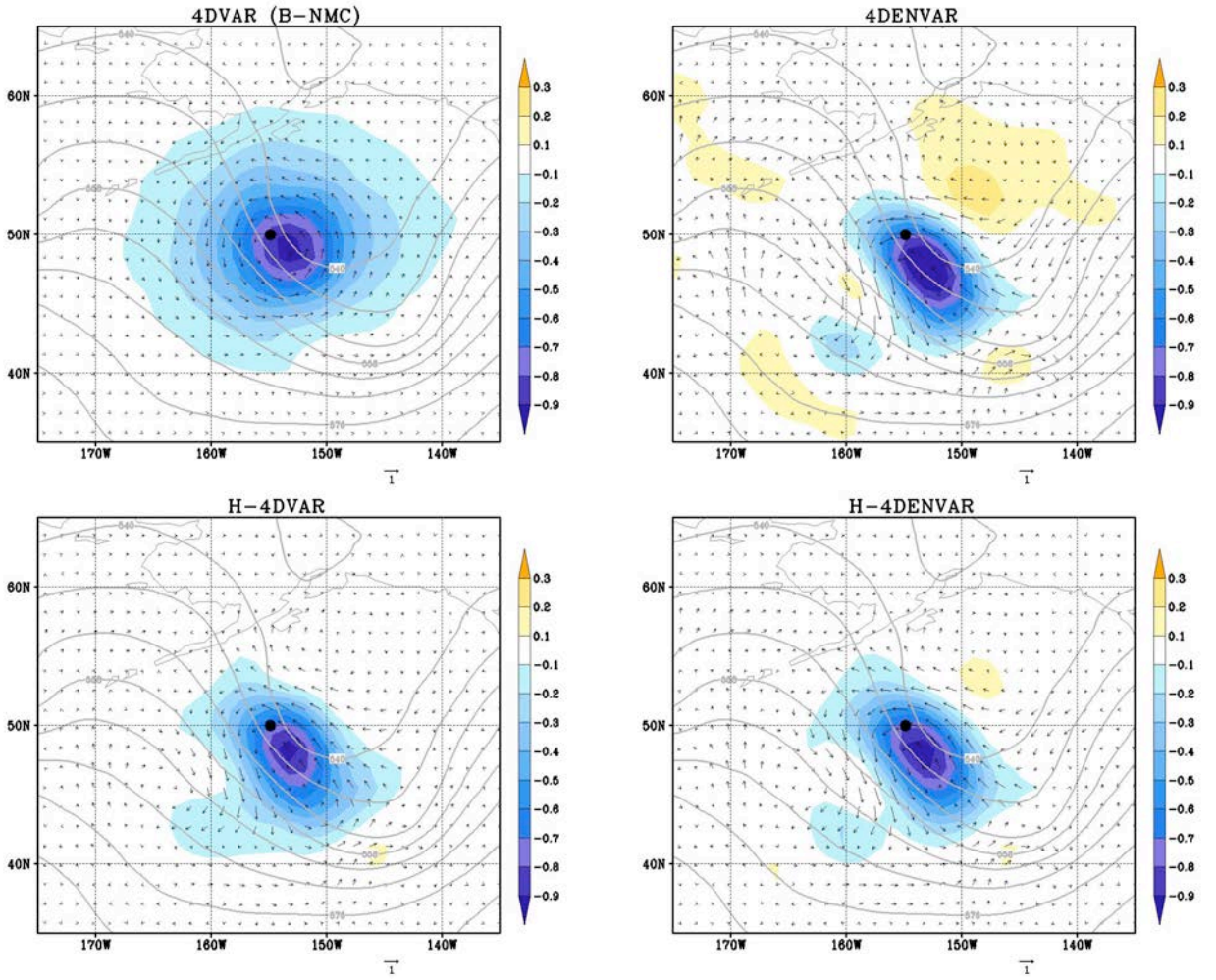

Figure 1: 500 hPa Temperature (shaded, K) and vector wind (vectors, m s^{-1}) analysis increment resulting from the assimilation of a single 500 hPa temperature observation (-2K innovation, 1K error) taken -3h from analysis time (location of observation denoted with black dot) for 4DVar (upper left), H-4DVar (lower left), 4DEnVar (upper right), and H-4DEnVar (lower right). The solid contours are the background 500 hPa geopotential height (dm) valid at the center of the assimilation window (analysis time). The reference arrow is representative of 1 m s^{-1}.

