2A.10

A Hybrid Data Assimilation (WRF-VAR and Ensemble Transform Kalman Filter) System Based Retrospective Tests

Meral Demirtas[(1)*], Dale Barker[(2)], Yongsheng Chen[(3)], Josh Hacker[(1)], Xiang-Yu Huang[(1)], Chris Snyder[(1)] and Xuguang Wang[(4)]

[1] National Center for Atmospheric Research (NCAR)[**], Boulder, CO, USA
[2]The UK Meteorological Office, Met Office, Exeter, United Kingdom
[3] York University, Toronto, Canada
[4] The School of Meteorology, University of Oklahoma, Norman, OK, USA

1. INTRODUCTION

Data assimilation provides techniques for blending in observations and prior model forecasts to create initial conditions for numerical weather prediction (NWP). Most of three dimensional variational (3D-VAR) data assimilation systems use static background error covariance that lacks flow-dependent information. There is a need for an ensemble system to estimate short-range background error co-variances to provide flow information for 3D-VAR systems (Barker, 1999).

In approximately the past decade, various data assimilation (DA) techniques have been developed to account for the uncertainties in initial conditions. One approach is the Ensemble Transform Kalman Filter (ETKF) (Bishop et al. 2001, Wang and Bishop 2003). The ETKF is a form of Kalman filter (Kalman, 1960) that provides forecast error covariance matrix estimation from the covariance matrix of the ensemble forecast perturbations. This technique transforms forecast perturbations into analysis perturbations through a transformation matrix approach. The ETKF solves for updated perturbations, given a current set of observations, in the ensemble subspace. It can also be used to update the mean in a full data assimilation system, but in the hybrid approach discussed here the mean is updated variationally.

Ensemble based DA systems suffer from problems with under-sampling because the number of ensemble members is much smaller than the number of degrees of freedom in an NWP model. Among other problems, this can cause an under-estimation of variance. Wang and Bishop (2003) introduced an adaptive inflation factor scheme to alleviate this problem. It aims to match the spread of the ensemble with the error of the ensemble mean forecast while accounting for the error in observations. The inflation factors may become large due to variance under-estimation and model errors (Bowler et al. 2008).

[*]*Corresponding author address:* Dr. Meral Demirtas, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000; Email: demirtas@ucar.edu

[**]National Center for Atmospheric Research is sponsored by the National Science Foundation.

Approaches that combine characteristics of ensemble-filter based DA schemes with variational DA techniques are often referred to as hybrid DA techniques. Some studies have suggested that an augmentation of an error covariance matrix with an ensemble-based error covariance matrix can provide an improved DA system (Barker 1999, Lorenc 2003, Etherton and Bishop, 2004). Lorenc (2003) has shown that by extending the control variables of an existing DA system, an ensemble based covariance model can be constructed to enhance the static background error covariance. The WRFVAR-ETKF based hybrid data assimilation scheme was introduced by Wang et al. (2008a). This hybrid technique updates the ensemble mean through WRF-VAR system using both the static and flow-dependent covariance estimates.

The remaining part of this paper is organized as follows. The description of WRFVAR-ETKF based hybrid system is given in Section 2. The experiment design is explained in Section 3. Outlines of some preliminary results are presented in Section 4. Section 5 outlines some conclusions drawn from this work.

2. THE WRFVAR-ETKF BASED HYBRID DATA ASSIMILATION SYSTEM CONFIGURATION

At the Data Assimilation Testbed Center (DATC), we used the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2005) and the WRF Variational Data Assimilation (WRF-VAR) system (Barker et al. 2004) in conjunction with the ETKF (Wang and Bishop, 2003) and hybrid technique (Wang et al. 2008a) to implement the hybrid WRFVAR-ETKF system.

The hybrid WRFVAR-ETKF system offers some tunable parameters to regulate contributions from the ensemble and 3D-VAR and a horizontal length scale for covariance localization (Hamill et al. 2001) that is designed for the ensemble covariance. We have set contributions from ensemble and static background error covariance estimates to be equal. The mean updated in WRF-VAR can be localized, and here horizontal scale of the covariance localization is 1500km. Further details of tunable parameters can be found in Wang et al. (2008b). Note that perturbations from the ETKF cannot be formally localized because they are computed in the ensemble subspace.

