

# Preliminary results of WRF 4D-Var

Xiang-Yu Huang<sup>\*†</sup>, Qingnong Xiao, Wei Huang, Dale Barker, John Michalakes, John Bray,  
Zaizhong Ma, Yongrun Guo, Hui-Chuan Lin, Ying-Hwa Kuo

National Center for Atmospheric Research, Boulder, Colorado

## 1. Introduction

The 4-dimensional variational data assimilation (4D-Var) idea (Le Dimet and Talagrand, 1986; Lewis and Derber, 1985) has been pursued actively by research community and operational centers over the past two decades. The 5<sup>th</sup> generation Pennsylvania State University – National Center for Atmospheric Research mesoscale model (MM5) based 4D-Var (Zou *et al.* 1995; Ruggiero *et al.* 2006), for example, has been widely used for more than 10 years. There are also successful operational implementations of 4D-Var (e.g. Rabier *et al.* 2000).

The 4D-Var systems have a number of advantages over 3-dimensional schemes including the abilities to:

- 1) Use observations at the almost exact times (to the width of the observation windows, see the discussion in the next section) that they are observed, which suits most asynoptic data,
- 2) Implicitly use flow-dependent background errors, which ensures the analysis quality for fast developing weather systems, and
- 3) Use a forecast model as a constraint, which ensures the dynamic balance of the final analysis.

The last mentioned advantage also implies that the current Weather Research and Forecasting model (WRF) based 3-dimensional variational data assimilation system (WRF 3D-Var), which is developed from MM5 3D-Var (Barker *et al.* 2004), should be enhanced with a 4-dimensional capability, using the WRF forecast model as a constraint, in order to provide the best initial conditions for the WRF model. The 4D-Var capability within the unified WRF 3/4D-Var (WRF-Var) system has been under extensive development since 2004. It uses the WRF model and WRF 3D-Var as its basic components (Huang *et al.* 2005). The WRF 4D-Var prototype was built in 2005 and has under continuous refinement since then.

Over the past year, major software engineering work has been carried out to parallelize WRF 4D-Var and to merge the 4D-Var code with the most updated WRF model and the WRF 3D-Var code. Many single observation experiments have been carried out to validate the correctness of the 4D-Var formulation. A series of real data experiments have been conducted to assess the meteorological performance of the 4D-Var system. This paper summarizes the preliminary results of these experiments.

## 2. 4D-Var For WRF

The WRF implementation of 4D-Var follows closely the incremental 4D-Var formulation of Courtier *et al.* (1994), Veersé and Thépaut (1998), and Lorenc (2003). The data flow and program structure of WRF 4D-Var is given in Fig. 1.

The input to WRF 4D-Var is as the following. The observations are grouped into  $K$  windows,  $\mathbf{y}_k$  ( $k=1, K$ ). A short-range forecast is used as the background,  $\mathbf{x}^b$ . Assume that the background error covariance matrix,  $\mathbf{B}$ , and the observation error covariance matrix,  $\mathbf{R}$ , are known. To integrate the WRF model over a time interval lateral boundaries, WRFBDY, are required. The 3D-Var solution can be obtained by setting  $K=1$  and removing WRF model related components.

The 4D-Var includes outer-loops and inner-loops. The outer-loops deal with nonlinear aspects of the assimilation problem while the inner-loops run a minimization algorithm for a quadratic problem. Using superscript  $\mathbf{n}$  for the outer-loop index the analysis vector,  $\mathbf{x}^n$ , is the final output of 4D-Var.

For the inner-loops, the minimization starts from a guess vector,  $\mathbf{x}^{n-1}$  (the analysis vector from the previous outer-loop). For the first outer-loop,  $\mathbf{n}=1$ ,  $\mathbf{x}^b$  is normally taken as the guess vector,  $\mathbf{x}^0$ . It should be stressed that in the incremental formulation the background vector and the guess vector should not be mixed. They are the same only during the first outer-loop.