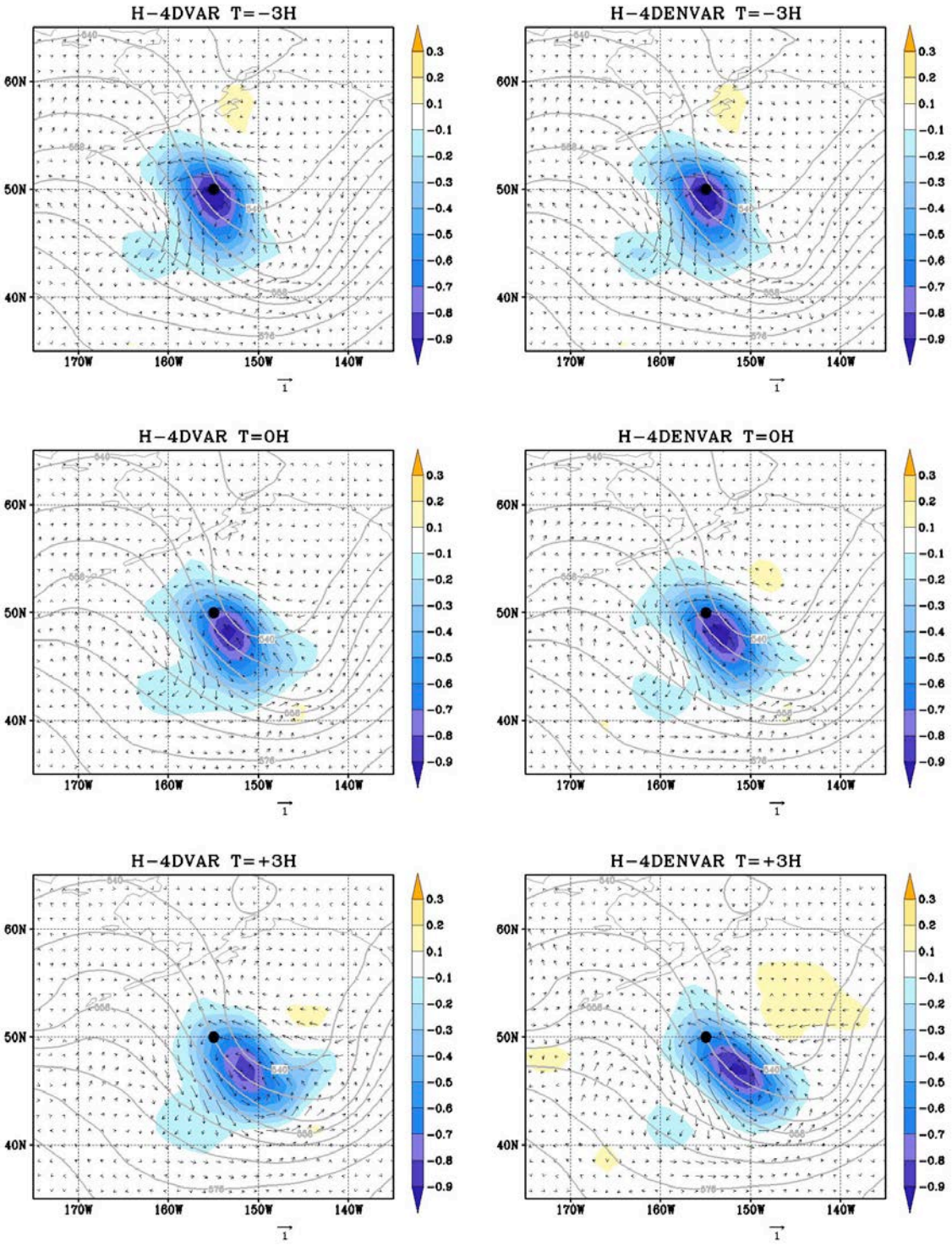

Figure 2: As in Fig. 1, except for the analysis increment valid at the beginning (t-3h, top), center (t=0h, middle), and end (t+3h, bottom) of the assimilation window for the H-4DVar (left) and H-4DEnVar (right) cases. The solid contours are the background 500 hPa geopotential height (dm) valid at beginning (top), center (middle), and end (bottom) of the assimilation window.

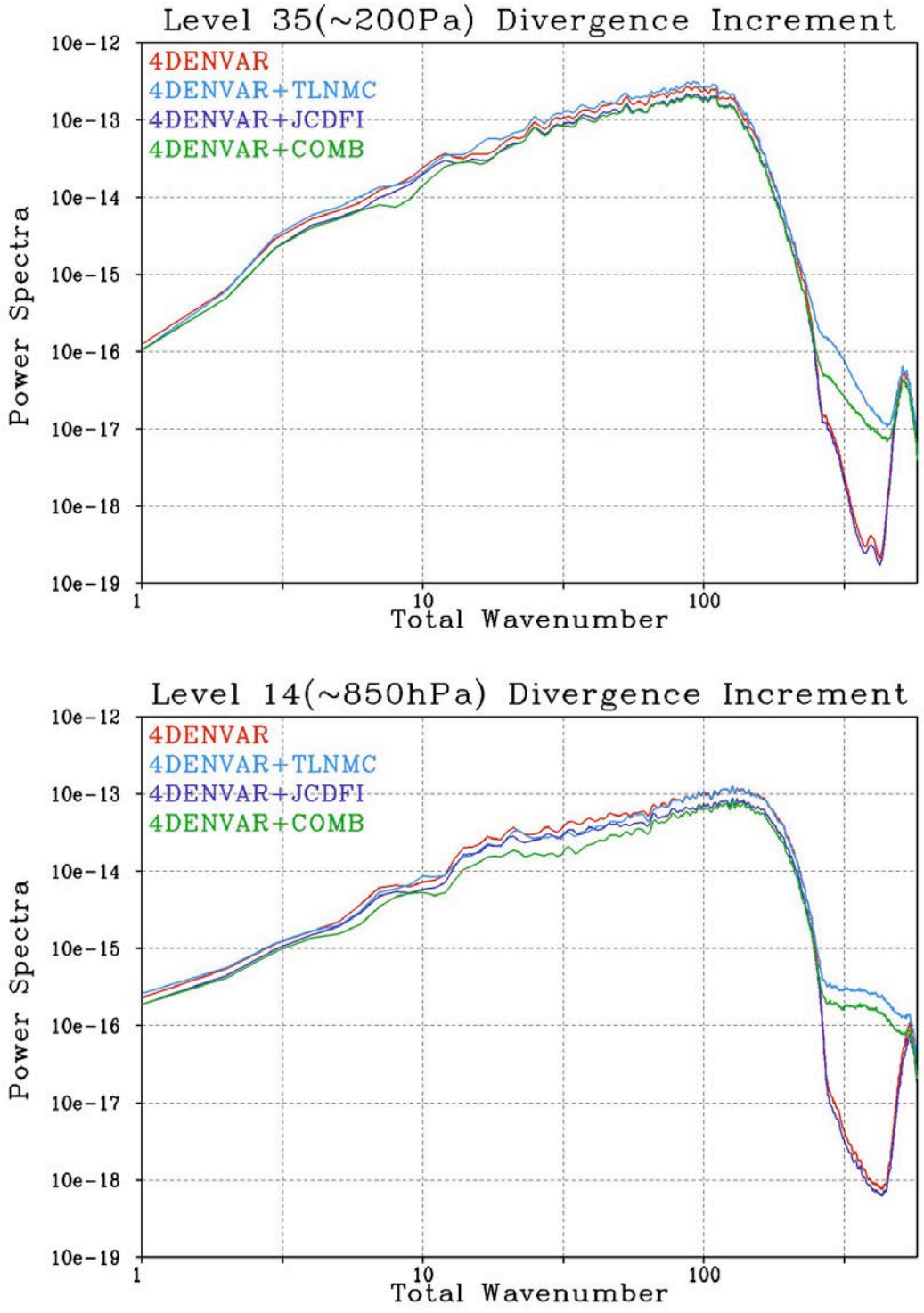

Figure 3: Power spectra of the divergence increment for a single analysis valid at 06 UTC 15 July 2010 on model level 35 (top, approximately 200 hPa) and 14 (bottom, approximately 850 hPa) utilizing 4DEnVar (non-hybrid) with no constraints (red), tangent linear normal mode constraint (blue), weak constraint digital filter (purple), and combined normal mode and weak constraint (green).