For each cycle, the ensemble mean of the forecasts is computed. Ensemble perturbations are derived from ensemble forecasts. The WRF-VAR is run in hybrid mode using the ensemble mean as the first guess, ensemble perturbations to provide flow-dependent information, and static background error covariance of the domain generated previously.

The ETKF system is used for accounting uncertainties in ICs and updating ensemble perturbations. Updated ensemble perturbations are then added to updated ensemble mean to generate updated ensemble ICs. The ensemble lateral boundary conditions (LBCs) are updated by incorporating updated ensemble ICs. The WRF system using the updated ensemble ICs and LBCs is run for generating next cycle's ensemble forecasts.

A number of sensitivity studies have been performed as a part of preliminary testing and optimization. It revealed erroneous observations leading to anomalously high inflation factors and computational instability issues. An additional step has been implemented to generate filtered (quality controlled) observations by eliminating observations largely deviated from the ensemble mean. This new step has resulted in stable runs. A flow diagram of the hybrid WRFVAR-ETKF system implementation has been depicted in figure 1.

3. EXPERIMENT DESIGN

We have set up a test domain with horizontal grid spacing of 45 km, 57 vertical levels, and the model top placed at 50 hPa. NCEP Global Forecast System (GFS) analyses, Agriculture Meteorology Modeling Systems (AGRMET) land surface analyses and Navy SST are used for initial and boundary conditions. Only conventional observations are assimilated.

We have designed a retrospective testing which consisted of 10-member ensemble with 3-hourly cycling for a 30-day period (15th August – 15th September 2007). The very first cycle's ensemble ICs and the ensemble LBCs are produced by adding spatially correlated Gaussian noise, which are provided by the background covariance model of WRF-VAR (Torn et al. 2006), to the GFS forecasts.

The runs outlined below help to evaluate whether the hybrid WRFVAR-ETKF system is more skilful and efficient than the standard WRF-3DVAR system. We performed a 30-day period retrospective runs as listed below:

i. Base runs: No variational data assimilation, only WPS, REAL and WRF.
ii. Computation of static background covariance matrix data for standard WRF-3DVAR
iii. Three hourly full cycling tests:
    - CYC1: Hybrid (the ETKF part with modest inflation factor generation mechanism).
    - CYC2: Hybrid (the ETKF part with a higher inflation factor generation mechanism).
    - CYC3: Using only standard 3DVAR

4. HIGHLIGHTS OF PRELIMINARY RESULTS

In this study, we tested two methods for ETKF inflation factor calculation. The first version provided modest inflation factors. The second version, coming directly from Wang and Bishop (2003), provided higher inflations factors and more spread. The impact of employing modest and high inflation factors has been evaluated by looking at the 500-hPa-height standard deviation. The 500-hPa height standard deviation of the ensemble for CYC1 and CYC2 indicate that CYC2 has better spread compared to modest CYC1 as shown in figure 2a and 2b respectively.

A simple verification against conventional observations was performed. For the brevity of this paper, we present only results based on CYC2 and CYC3 tests. Figures 3a and 3b depict the 24-h and 48-h forecast errors of wind (U and V), temperature (T), and water vapor mixing ratio (Q) for CYC2 (shown in red color) and CYC3 (shown in blue color) experiments respectively. The CYC2 (WRFVAR-ETKF) hybrid cycling run gives better RMSE scores especially for horizontal wind compared to CYC3 (WRF-3DVAR). It is noted that improvement for temperature and specific humidity fields are slight. We also verified our results against ECMWF analysis data set (T106) and found similar results as shown in figure 4a and figure 4b. The results indicate that there is no noticeable improvement at the model top levels. This may be due to some inherit problems of WRF model's representation of upper troposphere.

5. SUMMARY AND CONCLUSIONS

We have implemented a hybrid WRFVAR-ETKF system that combines ETKF flow-dependent error co-variances with static background error co-variances in the WRF-3DVAR system and tested it for a 30-day period with 3-hourly cycling.

This study used more conventional observation types compared to Wang et al. (2008b). This allowed the ETKF to account for errors in an analysis that uses a wider variety of observation types. Three-hourly cycling made better use of asynoptic observations. We also implemented an additional observation quality control step to prevent erroneous observations that can cause unrealistic high inflation factors. Preliminary simple verification results indicate a positive benefit of using the WRFVAR-ETKF hybrid system. This study also shows that the hybrid system can be effective even with a small ensemble size.