Mathematically 4D-Var minimizes a cost function  $J$ , using its gradient  $J'$  with respect to the control variable,  $\mathbf{v}^n$ :

\* Corresponding author address: NCAR/MMM, P.O. Box 3000, Boulder, CO 80307-3000. Email: huangx@ucar.edu

† On leave from Danish Meteorological Institute, Copenhagen, Denmark

$$J'(\mathbf{v}^n) = \mathbf{v}^n + \sum_{i=1}^{n-1} \mathbf{v}^i + \mathbf{U}^T \mathbf{S}_{V-W}^T \sum_{k=1}^K \mathbf{M}_k^T \mathbf{S}_{W-V}^T \mathbf{H}_k^T \mathbf{R}^{-1} \left\{ \mathbf{H}_k \mathbf{S}_{W-V} \mathbf{M}_k \mathbf{S}_{V-W} \mathbf{U} \mathbf{v}^n + H_k [M_k(\mathbf{x}^{n-1})] - \mathbf{y}_k \right\} \quad (1)$$

and the preconditioning is through a variable transform,

$$\mathbf{v}^n = \mathbf{U}^{-1}(\mathbf{x}^n - \mathbf{x}^{n-1}) \quad (2)$$

where  $\mathbf{U} = \mathbf{B}^{1/2}$  (Barker *et al.* 2004); superscripts -1 and T denote inverse and adjoint of a matrix or a linear operator;  $H_k$ ,  $\mathbf{H}_k$  and  $\mathbf{H}_k^T$  are nonlinear, tangent linear and adjoint observation operators over observation window  $k$ , which transform atmospheric variables between the gridded analysis space and observation space;  $M_k$ ,  $\mathbf{M}_k$  and  $\mathbf{M}_k^T$  are nonlinear, tangent linear and adjoint models, which propagate in time the guess vector  $\mathbf{x}^{n-1}$ , analysis increments  $\mathbf{U} \mathbf{v}^n$  and analysis residual,  $\{\cdot\}$  in Equation (1), respectively;  $\mathbf{S}_{W-V}$ ,  $\mathbf{S}_{V-W}$ ,  $\mathbf{S}_{W-V}^T$  and  $\mathbf{S}_{V-W}^T$  are the 4D-Var specific operators which transform variables (e.g. between T and  $\theta$ ) and grids (between A-grid and C-grid) between VAR and WRF<sup>+</sup>.

WRF<sup>+</sup>, VAR and COM are the three major components of WRF implementation of 4D-Var from a program structure point of view (Fig. 1):

## I. WRF<sup>+</sup>

WRF<sup>+</sup> comprises 4 models (WRF\_NL, WRF\_TL, WRF\_AD and WRF\_SN) under the same framework and compiled together as a single executable. The WRF model is referred to here as WRF\_NL. Significant time was spent on selecting a simplified subset of WRF\_NL to form a simplified nonlinear model, WRF\_SN, which contains the full dynamics of WRF\_NL plus a minimal set of physics. WRF\_SN has been shown to produce reasonable short-range forecasts compared to WRF\_NL (Xiao *et al.* 2005). The Transformation of Algorithms in Fortran (Giering and Kaminski, 2003) is used to construct the tangent linear model, WRF\_TL, and its adjoint, WRF\_AD, from WRF\_SN. Most of the generated code passed the standard gradient tests and TL/AD tests following Zou *et al.* (1997). Sensitivity studies using WRF\_AD have been carried out and reported by Xiao *et al.* (2005). Results in Section 4 may also be used as a check for the accuracy of WRF\_TL.