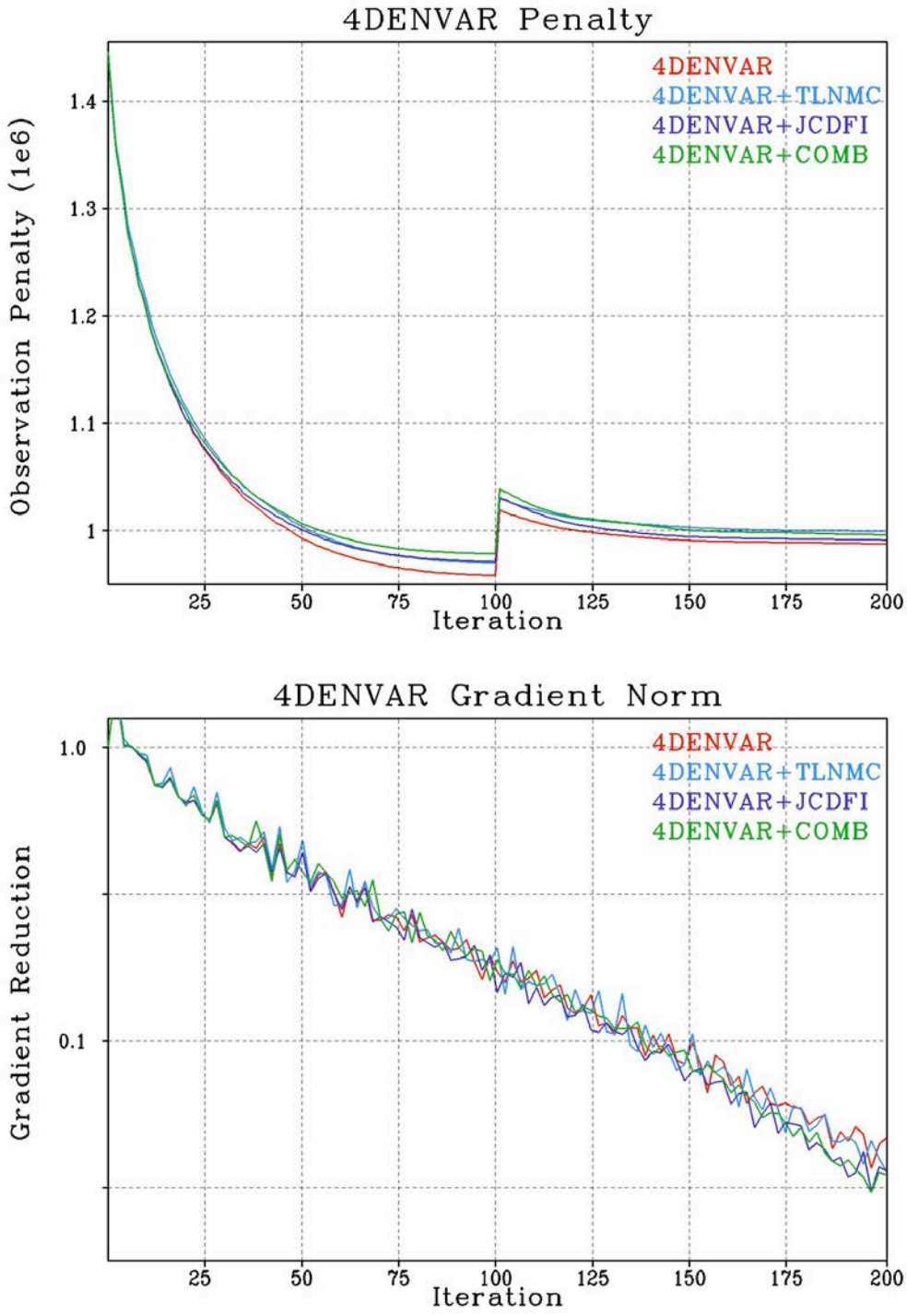

Figure 4: Observation penalty (top) and reduction of the gradient norm (bottom) by iteration for an analysis valid at 06 UTC 15 July 2010 utilizing 4DEnVar (non-hybrid) with no constraints (red), tangent linear normal mode constraint (blue), weak constraint digital filter (purple), and combined normal mode and weak constraint (green).