The ETKF scheme does not have covariance localization to mitigate spurious long-distance covariance, and the single inflation factor is applied domain-wide. A hurricane/tropical cyclone can therefore be subject to large (unphysical) perturbations, and the ensuing model run can become numerically unstable. We observed some instability issues during the active period of hurricane Dean, the WRF intermittently failed to complete a forecast. When model failures prevented computation of a flow-dependent covariance, the forecasts were restarted from the GFS-WRFVAR based ensemble IC and LBC generation explained in the Section 3.

Recent studies suggest that some forms of localization can be applied on the ETKF (Bowler et al. 2009 and Bishop and Hodyss 2009) to alleviate spurious correlations. Such approaches may help to reduce computationally unstable cases.

In future studies, additional isolated runs are needed to evaluate various tunable hybrid parameters. For example, the impact of increased weighted contribution from ensembles could be tested with additional sensitivity tests. It would also be interesting to investigate the impact of smaller/larger horizontal length scale for covariance localization. We only used static background error covariances produced from a series of single WRF forecasts (NMC method); investigating the benefit of tuning the static background error covariance matrix with ensemble-mean based forecasts (Wang et al. 2008b) could improve effectiveness of the static background error covariance. Using higher horizontal resolution may help to understand how the hybrid can contribute the improvement of mesoscale forecasts.

We may need to do some additional work to optimize the ETKF inflation factor further. Wang et al. 2007 has added a new adaptive factor to estimate the fraction of the ensemble-mean error variance projected onto the ensemble subspace in the ETKF. It is designed to alleviate the systematic underestimation on the error variance -which may be caused by small ensemble size. Bowler et al. (2008) suggested using a revised version of inflation factor generation mechanism that aims to avoid introducing oscillations to ensemble spread by using a geometric mean of the individual cycle inflation factors.

This study opens new avenues for future work to examine separate contribution of the various settings of the WRFVAR-ETKF hybrid system and their impact on sensitivity studies and real-time/operational weather forecasting. Our recent study suggests that a hybrid WRFVAR-ETKF system can be a good candidate for operational use.

Acknowledgements: The authors would like to thank the NCAR/CISL for providing supercomputer resources/facilities for our extensive runs. The Air Force Weather Agency (AFWA) funded this work. The views expressed are those of authors and do not necessarily represent official policy or position of the AFWA.

5. REFERENCES

Barker, D. M., 1999: Var scientific development paper 25: the use of synoptic-dependent error structure in 3DVAR. *UK MET Met. Office Technical Reports*, available from the UK Met. Office, Fitzroy Road, Exeter, Devon, EX1, 3PB, UK

Barker, D. M., W. Huang, Y. -R. Guo, and Q. N. Xiao, 2004: A Three-Dimensional (3DVAR) Data Assimilation System For Use With MM5: Implementation and Initial Results. *Mon. Wea. Rev.,* **132**, 897-914.

Bishop, C. H. and Hodyss, D. 2009: Ensemble covariances adaptively localized with ECO-RAP. Part 1: tests on simple error models. *Tellus*, **61A**, 84–96.

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

Bowler N. E., Arribas A, Beare S. E., Mylne K. R., Shutts G. J. 2009: The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.,* **135**, 767–776

Bowler, N. E., Arribas A, Mylne K. R., Robertson K. B., Beare S. E. 2008: The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc*. **134**, 703–722.

Etherton, B. J. and C. H. Bishop, 2004: Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev.*, **132**, 1065-1080.

Hamill, T. M., Jeffrey S. Whitaker and Chris Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.,* **129**, 2776-2790.

Kalman, R, 1960: A new approach to linear filtering and predicted problems. *J. Basic Eng.,* **83**, 95-105.

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.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, 2005: A Description of the Advanced Research WRF Version 2, NCAR Tech. Note, NCAR/TN–468+STR, 88 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO, 80307]. On-line at: http://box.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited area ensemble Kalman filters. *Mon. Wea. Rev.,* **134**, 2490-2502.