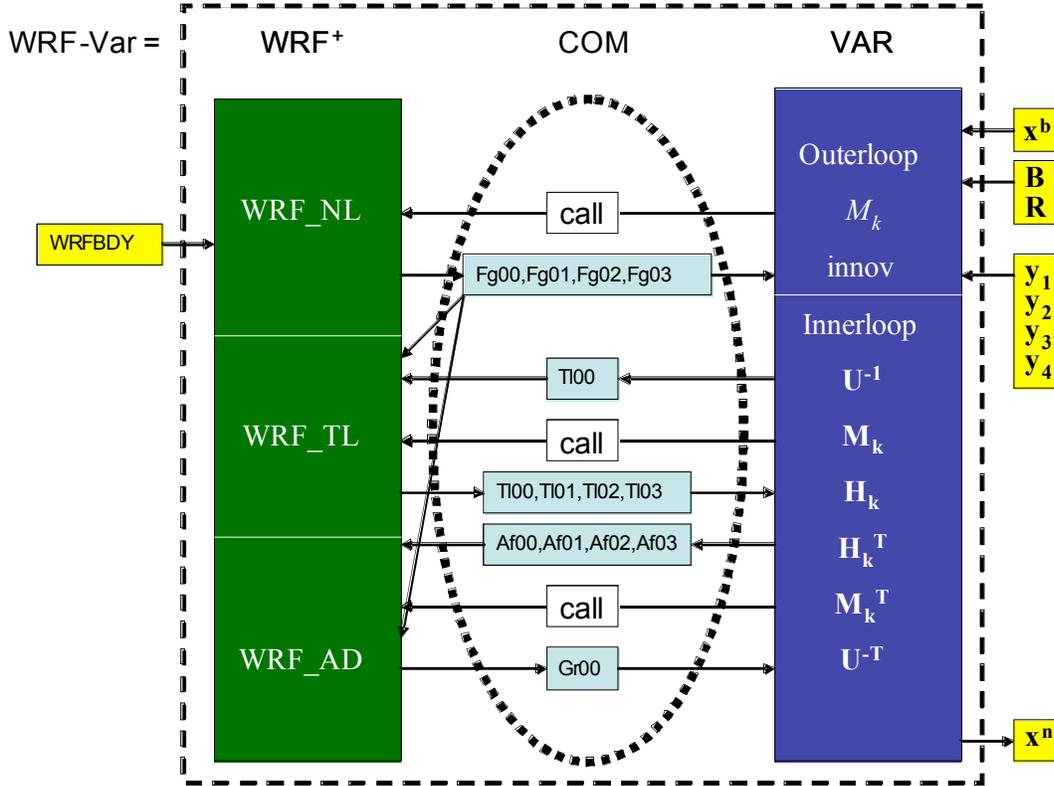


Fig. 1. The data flow and program structure of WRF 4D-Var.

## II. VAR

VAR contains all the components of WRF 3D-Var (Barker *et al.* 2004) plus the 4-dimensional related enhancements. Among the enhancements are the grouping of observations (break  $\mathbf{y}$  into  $\mathbf{y}_k$ ) and their related calculations (replace  $H$ ,  $\mathbf{H}$  and  $\mathbf{H}^T$  by  $H_k$ ,  $\mathbf{H}_k$  and  $\mathbf{H}_k^T$ ) according to the observation windows ( $k$ ); the calls to WRF\_NL, WRF\_TL and WRF\_AD; and the grid/variable transform operators.

## III. COM

As WRF<sup>+</sup> and VAR are separate components, communications between them are needed. COM manages this communication. The implementation of COM is hidden from the other two components, allowing the movement of data to be handled either through disk I/O or, for maximum efficiency, through memory.

## 3. The prototype

The 4D-Var prototype was built last year (Huang *et al.* 2005). The main features of the prototype are:

- 1) It runs as separate WRF<sup>+</sup> and VAR executables (wrfplus.exe and var.exe),
- 2) It uses calls to “system” to invoke wrfplus.exe from var.exe,
- 3) It uses disk I/O to handle the communication between WRF<sup>+</sup> and VAR, and
- 4) It can only run on a single CPU. (It runs as fast on a Mac G4 power book as it does on an IBM SP cluster, *e.g.* the NCAR bluesky.)

With the 4D-Var prototype it is possible to conduct single-observation experiments and case studies.

The first real data test of 4D-Var prototype has been made using a conventional data set. The background is a 6 h forecast valid at 0000 UTC 25 Jan 2000. Three-hour assimilation window is used, with 4 observation files centered at 0000, 0001, 0002 and 0003. A 3D-Var analysis, using the FGAT option (Huang *et al.* 2005), is also performed for comparison.

Figure 2 shows the cost functions ( $J$ ,  $J_o$  and  $J_b$ ) as functions of minimization simulations (iterations). It is shown that the convergence rate of the 4D-Var minimization is similar to that of 3D-Var and for this particular case 4D-Var reaches a lower minimum.

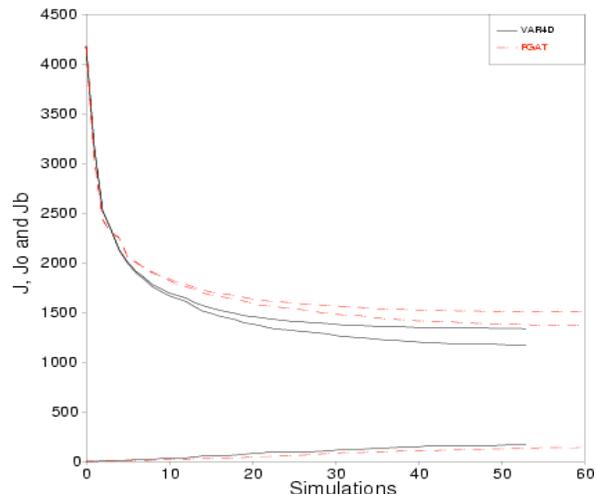


Fig. 2. The cost functions ( $J$ ,  $J_o$  and  $J_b$ ) as functions of minimization simulations (iterations). The full lines are for 4D-Var and the dashed lines are for 3D-Var (here the FGAT option of 3D-Var was used). For each experiment, the top two curves are for  $J$  and  $J_o$ , respectively and the bottom curve is for  $J_b$ .

