J2.4

INITIAL VERIFICATION OF THE MM5 3DVAR DATA ASSIMILATION SYSTEM

Dale M. Barker, Wei Huang, Yong-Run Guo and Francois C. Vandenberghe

MMM Division, National Center for Atmospheric Research,*
Boulder, Colorado

1. INTRODUCTION

In recent years, much effort has been spent in the development of variational data assimilation systems as the "next-generation" to replace previously used schemes e.g. the Cressman (MM5), FDDA (MM5), optimum interpolation (OI - NCEP, ECMWF, HIRLAM, NRL, etc) and analysis correction (UKMO) algorithms. Practical advantages of the variational approach to more traditional data assimilation techniques include

• Observations can be assimilated in a relatively "raw" form without the need for independent retrieval prior to the assimilation step, e.g. satellite radiances can be assimilated directly. This results in a more consistent treatment of all observations through use of a common definition of model error. In addition, errors in the raw observations are likely to be less correlated, leading to practical simplifications in the analysis algorithm.

• A degree of flow-dependence can be applied in the background error covariances used as input to the variational data assimilation system. In 3DVAR, the semi-geostrophic transform, anisotropic recursive filter and additional analysis variables have all been utilized for this purpose. In 4DVAR systems, the use of the forecast model (or its tangent linear (TL)) and a corresponding adjoint model automatically implies a synoptical-dependence in the background error covariances.

• Approximate dynamical balance can be implemented relatively easily via "weak constraints" e.g. the use of "unbalanced" control variables, additional terms in the cost function or (in 4DVAR) the balance of the forecast model itself.

• In 4DVAR, asynoptic data can be inserted within a model timestep of its validity time, thus reducing time-tendency errors of static data assimilation systems.

• Variational data assimilation systems can be formulated in a variety of spaces including grid-point, spectral, observation and EOF space. Each formulation has its own merits in particular circumstances.

Having expounded the advantages of variational data assimilation it is wise to also recognise its weaknesses. Although the variational analysis is frequently described as "optimal", this label is subject to a number of assumptions. Firstly, given both imperfect observations and prior (e.g. background forecast) information as inputs to the assimilation system, the quality of the output analysis will depend crucially on the accuracy of their respective *errors*. The specification of both observation and background error is a large subject in itself and one which is being studied at NCAR and many other establishments. Secondly, although the variational method allows for the inclusion of linearised dynamical/physical processes, in reality errors in the NWP system may be related in a highly nonlinear manner. This limits the usefulness of variational data assimilation systems, especially in the tropics and at meso- and convective-scales.

Despite the above weaknesses, the use of variational methods has led to significant improvements in forecast verification in both operational and research communities. The list of operational centres running VAR systems grows every year. Following the first operational 3DVAR at NCEP (Parrish & Derber, 1992), other centres have followed with variants of the basic algorithm - ECMWF (1996), CMC (1997), DAO (1997), Meteo-France (1998), NCEP Eta(1998), UKMO (1999), FSL(2000) and NRL(2000). Global incremental 4DVAR is now implemented at ECMWF (1998) and Meteo-France (2000) and has led to significant forecast improvements.

Given the short-term (1-2 year) goal of designing a variational data assimilation system for operational use at the U.S. Air Force Weather Agency (AFWA) and Taiwan's Civil Aviation Authority (CAA), our plan has been to initially concentrate on producing a respectable

---

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

Corresponding author address: Dr. Dale M. Barker, MMM Division, NCAR, P.O. Box 3000, Boulder, CO 80307-3000. Email: dmbarker@ucar.edu

($i.e.$ accurate, computationally efficient and robust) 3DVAR system. The reasons for this are not just the computational resource issue of running 4DVAR but also

• Many of the algorithms used in 4DVAR are found in the much less computationally expensive 3DVAR system (eg observation operators, minimization, preconditioning, multivariate, background error specification, data assimilation diagnostics). The only notable exception is the adjoint of the forecast model (and optionally its TL), required for 4DVAR. The 3DVAR system therefore provides a training ground for these crucial aspects of variational data assimilation.

• There are still many potential ways of improving existing 3DVAR systems without resorting to 4DVAR. These improvements include additional observation types, more sophisticated background/observation errors and improved balance constraints. These methods are referred to as "low-hanging fruit" in that they are sometimes computationally very cheap and also there is (and will be for the forseeable future) much observational data that is under-used or ignored completely in current data assimilation systems.

• Previous 3/4DVAR systems have been initially developed for the global assimilation problem. It is prudent to consider the particular aspects of the mesoscale (eg differing balance, impact of moisture, convection, etc) before following the path of other systems' development.

The layout of the rest of this report is as follows. In section 2, a summary of the main features of the MM5 3DVAR system is given. Section 3 describes results of preliminary tests. Finally, future plans for NCAR-MMM/MM5's data assimilation efforts are briefly described in section 4.

2. THE MM5 3DVAR SYSTEM

The goal of variational data assimilation is to find the analysis $\mathbf{x}$ which minimizes a prescribed cost function e.g.