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

Wang, X., D. M. Barker, C. Snyder, T. M. Hamill, 2008b: A hybrid WRFVAR-ETKF data assimilation scheme for the WRF model. Part II: real observation experiments. *Mon. Wea. Rev.,* **136**, 5132–514

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

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

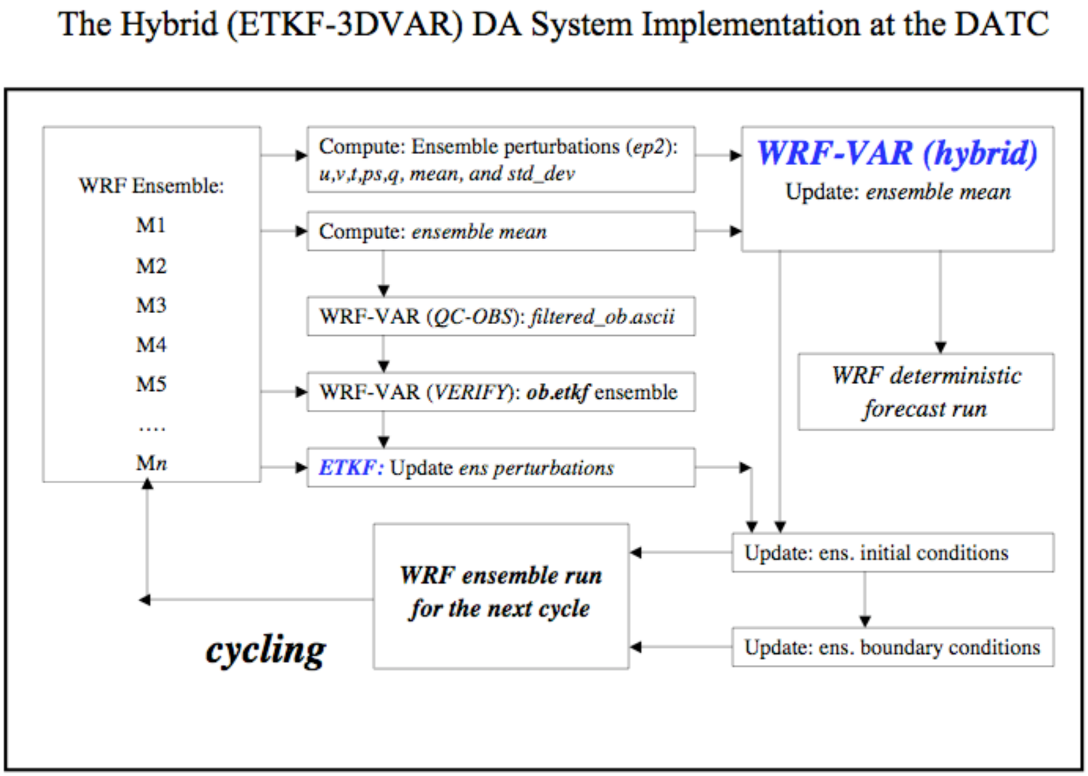

Figure 1: *A flow chart depicting the WRFVAR-ETKF-hybrid system implemented at the NCAR/DATC.*