There are two major limitations with the 4D-Var prototype. First, the problem size cannot be too large. The largest grid we have tested so far has  $91 \times 73 \times 17$  grid points, with 45 km horizontal spacing. Although this grid is quite small compared to most operational grids, the results from this grid are still useful for the future operational configuration we have planned, in which 4D-Var will be multi-incremental with the inner loop running on coarser resolution. Typically a factor of 3 between the outer-loop resolution and the inner-loop resolution is used, *i.e.*, if we use the above grid as inner loop, the outer-loop (and therefore the forecast model) will be on a grid with  $271 \times 217 \times 17$  grid points and 15 km horizontal spacing. Interpolation of nonlinear model trajectory (FG01, FG02, etc. in Fig. 1) from high resolution to low resolution, nest-up, and interpolation of the analysis increment  $\delta \mathbf{x}$  from low resolution to high resolution, nest-down, are needed in the multi-incremental formulation.

Second, it is very slow without parallelization. Using disk I/O may further slow down the computation significantly. With the 45 km grid, a 6-h assimilation window and a realistic observation data set ( $\sim 20000$  like those to be described in Section 5), a 4D-Var analysis needs to run 5 days on a Mac G4, if 100 iterations are required for finding the cost function minimum.

#### 4. 4D-Var structure functions

Analysis increments due to a single observation produced by a data assimilation system implicitly provide structure functions or effective background error covariance matrix  $\mathbf{B}$  (Thepaut *et al.* 1996). In order to compare the implicit structure function of 4D-Var and that of 3D-Var, many single observation experiments are carried out. An example of these experiments is shown in this section.

The background, a 6-h forecast valid at 0000 UTC 25 Jan 2000, is used for both 3D-Var and 4D-Var analyses. A single temperature observation at 0600 UTC is placed at (75 W, 30 N, 500 hPa). The case is constructed to demonstrate one of the potential problems related to 3D-Var when assimilating asynoptic observations. Although this case is constructed with a large time difference, the problem exists as long as the observation time differs from the analysis time.

The 3D-Var increments [the first panel (00h) of Fig. 3] show a Gaussian-like structure centered at the observation location. This is a graphic presentation of the background error covariance matrix,  $\mathbf{B}$ , or 3D-Var structure function. The increments are added to the background at the analysis time to produce the

3D-Var analysis. Two forecasts using WRF\_SN are then made, one from the background and the other from the analysis. The differences between the two forecasts are shown in Fig. 3. In this particular case, as the observation time and analysis time are 6 hours apart, it is clearly shown that the 6-h forecast from the analysis does not fit the observation anymore.

The 4D-Var increments have a temporal dimension. They are shown in Fig. 4. The increments at 06 h (the last panel of Fig. 4) give a graphic representation of the background error covariance matrix at 06h,  $\mathbf{MBM}^T$ , or 4D-Var structure function. In addition to providing a fit to the observation at the observation location, it has a clear flow-dependent nature. The increments at the analysis time (00 h, the first panel of Fig. 5) are small with a center upstream of the observation. The 4D-Var analysis is obtained by adding the increments at 00 h to the background. Again, two forecasts using WRF\_SN are made, one from the background and the other from the analysis. The differences between the two forecasts are shown in Fig. 5. The 6-h forecast from the 4D-Var analysis provides a good fit to the observation. It is also clear from Figs. 4 and 5, the linear approximations made in 4D-Var are reasonable for this case.