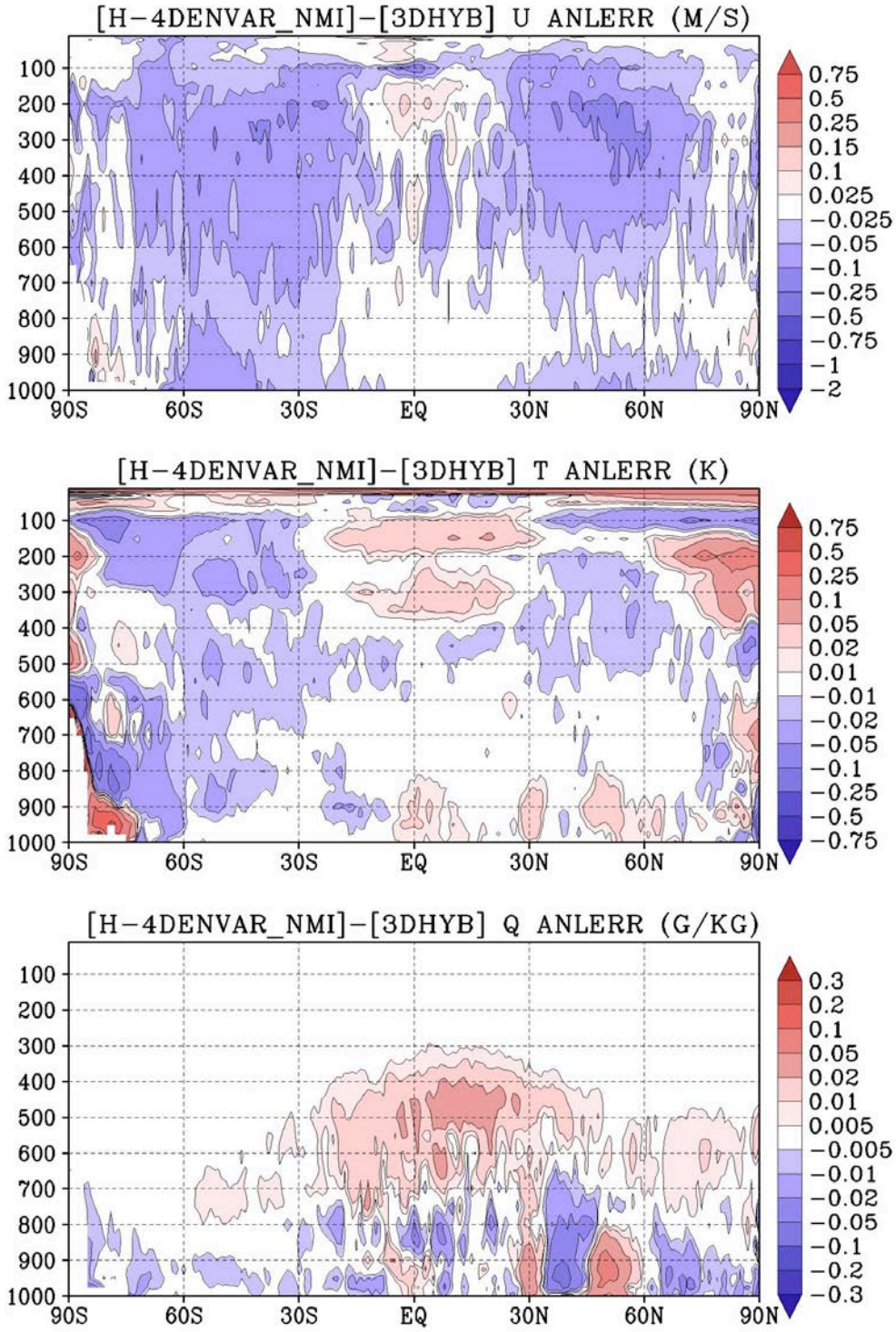

Figure 5: Difference in the time-averaged (August) zonal mean standard deviation of the analysis error for the H-4DENVAR_NMI experiment relative to 3DHYB for zonal wind (top, ms^{-1}), temperature (middle, K), and specific humidity (bottom, g kg^{-1}), covering the period spanning 00 UTC 1 August 2005 through 18 UTC 30 August 2005.

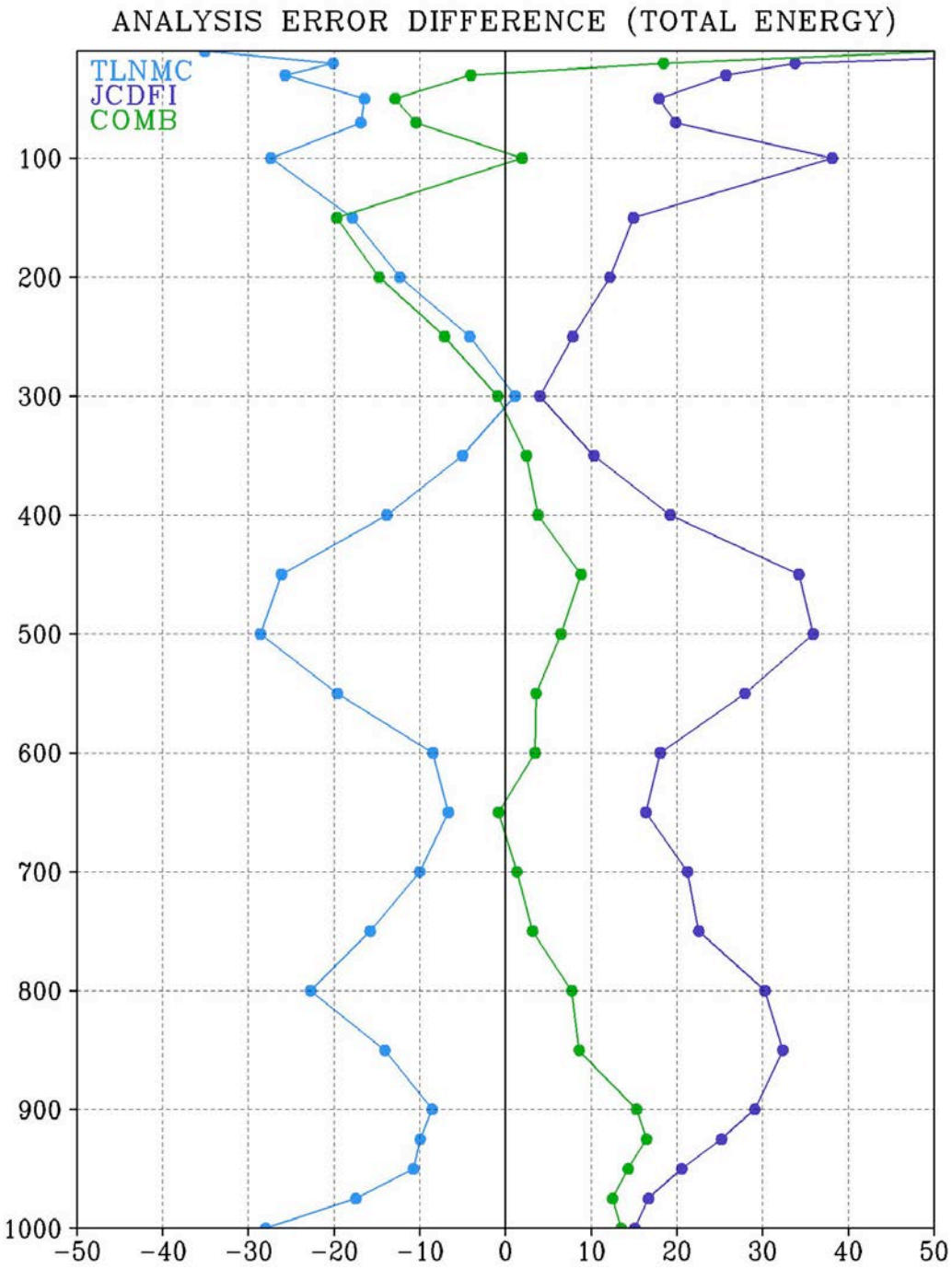

Figure 6. Difference in the total energy (J kg^{-1}) time-averaged global mean standard deviation of the analysis error for the H-4DEnVar_NMI (blue), H-4DEnVar_DFI (purple), and H-4DEnVar_COMB (green) experiments relative to H-4DEnVar, covering the period spanning 00 UTC 1 August 2005 through 18 UTC 30 August 2005.