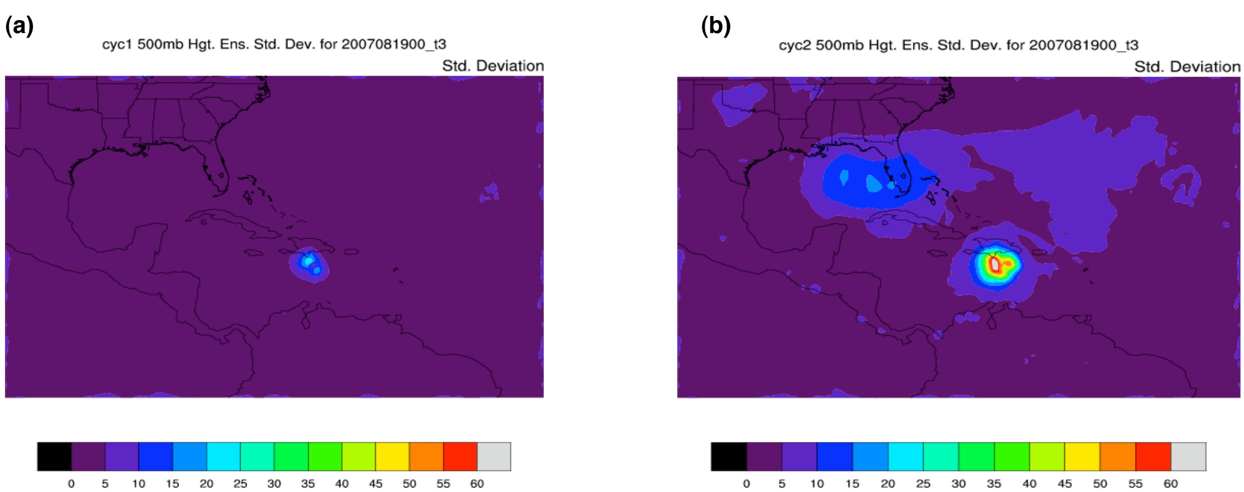

Figure 2a and 2b: *500hPa height (m) standard deviation for CYC1 and CYC2 tests respectively.*