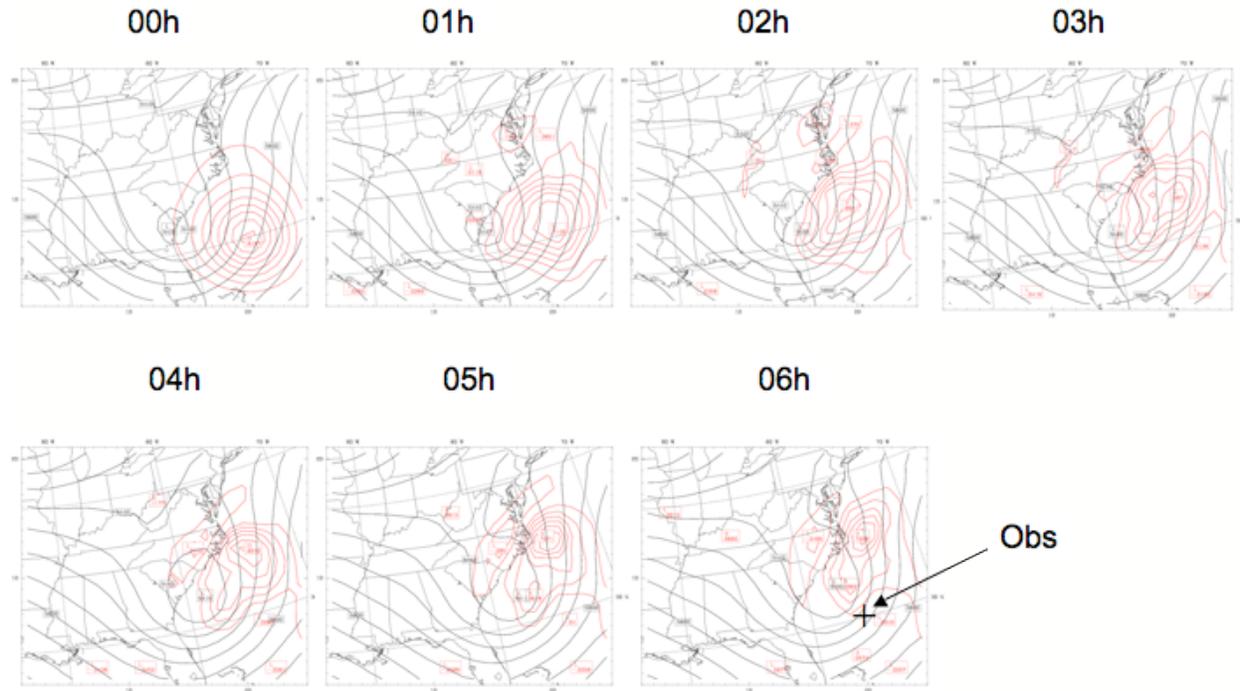


Fig. 3. 500mb  $\theta$  difference at 00,01,02,03,04,05,06h from two nonlinear runs, one from background and the other from FGAT.