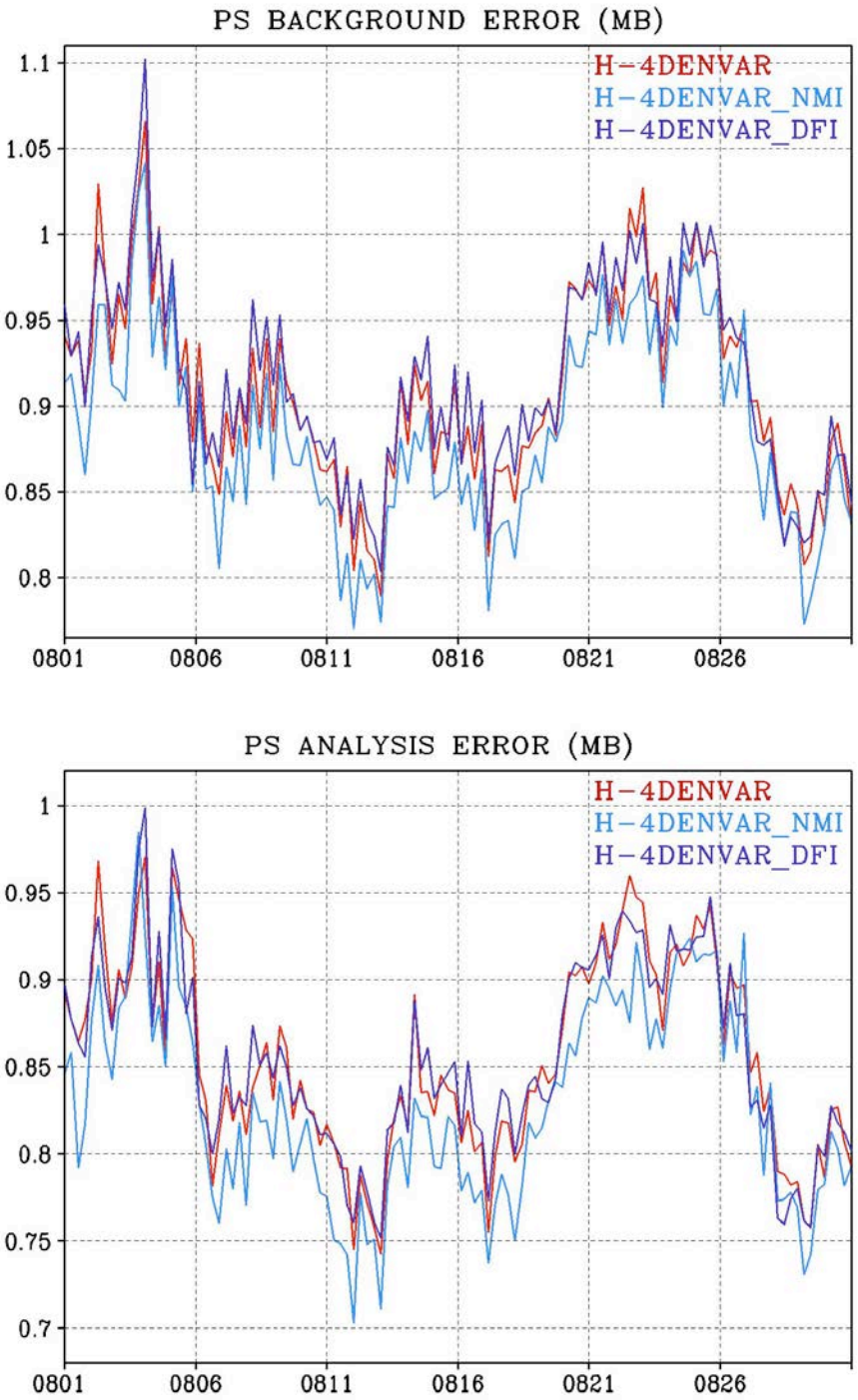

Figure 7. Time series of the global mean surface pressure error (hPa) for the background (top) and analysis (bottom) from the H-4DEnVar (red), H-4DEnVar_NMI (blue) and H-4DEnVar_DFI (purple) experiments.