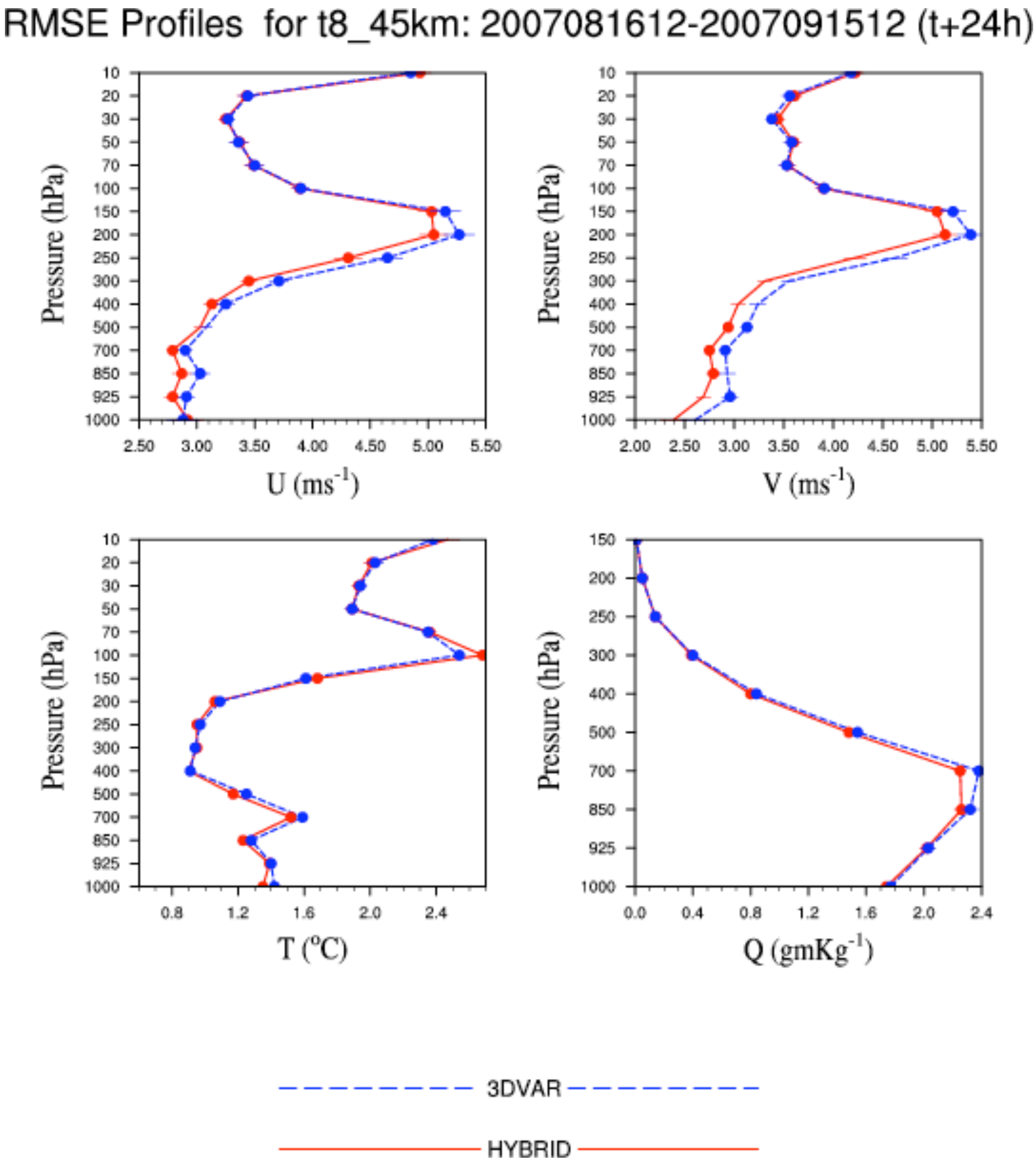

Figure 3a: *RMSE computed from radiosonde observations and twenty-four hour forecasts through the 30-day experiment period: U-wind (upper-left), V-wind (upper-right), temperature (lower-left) and specific humidity (lower-right). The horizontal bars indicate statistical significance. (Note that the CYC2 run is shown in red color and the CYC3 run is shown in blue color.)*