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}^b - \mathbf{x})^T \mathbf{B}^{-1} (\mathbf{x}^b - \mathbf{x}) + \frac{1}{2}(\mathbf{y}^o - \mathbf{y})^T (\mathbf{E} + \mathbf{F})^{-1} (\mathbf{y}^o - \mathbf{y}) \quad (1).$$

The various sources of information are the background vector $\mathbf{x}^b$ and its error covariance matrix $\mathbf{B}$, the observation vector $\mathbf{y}^o$, the observation instrument $\mathbf{E}$ error and the error $\mathbf{F}$ associated with the observation operator H which transform between analysis and observation space $\mathbf{y} = H\mathbf{x}$.

The fundamental equations of variational data assimilation have been written down many times in previous works (e.g. Lorenc 1986, Courtier *et al.* 1994). The particular features of the MM5 3DVAR system are summarized below. Many of the ideas presented here found their first application in the operational global/mesoscale 3DVAR system of the UKMO described in Lorenc *et al.* (2000).

• **Observation types** currently assimilated include surface, rawinsonde, aircraft and satellite cloud track winds. Total precipitable water (TPW) and oceanic surface wind speed retrievals from SSM/I data may be assimilated as may TPW retrievals from ground-based GPS data. A single Fortran90 derived data type is used to store all observations, with specific sub-types for each observation platform. Coded in this way, the introduction of new observation types is relatively straightforward and efficient use is made of core memory.

• **Background error covariances** are currently derived via the *NMC-method* of scaled, averaged forecast differences valid at the same time.

• The MM5 3DVAR **analysis domain** is not tied to the domain of the subsequent forecast (except in cycling mode). This option may be useful in future studies of cold-starting mesoscale 3DVAR from e.g. global analysis background forecasts where observations close to, but outside the forecast domain may be used to improve the analysis of lateral/top boundaries.

• The cost function minimization is performed in terms of **analysis increments**. This formulation restricts the impact of new observations, and any imbalance due to approximations in the analysis, to the (relatively small) increments $\mathbf{w}'$. The full analysis is $\mathbf{x}^a = \mathbf{x}^b + I\mathbf{w}'$. The operator I may include dynamical initialization, change of grid resolution/stagger, variable etc.

• The incremental cost function minimization is performed in **preconditioned control variable** space $\mathbf{v}$ defined by $\mathbf{w}' = U\mathbf{v}$. The transform U is chosen to satisfy the approximation $\mathbf{B} \simeq UU^T$ via spatial filtering and judicious choise of physical variables. The analysis space is gridpoint in the horizontal and a projection onto eigenvectors of the background error covariance matrix in the vertical. This projection allows efficient data compression at the expense of some spatial/temporal averaging of vertical error correlations (calculated using NMC-method data).

• **Horizontal background error correlations** are currently represented using an isotropic and homogeneous recursive filter algorithm supplied by Jim Purser (NCEP). Horizontal filter parameters are a function of vertical eigenvector, hence the background error covariance is non-separable.

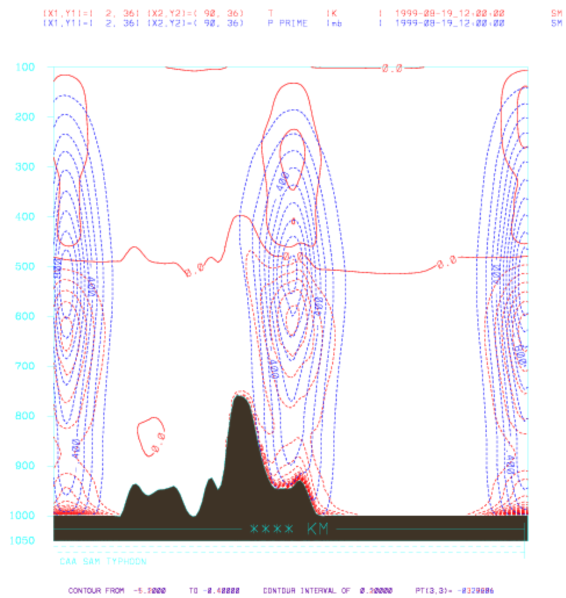

Figure 1. Test 3DVAR increment response to three pressure observations introduced in the mid-tropsphere (500hPa). Both temperature (full) and pressure (dashed) increments are shown. The synthetic observations are all given the same O-B value of -1.5hPa.

- The U transform includes a final **transform of variable** from streamfunction, velocity potential, unbalanced pressure and specific humidity (chosen as their errors are relatively uncorrelated) to forecast model variables. The code includes switches to experiment with alternative control variables, e.g. relative instead of specific humidity. The final increments $\mathbf{w}'$ are produced on the Arakawa-B grid and sigma-height vertical coordinate of the non-hydrostatic MM5V3. The increments are multivariate through application of a linearized geostrophic/cyclostrophic balance equation:

$$\nabla^2 p_b' = -\nabla \cdot \overline{\rho} \left[f\mathbf{k} \times \mathbf{v}' + \overline{\mathbf{v}} \cdot \nabla \mathbf{v}' + \mathbf{v}' \cdot \nabla \overline{\mathbf{v}} \right] \quad (2)$$

and use of the hydrostatic approximation. The linearization is performed around the background $\mathbf{x}^b$. The accuracy of this approximation is tested via statistical regression of p_b' on p using forecast difference data.