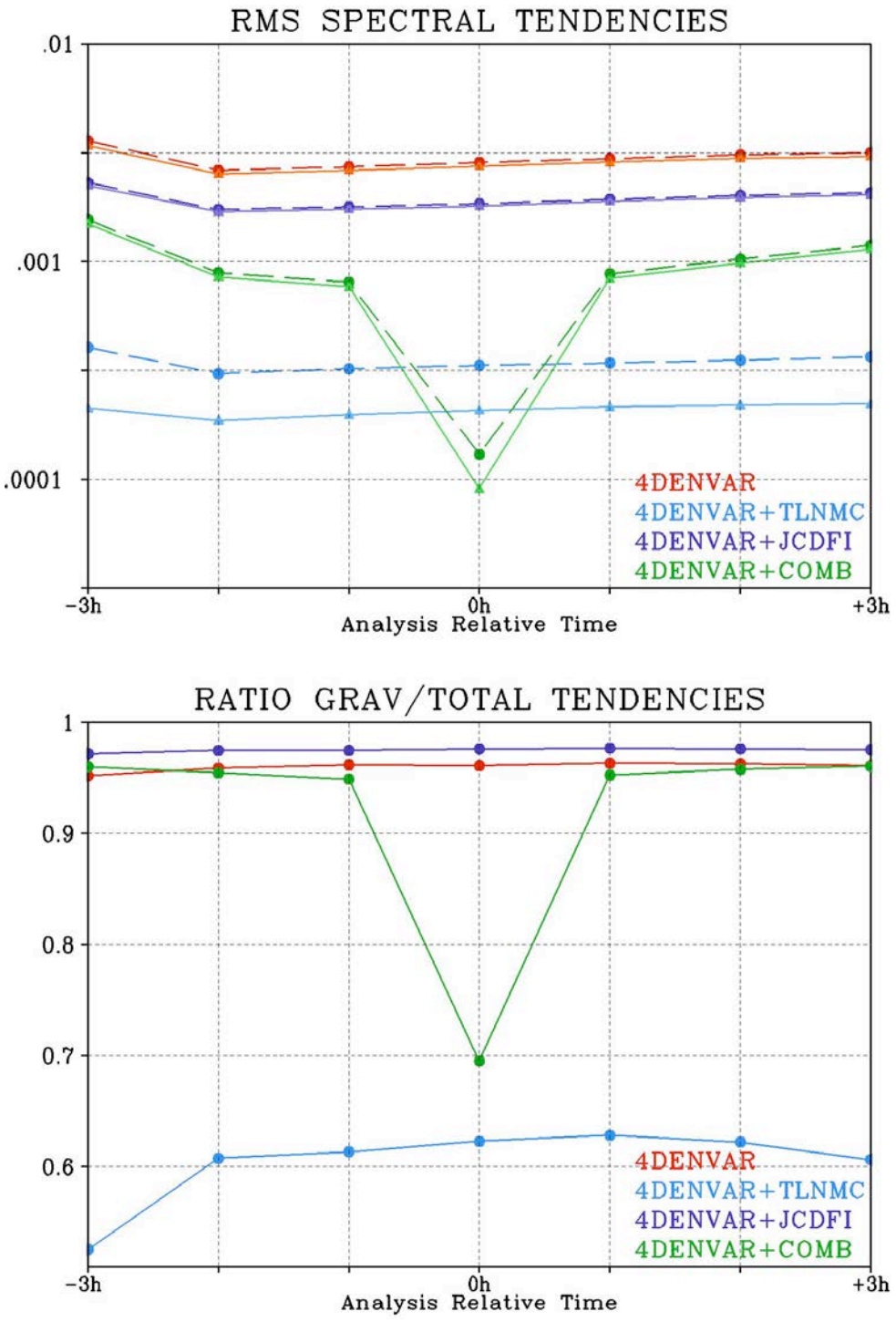

Figure 8. The root mean square sum of the incremental spectral tendencies (top) for the total tendency (dashed) and gravity mode tendency for the eight vertical modes (solid) for the 4DEnVar (red), 4DEnVar_NMI (blue), 4DEnVar_DFI (purple) and 4DEnVar_COMB (green) configurations for the single analysis valid at 06 UTC 15 July 2010 as a function of observation bin (analysis relative time). Also plotted is the ratio (bottom, gravity mode/total tendency) for the 4DEnVar (red), 4DEnVar_NMI (blue), 4DEnVar_DFI (purple) and 4DEnVar_COMB (green) configurations.

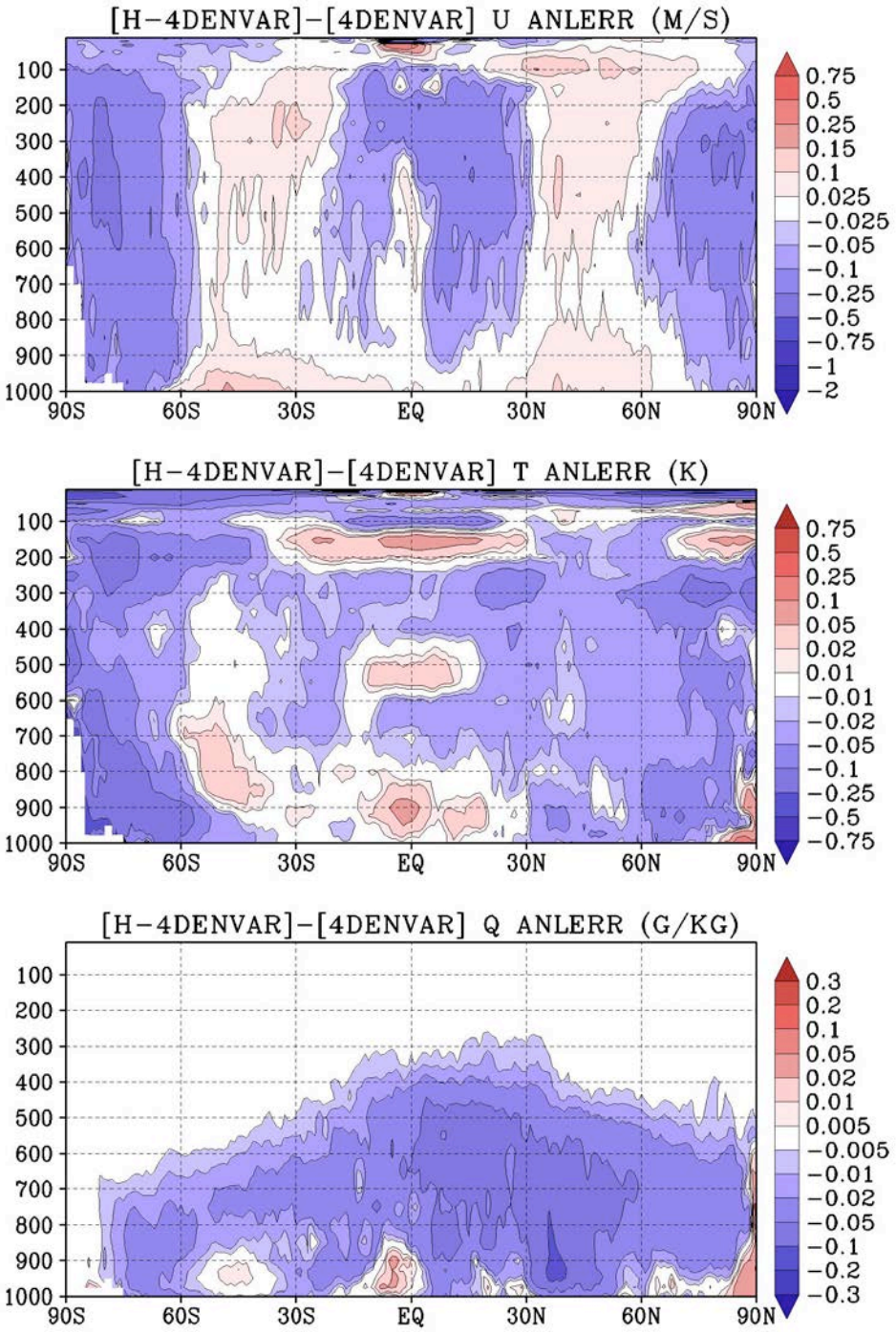

Figure 9. As in Fig. 5, but for the difference H-4DEnVar-4DEnVar.