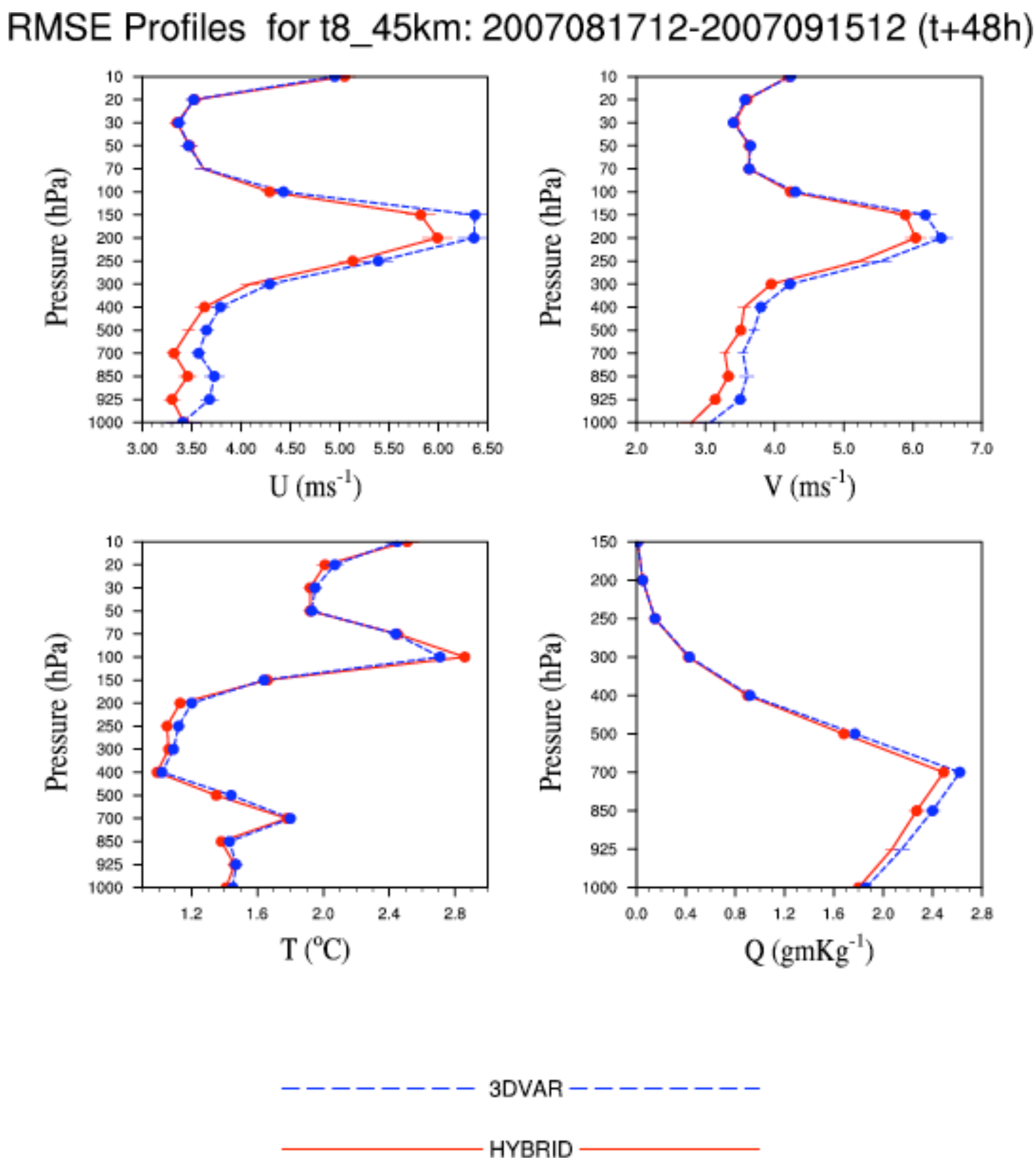

Figure 3b: *The same as 3a, but for forty-eight hours forecasts.*

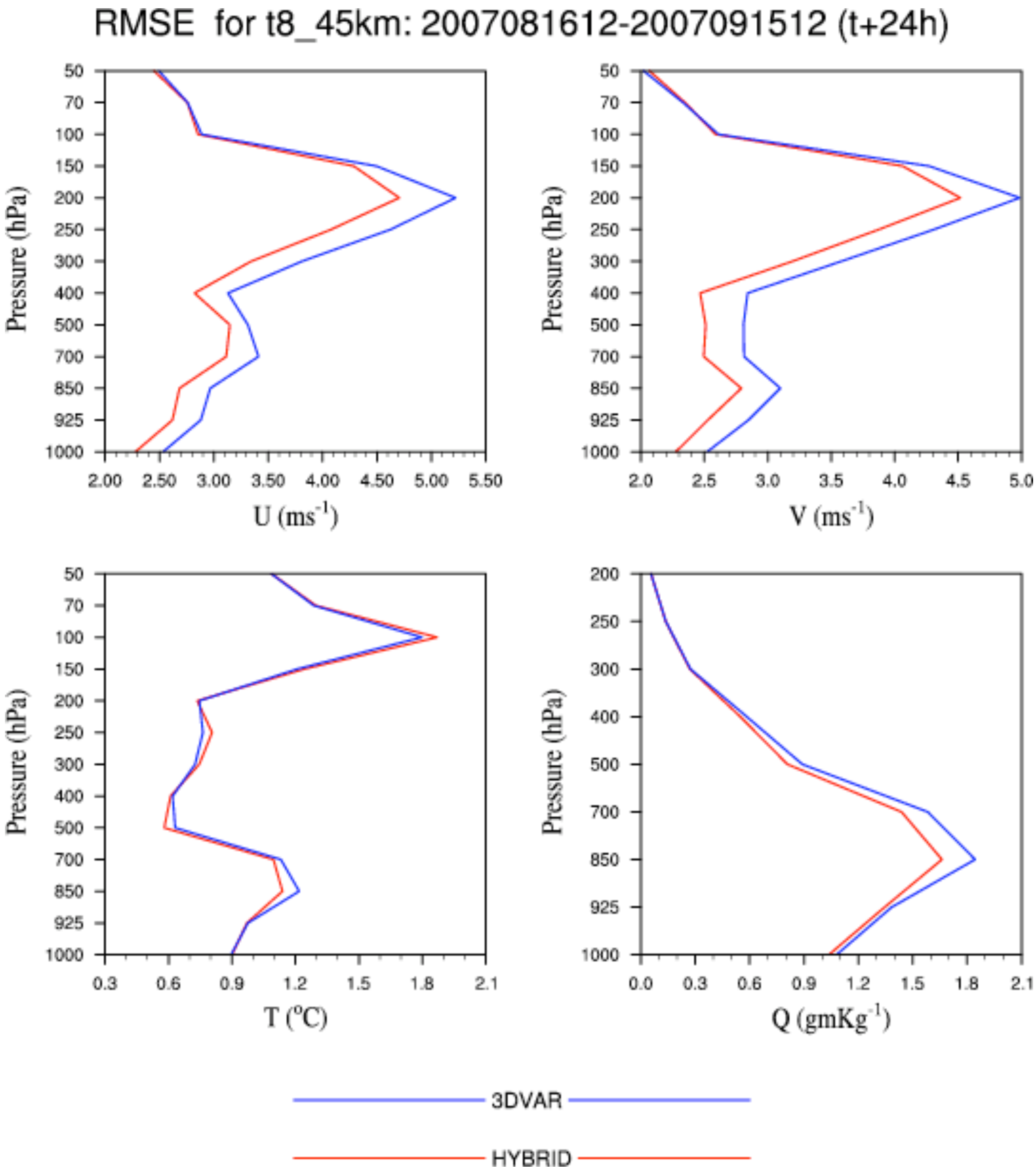

Figure 4a: *RMSE computed from ECMWF analysis data (T106) and twenty-four hour forecasts through the thirty-day experiment period: U-wind (upper-left), V-wind (upper-right), temperature (lower-left) and specific humidity (lower-right). The horizontal bars indicate statistical significance. (Note that the CYC2 run is shown in red color and CYC3 run is shown in blue color.)*