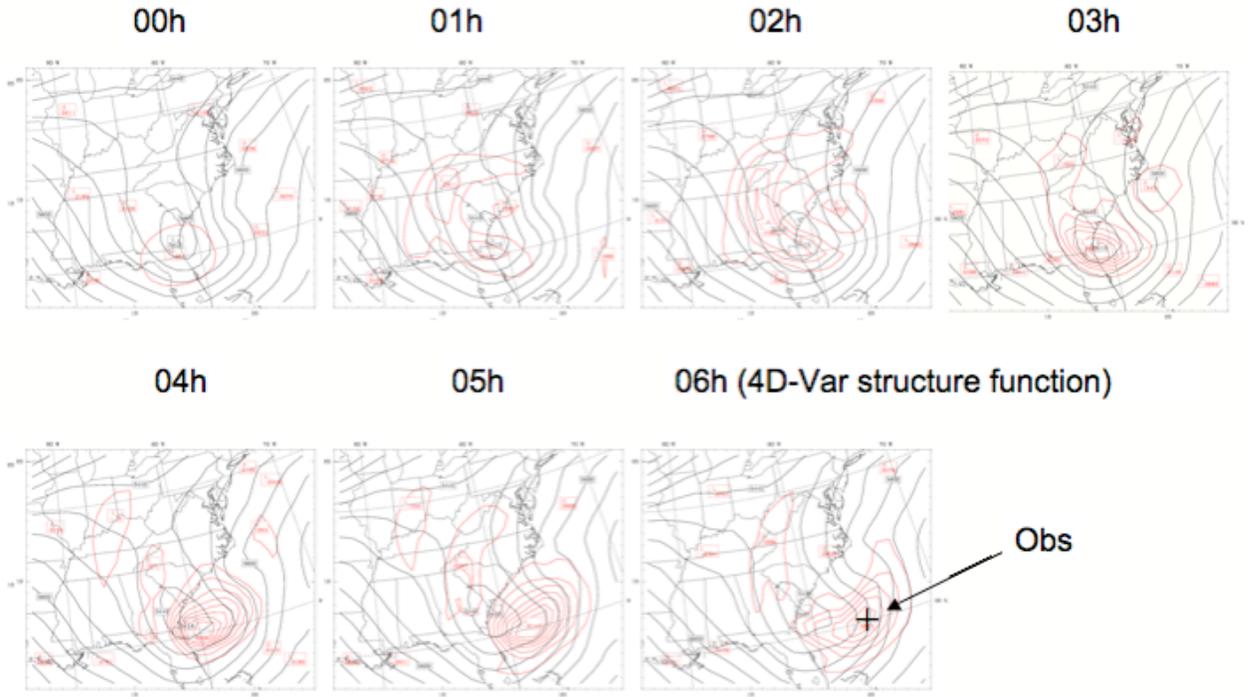


Fig. 4. 4D-Var 500mb  $\theta$  increments at 00,01,02,03,04,05,06h to a 500mb T ob at 06h.

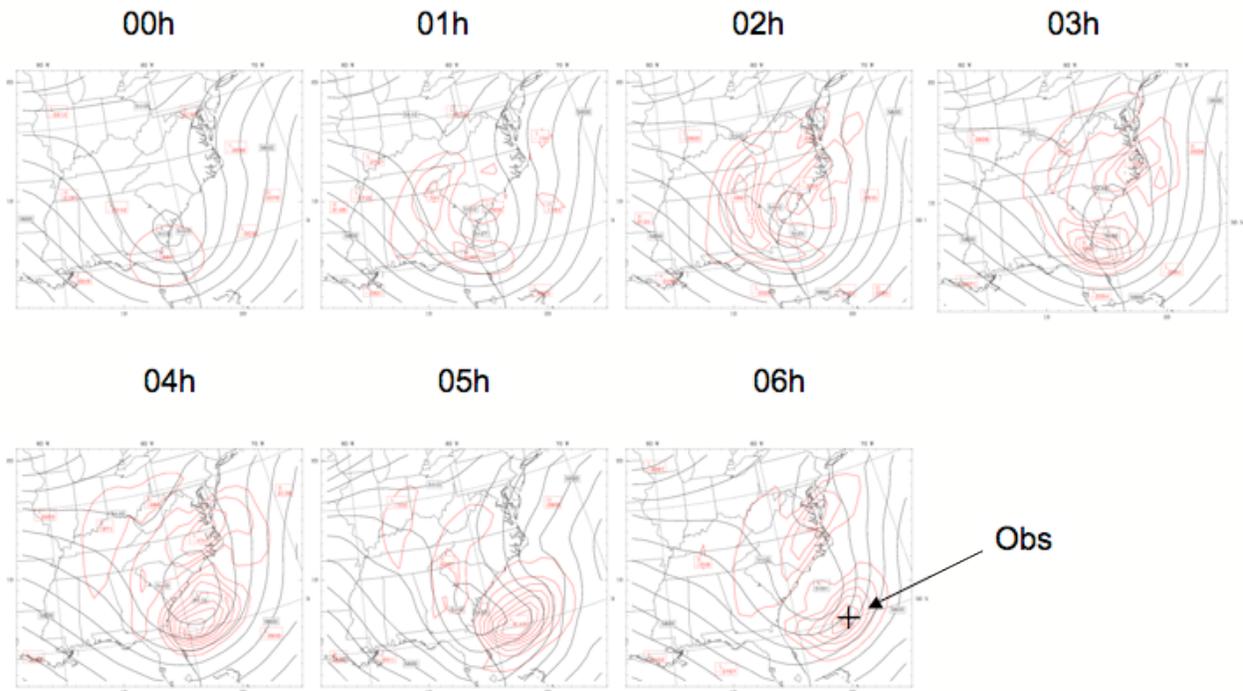


Fig. 5. 500mb  $\theta$  difference at 00,01,02,03,04,05,06h from two nonlinear runs, one from background and the other from 4D-Var.

## 5. Typhoon Haitang

To assess the 4D-Var performance and to test it in a near operational configuration, a series of experiments have been carried out on Typhoon Haitang, which hit Taiwan on 0000 UTC 18 July 2005 (Guo *et al.* 2006).