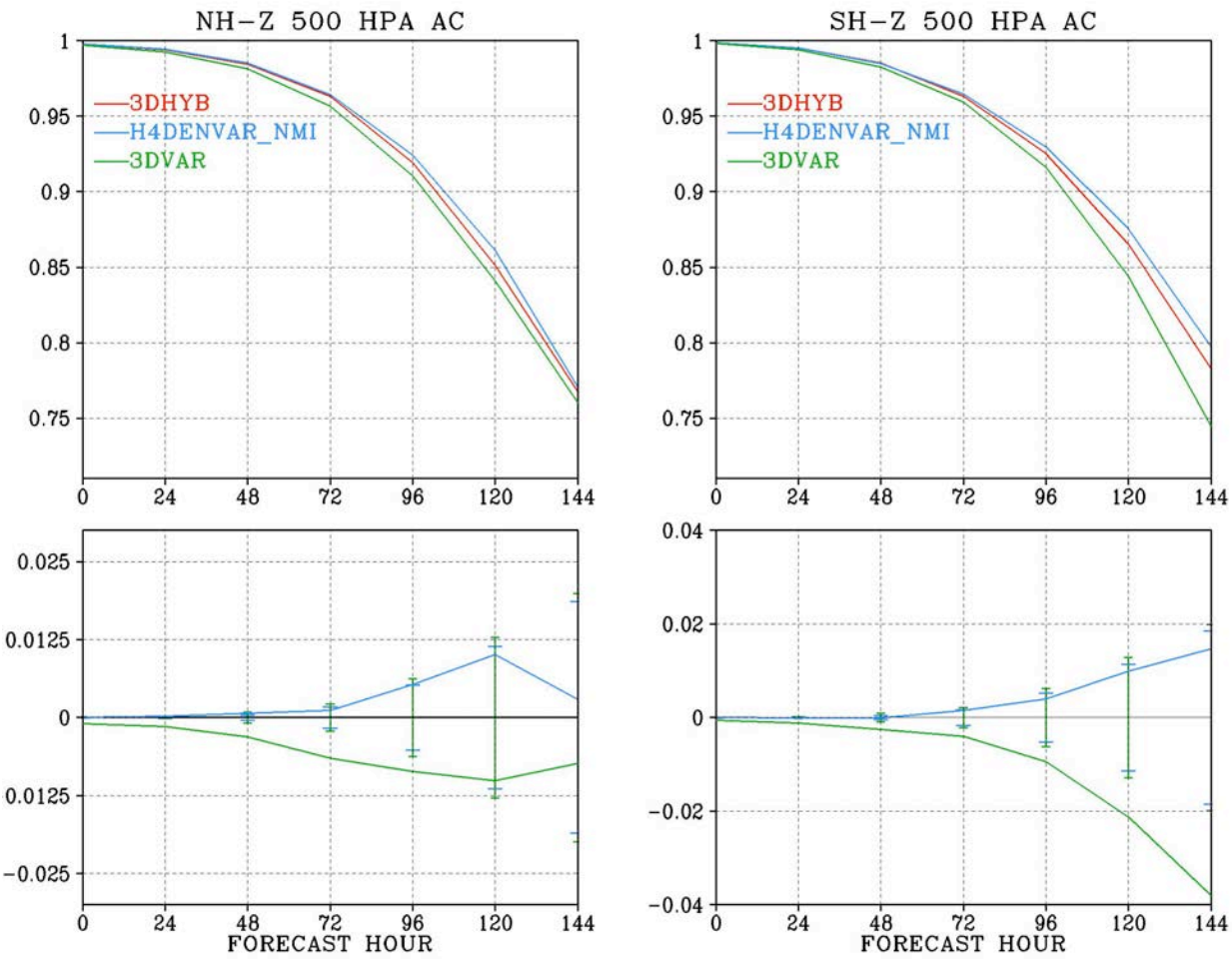

Figure 10. Time-averaged 500 hPa anomaly correlation (upper panels) for the Northern Hemisphere (left) and Southern Hemisphere (right) for the 3DHYB (red), H-4DEnVar_NMI (blue), and 3DVar (green) experiments for forecasts from the 00 UTC analyses as a function of lead time as well as the difference (lower panels) for H-4DEnVar-3DHYB (blue) and 3DVar-3DHYB (green). The 95% confidence threshold for a significance test (derived from a standard t-test) is also plotted in the lower panels for the H-4DEnVar (blue bars) and 3DVar (green bars) differences.