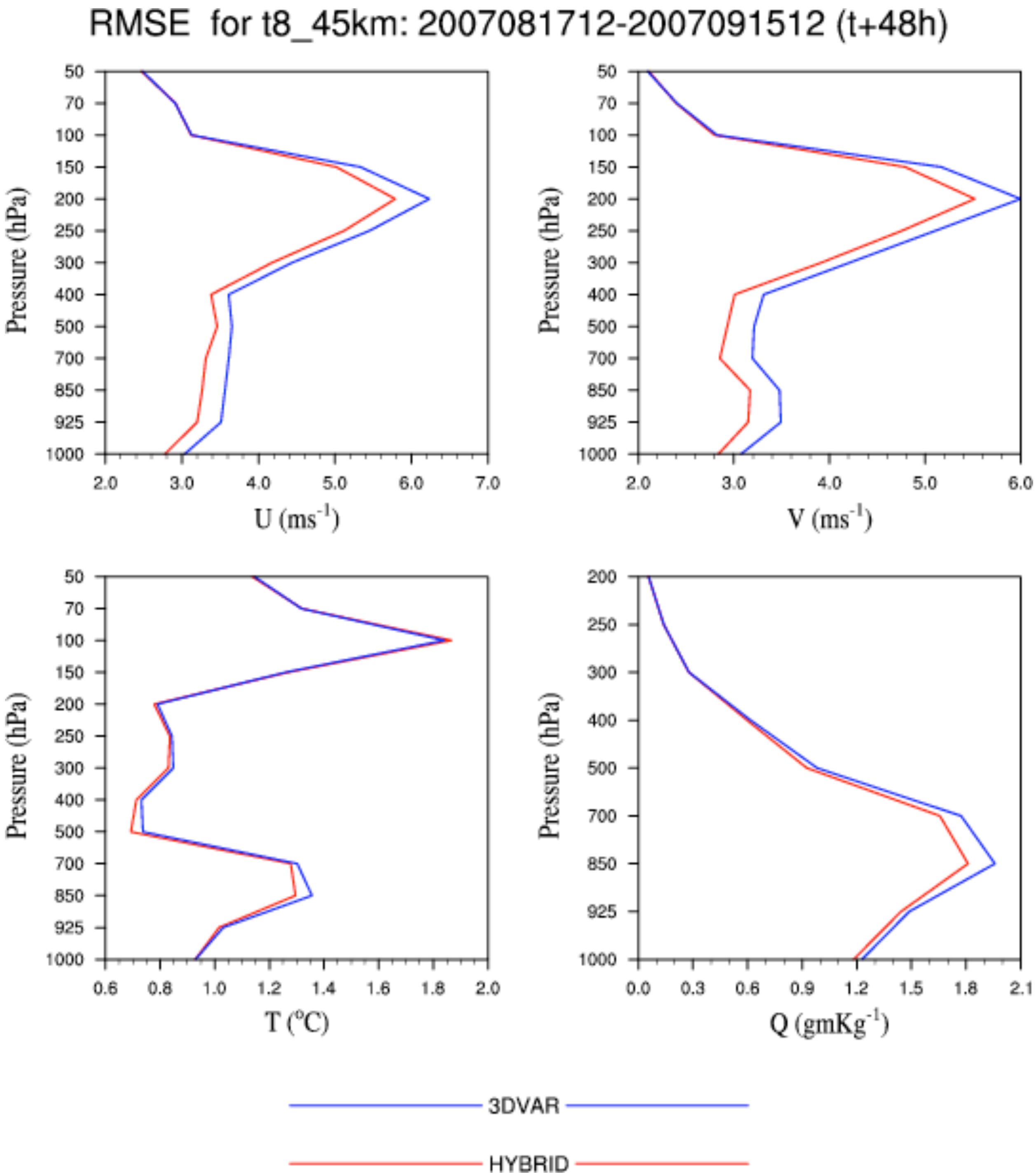

Figure 4b: *The same as 4a, but for forty-eight hours forecasts.*