Starting from 0000 UTC 16 July (denoted as 1600) and repeating every 6 h to 0000 UTC 18 July (1800), 5 parallel experiments are run at each analysis time:

FGS – forecast from the background [The background fields are 6-h WRF forecasts from National Center for Environment Prediction (NCEP) GFS analysis.]

AVN- forecast from the NCEP GFS analysis

3DVAR – forecast from 3D-Var

FGAT – forecast from FGAT [an option of 3D-Var, see Lee and Barker (2005) and Huang *et al.* (2005)]

4DVAR – forecast from 4D-Var

The same parameter set and physics options are used for all forecast runs. The grid has 91x73x17 grid points with a 45 km horizontal spacing and 4 min time step.

The observations include conventional data, satellite data and bogus data from the Central Weather Bureau of Taiwan. The numbers of different observation types assimilated by 4D-Var at 0000 UTC 16 July (between 0000 UTC and 0600 UTC) are given in Table 1. At other analysis times, there are also GPS refractivity ( $N$ ) data and QuikScat wind ( $QS-u$ ,  $QS-v$ ) data (e.g. 212  $N$ , 2594  $QS-u$  and 2605  $QS-v$  at 0600 UTC 16 July).

Table 1. The numbers of different observation types assimilated by WRF 4D-Var at 0000 UTC 16 July.

Obs type	$u$	$v$	$T$	$p$	$q$	$\Delta Z$
TEMP	727	724	869		697	
SYNOP	119	218	237	226	236	
SATOB	3187	3182				
AIREP	923	930	939			
PILOT	156	160				
METAR	167	191	216		200	
SHIP	69	70	77	79	73	
SATEM						511
BUOY	67	67		64		
BOGUS	1200	1200	788	788	80	

The 48-h forecast typhoon tracks, all started at 0000 UTC 16 July 2005, are plotted in Fig. 6,

together with the observed track. The background sea level pressure field is also shown in the figure. For this case, the forecast from FGS is worst judging from the track. The forecasts from AVN, 3DVAR and FGAT are of the same quality. The forecast from 4DVAR produces the best track.

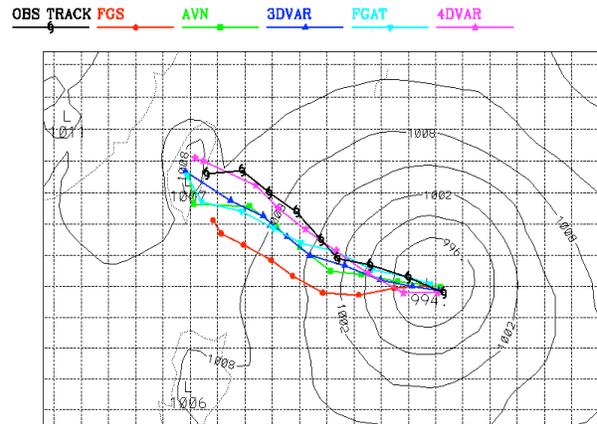


Fig. 6. 48-h forecast typhoon tracks from FGS, AVN, 3DVAR, FGAT, 4DVAR, together with the observed track. Forecasts are all made from 0000 UTC 16 July 2005. The background sea level pressure field from FGS is also shown in the figure.

Similar track plots have been obtained for forecasts started at other analysis times. To make the comparison easier, the track errors in km averaged over the 48-h forecast range are listed in Table 2. The best forecast at each analysis time is highlighted. It is evident that the 4D-Var produces superior forecast for Haitang track over this period.

Table 2. The track error in km averaged over 48 h for each forecast.

Time	FGS	AVN	3DVAR	FGAT	4DVAR
1600	159	85	72	77	<b>66</b>
1606	108	83	<b>67</b>	97	79
1612	93	100	95	<b>82</b>	137
1618	116	67	103	<b>52</b>	54
1700	80	66	68	62	<b>52</b>
1706	83	80	80	67	<b>65</b>
1712	111	104	<b>90</b>	112	128
1718	113	113	133	129	<b>93</b>
1800	116	221	192	<b>103</b>	111
Sum	109	102	100	<b>87</b>	<b>87</b>

Up to now, only the typhoon track forecasts have been investigated. Other aspects of the analyses and

forecasts will also be studied and reported in the near future. Observing system experiments will be carried out to assess the impact of different observation types, in particular that of bogus (Guo *et al.*, 2006).