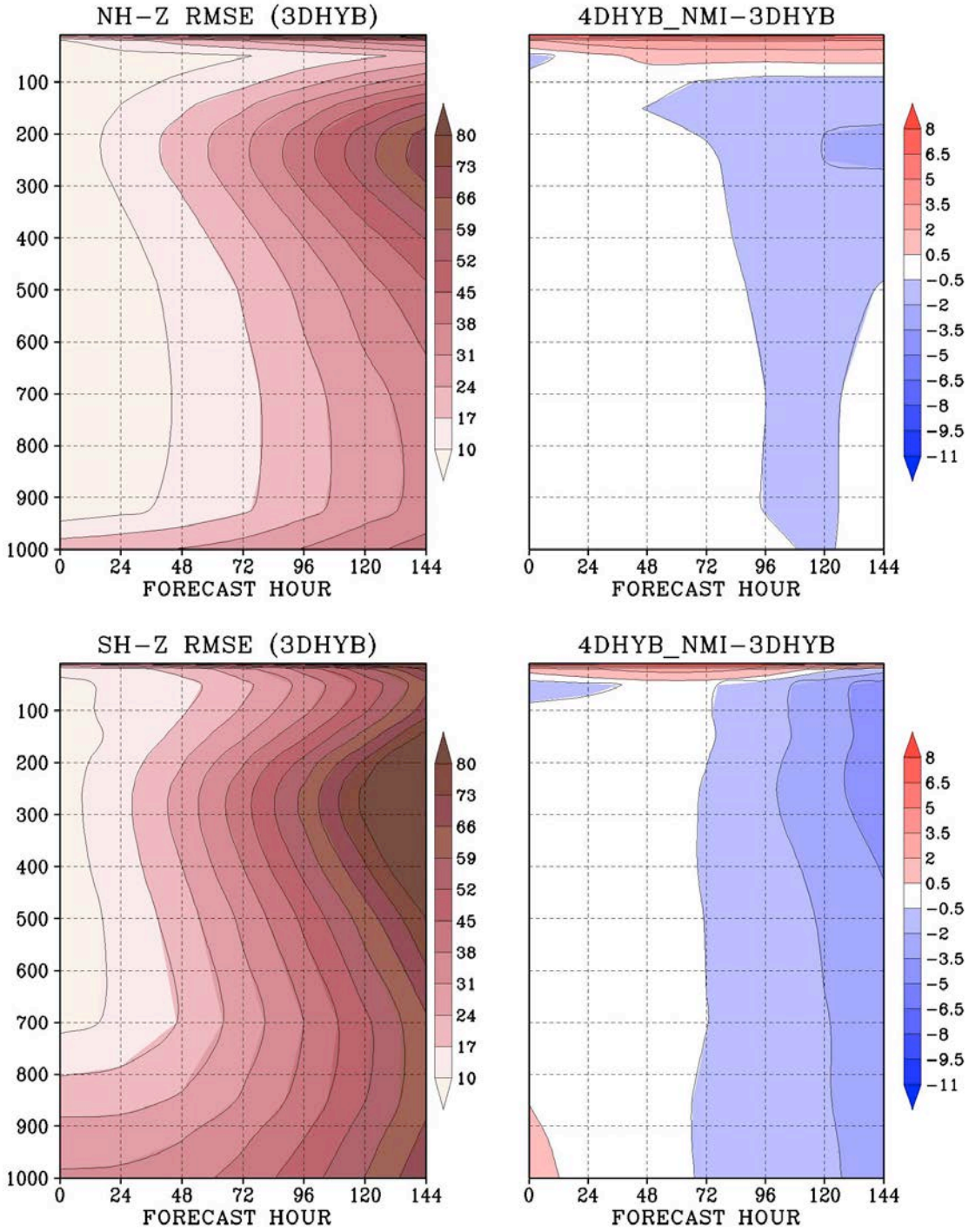

Figure 11. Time-averaged root mean square geopotential height errors (m) for forecasts from the 00 UTC analyses in the 3DHYB experiment as a function of lead time for the Northern Hemisphere (upper left) and Southern Hemisphere (lower left) verified against the ECMWF nature run forecasts verifying between 27 July 2005 and 01 September 2005. The difference between the two experiments (H-4DEnVar_NMI-3DHYB) for the Northern Hemisphere (upper right) and Southern Hemisphere (lower right) is also plotted.

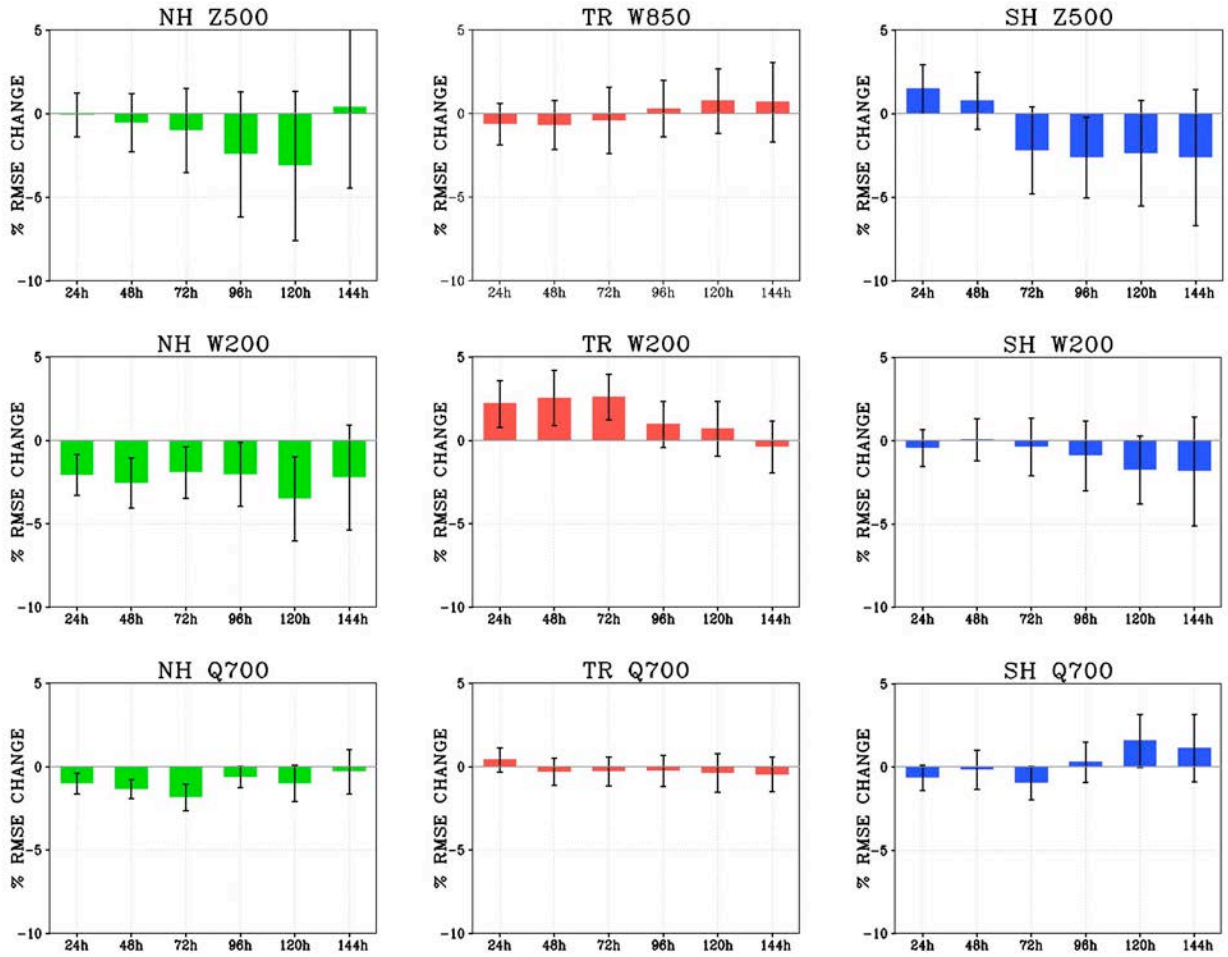

Figure 12: Percent change in root mean square error from 4DHYB_NMI minus 3DHYB for the period covering 27 July 2005 and 01 September 2005 in the northern hemisphere (green), tropics (red) and southern hemisphere (blue) for selected variables as a function of forecast lead time. The forecast variables include 500 hPa geopotential height (upper left, upper right), 850 hPa vector wind (upper middle), 200 hPa vector wind (middle row), and 700 hPa specific humidity (bottom row). All verification is performed against the ECMWF nature run. The error bars represent the 95% confidence threshold for a significance test.