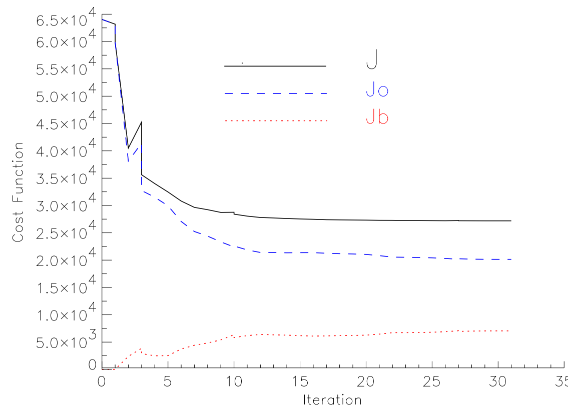

Figure 2. Minimization of the cost function (and its components) for the MM5 3DVAR run with an analysis time of 12Z on the 19th August 1999. The domain is the 135km East Asia "Domain 1" of the Taiwanese CAA system.

3. PRELIMINARY RESULTS

Extensive tests are currently underway to assess the performance of the MM5 3DVAR system relative to the LITTLER package previously used with MM5. Within 3DVAR itself, the accuracy of the adjoints of both observation operators and control variable transforms are continually monitored, as is the invertibility of change in physical variable.

Background error covariances are tested via single/multiple observation tests. An example is shown in Figure (1). In this test, three mid-tropospheric pressure observations have been assimilated in 3DVAR. The vertical structure of the increments are defined via the eigenvectors of the vertical component of **B**. The horizontal structure is defined via the recursive filter. Also shown in Figure (1) is the temperature increment response to the pressure increments. This indicates a warm anomaly above and cool anomaly below the maximum (negative) pressure increment. The temperature increments are derived assuming hydrostatic balance.

Full 3DVAR analyses containing a wide variety of observations have been produced. The successful minimization of one case is shown in Figure (2). This example was run on the Taiwanese CAA 135km Domain 1 which covers most of East Asia. Convergence (defined by a reduction in initial cost function gradient of two orders of magnitude) is achieved in 31 iterations. The observations assimilated were 667 synops, 65 metars, 87 ships, 297 aireps, 187 soundings and 14748 satellite cloud track winds (SATOBs). On a DEC Alpha workstation, 3DVAR required just 145s CPU for this case.

Initial verification of forecasts run from the 135km domain as well as the nested domains at 45km and 15km has begun. Results from these tests will allow us to tune the system to be ready for release to operational and research communities in the next 12 months.

4. FUTURE WORK

The development of the MM5 3DVAR system to include the above features has continued at a fast pace since the freezing of an initial univariate version of the code in December 1999. The system already contains many of the features present in successful operational implementations of 3DVAR. However, continued testing and tuning is certainly required before the code is ready for release to both operational centres (the US Air Force and Taiwanese CAA) and the general MM5 user community. In particular, the background error covariances must be tuned and the system must be modified to run on parallel architectures. All this work is currently underway.

A particular goal in the development of the MM5 3DVAR system has been to maintain flexibilities which will enable aspects of the MM5 3DVAR system to be used in future WRF data assimilation systems e.g. 3/4DVAR, coupled variational-ensemble systems. This effort will continue as the WRF data assimilation system matures and additional capabilities e.g. satellite radiances, anisotropic recursive filters, extension to 4DVAR are included.

REFERENCES

Courtier, P., Thepaut, J.-N. and Hollingsworth, A., 1994: *A Strategy For Operational Implementation of 4D-VAR, Using An Incremental Approach. Quart. J. Roy. Meteor. Soc.*, , **120**, 1389-1408.

Lorenc, A.C., 1986: *Analysis Methods For Numerical Weather Prediction. Quart. J. Roy. Meteor. Soc.*, , **112**, 1177-1194.

Lorenc, A. C., Ballard, S.P., Bell, R. S., Ingleby, N. B., Andrews, P. L. F., Barker, D. M., Bray, J. R., Clayton, A. C., Dalby, T., Li, D., Payne, T., J. and Saunders, F. W., 2000: *The Met. Office Global 3-Dimensional Variational Data Assimilation Scheme. Quart. J. Roy. Meteor. Soc.*, , **126**, 2991-3012.

Parrish, D.F. and Derber, J.C, 1992: *The National Meteorological Center's Spectral Statistical Interpolation Analysis System. Mon. Wea. Rev.*, **120**, 1747-1763.