## 6. Summary and future plans

In this paper, a brief overview of the 4D-Var capability within WRF-Var is given and the progress made in the past year is summarized. Preliminary results indicate that the 4D-Var works properly and can be used to assimilate many observations of different types.

In the experiments discussed in this paper, only one outer-loop is used in analyses and the NCEP background error (BE) statistics (CV3) is used. As discussed in Guo *et al.* (2006), further improvements could be obtained by using more outer-loops, a better BE and well-tuned parameters.

We are in the process of building the 4D-Var basic system, featuring parallelization and using memory for communication between WRF<sup>+</sup> and VAR. The 4D-Var basic system is expected to be computationally efficient as the program structure of WRF model and WRF-Var will be maintained. Extensive data assimilation experiments will be carried out with the basic system.

Gravity waves are generated due to imbalances during the minimization iterations. To control these waves a penalty term based on the Digital Filter Initialization, JcDFI (Wee and Kuo, 2004), will be implemented.

Other planned developments include a penalty term for lateral boundary control, more physics in the tangent linear and adjoint models, and the multi-incremental capability.

## Acknowledgments

The 4D-Var development for WRF has been primarily supported by the Air Force Weather Agency and the Korea Meteorological Administration.

## References

Barker, D.M., W. Huang, Y.-R. Guo, A.J. Bourgeois, Q.N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897-914.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation

of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367-1387.

Giering, R., and T. Kaminski, 2003: Applying TAF to generate efficient derivative code of Fortran 77-95 programs. *PAMM*, **2(1)**, 54-57.

Guo, Y.R., H.-C. Lin, X. X. Ma, X.-Y. Huang, C.T. Terng, and Y.-H. Kuo 2006. Impact of WRFVar (3DVar) Background Error Statistics on Typhoon analysis and Forecast. WRF users' workshop, Boulder, Colorado, 19-22 June 2006.

Huang, X.-Y., Q. Xiao, W. Huang, D. Barker, Y.-H. Kuo, J. Michalakes, Z. Ma, 2005: The weather research and forecasting model based 4-dimensional variational data assimilation system. Reprints, WRF/MM5 users' workshop, Boulder, Colorado, 27-30 June 2005.

Lee, M.-S., and D. Barker, 2005: Preliminary Tests of First Guess at Appropriate Time (FGAT) with WRF 3DVAR and WRF Model, *J. Korean Met. Soc.* **41**, 495-505.

Le Dimet, F. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretic aspects. *Tellus*, **38A**, 97-110.

Lewis, J. and J. Derber, 1995: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**, 309-327.

Lorenc, A.C. Modelling of error covariances by 4D-Var data assimilation. *Quart. J. Roy. Meteor. Soc.*, **129**, 3167-3182.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf and A. Simmons. 2000: The ECMWF operational implementation of four dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143-1170.

Ruggiero, F. H., J. Michalakes, T. Nehrkorn, G. D. Modica, and X. Zou, 2006: Development of a New Distributed-Memory MM5 Adjoint. *J. Atmos. Ocean. Technol.*, **23**, doi: 10.1175/JTECH1862.1, 424-436.

Thépaut, J.-N., P. Courtier, G. Belaud and G. Lemaître, 1996: Dynamic structure functions in a four-dimensional variational assimilation: A case study. *Quart. J. Roy. Meteor. Soc.*, **122**, 535-561.

Veersé, F. and J.-N. Thépaut, 1998: Multi-truncation incremental approach for four-dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1889-1908.

Wee, T.-K., and Y.-H. Kuo, 2004: Impact of a digital filter as a weak constraint in MM5 4DVAR. *Mon. Wea. Rev.*, **132**, 543-559.

Xiao, Q., Z. Ma, W. Huang, X.-Y. Huang, D. M. Barker, Y.-H. Kuo, and J. Michalakes, 2005: Development of the WRF tangent linear and adjoint models: Nonlinear and linear evolution of

initial perturbations and adjoint sensitivity analysis at high southern latitudes. WRF/MM5 users' workshop, Boulder, Colorado, 27-29 June 2005.

Zou, X., Y.-H. Kuo, and Y.-R. Guo, 1995: Assimilation of atmospheric radio refractivity using a nonhydrostatic mesoscale model. *Mon. Wea. Rev.*, **123**, 2229-2249.

Zou, X., F. Vandenberghe, M. Ponca and Y.-H. Kuo, 1997: Introduction to adjoint techniques and the MM5 adjoint modeling system. NCAR Tech. Note NCAR/TN-435-STR, 110pp.